Simulation of Intrinsic Parameter Fluctuations in Nano-CMOS Devices

Gareth D. Roy

Submitted to the University of Glasgow,

Department of Electronics and Electrical Engineering,

in fulfilment of requirements for the degree of Doctor of Philosophy.

October 2005

Abstract

As devices are scaled to gate lengths of sub 100 nm the effects of intrinsic parameter fluctuations will become increasingly important.This work presents a systematic simulation study of intrinsic parameter fluctuations, consisting of random dopant fluctations, line edge roughness and oxide thickness fluctuations, in a real 35 nm MOSFET developed by Toshiba. The simulations are calibrated against experimental data for the real device and it is found that discrete random dopants have the greatest impact on both the threshold voltage and leakage current fluctuations with a σV_T of $33.2mV$ and a percentage increase in the average leakage current of 50%. Line edge roughness has the second greatest impact with a σV_T of $19mV$ and percentage increase in the average leakage current of 45.5%. The smallest impact is caused by oxide thickness variations resulting in a σV_T of $1.8mV$ and a 13% increase in the average leakage current. The combined effects of pairs of fluctuations is also studied, showing that these sources of intrinsic parameter fluctuations are statistically independent and a calculated σV_T of $39mV$ is given for all of the sources combined. This value is on par with that reported in literature for the 90 nm technology node.

Acknowledgements

I would like to thank my Mother and Father for all of the aid and support that they have offered me as I have worked to complete this thesis, without their love and strength I would never have gotten finished. I'd also like to thank my PhD supervisor, Asen, without whose help this document would never have been completed and who has been a constant guide throughout the years of this work. I would also like to thank my brother, Scott, who acted as my second supervisor, who gave aid and support throughout this work, again without his assistance I would never have finished. I'd also like to thank those in the device modelling group for all their aid and support, particularly Jeremy for his help with the lattice placement algorithms, Andy for his guidance with the "atomistic" simulator, Campbell and Karol for their sanity checking, my office mates Craig and Fikru for listening to me rant and adding their support and Binjie for doing useful things with my results.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 1965 Gordon E. Moore made the observation that there had been an exponential growth in the number of transistors used within integrated circuits in the past and postulated that this trend would continue in the future [2]. Years later Moore reiterated this statement [3] which has become popularly known as Moore's Law. In simple terms Moore's Law results in doubling the transistor performance and quadrupling the number of devices every three years [4]. In order to accomplish this, the dimensions of MOSFETs have been aggressively scaled in most cases using the rules proposed by Robert Dennard in 1974[5]. A wealth of papers address the different approaches of MOSFET scaling to sub-$0.1\mu m$ dimensions and the related issues [6, 7, 8, 9]. The path of this aggressive scaling has been laid out in many generations of what has become known as SIA's International Technology Roadmap for Semiconductors (ITRS) [10]. This document has become a strategic guide for the semiconductor industry in its continuous efforts to shrink MOSFETs to nanometer dimensions. It reviews all

of the MOSFET design parameters required to keep pace with Moore's Law and the technological barriers that need to be overcome.

Together with the postulation of Moore's Law and the formulation of the scaling rules the ultimate limits of semiconductor device miniaturisation have been continuously examined [11, 12]. With the progressive scaling of MOSFETs to nano dimensions these limits have become more apparent and problematic [13]. High channel doping and tunnelling thin gate oxides already restrict the scaling of the conventional MOSFET [14]. Intrinsic parameter fluctuations associated with the discreteness of charge and matter become increasingly important [15, 16, 17] in conjunction with the continuous reduction in supply voltage [6]. The scaling rules that have worked for so long have begun to fail. Many of these problems are intrinsic to the nature of semiconductor devices and cannot be removed by better processing steps or improved equipment [18]. The problems associated with the scaling of conventional MOSFETs have prompted alternate transistor designs such as Ultra Thin Body (UTB) Silicon on Insulator (SOI) MOSFETs [19], Double Gate FET [20, 21], FinFET [22, 23] and others [4]. However the introduction of new device architectures have been slow. For technological and economical reasons MOSFETs [24, 25, 26] are still the workhorse of the industry. They will continue to co-exist even after the introduction of UTB SOI MOSFETs at the 45 nm technology node and double gate MOSFETs at the 32 nm node. Therefore understanding and predicting the effects that intrinsic parameter fluctuations have on the behaviour and performance of conventional MOSFETs is an important problem.

Numerical simulations have become an important tool for understanding the

physics and operation of MOSFETs and for improving their design [27]. The corresponding simulation tools can also be extended to understand the effects that intrinsic parameter fluctuations have upon device characteristics and performance. Conventionally in numerical device simulations the MOSFET is treated as a perfect device with smooth interfaces and boundaries and a continuous doping profile [28]. The inclusion of different sources of intrinsic parameter fluctuations cause statistical variations within an ensemble of devices. This shifts the paradigm of numerical device simulation [29]. It is no longer enough to simulate a single perfect device. In order to take into account, statistically, the impact of various sources of intrinsic parameter fluctuations an ensemble of microscopically different devices must be simulated. The mean values, the variances and in some cases the higher moments of the statistical distribution of important design parameters such as threshold voltage, tranconductance, subthreshold slope and off current should be studied to provide an understanding of how a device will behave in a circuit environment.

The purpose of this thesis is to study, using numerical simulations, different sources of intrinsic parameter fluctuations in a real 35 nm MOSFET typical for the late stages of the 90 nm technology node. It will focus on the introduction of different sources of intrinsic parameter fluctuation into a 3D device simulation, and will study the effect that each source has on the behaviour and performance of a MOSFET. It will also study the combined effect of different sources of intrinsic parameter fluctuation and the possible correlation between them. We will evaluate the suitability of different simulation techniques for studying intrinsic parameter fluctuations, will highlight problems associated with the inclusion of

statistical variations within standard simulation methods and will discuss the potential solutions to the problems.

1.1 Outline

The remainder of this thesis is organised as follows, Chapter 2 entitled "Atomistic Simulation" reviews the modelling efforts to study intrinsic parameter fluctuations. It outlines the main sources of intrinsic parameter fluctuation including discrete random dopants, line edge roughness and oxide thickness fluctuations and discusses their origin. It then explains the effects of the different sources of intrinsic parameter fluctuations on the MOSFET parameters and characteristics. It concludes by discussing various techniques which can be employed to simulate intrinsic parameter fluctuations comparing their strengths and weaknesses.

Chapter 3, entitled "Development of the Device Simulator" contains an outline of the simulation techniques used in this study of intrinsic parameter fluctuations. It gives a general overview of the Glasgow Drift Diffusion simulator and the efforts to port the simulator to new computational resources which became available for the purpose of this work. It then describes in detail the implementation of each individual source of intrinsic parameter fluctuation currently available in the simulator. The chapter concludes with a description of the extensions and improvements that have been made to the device simulator and the models used to improve its accuracy and efficiency.

Chapter 4 entitled "Resolving Discrete Charges" contains a description of the approach that has been implemented to allow the resolution of individual discrete

dopants in 3D drift diffusion simulations. It outlines the problems associated with the inclusion of individual discrete dopants in classical drift diffusion simulations and illustrates these problems in the study of an "atomistic" resistor. It then outlines two possible methods to solve this problem, the first being the use of charge smearing using various charge assignment schemes, and the second being the introduction of quantum corrections for both the electrons and the holes in the solution domain.

Chapter 5 entitled "Real Device Simulations" contains the results of the simulation of intrinsic parameter fluctuations in a real 35 nm MOSFET. It begins by outlining the methods of importing the doping profile from a commercial TCAD process simulator and describes the calibration of the Glasgow device simulator to match the published characteristics of the 35 nm device. Next we present the results for the simulation of this device with three different sources of intrinsic parameter fluctuation; random dopants, line edge roughness and oxide thickness fluctuations. The chapter concludes with the presentation of results of simulations which include pairs of intrinsic fluctuation sources, their effect on the device characteristics and their potential correlation.

Chapter 6 contains the conclusions drawn from this work outlining what has been accomplished and outlining possible future work that can extend the simulation and the understanding of the effects of various sources of intrinsic parameter fluctuations in nano CMOS devices.

Chapter 2

Atomistic Simulation

The International Roadmap for Semiconductors (ITRS) [10] states that by 2006 the MOSFETs in mass production will have a printed gate length of 40 nm and a physical channel length of 28 nm. Near the end of the ITRS the mass production devices will have a printed gate length of 13 nm leading to a physical gate, after processing, of 9 nm. In recent years devices with 40-50nm physical gate lengths have been manufactured [30, 31] and within the research environment devices with 35 nm (Toshiba [32]), 20 nm (Intel [24]), 15 nm (AMD [25]) and 5 nm ([26]) have been reported. As MOSFETs are aggessively scaled to such small dimensions, intrinsic defects caused by production methods, as well as variations in the atomic structure of these devices, will play an important roll influencing their performance and introducing variations in the device characteristics. Due to such atomic scale variations in the microscopic structure of the devices the conventional method for simulating MOSFETs, assuming continuous doping profiles, straight boundaries and smooth interfaces, is no longer valid. The in-

Year	2001	2002	2003	2004	2005	2006	2007	2010	2013	2016
Printed Gate Length (nm)	90	75	65	53	45	40	35	25	18	13
Physical Gate Length (nm)	65	53	45	37	32	28	25	18	13	9
Vdd (V)	1.2	1.1	1.0	1.0	0.9	0.9	0.7	0.6	0.5	0.4
Equivalent Physical oxide thickness (nm)	1.3 - 1.6	1.2 - 1.5	1.1 - 1.6	0.9 - 1.4	0.8 - 1.3	0.7 - 1.2	0.6 - 1.1	0.5 - 0.8	0.4 - 0.6	0.4 - 0.5
Tox (nm) Electrical Equivalent	2.3	2.1	2.0	2.0	1.9	1.9	1.4	1.2	1.0	0.9
Tox No. Of Atomic Layers	5 - 6	4 - 5	4 - 5	3 - 4	3 - 4	3 - 4	2 -3	2 - 3	1 - 2	1 - 2
Channel Concentration for Wdepletion 1/4Leff (cm-3)	4.0e18	6.0e18	8.0e18	1.1e19	1.4e19	1.6e19	3.3e19	5.0e19	1.3e20	5.0e20
Uniform Channel Concentration for Vt = 0.4V (cm-3)	0.8 - 1.5e18	0.8 - 1.5e18	1.5 - 2.5e18	1.5 - 2.5e18	1.5 - 2.5e18	2.0 - 4.0e18	2.5 - 5.0e18	5.0 - 9.0e18	0.9 - 1.8e19	1.5 - 3.0e19
Calculated No. of Dopants in channel region per micron.	2340 - 4387	1653 - 3100	2227 - 3712	1498 - 2497	1152 - 1920	1176 - 2352	1125 - 2250	1170 - 2106	1170 - 2340	945 - 1890
Calculated No. of Dopants in channel where W = 2Leff.	304 - 570	175 - 328	200 - 334	110 - 184	73 - 122	65 - 131	56 - 112	42 - 75	30 - 60	17 - 34

Table 2.1: Predicted values from the International Roadmap of Semiconductor Devices for the years 2001 to 2016, showing the desired attributes of conventional MOSFETs. No. of Dopants in the channel are derived numbers based on the doping concentration and a analytical square channel.

fluence of various sources of intrinsic fluctuation need to be properly taken into account.

Table 2.1 outlines predicted MOSFET design parameters for high performance MOSFETs according to the International Roadmap for Semiconductors [10]. With scaling of the device size the oxide thickness, channel doping concentration and physical gate length will become critical parameters in MOSFET design and all of these will become an important source of intrinsic parameter fluctuations. An example is the influence of discrete random dopants where variation in their number as well as their location within a MOSFET, make each device microscopically different [16]. These microscopic differences will alter the behaviour of the devices affecting threshold voltage and current flow. Such intrinsic variations are unavoidable and may be reduced but not removed from a device design if doping in the channel and in the source / drain regions is

present. For instance a MOSFET produced in 2016 using the information found in Table 2.1 will have a physical channel of only 9nm, which is of comparable scale to the experimental NEC device [26] of 5nm. If the width of the transistor is two times the effective channel length (18 nm) there will be approximately 17 to 34 dopant atoms in the active channel region, with the channel itself being only approximately 18 Si atoms in length. When devices of this size are mass produced, the "atomistic" nature of the device will strongly influence their behaviour. The movement of carriers within the channel of this device will follow random conducting paths created by potential fluctuations associated with the random discrete dopants. This will alter the characteristics of each device altering, for example, such parameters as the threshold voltage. The possibility of trapping individual carriers on defect states within the channel could also lead to noise, degradation and potentially a complete failure of the device.

Fluctuations will also occur due to variation in the thickness of the gate oxide material. From Table 2.1 the thickness of the oxide will become only a few atomic layers by 2007. The typical roughness at the Si/SiO_2 interface is on the scale of 1 to 2 atomic layers [33] which implies that within future devices thickness fluctuations will cause variations of up to 50% of the thickness of the oxide. These variations will act as an additional source of intrinsic parameter fluctuation. To avoid problems with the high gate leakage current through such thin oxides layers, high-κ dielectric materials will replace the silicon dioxide in the gate stack of the transistor after the 65 nm technology node. The high-κ materials allow the use of thicker dielectric layers without compromising the gate capacitance. This will introduce new sources of intrinsic parameter variations

associated with local variations in the structure and the composition of high-κ materials.

The molecular aggregates found within the photoresist used in the manufacture of MOSFETs will introduce unavoidable line edge fluctuations within the gate pattern [34]. These fluctuations will locally reduce or increase the effective length of the channel. The current size of these fluctuations is of the order of 3-6 nm. Bearing in mind that the devices aimed for production in 2016 will have 9 nm channel lengths, such LER could double the size or completely destroy the channel if no improvements are made in the nature of the molecular resist and the corresponding LER.

This chapter gives more detail on the most common sources of intrinsic parameter fluctuations, which have the greatest effect on the behaviour of the present and next generation MOSFETs. It also outlines the effects that these fluctuations have upon circuits and, using as a particular example the implication for SRAM cell performance. The chapter concludes by outlining the various methods that are employed in MOSFET simulation including the effect of fluctuations within these devices, comparing both their strengths and weaknesses.

2.1 Intrinsic Parameter Fluctuations

In the past the mismatch in the characteristics of transistors in semiconductor chips was mainly associated with variations in the process parameters resulting in macroscopic variations in the layer thickness, geometry and average doping profile within the device. "Intrinsic" parameter fluctuations considered in this

work are associated with the granular nature of matter and charge, and will occur even with perfect processing conditions. The three main sources of "intrinsic" parameter fluctuations which will be considered here are random discrete dopant distributions, line edge roughness and oxide thickness variations. Random dopant fluctuations are caused by variation in the position and number of the dopant atoms within the channel which result in potential fluctuations which shape the flow of carriers within the channel. The random position of the dopants allow percolation paths to form as the device turns on, and this can cause devices to turn on more easily if there are many conducting paths, or with greater difficulty, if there are few conducting paths. Line edge roughness causes alterations in the local gate length. This in turn guides the implantation of dopant atoms resulting in random variations in the p-n junction positions and the channel length causing drain current variations between devices. Fluctuations associated with oxide thickness variations occur because of local alterations to the capacitance of the MOS capacitor affecting locally the inversion layer conditions. The variations in the oxide thickness are related to the interface roughness which also causes fluctuations in the mobility due to the variation in the surface roughness scattering from device to device. Other sources of fluctuation, such as strain effects and dielectric composition fluctuations will also be considered in general but are not an integral part of the research in this project.

2.1.1 Random Dopants

Modern MOSFETs have complex doping profiles. The impurity atoms are implanted with a sufficient energy to penetrate into the silicon and are activated

using annealing which allows them to replace Si atoms in the silicon lattice. Figure 2.1 shows a cartoon of a conventional MOSFET of gate length 4.2 nm with the correctly seated crystalline lattice superimposed onto the structure of the MOSFET. In the implantation process the trajectory of the individual impurities is random since they suffer a large numbering of scattering events before coming to rest. Therefore the final position of the individual impurities after the implantation is uncorrelated although as a whole the doping has a particular distribution determined by the implanted species and the implantation conditions. In modern MOSFETs there would be a series of implantations in order to fabricate the device; well implantations, implantations for control of the threshold voltage and short channel effects, source and drain implantation including the source and drain extensions. There may also be pocket implantations around the extensions to further improve the short channel effects. The diffusion of the implanted dopants during the annealing and the other high temperature fabrication steps is also a random process often modelled using a kinetic Monte Carlo approach, which stochastically traces the trajectory of each dopant atom.

The overall effect of the implantation and the annealing is a random dopant distribution for each device at time of manufacture. No two devices are identical. The unique microscopic doping distribution in each device alter the characteristics from device to device. An example of this random dopant distribution is shown in Figure 2.2 where the individual discrete dopants of a 50×50nm square MOSFET are shown. The effects of discrete random dopants on the threshold voltage were first predicted in the early seventies [12, 11], and later confirmed experimentally [16, 35, 36, 37, 15, 38, 39]. This effect has been studied

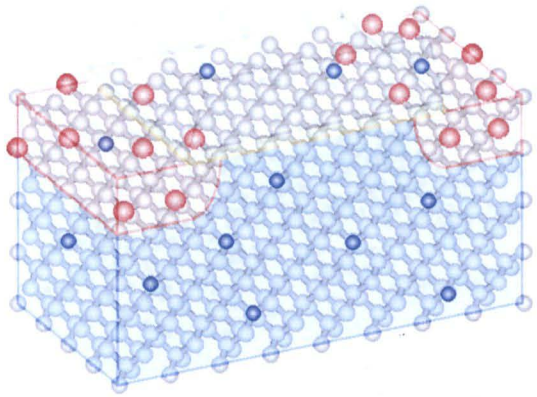

Figure 2.1: Schematic diagram of a 4.2 nm conventional MOSFET showing the crystaline lattice and the placement of individual discrete random dopants

using analytical models [35, 40, 37, 7, 41], 2D [41, 42, 43] and 3D simulations [18, 44, 28].

The random placement of dopant atoms within the simulation is in most cases based on a continuous doping profile [44, 42, 45, 46, 47, 48].

A better, more realistic way, is to use an atomic scale process simulator which traces the trajectories of the individual dopants during the fabrication in a more *ab initio* fashion [49, 50]. In the case of generating random dopant distributions from a continuous doping profile, the expected number of dopants in the "atomistic" simulation region can be calculated by integrating the continuous doping profile. The actual number of dopants is then selected from a Poisson distribution with the above calculated mean and then the dopant placed in the simulation domain following the probability determined by the continuous

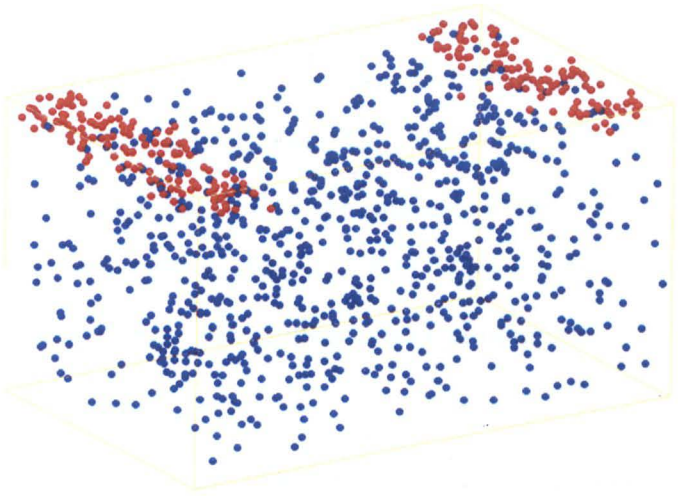

Figure 2.2: The postion of discrete random dopants in an "atomistic" 50 × 50 nm MOSFET device, red dopants are donors within the source and drain region while the blue dopant positions represent the acceptors in the channel and substrate region.

doping distribution using a rejection technique.

Using 3D simulation it has been shown that the random discrete dopants introduce fluctuations in the threshold voltage, V_T, of a MOSFET and a lowering of the average threshold voltage of an ensemble of devices [18]. Although in the scaling process the total doping density within the channel is increased (as required by standard scaling rules to suppress short channel effects), the overall number of discrete random dopants within the channel decreases due to the physical size reduction. Because of this reduction in the average number of dopant atoms, N, there will be an increase in the magnitude of the fluctuations, ΔV_T determined by $\sqrt{N}$. Numerical simulation has also shown that the doping concentration dependence of random dopant induced fluctuations is stronger than that obtained from analytical models [28]. This is due to the fact the the

analytical models only take into account variations in the number of dopants and not the variation in their relative positions. This positional dependence is of great importance as it has been found that the dopants closest to the interface are responsible for a large fraction of the intrinsic fluctuation. This observation has recently been confirmed analytically [51]. This implies that devices which have steep HALO doping or low doped epitaxial channels [52] have significant fluctuation resistance.

The inclusion of quantum mechanical effects in the simulations result in an increase in the random dopant induced threshold voltage fluctuations [53]. The quantum mechanical effects which were taken into account in [53] using the Density Gradient approximation, increase the fluctuations mainly due to the increase of the equivalent oxide thickness (EOT). The quantum mechanical effects increased the threshold voltage fluctuations more than 50% in MOSFETs with oxide thicknesses of less than 1.5 nm. If the random dopants in the poly-silicon gate is also taken into account this figure rises to 100% in devices with ultra thin gate oxides. In the atomistic simulations the quantum mechanical shift in threshold voltage associated with the inclusion of quantum mechanical effects is partially compensated by the random dopant induced threshold voltage lowering.

As noted above, the random dopants in the poly-silicon in combination with the poly-depletion effects have an increased effect on the fluctuations in an ensemble of MOSFET devices. It has been shown [54] that when both gate depletion and random dopants within the gate are included in 3D simulations, they have a great influence on MOSFET devices with oxide thicknesses within the

range of 1-2 nm, which according to the ITRS [10] will be in production from 2003 onwards. While both the random gate dopants and the poly-depletion influence the fluctuations, the dominant effect is the increase of the equivalent oxide thickness. The poly Si grain boundaries are an additional factor effecting the intrinsic parameter fluctuations.

2.1.2 Line Edge Roughness

One of the most important steps in the fabrication of a MOSFET is the patterning of the gate. The gate edge, in most cases, is used as a mask for the implantation of dopants in the source and drain regions and controls the position of the metallurgical p-n junctions. In order to create the gate pattern using modern lithographic tools an organic photo resist is used, which is selectively exposed to UV radiation in selected areas. These resists come in two flavours, either positive resists where the exposed region is dissolved on development or negative resists where the unexposed region is removed. Both these types of resists suffer from line edge roughness (LER) which is limited by the molecular nature of the resist and not by the lithographic means by which the pattern is created [55]. Line edge roughness is caused by polymer aggregates that form within the resist material, these aggregates can reach sizes of up to 30nm within commercial resists [56]. In the past LER has caused little worry since the critical dimensions of MOSFETs were orders of magnitude larger than the roughness, but the size of the largest polymer aggregate is now of the same order of magnitude as the channel length of modern MOSFETs. These polymer aggregates have a different, longer solubility rate than the surrounding resist material due

to an increased density within the aggregate. This means that when an area of resist material, whether positive or negative, is being removed the aggregates are removed at a different rate than the surrounding material, which either leave the shape of the polymer aggregate in the sidewall or leaves the aggregate itself causing a rough line [17]. This is illustrated in Figure 2.3 where 2.3 a) shows the effect on the line edge when a negative resist is used and 2.3 b) shows the effect when a positive resist is used.

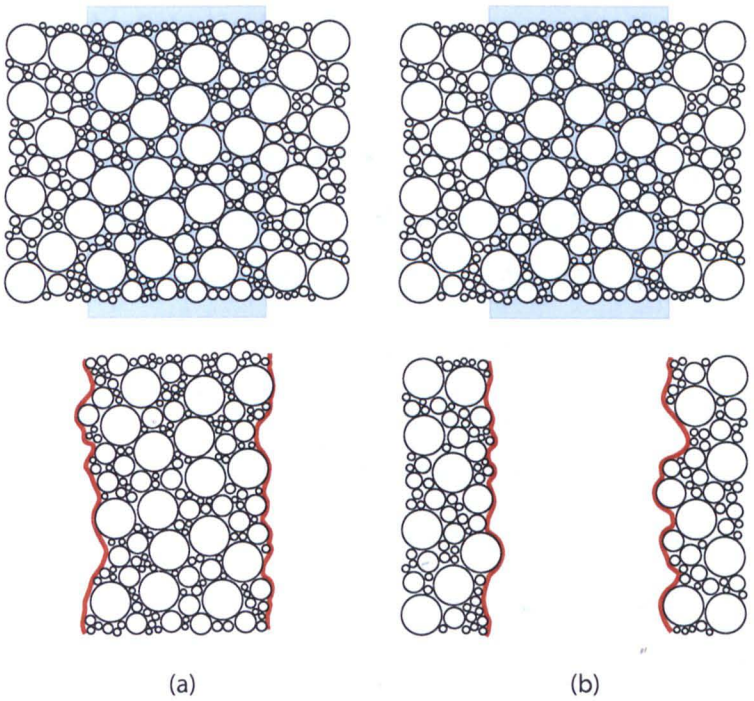

Figure 2.3: The effect of polymer aggregates in the laying down of a line pattern. a) shows the effect of a negative resist removing the region that is unexposed while b) shows the effect of a positive resist where the exposed region is removed.

Line edge roughness can be characterised with its rms amplitude, Δ, and a

correlation length, Λ assuming different types of auto-correlation functions. The typical magnitude of the rms variations is on the order of 2-3Δ with current values for Δ in the range of 2 nm [57], these figures come from data collected from ASET[58], IBM[59], IMEC[60], SANDIA[34], SONY[61] and the ITRS Roadmap[10]. Less is know about the correlation length, with figures stated in the roadmap from 10nm to 50nm. Work carried out at Glasgow has shown that the typical values of the correlation length are in the range of 30nm [57].

Line edge roughness cause variations in the effective length of the active channel along the width of the device. This variation results in areas with shorter channel length which will cause preferential conducting paths that allow the device to turn on earlier, or regions which are wider than the predicted channel length that require higher gate potentials locally to produce the nominal drain current. The ITRS roadmap [10] shows that in 2006 the predicted drawn gate length will be 40nm with a physical gate length of 32nm. If the worst case scenario of a local variation from the mean line of 3Δ or 6nm on both edges of the gate then the physical channel of this device could vary in size from 44nm to 20nm. At the extreme end of the roadmap in 2016 the devices are expected to have a gate length of 13nm with a physical channel length of 9nm. If we again apply the worst possible case of today's line edge roughness to this device then the channel length would vary from 21nm to 0nm, in this case the effects of line edge roughness would act to remove the channel entirely leaving part of the MOSFET channel short-circuited. It is possible that different types of resist with smaller molecular weights and smaller aggregate sizes will become available reducing the line edge roughness, but such resists presently do not exhibit

sufficient contrast to be commercially viable.

Attempts have been made to analytically model the effects of the line edge roughness upon the leakage current [62]. They however rely heavily on fitting parameters and do not fully capture the complex 3D geometries involved. Due to the significant computational effort of a full 3D simulation, a simplified statistical method [59, 63, 64] based on 2D simulations of MOSFETs with statistical variations in the channel length but with fixed channel width has been used historically to study the effects of LER. The use of 3D numerical simulations in studying line edge roughness began with a simple "square wave" approximation for the gate edge [65, 66]. These studies found a more than 30% discrepancy in 3σ estimates of the off-current predicted by the statistical 2D simulations. A more realistic statistical approach to 3D modelling of the impact of the LER of the gate edge on the intrinsic parameter fluctuations was developed in [57, 67]. The LER is characterised by an rms amplitude, Δ, and a correlation length, Λ and statistical samples of the gate edges in the simulation are generated using a 1D Fourier synthesis approach assuming either Gaussian or exponential correlation functions.

The effects of line edge roughness on the MOSFET characteristics are outlined in [57], showing that at the present status of lithography the leakage current in MOSFETs is highly sensitive to line edge roughness if the gate length is less than 50 nm. Devices with a larger W_{eff} exhibit reduced LER induced variations because of self averaging. As L_{eff} decreases the percentage deviation increases rapidly becoming greater than 100%. It was also shown that, similarly to random discrete dopants, the line edge roughness cause not only fluctuations

in the threshold voltage but also threshold voltage lowering. With a value of Δ = 2 nm the threshold voltage lowering in a 30 nm gate length transistor was approximately 4 mV while for Δ = 3 nm it increases to approximately 9 mV.

2.1.3 Oxide Thickness Fluctuations

The oxide growth upon a silicon lattice involves a chemical reaction between the silicon lattice and oxygen in the form of O_2 gas or as water vapour, to form a SiO_2 layer. The precise control of the oxide thickness requires both the temperature and the pressure to be strictly controlled. The growth of ultrathin oxides must be carried out in a vacuum to avoid the growth of a thicker native oxide. In mass produced MOSFETs the gate oxide thickness has already reached 1.5 nm [30, 68] with research devices having sub 1 nm thicknesses [69]. Atomic scale roughness of the Si/SiO_2 introduces significant intrinsic parameter fluctuations in contemporary MOSFETs. When the oxide thickness is of the scale of a few atomic layers the interface roughness steps on the scale of 1-2 atomic layers [70] cause each device to be microscopically different in terms of interface roughness and oxide thickness patterns. This in turn results in variation in transport and mobility [71], gate tunneling current [72] and real [73] or apparent [74] threshold voltage.

Figure 2.4 a) illustrates schematically the atomic nature of the MOSFET interface. The regular structure of both the polysilicon and silicon interfaces can be seen as well as the amorphous nature of the oxide. Figure 2.4 b) shows how this interface can be translated in simulations assuming only one atomic step at both interfaces.

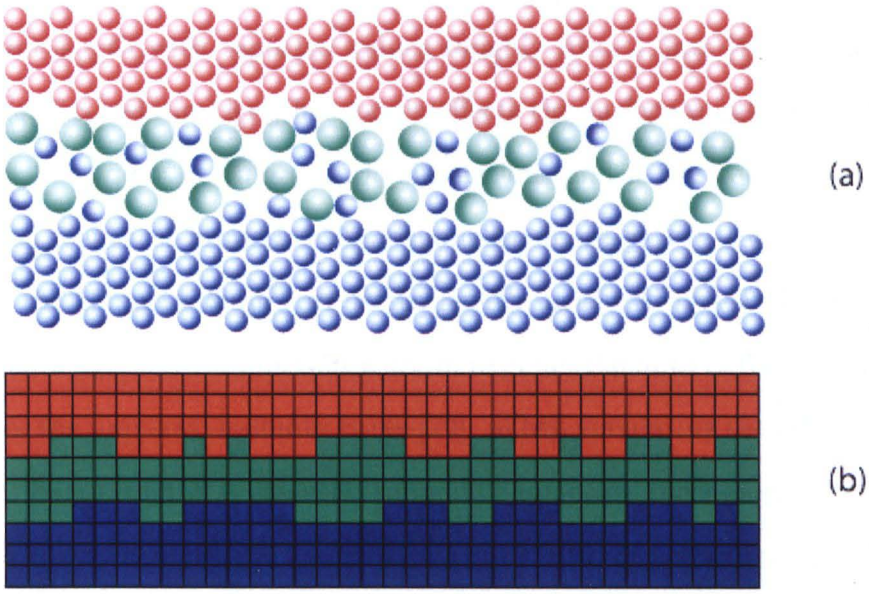

Figure 2.4: The interface between Si, SiO_2 and polysilicon in a MOSFET device. a) shows an atomic abstraction of the lattice structure of the polysilicon and silicon wheres a the molecular nature of the oxide is also displayed. b) shows the interface in an atomic layer abstraction showing the rough interface as the oxide interacts with the silicon.

Similarly to line edge roughness, the interface roughness pattern can be modelled as a random surface characterised by an rms amplitude, Δ, and a correlation length, Λ. Values for the rms amplitude reported in literature vary from 0.21nm to 0.48nm [14, 75, 76, 77] which is on the order of one atomic layer (or 0.27 nm) which is the value that will be used in the simulations in this work. There is still a discussion in the literature as to value is to be used for the correlation length with reported values roughly from 1.3nm [14] to 26nm [75]. The value of correlation length strongly effects the surface roughness dominated transport behaviour in Monte Carlo simulations and it was found that a short correlation length of around 1.55nm modelled more closely the mobility compared to the 18nm to 25nm correlation length measured by AFM [78]. Shorter correlation lengths in conjunction with an exponential auto-correlation function reproduce the universal mobility curve [79]. When simulating the effects of oxide thickness fluctuation it is essential to include quantum mechanical confinement effects [80], which move the inversion layer away from the rough interface and smooth the spatial inversion charge variations [81] when compared with classical simulations.

Fluctuations in the oxide thickness already play a role in the behaviour of the present devices. The ITRS Roadmap [10] predicts EOT of 0.6 nm or below for the devices that will be in mass production after 2006. The oxide thickness fluctuations in such devices enhance the gate leakage and can result in gate leakage fluctuations. As MOSFETs are scaled towards the end of the roadmap the correlation length of the interface will become comparable to the channel length and will cause significant fluctuations in the threshold voltage between

devices [82]. In [82] the rough interface is modelled as a random surface with a Gaussian power spectrum and then digitized to allow for a step at the interface of one atomic monolayer. This digitized random surface is then used to model the effects of oxide thickness fluctuations.

The reduction in the oxide thickness to two monolayers in the next generation devices result in unacceptably high gate leakage and calls for the replacement of the SiO_2 or $SiON$ with high-κ dielectrics such as Al_2O_3, HfO_2 and $ZrSiO_4$. These materials have much higher permittivites than SiO_2 which allows the dielectric to have the same SiO_2 EOT for larger physical thickness. However the growth of high-κ materials continues to be problematic. A thin layer of sub oxide commonly present between the silicon and high-κ material which effects quality and smoothness of the interface in addition to the associated increase in the equivalent oxide thickness also causes problems.

2.1.4 Other Possible Sources of Intrinsic Parameter Fluctuations

Moving to high-κ materials introduces new potential sources of intrinsic parameter fluctuations. The interface even in the presence of the sub oxide is rough. In addition separation between the high-κ materials typically are not perfectly homogenous, amorphous and crystalline phases and poly-crystalline interfaces are often observed [83]. This structural non-uniformity can change the local electrical characteristics of the gate oxide causing fluctuations in the dielectric properties and the local gate capacitance. Figure 2.5 illustrates the potential sources of "intrinsic" parameter fluctuations associated with the high-

κ gate stack. Similar to the previous sources of intrinsic parameter fluctuations the random structure of the high-κ material introduces variations in the microscopic structure and the corresponding variations in their electrical parameters. The high-κ related fluctuations are becoming an important area of research with the industrial consensus that high-κ material has to be introduced at the 45 nm technology node.

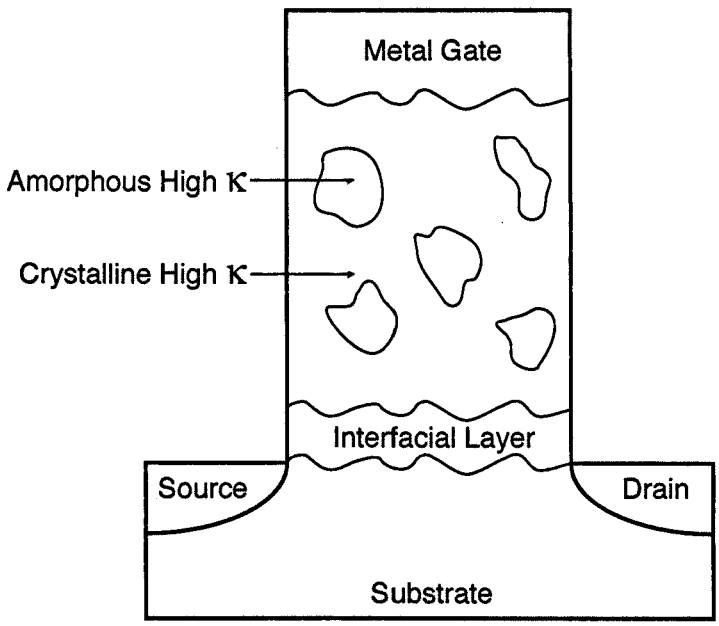

Figure 2.5: The gate stack of a MOSFET using high-κ dielectrics, showing the difference between amorphous and crystalline high-κ material which will alter the dielectric of the material on a local level [84].

Many of the processing steps in the current CMOS technology introduce a considerable strain in the silicon material, affecting its local and macroscopic electrical properties. Process induced strain is currently used as a source of

mobility enhancement in the MOSFET channel improving the device current. However the inclusion of strain induced by manufacturing can lead to problems, becoming another source of MOSFET fluctuation which must be studied and accounted for in the simulations. An example is the strain associated with the polysilicon gate. The line edge roughness of the gate results in non-uniform strain in the channel and random variations in the channel transport properties, which result in MOSFET current variations.

The polysilicon gate is another source of intrinsic MOSFET parameter fluctuation. Similarly to the source and drain regions, the poly gate is heavily doped with random positions of the dopant atoms. Variations in the local depletion at the poly-oxide interface effects the electrical characteristics of the transistor. Another problem associated with the polysilicon gate, particularly in p-channel MOSFETs is the boron penetration though the oxide into the channel region of the device. This happens predominately at the grain boundaries where the diffusion is enhanced and results in random variations of the channel doping and the electrical behaviour of the transistor.

An important additional source of fluctuations are the traps and defects within the MOSFET. In current devices single traps do not result in significant changes in the device current [85]. However with the scaling in the next generation devices the relative effect of these traps will increase. A single trap within the gate oxide or the upper regions on the channel will have a profound effect on the behaviour of a device altering the current and introducing fluctuations that has to be taken into account in the simulations.

2.2 Circuits

The combined effects of all the intrinsic parameter fluctuations noted in Section 2.1 will have significant impact on the functionality, yield and reliability of corresponding circuits and systems. As the transistor count increases and the supply voltage scales according to the Roadmap, fluctuation noise margins will shrink, impacting the reliability of circuits and systems.

The effects of the random dopant induced intrinsic parameter fluctuations on CMOS has been studied initially using an analytical model [86]. This quasi 3D approach considers the channel of a MOSFET as a series of cubes which are treated as simple MOS capacitors. The overall characteristics of the device are deduced by considering the statistics of an ensemble of the discrete elements with different doping concentrations, and tracing possible percolation paths through the ensemble. The size of the cubes is chosen such that each volume l^3 contains one impurity. This then allows the study of the effect random discrete dopants have on the characteristics of a device. The method has been used to study the effect that random dopant induced MOSFET parameter variations have upon SRAM cells [87]. A model for the SRAM static noise margin was developed and the effect of the random dopant induced threshold voltage fluctuations in the cell analysed. This provided quantitative predictions of the effect that fluctuations have on the distribution of the static noise margin for 6T SRAM cells, shown in Figure 2.6. Devices with gate lengths from 250 nm down to 50 nm have been considered in the above study. The results showed that the 6σ deviations in the static noise margin would exceed the allowable static noise margin in sub 100 nm devices, with the 70 nm device having a SNM of 82mV and a σSNM of

14 mV, and the 50 nm device having a SNM of 60 mV and a σSNM of 16 mV. This measure of 6σ is equivalent to 1 device failing in every 10^9 devices. The results show that problems will arise in the manufacture of 6T SRAM cells as the size of MOSFET devices are aggressively scaled and SRAM cell counts increase.

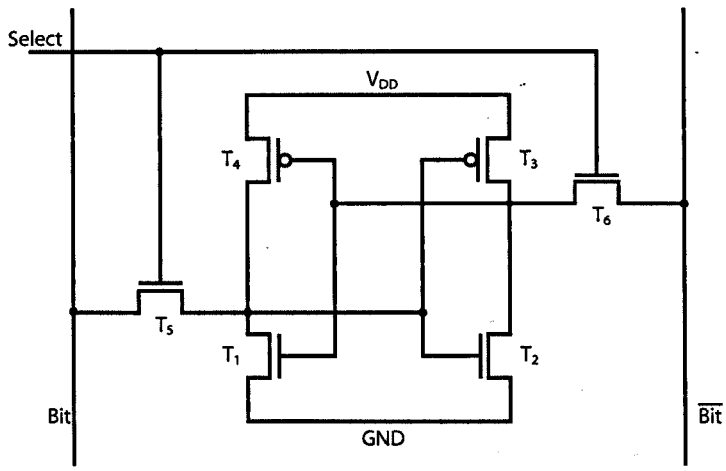

Figure 2.6: A 6 Transistor SRAM cell, showing the input word select line as well as the Bit read lines.

The effects of extrinsic and intrinsic parameter fluctuations on the power dissipation and critical path delay in CMOS devices has been studied in [88]. The critical path delay is modeled as a number of identical two input static CMOS NAND gates, which have a fan-out of three. From this model both the critical path delay and the overall power dissipation of the system is calculated.

It has been shown for a feature size of 50 nm corresponding to a physical gate length of 40 nm and 5 gates in the critical path, a number obtained from the historical trends for scaling, that the extrinisic parameter fluctuations cause

a 12%-29% increase in critical path delay and a 22%-46% increase in power dissipation. The extrinsic parameter fluctuations are generated from uncorrelated Gaussian distributions. There is, however, little experimental evidence for this and in general it is unknown if the different sources of fluctuations are uncorrelated or not. The inclusion of intrinsic parameter fluctuations due to random discrete dopants results in an increase in critical path delay variations to 18%-32% and in the power dissipation by 31%-53%. These figures show that parameter fluctuations in MOSFET devices will indeed have a significant effect at the circuit and system level.

The quasi 3D model used in [88] has been compared to full scale 3D Drift Diffusion simulations for 100 × 100 nm, 70 × 70 nm and 35 × 35 nm devices [89]. This comparison shows that the quasi-3D approach is useful in predicting variations in larger devices but in the case of the 35nm device the threshold voltage variations are significantly exaggerated and a rise in the ensemble average threshold voltage is observed, compared with the full 3D simulations. A different approach for studying the impact of intrinsic parameter fluctuations on circuits and sytems using the results from full 3D simulation is outlined in [90] and in [91]. In this approach the I_D/V_G characteristics resulting from full 3D MOSFET simulation for an ensemble of 200 devices are used to create an ensemble of 200 compact models of the simulated transistor. The compact model ensemble contains the statistical fluctuation information from the set of full 3D simulation. Randomly chosen devices from the compact model ensemble can then be included in circuit simulation to study the effect that intrinsic parameter fluctuations have on the circuits behaviour. A study of a 6T SRAM cell built

from 35nm transistors using the above approach is reported in [92]. It was found that for a cell ratio of 1, $SNM = 41\text{mV}$ with a standard deviation $\sigma SNM = 19.5\text{mV}$. This means that for square transistors commonly found in SRAM cells the 6σ would be twice as large as the SNM and therefore it would be impossible to achieve the required yield of working SRAM cells. In general when devices are aggressively scaled the effects associated with intrinsic parameter fluctuations upon the circuits becomes larger and a lot of care should be exercised in the design to avoid low yield.

2.3 Simulation of Intrinsic Fluctuations

There are different approaches which can be applied in the simulation of MOSFETs. There are however some basic requirements for a simulation engine that can be used to model intrinsic parameter fluctuations such as random dopants, line edge roughness and oxide thickness variations. To properly treat discrete random dopants it is essential to resolve the position of each individual dopant which have increasing impact on the device behaviour. In the modelling of the effects associated with line edge roughness and interface roughness the statistical description of the gate edge and the interface has to be introduced properly. All the above sources of intrinsic parameter fluctuations are three dimensional in nature and therefore in order to correctly capture their effects full scale 3D simulation should be carried out.

A second consideration is that it would be impossible to simulate the whole population of microscopically different devices. For instance in the case of ran-

dom dopants it is impossible to simulate all devices with microscopically different dopant distributions. The solution is to simulate a statistically large enough sample of devices, which allows us to extract with enough accuracy the parameters characterising the statistical distribution. For example a sample size of 200 devices provides a 5% accuracy for the standard deviations [18] but leads to a significant amount of computational time. Therefore our simulation technique used to study intrinsic parameter fluctuations should be fast and efficient allowing the simulation of a large ensemble of devices within a relatively short time. Statistical enhancement techniques based on the identification of rare events will be needed to access the tails of the statistical distributions.

Finally, some of the most important parameters for the designers of CMOS circuits are the threshold voltage, and the leakage current. It is assumed, at least in the case of random dopants, that the on-current will have less fluctuations due to the screening of the individual dopant by the inversion layer charge and therefore simulation of the on-current is less important. These three sets of requirements create a basis on which to judge the different simulation techniques that can be used to simulate intrinsic parameter fluctuations in modern semiconductor devices.

There is a well established hierarchy of simulation techniques that can be used in the simulation of modern semiconductor devices [93] illustrated in Figure 2.7, where the vertical axis represents the computational complexity and accuracy and the horizontal axis represents the computational time. At the bottom of the hierarchical ladder are the compact models, which contain very little information about the physics behind the system and in general mimic the behaviour of the

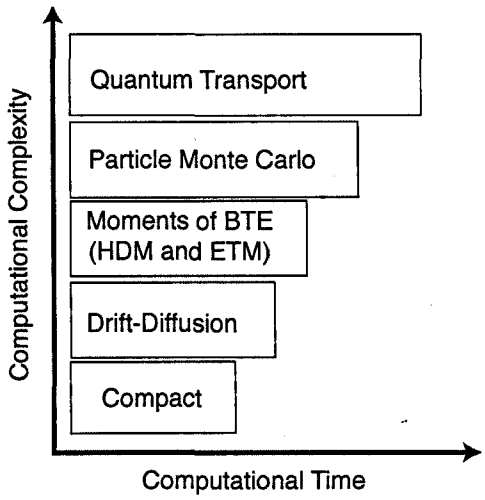

Figure 2.7: Graph of computation time versus computational complexity for various simulation techniques based on there position in the hierarchy of techniques.

device using analytical approximations and empirical fitting parameters. The next level of simulation techniques is the drift-diffusion approximation to the Boltzmann Transport Equation (BTE). The drift-diffusion approach takes into account only the first two moments of the BTE which are current continuity and the momentum conservation equations. These are coupled to the Poisson's equation for electrostatic potential. The drift diffusion approximation includes a local relationship between the velocity and the electric field and cannot represent properly non-equilibrium transport effects.

The next level, above the drift-diffusion approach, is to use the Hydrodynamic approximation to the BTE. In this case the third moment of the BTE , the energy conservation equation, is included and the momentum conservation equation becomes more complex. This allows the treatment of non-equilibrium

effects which includes a non-local relation between the electric field and the velocity responsible for velocity overshoot effects. The direct solution of the BTE takes us to the next level of simulation techniques. This can be done, for example, using spherical harmonic expansions for the distribution function in conjunction with numerical techniques or a scattering matrix approach. However the direct solution techniques are unstable and computationally expensive, and so prohibitive for practical simulation. An alternative Monte Carlo technique can be used to solve the BTE. In this technique an ensemble of particles evolves through real space acceleration (free flight) and randomly chosen scattering events. This approximates the solution of the BTE in a less computationally expensive manner.

The highest level is the realm of Quantum mechanical approaches which use such things as coupled Poisson and time-independent Schrödinger equation; density matrix, Wigner distribution function. These are all highly computationally expensive. Another technique which is gaining popularity is the use of Non-Equilibrium Green's Functions (NEGF) which allow the inclusion of scattering in the quantum transport formulation. The rest of this section will briefly outline the various techniques and their strengths and weaknesses in regards to their use for simulation of intrinsic parameter fluctuations.

2.3.1 Drift Diffusion

Drift Diffusion (DD) represents the simplest model used in multi-dimensional numerical simulations based upon the lower order moments of the BTE [27]. In this model the electron and hole current density are approximated using two

components [94], a drift component driven by the electric field and a diffusion component driven by the carrier density gradient. The sum of these components give the density current J which in the case of electrons and holes is given in equation (3.2) and (3.3).

$$J_n = qD_n\nabla n - q\mu_n n\nabla\psi \tag{2.1}$$

$$J_n = qD_p\nabla p + q\mu_p p\nabla\psi \tag{2.2}$$

In equation (3.2) and (3.3), μ is the mobility of electron or holes, D is the diffusion coefficient of the electrons or holes and ψ is the electrostatic potential. This approximation of the BTE ignores carrier heating and is strictly speaking only valid at fields where the carrier velocity is directly related to the electric field. [93]. However the validity of the drift diffusion approximation can be empirically extended by introducing field dependent mobility models which allows for higher field to be used in simulation. Even with this extension the DD system only works in quasi equilibrium where the electric field varies slowly and the velocity is locally related to the electric field. The DD approach is perfectly adequate to analyse fluctuations in V_T and the subthreshold slope which are predominately by the device electrostatics. A major benefit of the drift diffusion approach is low computational cost and therefore is suitable for carrying out large scale statistical simulations needed to characterise the impact of various sources of intrinsic parameter fluctuations. The DD equations have been in use for over 40 years and because of this there is a wealth of well developed numer-

ical techniques which can be used in 3D. The 3D simulations are an order of magnitude more expensive compared to 2D simulation routinely used in TCAD but are essential when studying intrinsic parameter fluctuations.

2.3.1.1 Quantum Corrections

Although full Quantum mechanical simulations are prohibitive in terms of computational time it is possible to include quantum mechanical effects into otherwise classical simulation using so called quantum corrections. The quantum corrections allow the quantum confinement effects and some aspects of tunnelling to be included in our simulation. The quantum effects play an increasingly more important role as devices are aggressively scaled into the nanometer regime. The two most well know methods for including quantum correction in classical device simulation are the Density Gradient approach and the Effective Potential approach.

The Density Gradient approach introduces a quantum potential which provides an additional driving force in the expression for the current density [95]. This quantum potential is proportional to the second derivative of the carrier density. The quantum potential drives the electrons away from steep variations in the classical potential. In the inversion layer it pushes the electrons away from the interface replicating the quantum distributions of the inversion layer charge. In "atomistic" simulations it prevents the artificial charge trapping in the Coulomb wells associated with individual discrete dopants.

In the Effective Potential technique [96] carriers are represented by a minimum dispersion Gaussian wavepacket. The effective potential is related to the self

consistent potential through an integral convolution. This potential smoothing associated with the convolution operation represents the quantum mechanical effects that shift the carrier concentration away from the interface and reduce the sharp peaks in potential coming directly from the solution of the Poisson equation.

The quantum corrections significantly improve the accuracy of the drift diffusion simulation of nanoscale MOSFETs when quantum confinement effects influence the threshold voltage and the overall device performance. They also become essential when resolving the influence of individual discrete dopants.

2.3.2 Higher Order Moments of the BTE

With the scaling of MOSFETs to smaller dimensions there is a growing disparity between the rapid reduction in the physical dimensions and the much slower reduction of the supply voltage resulting in an increase in the fields within MOSFETs. In DD we take into account the high field by introducing phenomenological field dependant mobility models which locally relate the carrier velocity to the magnitude of the field component in the direction of current flow. This approach ignores non-local transport phenomena where the carrier velocity at a particular point is determined by the field distribution along the proceeding current path. Another approach which overcomes these limitations is to use the higher order moments of the BTE. A particular example is the Hydrodynamic model, in which the current relationship includes an additional driving term which is proportional to the gradient of the electron temperature [97] and a relatively simple energy balance equation is added to the system. The main

problem with inclusion of higher moments of the BTE is how to close the system of equations. Every lower moment of the BTE is coupled to the higher moments and the closure involves simplifying assumptions . Another problem related to the inclusion of higher order moments is the numerical stability of the solution. This leads to extra computational effort and increased code complexity. These difficulties rule out the use of higher order moments of the BTE.

2.3.3 Monte Carlo

An alternative way of simulating carrier transport which does not involve discretisation of the the BTE or its moments is the Ensemble Monte Carlo approach. The Ensemble Monte Carlo method is a stochastic technique which uses random numbers to obtain a statistical approximation to the exact solution of the BTE [93]. The Monte Carlo method traces the classical trajectories of carriers in a simulated device scattering each particle after a free flight period determined stochastically by the cumulative scattering rates. The scattering is calculated quantum mechanically and include electron-phonon interaction, electron-donor interaction and electron-electron interaction. The Monte Carlo simulation is thus a series of free-flights interspersed with scattering events. This is illustrated in Figure 2.8 where the scattering at the end of each free flight changes stochastically the momentum and possibly even the energy of the particle. This particle movement is then coupled self consistently to the solution of Poisson's equation to allow the updated force driving the particle to be calculated. In order for these simulations to be effective, an ensemble of "superparticles" is usually used which statistically represents the actual carriers within

the system.

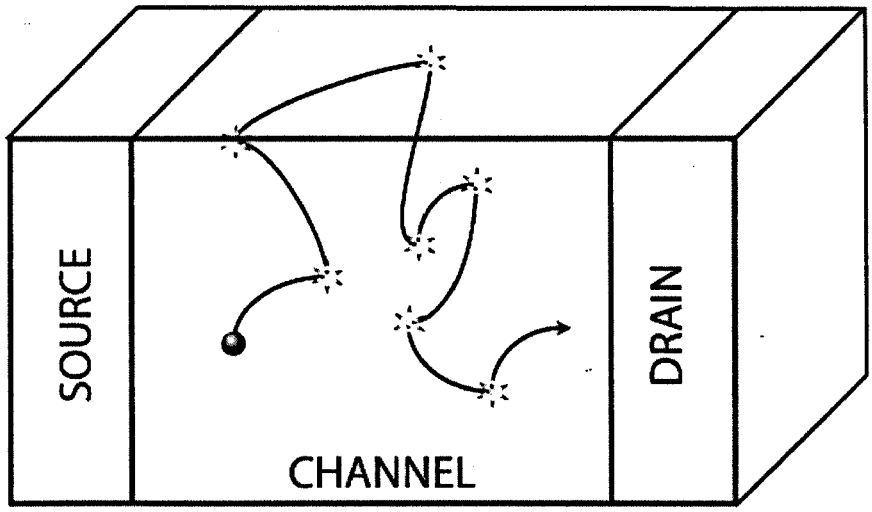

Figure 2.8: A particle moving between the source and drain region being scattered by various processes at random points.

The Ensemble Monte Carlo approach is widely used to simulate semiconductor devices in general and in particular for simulation of MOSFETs [98, 99, 100, 101, 102]. However this method does have some drawbacks. Because of the large number of particles that must be simulated, the large number of random numbers that is required and the self-consistent coupling with Poisson's equation the Monte Carlo method is very computationally expensive. Also associated with the limited number of "superparticles" is the problem of large statistical noise, particularly in the transient MOSFET simulation. This requires long simulation times to allow reliable statistical averaging. This method is the best available method at present to simulate high current flow in small MOSFETs at the extreme non-equilibrium conditions of their operation. It is also the only method which can capture variation in the device transport due to variations in the doping and interface configuration in small MOSFETs. However because simulation

times are so large it would be difficult to use this method on a statistical scale to study the effects of intrinsic parameter fluctuations due to the large number of device configurations that need to be studied.

2.3.4 Non-Equilibrium Green's Functions

At the top of the hierarchy of simulation techniques are the quantum transport techniques which come in different flavours. Attempts to model the transport within a solid using a full quantum approach are computationally expensive and impractical for realistic MOSFET simulations. A technique that has gained popularity in the field of quantum transport is the Non-Equilibrium Green's Functions (NEGF) approach. NEGF use a form of mathematical method known as Green's functions to obtain the solution to a time independent Hamiltonian. This Green's function, at a given energy, has two inputs which can be connected to two positions in real space allowing us to simulate areas of a MOSFET. The Green function takes into account the influence of a perturbation happening at a one input on the other input and has (in theory) the ability to model the physical properties of the system such as electron density, current density and the density of states.

One of the main drawbacks with the use of NEGF's is that it is exceedingly computationally expensive. To solve Green's functions require the ability to invert large matrices which requires both large memories and large computational time. For a simple 2D simulation of a 10 nm by 20 nm double gate transistor, the matrix would contain 20000×20000 elements, and for a 3D simulation this size of a device would be almost impossible to simulate. Because of these limi-

tations most of the work being carried out using Green's functions is 1D, with a number of 2D simulators now available, and 3D simulators in the development stage. This form of simulation technique does not lend itself to study the effects that intrinsic parameter fluctuations have upon the operation of standard MOSFETs.

In this chapter we have reviewed the various sources of intrinsic parameter fluctuations found in the literature. Three key sources of intrinsic parameter fluctuations which are critical for the operation of conventional MOSFETS, and are amenable to numerical simulation and analysis have been identified. These sources are individual random discrete dopants, line edge roughness of the gate edge and oxide thickness fluctuations. In order to carry out numerical simulation of intrinsic parameter fluctuations key factors have been identified. Full 3D simulations are mandatory due to the inherently 3D nature of the intrinsic parameter fluctuation sources. Computational efficiency is paramount because statistical samples containing 100-200 devices are to be simulated to allow statistical analysis of their characteristics with sufficient accuracy. The simulation technique should also correctly capture the physics in the area of interest which in in this case includes the subthreshold slope region of operation and threshold voltage. Of the techniques reviewed the only simulation technique that satisfies the above three requirements is the Drift Diffusion approach. The other simulation techniques like the hydrodynamic and ensemble Monte Carlo approaches that capture more accurately the transport in deep submicron MOSFETs are too computationally expensive. In the next chapter we will look at the Glasgow Drift Diffusion simulator in more detail and lay out the process where by the

study of intrinsic parameter fluctuations can be made.

Chapter 3

The Development of the Device Simulator

Simulation approaches with different levels of complexity, such as Drift Diffusion (DD) , Monte Carlo (MC) and Non-Equilibrium Green Functions (NEGF) can all be used to study the operation of nanoscale MOSFETs and the effects of various sources of intrinsic parameter fluctuations. Each of these simulation techniques has its own strengths and weaknesses when used to study intrinsic parameter fluctuations. DD for instance is the least computationally expensive however it cannot reliably capture on-current fluctuations reliably and is therefore limited to studying fluctuation effects below threshold where the characteristics of the device are mainly controlled by electrostatics. MC, which requires far more computational effort, can capture the impact that sources of intrinsic fluctuations like random dopants and interface roughness have on carrier transport and therefore on the on-current. However the correct introduction of quantum effects in MC

simulations is problematic. NEGF, which correctly captures quantum transport effects faces problems with the introduction of scattering and requires an almost prohibitive amount of computational effort.

The systematic study of the effects of intrinsic parameter fluctuation requires a statistical approach as every microscopic device configuration associated with one or more sources of intrinsic parameter fluctuation cannot be simulated. For instance there are thousands of dopant atoms within a 35nm MOSFET which can be located at millions of different lattice positions. To simulate every microscopic doping distribution would be practically impossible. We must therefore simulate a statistical sample and study the distribution of results in order to understand the effects that a particular fluctuation source have over the combination of different sources. The number of devices that must be simulated is therefore smaller than the complete space of microscopically different devices. For example in order to estimate the standard deviation to within a 5% level of confidence a sample of 200 devices has to be simulated. The time used to carry out this number of simulations using MC or NEGF is prohibitive and therefore the more efficient DD approach is used in this study. This chapter describes the development of the existing device simulator used at the University of Glasgow in order to improve the accuracy and realism with which different sources of intrinsic parameter fluctuations are simulated.

3.1 The "Atomistic" Device Simulator

The 3D DD "atomistic" MOSFET simulator used in the Device Modelling Group at the University of Glasgow was developed to study the effects of different sources of intrinsic parameter fluctuations upon the operation of MOSFETs. It has already been used for initial studies on the effect of random discrete dopants, interface roughness and line edge roughness on MOSFETs with simplified and idealised structures [18, 29, 103, 82, 104, 105, 53, 106].

However it has not originally developed to study intrinsic parameter fluctuations in devices with a realistic structure. The original simulator was, for example, unable to simulate devices with realistic doping profiles imported from commercially available process simulations. At the same time transistors with complex doping profiles including HALO channel doping, pockets and source/drain extensions became the norm in modern MOSFET design. The development of capabilities to simulate modern CMOS transistors with realistic structures and complex doping profiles is one of the main objectives of this work.

3.1.1 General Overview

The approach used to simulate 3D "atomistic" MOSFETs is outlined in [29]. The 3D DD "atomistic" MOSFET simulator uses a self consistent solution of Poisson's equation (Equation 3.1) and the steady state current continuity equations for electrons (Equation 3.2) and holes (Equation 3.3). In equation 3.1, ψ is the potential, ϵ is the permitivity, q is the charge of an electron, n and p are the electron and hole concentrations, N_A^- is the acceptor concentration and N_D^+

is the donor concentration.

$$\nabla(\epsilon\nabla\psi) = q(n - p + N_A^- - N_D^+) \tag{3.1}$$

In equation 3.2, D_n is the diffusion coefficient for electrons and μ_n is the mobility of electrons, and in equation 3.3, D_p is the diffusion coefficient for holes and μ_p is the mobility of holes, with the other symbols having the same meaning as in equation 3.1.

$$J_n = qD_n\nabla n - q\mu_n n\nabla\psi \tag{3.2}$$

$$J_n = qD_p\nabla p + q\mu_p p\nabla\psi \tag{3.3}$$

The drift diffusion approximation does not properly capture non-equilibrium carrier transport and therefore underestimates the on-state drain current, however it is perfectly adequate for accurate calculations of the threshold voltage based on current criterion. The "atomistic" simulator uses a uniform mesh with a step size of h (which is usually 1 nm or 0.5 nm in our simulations), which is used in the discretisation of the Poisson and the current continuity equations. Both the non-linear Poisson equation and the current continuity equations are discretised using finite difference methods with the current continuity equations using the Sharfetter-Gummel discretisation scheme [27]. This coupled set of equations is solved by Gummel iterations [107]. The solution of the non-linear Poisson equation is obtained by using a one step Newton-SOR scheme with a

black/red ordering which allows the solver to be easily executed in parallel. The current continuity equations are solved using a BiCGSTAB solver which is required due to the inclusion of complicated mobility models and the more complex Sharfetter-Gummel discretisation scheme, which means that the corresponding non-symmetrical and non-diagonally dominant system of linear equations becomes difficult to solve. These numerical solvers are used self consistently allow the electrostatic properties of a MOSFET device to be studied and, with the inclusion of sources of parameter fluctuation, allow the study of the effects of intrinsic parameter fluctuations.

3.1.2 Quantum Corrections

In addition to the simple classical DD equations discussed in the previous section, the device simulator also has the ability to introduce Quantum Corrections. With the reduction of device dimensions it is likely that quantum mechanical effects will have an increasingly important influence on the device operation[80]. In order to correctly simulate the characteristics of modern MOSFET devices it proves to be essential to include appropriate quantum corrections. The major quantum effects are associated with quantum confinement in the inversion layer. Quantum confinement shifts the threshold voltage, as it reshapes the inversion charge distribution moving the maximum away from the interface and therefore increases the equivalent oxide thickness.

The Density Gradient approach introduces an additional term into the classical equation of state for electrons which alters it based on the gradient of the concentration of the carriers as shown in equation (3.4). Here n is the electron

concentration and $b_n = \frac{\hbar^2}{12qm_n^*}$, where $\hbar$ is Plancks constant divided by 2π, q is the charge of a single electron and m_n^* is the effective mass of an electron. ϕ_n is the quasi Fermi level for electrons, ψ represents the potential, k_b represents Boltzmann's constant, T is the temperature and n_i is the electron concentration of intrinsic silicon.

$$2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = \phi_n - \psi + \frac{k_b T}{q} ln\left(\frac{n}{n_i}\right) \tag{3.4}$$

Equation (3.4) can then be used to modify the current continuity equation (3.5) which also includes a "quantum" driving force dependant on the carrier concentration. Here D_n is the diffusion constant of electrons and μ_n is the mobility of electrons with all other symbols meaning the same as in (3.4).

$$J_n = qD_n \nabla n - q\mu_n n \nabla \psi + 2nb_n \mu_n \nabla \left(\frac{\nabla^2 \sqrt{n}}{\sqrt{n}}\right) \tag{3.5}$$

The additional "quantum" driving force pushes the carriers away from the interface in order to reproduce the quantum charge distribution in the inversion layer. In DD simulations which involves discrete random dopants the same force pushes carriers away from the sharply resolved potential wells associated with the individual dopants and avoids the artificial charge trapping.

The Density Gradient equation (3.4) is added to the system of drift diffusion equations. Similar in form to the Poisson equation it is first discretised into a non-linear system of algebraic equations and solved using the same iterative techniques. Simulation experiments show that the inclusion of the Density Gradient equations directly into the Gummel cycle make the convergence of the

Gummel iteration problematic. Therefore a modified Gummel procedure was developed, illustrated in Figure 3.1. The procedure involves the solution of the Density Gradient equation self consistently with Poisson's equation. This pair of equations is then solved self consistently with the current continuity equations. This method increased the overall stability of the Gummel method in the presence of DG corrections and improved the convergence of the simulation greatly.

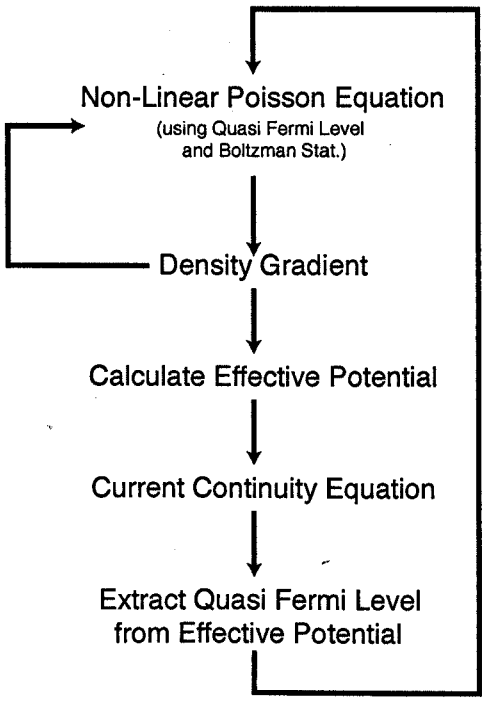

Figure 3.1: Flow diagram showing the process by which Density Gradient is applied to the simulation. The Density Gradient equations are solved self consistently with Poission's equation, which pair of equations are solved self consistently with the current continuity equations.

3.2 Porting of Code

The original Glasgow "atomistic" device simulation code was developed in Fortran66 and then in Fortran77, and was running under Sun Microsystems Solaris operating system. At the beginning of this research a new 32 cpu IBM p640 system running the AIX operating system was purchased offering more computational power. The first task associated with the enhancement of the existing code was to port the "atomistic" device simulator to this new platform. The new IBM system uses the XL suite of compilers which contained some syntactical differences compared to the Sun compilers. Many of the syntactical changes were minor however the most serious problem was caused by a clash in the methods used to allow the numerical solvers to run in parallel. The IBM XL compilers use a different form of the OpenMP Parallel Do instruction which was incompatible with the Sun version.This prompted the re-writing of the parallelisable DO loop instructions, the accomplishment of this produced a version of the simulator which can take advantage from the parallelism and improved speed available on the new IBM system.

As part of the porting process it was decided also to upgrade the code from Fortran77 to the Fortran90/95 standard which would allow both the use of freeform coding but also the use of intrinsic array operators. The expectation was that this would improve the efficiency of the code. However it was found that the migration to Fortran90/95 introduces unresolvable convergence problems particularly when the Density Gradient quantum corrections are turned on. A careful analysis of the execution pattern of the code produced by the available, at the time, version of the Fortran90/95 compiler showed it had issues with

memory leakage, memory protection and optimisation. It was found that in the migration from Fortran77 to Fortran90/95 the common blocks used for large array storage had been altered to utilise Fortran90/95 modules. Although the intention was to create a more streamlined code that better interacts with the memory, corruption of stored data became apparent. Also, the use of intrinsic array operations intended to improve the efficiency of the code, actually caused the opposite effect. It was found that in the assembly code which was generated array access was happening in the wrong order. In most modern operating systems the memory is laid out in flat arrays, which is accessed in row major order. However the assembly code generated by the Fortran90/95 compiler was accessing this memory in column major order as commonly used in Fortran. However this in turn prompted out of order access of the memory resulting in cache misses and slowing the code when intentionally accessing the memory. These issues, in particular the memory corruption which seemed to cause the lack of convergence, were resolved by returning back to Fortran77. This sorted out many of the issues that had been present in the Fortran90/95 version of the code.

3.3 Implementation of Intrinsic Fluctuation Sources

Three sources of intrinsic parameter fluctuations are considered in the "atomistic" device simulator, Line Edge Roughness (LER), Oxide Thickness Fluctuations (OTF) and Random Discrete Dopants (RD). LER is introduced into real fabricated devices though the lithographic and etching processes in part due to

the molecular structure of the photoresist. OTF is associated with the atomic structure of the Si/SiO_2 interface. The growth of amorphous SiO_2 on the crystalline Si substrate results in atomic scale steps, possibly present to accommodate strain mismatch. RD fluctuations are associated with the discrete nature of the dopant atoms. In the fabrication of MOSFETs impurities are introduced by ion implantation. The paths of the implanted impurity atoms are random and therefore they come to rest in random positions corresponding variations in the microscopic structure of the individual devices causes variations in their operation.

In order to study the effects of the above sources of intrinsic parameter fluctuations upon the device characteristics the individual sources must be properly modelled and introduced into the simulator. This section discusses the models used to introduce LER and OTF in the existing code and also the improvements that have been made to the method for introducing RD.

3.3.1 Line Edge Roughness

Line Edge roughness is introduced into the "atomistic" Drift Diffusion simulator as outlined in [57, 67]. The method used to generate random junction patterns is based on a 1-D Fourier synthesis which generates gate edges from a power spectrum corresponding to a Gaussian or exponential autocorrelation function. The parameters used to describe this gate edge are the correlation length, Λ , and the rms amplitude, Δ. The rms amplitude is the standard deviation of the x-coordinate of the gate edge if we assume that the gate edge is, on average, parallel to the y-direction. In most cases the value quoted as LER is 3 times

the rms amplitude or 3Δ. The algorithm for generating a random line creates a complex array of N elements whose element amplitudes are determined by the power spectrum obtained from either a Gaussian or exponential autocorrelation function. S_G in Equation 3.6 is the power spectrum for a Gaussian autocorrelation function where $k = i(2\pi/Ndx)$, dx is the discrete spacing used for the line and $0 \leq i \leq N/2$. Equation 3.7 shows a similar power spectrum S_E, this time for an exponential autocorrelation function.

$$S_G(k) = \sqrt{\pi}\Delta^2\Lambda e^{-\frac{k^2\Lambda^2}{4}} \tag{3.6}$$

$$S_E(k) = \frac{2\Delta^2\Lambda}{1 + k^2\Lambda^2} \tag{3.7}$$

The phases of each of the elements is selected at random which makes each line unique, however only $(N/2) - 2$ elements are independent while the rest are selected through symmetry operations so that after an inverse Fourier transform the resulting height function, $H(x)$ will be real. An example of these generated random lines is shown in Figure 3.2 which shows two random lines, one generated with a Gaussian auto correlation function and one with an exponential autocorrelation function. In both cases typical values of Δ and Λ are used. The line generated from the Gaussian autocorrelation function is smoother due to lack of high frequency components which are characteristic of the corresponding exponential power spectrum.

An example of applied line edge roughness is shown in Figure 3.3 and Figure 5.16 taken from [67]. Figure 3.3 shows the gate edge doping profile and the p-n

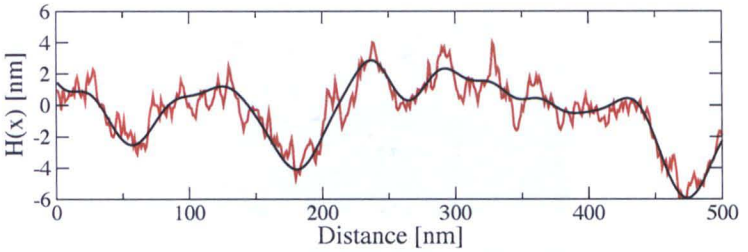

Figure 3.2: Two random lines generated from method outlined in [57, 67]. One random line is generated using a Gaussian power spectrum while the second random line is generated from a exponential power spectrum. The exponential random line can be seen to contain higher frequency components than the Gaussian random line

junction of a 30×200 nm MOSFET. In our simulated example the shape of the surface p-n junction replicates the gate edge doping profile and follows through the depth of the device a Gaussian doping profile. The lateral p-n junction edge is not smeared in our simulation due to the effects of 3-D ion implantation or thermal processing. Such an approach is not as simple as it appears as the typical LER correlation length of a device scaled below 50nm will be larger than the junction depth of such a device.

Figure 5.16 shows the potential profile of a 30×200 nm MOSFET with applied line edge roughness. This MOSFET has a gate voltage, V_G equal to the threshold voltage, V_T at an applied drain voltage, V_D equal to 0.01V. The potential in this MOSFET approximately follows the metallurgical p-n junction

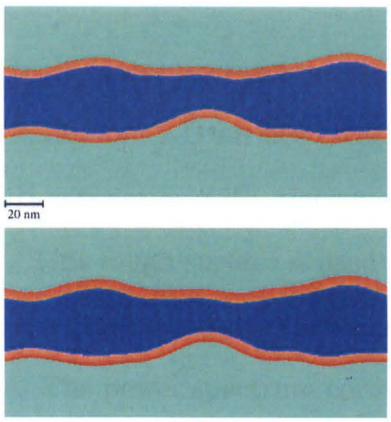

Figure 3.3: The simulated p-n junction and gate edge in a 30×200 nm MOSFET showing that the p-n junction exactly follows the gate edge produced, taken from Asenov *et al.* [67]

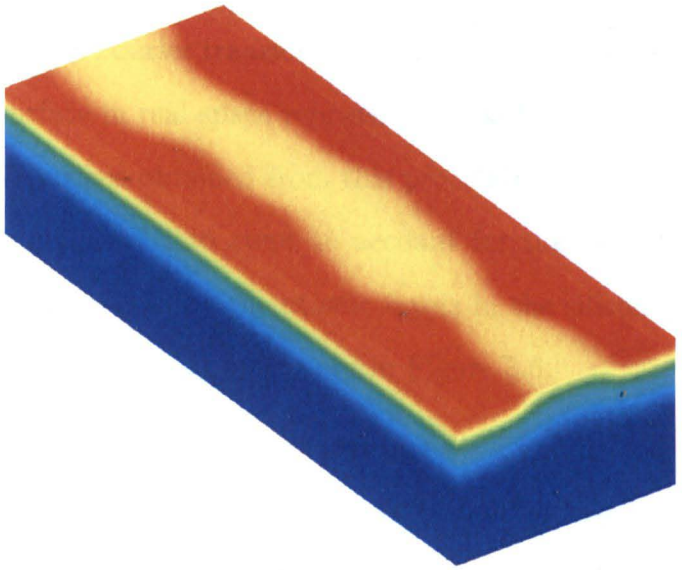

Figure 3.4: The potential profile in a 30×200 nm MOSFET with applied line edge roughness with a value of $V_G = V_T$ and $V_D - 0.01V$, taken from Asenov *et al.* [67]

3.3.2 Oxide Thickness Fluctuations

The method by which oxide thickness fluctuations are introduced into our simulation is outlined in [106]. A random 2-D surface is constructed using a Gaussian or exponential autocorrelation function with a given correlation length, Λ , and the rms amplitude, Δ [14]. This rough surface is used to represent the boundary between the dielectric and the gate material or between the silicon and the dielectric within our simulations. The power spectrum corresponding to the Gaussian or exponential autocorrelation function can be obtained by a 2-D Fourier transformation. To generate a rough interface an $N \times N$ complex matrix is formed in the Fourier Domain, the magnitudes of its elements are found through the appropriate power spectrum. The phases of these elements are chosen at random to ensure a unique surface. Certain conditions must be met [70] to ensure that when an inverse Fourier transformation is applied to this matrix the corresponding 2-D surface in real space represents a real function. Figure 3.5 shows the random 2-D surface obtained from this procedure. This generated random surface is then quantized into steps to account for the discrete nature of the interface roughness created by the atomic layers in the silicon lattice substrate [108]. This digitised surface is illustrated in Figure 3.6.

An example of these oxide thickness variations is found in [106] and is outlined below. Figure 3.7 depicts the typical Si/SiO_2 interface used in the simulation of a $30 \times 30nm^2$ MOSFET generated by the method outlined above. The interface was reconstructed using a Gaussian autocorrelation function and roughness was only applied to the Si/SiO_2 boundary and not the SiO_2 polysilicon gate boundary. The potential distribution at threshold voltage is shown in the bot-

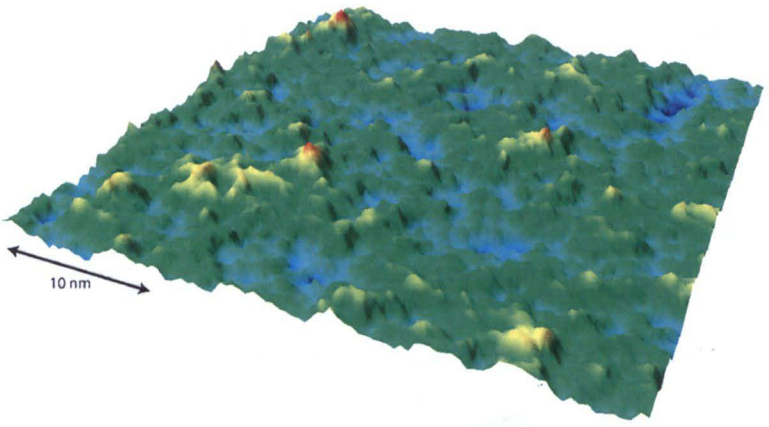

Figure 3.5: 2-D rough surface generated to act as a template for interface roughness. This 2-D surface was generated in a similar fashion to the random line shown in Figure 3.2 using the inverse Fourier transform of a complex vector created from a Gaussian autocorrelation function. The plot shows a 50×50 nm section of the surface with a height of $\pm 0.3nm$.

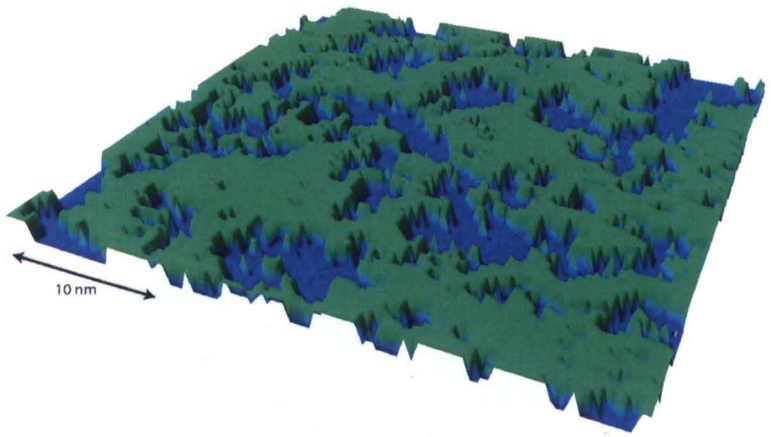

Figure 3.6: 2-D digitised random surface created from the Gaussian random surface shown in Figure 3.5. This surface can then be used to create a random interface by altering the material profile at the interface.The plot shows a 50×50 nm section of the surface digitised on two levels, the blue representing 0 and the green representing 1.

tom image of Figure 3.7. The oxide thickness fluctuations introduce potential fluctuations at the surface which are similar in effects to the inclusion of discrete random dopants. In this particular simulation quantum corrections were included using the Density Gradient method and an equiconcentration surface corresponding to the electron charge density of $1 \times 10^{17} cm^{-3}$ is shown in the middle image of Figure 3.7. This equiconcentration level illustrates the effects of quantum confinement in both the vertical and lateral directions.

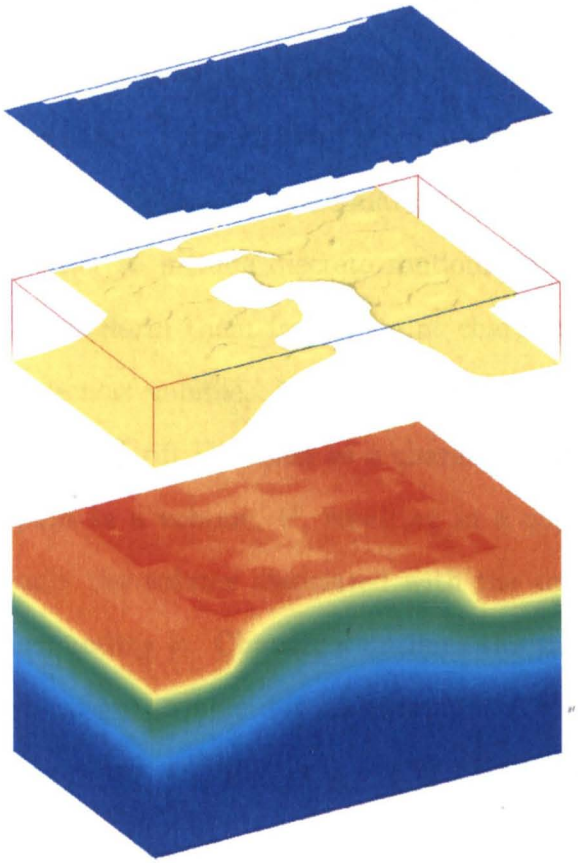

Figure 3.7: Typical Si/SiO_2 interface used in the simulation of a $30 \times 30nm^2$ MOSFET, the top image represents the rough surface, the middle image the equiconcentration contour obtained from DG simulations and the bottom image the potential profile, taken from Asenov *et al.* [106]

3.3.3 Random Discrete Dopants

There are two important aspects associated with the inclusion of discrete random dopants in a Drift Diffusion simulatior. The first is how discrete dopants could be represented bearing in mind the continuum nature of the Drift Diffusion approach. The second is how the random position of the dopants will be chosen in any individual device that is to be simulated.

In the Poisson equation associated with the Drift Diffusion approach the charge in the system is presented as a charge density. The current is also represented by the flow of continuous charge through the device associated with an ensemble of carriers and not as the motion of individual carriers such as in the Monte Carlo approach. In order to include discrete random dopants in the Poisson equation we have to transform them from a point charge to a charge density associated with a particular volume. Normally a discrete dopant is modelled in the continuum world of Drift Diffusion as a charge density produced by the spreading of the dopant in a volume, V, resulting in a doping density in this volume of $\frac{1}{V}$. Normally the volume associated with the single point charge is the mesh spacing surrounding it. Chapter 4 outlines the various methods that we have employed to assign a single dopant charge to the mesh using various charge assignment techniques, and discusses the relative merits of each.

The second problem that we must address is how to populate our device with discrete random dopants. The original method used in the simulator employs a simple rejection technique. In this approach, for each node of the discretisation mesh representing a device, a decision is made whether to introduce a dopant based on the total number of dopants in the device and the continuous doping

distribution at this node [29]. This was accomplished by generating a random number and an probability associated with each mesh point based on equation (3.8), where the probability is created from the volume surrounding the mesh node and the doping density. In equation (3.8) ρ is the probability of a charge being assigned, dx, dy, dz are the x, y, z components of the mesh cell and N_d is the doping density associated with the mesh node.

$$\rho = (dx * dy * dz)N_d \tag{3.8}$$

If the random number was greater than the generated probability then a dopant would be placed in the mesh node, otherwise only a very low doping, $N_D = 1 \times 10^{15} cm^{-3}$,would be introduced to avoid numerical problems. This method of introducing discrete random dopants artificially couples their position to the mesh nodes and for a large mesh spacing introduces artificial correlation in the discrete dopant distribution.

A better approach to generating random discrete dopant positions has been outlined in [109]. It allocates the dopants to sites of the Si crystal lattice covering the simulated device. This approach has become feasible due to the reduced sizes of modern MOSFETs, which are below 100nm. It becomes possible and practical from a computational point of view to cover the whole device with the actual crystalline lattice and to randomly populate the sites of the lattice with dopants.

In order to construct the silicon lattice we first create a basis of 8 silicon atoms, depicted in Figure 3.8 where the blue spheres are part of the basis and the red spheres are the remaining silicon lattice. The positions of these atoms relative to

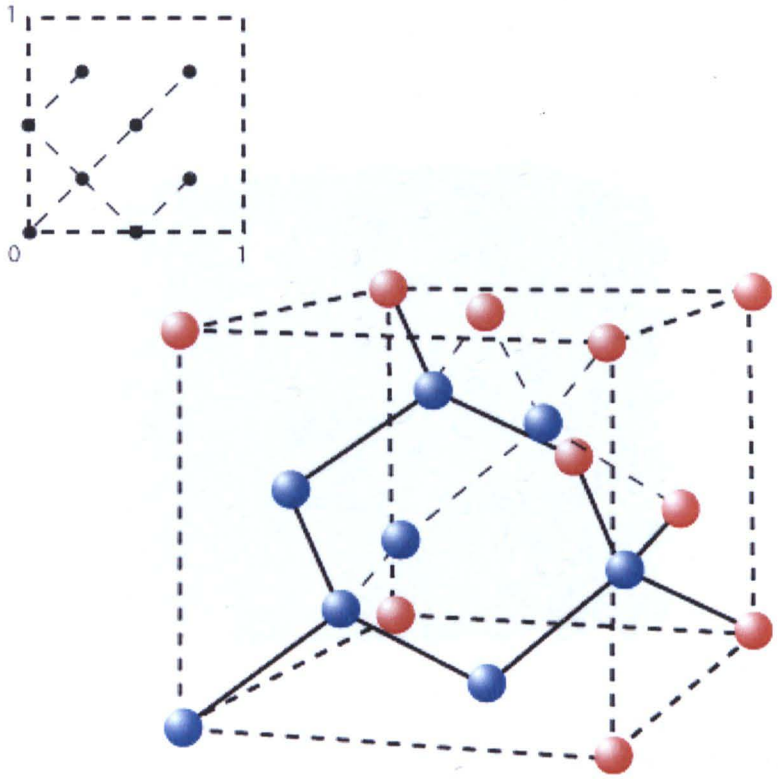

Figure 3.8: Silicon lattice basis consisting of 8 silicon atoms at points $(0,0,0), (0,\frac{1}{2},\frac{1}{2}), (\frac{1}{2},0,\frac{1}{2}), (\frac{1}{2},\frac{1}{2},0), (\frac{1}{4},\frac{1}{4},\frac{1}{4}), (\frac{1}{4},\frac{3}{4},\frac{3}{4}), (\frac{3}{4},\frac{1}{4},\frac{3}{4}), (\frac{3}{4},\frac{3}{4},\frac{1}{4})$ offsets are normalized to a unit cube.

a unit cube are $(0,0,0), (0,\frac{1}{2},\frac{1}{2}), (\frac{1}{2},0,\frac{1}{2}), (\frac{1}{2},\frac{1}{2},0), (\frac{1}{4},\frac{1}{4},\frac{1}{4}), (\frac{1}{4},\frac{3}{4},\frac{3}{4}), (\frac{3}{4},\frac{1}{4},\frac{3}{4}), (\frac{3}{4},\frac{3}{4},\frac{1}{4})$. This basis determines the unit translational vector which size replicates the lattice constant of Si which is 0.543 nm. As this volume is replicated and translated by multiples of the translational vector it fills all of the simulation domain creating the lattice sites corresponding to a standard bi-cubic silicon lattice, depicted in Figure 3.9. The illustrated section of a generated lattice is generated by $5 \times 5 \times 5$ unit volumes (cells) and contains 1000 silicon atom sites.

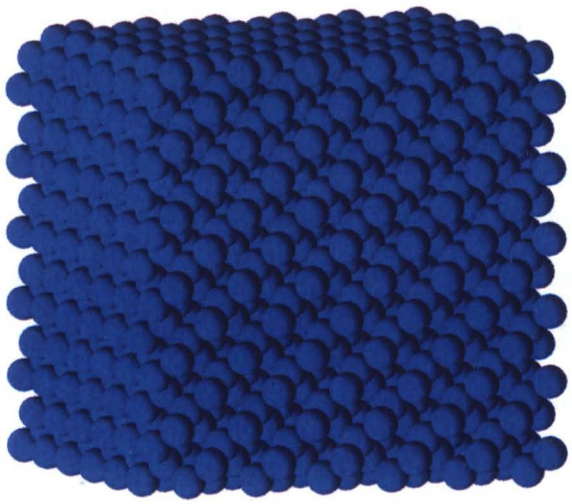

Figure 3.9: Section of a generated silicon lattice created by the replication of the basis shown in Figure 3.8. This section is 5x5x5 silicon sites and contains 1000 silicon atoms.

The process by which Si atoms are replaced by dopants is similar to the method outlined before. A simple rejection technique is used but in this case instead of stepping over all of the mesh, this method steps over all of the possible lattice sites and selects whether or not a dopant atom is to be placed there depending on a probability generated by the ratio of the doping concentration and the Si concentration at that site.

Previously the actual number of dopants in a particular region was chosen from a Poisson distribution with a mean equal to the calculated average dopant number, this average value comes from integrating the continuous doping within a given region [44]. In this case the number of dopants over an ensemble of devices is guaranteed to follow a Poisson distribution. It is expected that this lattice placement technique should also generate a Poisson distribution due to the independent nature of dopant placement.

To validate our method for allocating dopants to the sites of the silicon lattice we simulated a slab of silicon of size $30 \times 30 \times 24$ nm with a doping concentration of $1 \times 10^{19} cm^{-3}$ and the simulated 200 devices with different microscopic doping configurations. The mean number of dopants obtained from these 200 simulations was 224.45 with a standard deviation of the distribution of devices of 15.43. The expression for the Poisson distribution [110] is shown in equation (3.9), where r is the number of events, n is the number of possible events and ρ is the probability of an event occurring. In our test case the number of possible events is equal to the average doping concentration that we have assigned to our simulation which is $1 \times 10^{19} cm^{-3}$ and the probability of the events occurring is equal to the volume $30 \times 30 \times 24$ nm.

$$P(r;\mu) = \frac{e^{-\mu}\mu^{r}}{r!} \quad where \quad \mu = n\rho \tag{3.9}$$

The mean of the Poisson distribution is given by equation 3.10, where μ is as described in equation (3.9). This means that the mean for a Poisson distribution is the calculated value of μ which in this case is 216. This difference between this mean and the simulated mean is 3% of the mean which gives confidence that we

are generating the correct number of dopants with our lattice placement method.

$$\overline{r} = \sum_{r=0}^{\infty} rP(r;\mu) = \mu e^{-\mu} e^{\mu} = \mu \tag{3.10}$$

The derivation of the standard deviation of a Poisson distribution is given by equation (3.11) and (3.12). As before $\mu = n\rho$ and all other symbols have the same meaning as described previously.

$$\overline{r^2} = \sum_{r=0}^{\infty} r^2 P(r;\mu) = \mu^2 e^{-\mu} e^{\mu} + \mu e^{-\mu} e^{\mu} = \mu^2 + \mu \tag{3.11}$$

$$\sigma^2 = \overline{r^2} - (\overline{r})^2 = \mu \quad or \quad \sigma = \sqrt{\mu} \tag{3.12}$$

The importance of the μ parameter can be seen by equation (3.12), not only is μ the mean of a Poisson distribution it is also the variance of a Poisson distribution, the calculated value of the standard deviation is then 14.69 while the measured value is 15.43, which is a 5% difference, this again gives confidence that our method for allocating dopants is valid and giving the correct distribution of dopant atoms.

3.4 Mobility Models

Another extension made to the device simulator was the inclusion of more physical and accurate mobility models. In the original simulator only a simple constant mobility model was available to represent the transport in the MOSFET

channel. Such simple mobility models do not take into account the mobility reduction in the highly doped source and drain regions, nor the effects of the lateral and perpendicular electric fields in the channel. In order to include the doping concentration and field dependence of the mobility we have adopted the well regarded mobility model published by D. M. Caughey and R. E. Thomas [111].

This more accurate approach to modelling the mobility is required when studying realistic devices in order to match in the simulations the measured characteristics of these devices. The simulation of intrinsic parameter fluctuations in realistic MOSFETs should start with a calibration of the simulator against the measured characteristics of a wide self averaging device which in turn requires accurate and flexible mobility models. For the 35 nm Toshiba MOSFET which is the focus of this work the Caughey and Tomas mobility model works well.

3.4.1 Concentration Dependant Mobility

The first part of the Caughey-Thomas mobility model describes the doping concentration dependence of the mobility. In the case of electrons the doping concentration dependent mobility is given by equation 3.13 where $\mu_{n_{min}}$ and $\mu_{n_{max}}$ are the minimum and maximum mobilites, N_{refn} is a reference concentration, N_{total} is the doping concentration at a particular position in the device and α_n is a fitting parameter.

$$\mu_{0n} = \mu_{n_{min}} + \frac{\mu_{n_{max}} - \mu_{n_{min}}}{1 + (\frac{N_{total}}{N_{refn}})^{\alpha_n}} \tag{3.13}$$

The values of the parameters used in our simulation are shown in Table 3.1. They are the same as the values used in the commercial simulator Taurus used initially in the calibration process.

Parameter	Value
$\mu_{n_{min}}$	$55.24cm^2/Vs$
$\mu_{n_{max}}$	$1429.23cm^2/Vs$
N_{refn}	$1.072 \times 10^{17} atoms/cm^3$
α_n	0.73

Table 3.1: List of the parameters used in the combined analytical mobility model used in the simulation of the 35nm MOSFET device, along with the values used in the simulation of the 35nm device

Figure 3.10 compares the doping concentration dependance of the mobility calculated using (3.13) with two sets of experimental data [112, 113]. The selected parameters for the Caughey-Thomas analytical model produce a good agreement over the whole range of doping concentrations relevant to the simulated MOSFETs in this work.

3.4.2 Field Dependant Mobility

The concentration dependent mobility obtained from (3.13) is further modified to take into account the lateral and vertical field dependance of the mobility. The vertical field dependance is introduced by Equation (3.14).

$$\mu_{S,n} = G_{surfn} \frac{\mu_{0n}}{\sqrt{1 + \frac{E_{\perp n}}{E_{cn}}}} \tag{3.14}$$

In this equation G_{surfn} is a parameter used to alter the mobility taking into

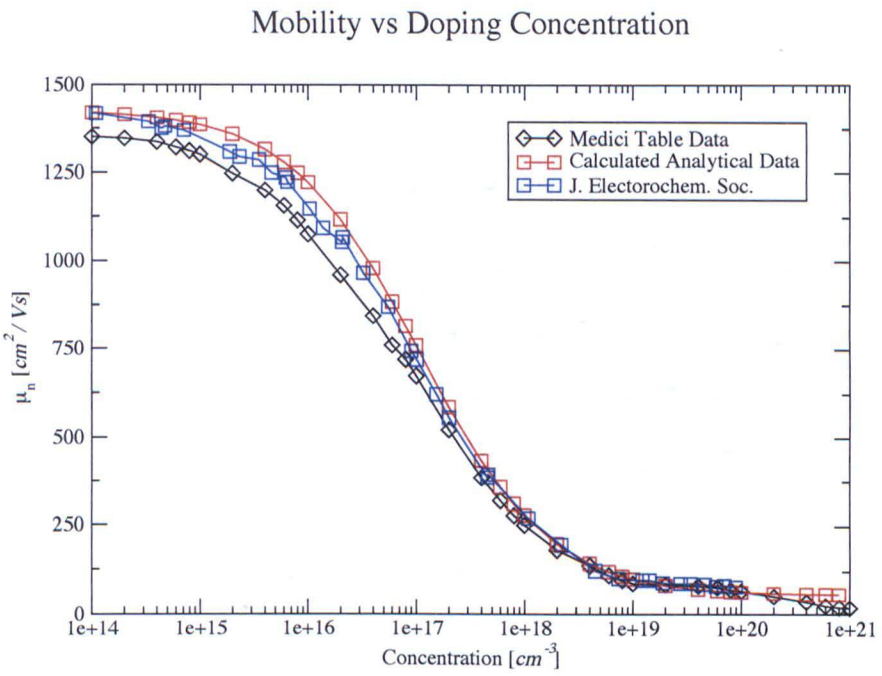

Figure 3.10: Plot of mobility versus doping, showing the calculated analytical curve based on the Caughey-Tomas mobility model as well as two sets of experimental data [112, 113].

account the interface roughness scattering, E_{cn} is a reference electrical field, μ_{0n} is the concentration dependent mobility obtained from (3.13) and $E_{\perp n}$ is the perpendicular electric field.

Equation (3.15) introduces the lateral field dependance of the mobility where $\mu_{S,n}$ is the adjusted mobility taking into account the perpendicular field dependence according to (3.14).

$$\mu_n = \frac{\mu_{S,n}}{[1 + (\frac{\mu_{S,n} E_{\|n}}{v_n^{sat}})^{\beta_n}]^{\frac{1}{\beta_n}}} \tag{3.15}$$

$E_{\|n}$ is the electric field parallel to the interface and v_n^{sat} is the electron saturation velocity and β_n is a fitting parameter which is set to 2. The values of the parameters used in our simulation are shown in Table 3.2. They are the same as the values used in the commercial simulator Taurus used initially in the calibration process.

Parameter	Value
G_{surfn}	1
E_{cn}	$4 \times 10^5 V/cm$
v_n^{sat}	$1.1 \times 10^7 cm/s$
β_n	2

Table 3.2: List of the parameters used in the Caughey-Tomas mobility model for electric field dependence used in the simulation of the 35 nm MOSFET device, along with the values used in the simulation of the 35 nm device

Figure 3.11 shows results describing the lateral field dependance of the mobility according to equation (3.15) with experimental mobility field characteristics reported by Canali [1] for two orientations of silicon. There is some disagreement between the model and the experimental data at high electric fields. This

however is not if great concern to our DD simulations which cannot represent non-equilibrium transport and are mainly valid in the subthreshold region when the current is dominated by diffusion.

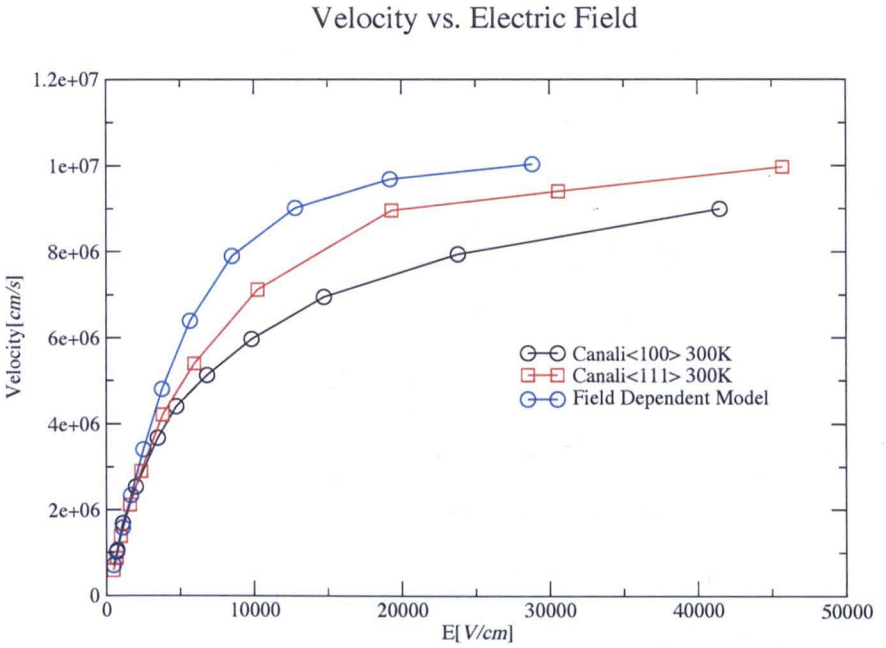

Figure 3.11: Velocity versus Electric Field comparing the Field Dependent mobility model to the measured data given in the Canali work [1].

While the Caughy-Tomas mobility model is a great improvement over the original method of calculating the mobility it does have a number of restrictions in this case. This model cannot be used in the "atomistic" regime, because of the large doping concentrations associated with discrete random dopants and the high electric fields associated with the finely resolved potential well this mobility model gives incorrecect values for mobility at these positions. These incorrect parameters cause problems when trying to gain a solution to the cur-

rent continuity equations and therefore it cannot be used in the "atomistic" regime. To overcome this problem the mobility for a uniform device is used in the "atomistic" regime, this means that a complete solution is found for each voltage point within the uniform device and the associated mobility is used for each point in the "atomistic" device. While this solution is not perfect it does give us a better estimate of the mobility within an "atomistic" device.

In this chapter we have looked at the methods and models that have been used in this work. A 3D Drift Diffusion simulator is used due to its relative speed to allow us to carry out a statistical analysis of multiple sources of fluctuations. The models to implement line edge roughness and oxide thickness fluctuations have been outlined and an improved method for placing discrete random dopants using the silicon lattice was introduced. The implementation of an improved mobility model following the method outlined by Caughy and Tomas was also outlined as well as the method by which it is included in "atomistic" simulations.

Once a more correct method for including "atomistic" doping was created and a more physical model for the mobility within a device was introduced it was possible to begin the simulations of realistic devices. With these improvements in place it was possible to begin calibration in respect of the real 35 nm MOSFET published by Toshiba [32]. However not all the problems associated with the simulation of a real device in the presence of random discrete dopants where resolved by the development of the code as outlined in this chapter. New techniques had to be developed to overcome some problems associated with the fine resolution of individual discrete dopants in simple DD simulations. These problems and the developed techniques are covered in the next chapter.

Chapter 4

Resolving Discrete Charges

In order to correctly model a realistic MOSFET it is necessary to expand on and improve the models used for simulation. To improve the method of generating a device, a process of introducing discrete random dopants at lattice positions was introduced. More accurate and realistic mobility models, dependent on the electric field and doping concentration, were also introduced. The purpose behind these additions was to create the best possible model so that results for a realistic device could be obtained and interpreted correctly.

In the early stages of MOSFET simulation a problem was discovered with the classical representation of dopant atoms in small MOSFET simulations. In classical Drift Diffusion simulations the electron concentration closely follows the electrostatic potential obtained from the solution of Poisson's equation, due to the use of Boltzmann or Fermi-Dirac statistics. When discrete random dopants are introduced into the continuous Drift Diffusion model the mesh based solution of the Poisson equation produces sharply resolved potential well or peaks, which

at high mesh densities capture more closely the singularities in the Coulomb potential of a point charge. When Boltzmann or Fermi-Dirac statistics are applied to these sharply resolved potential wells, or peaks a large amount of the available mobile charge becomes localised (trapped) in the surrounding region. This artificial localisation of mobile charge leads to a mesh spacing dependent reduction of conductance within a heavily doped "atomistic" region and therefore artificial increases the resistance of, for example, the source and drain regions of a MOSFET.

This chapter is dedicated to the problems associated with charge localisation in "atomistic" device simulation. An example of a simple resistor is used to study and understand the associated problems. It outlines two possible solutions to the problem, the first being the use of various charge assignment schemes to "smear" the charge density associated with an individual dopant atom across a series of mesh cells, reducing the concentration of charge at a specific node and so reducing the sharpness of the potential well, or peak generated by Poisson's Equation[49]. The second solution is to include Quantum Corrections in our case based on the Density Gradient approach[114, 115]. Due to confinement in the sharply resolved Coulomb potential well the ground state for electrons shifts up within the well, close to the intrinsic conduction band edge (typically $50meV$ below intrinsic Ec)[116, 117, 118], and this substantially reduces the amount of erroneous charge trapping.

The trapping related mesh dependance is carefully investigated in this chapter. When the mesh size and the volume of a mesh cell is reduced the charge density associated with the assignment of a single discrete dopant increases creating a

more sharply resolved potential well. This in turn causes more charge localization leading to increased resistance and introducing a strong mesh dependence in the simulation device characteristics.

The effects of introducing quantum corrections to the MOSFET simulation are presented and discussed. Finally a case for including Density Gradient quantum corrections for holes as well as for electrons in a "unipolar" n-MOS simulation is made. It is shown that even though the holes do not participate in the transport the trapping of holes in sharply resolved Coulomb wells in the substrate effects the shape and size of the depletion region and therefore the threshold voltage of the simulated device.

4.1 Problems in Classical Simulations

The inclusion of discrete random dopants in classical Drift Diffusion simulations introduces several problems when mesh based discretisation is used to solve Poisson's equation [50]. The use of Boltzmann or Fermi-Dirac statistics lead to a "trapping" of carriers by the sharply resolved potential wells or peaks associated with the mesh resolved Coulomb potential of individual donors or acceptors. The effect of this is that an area of a device being simulated "atomistically" using discrete random dopants will have an artificially lower conductance and higher resistance compared to the case of a continuously doped simulation[114, 115]. This leads to problems particularly in the case of MOSFET simulations where the transistors will have an artificially increased access resistance associated with the introduction of "atomistic" source and drain doping and therefore reduced

current.

A second problem caused by the inclusion of discrete random dopants in Drift Diffusion is the strong mesh dependance of the solution. Normally a discrete dopant is introduced into the continuum world of the Drift Diffusion formalism as a charge density inversely proportional to the volume of the mesh cell. This implies that as the mesh is decreased the charge density increases and becomes more localised. This results in a better resolution of the singular Coulomb potential of a point charge associated with the dopant and in sharper peaks in the solution. As a result more mobile charge becomes localised and this reduces the free charge participating in the transport and contributing to the device current. Therefore a simulation of a MOSFET that has a finer mesh size will result for example in a larger source / drain access resistance.

Various solutions have been proposed to correct the problems associated with the inclusion of "atomistic" dopants. These include charge assignment schemes where the charge associated with an individual atom is spread over the surrounding mesh cells[49], the splitting of the Coulomb potential into long and short range components[119] and the use of quantum corrections to alter the shape of the electron concentration profile[114, 115, 120, 121]. In the case of splitting the potential into long and short range components based on screening considerations, there is the possibility of "double counting" the effects of mobile charge screening and also that the method uses a somewhat arbitrary cut-off point. There are two different methods proposed to use quantum corrections, one using the density gradient formalism will be discussed in more detail later in this chapter and the other relies on quantum perturbation theory to create

an electron concentration profile which is then curve fitted to a more simplified equation to be used in simulations.

4.1.1 Charge Localisation

As stated earlier the use of Boltzmann or Fermi-Dirac statistics in classical drift-diffusion simulations means that the electron concentration follows the electrostatic potential. In the case of n-type material where the majority carriers are the electrons, a dopant atom will produce a sharply resolved potential well in the conduction band edge as seen by an electron. As a result a significant amount of electron charge can become "trapped" or localised in the sharply resolved Coulomb potential well created by a discrete dopant assigned to a fine mesh. Such trapping is physically impossible since, in quantum mechanical terms, confinement keeps the ground electron state high in the well. Figure 4.1 , which depicts a 1D Poisson-Schrödinger solution, shows the Coulomb potential well corresponding to a charge plane and the bound energy states, illustrating this point. It is clear that a large fraction of the potential in the well cannot be followed by the electron concentration due to such quantisation. This results in a quantum reduction of the electron concentration in the well when compared to the classical electron concentration which can also be seen in figure 4.1.

4.1.2 Impact on Device Simulation

The effects of charge localisation can have a significant impact on the results of classical MOSFET simulation. Here the inclusion of discrete random dopants

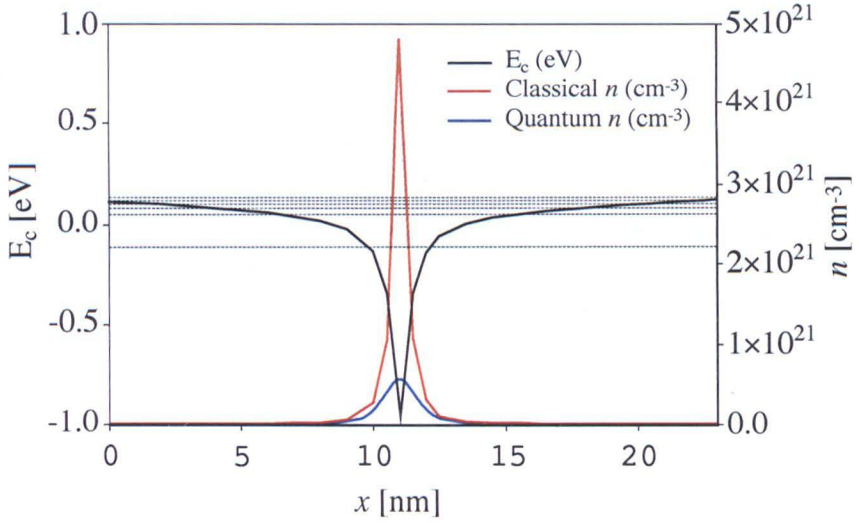

Figure 4.1: A 1D Poisson-Schrödinger solution, showing the Coulomb potential well and bound energy states. Also showing the classical electron concentration and the quantum electron concentration associated with this potential well

particularly in the source / drain regions can result in an increased access resistance of these source / drain regions current. This can be understood by inspecting the potential across the channel obtained from the "atomistic" Drift Diffusion simulations of a classical n-MOSFET. Figure 4.2 shows a 2D potential slice along the channel of a 35 × 35 nm square n-MOSFET. In this case only the substrate is "atomistically" doped, while the source and drain regions are continuously doped. In this case the electrons, as the majority carrier would move over the potential barrier and through the valleys produced by the potential spikes caused by the acceptors in the channel. There would be no charge localisation in the source /drain regions to effect the access resistance and the device current.

Figure 4.3 shows the inverse situation. In this case the substrate region is

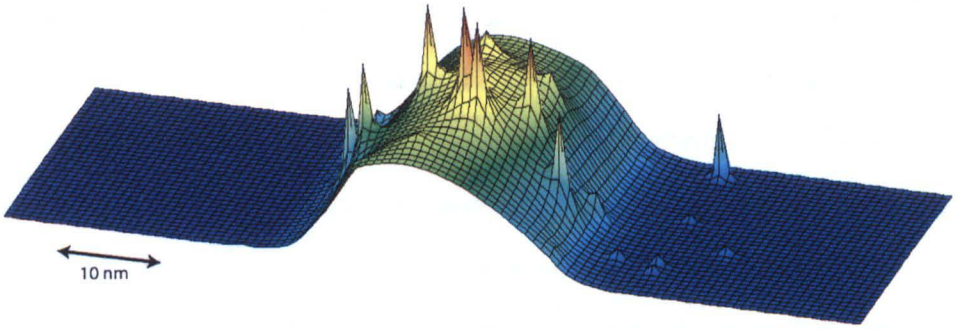

Figure 4.2: 2D Potential slice along the channel of a 35×35 nm MOSFET. The Source and Drain regions of the MOSFET are continuously doped whereas the substrate is "atomistically" doped. The potential plane is 35×107 nm in size and varies from approximately 0V (represented by the blue region) to approximately 0.7V (represented by the red).

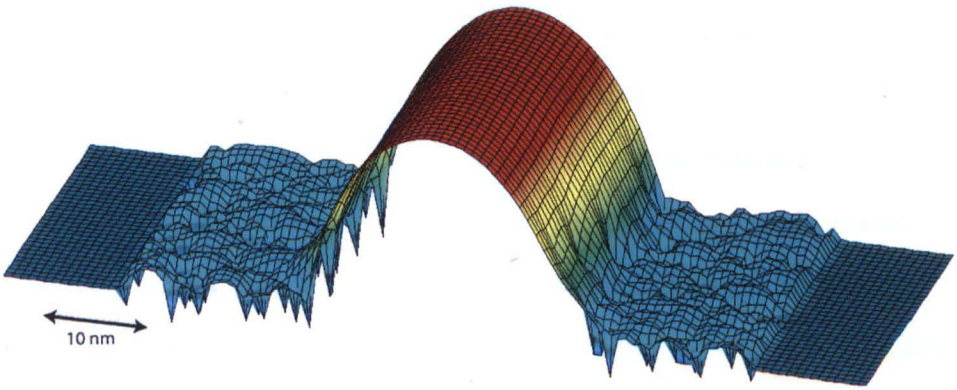

Figure 4.3: 2D Potential slice along the channel of a 35×35 nm MOSFET. The Source and Drain regions of the MOSFET are "atomistically" doped while the substrate region is continuously doped. he potential plane is 35×107 nm in size and varies from approximately 0V (represented by the blue region) to approximately 0.7V (represented by the red).

continuously doped while the source and drain regions are now "atomistically" doped. As can be seen there are now sharply resolved potential wells within the source/drain regions. When the electron concentration is calculated a high proportion of the majority carriers (in this case electrons) will be trapped in these wells. The electrons in the base of these potential wells now have to overcome a much larger potential barrier in order to participate in the current. This results in an increased resistance in the source/drain regions and in turn leads to a reduced current in the device.

4.2 "Atomistic" Resistor Study

To investigate the overall increase in resistance associated with charge trapping, a simple device structure was needed so that the more complex effects associated with the design and operation of a MOSFET would be excluded. A simple resistor was chosen as the best and simplest possible device suitable for this purpose. The resistor was modelled as a slab of n-doped silicon, which could be doped at various concentrations to study the effects of charge localisation. For the simulations presented here the doping concentration used was $1 \times 10^{20} cm^{-3}$ although the results were consistent at other doping levels. This particular concentration is representative of the doping concentrations in the source/drain regions of modern MOSFETs.

Initially the size of the "atomistic" resistor was chosen to be $50 \times 50 \times 50$ nm, using a mesh spacing of 1 nm which is also typically used in "atomistic" MOSFET simulation. With the introduction of quantum corrections the simulation

time for a mesh of this size became prohibitive so a smaller resistor with a size of $30 \times 20 \times 20$ nm was adopted keeping the mesh spacing of 1 nm. The resistor itself could be doped both continuously as well as "atomistically" which allowed a comparison between the continuous resistance and the "atomistic" resistance.

$$\rho = \frac{1}{qnu_n} \tag{4.1}$$

$$R = \frac{\rho L}{WH} \tag{4.2}$$

To begin with the analytical resistance for a resistor of the size and doping described above was calculated. To calculate the resistance of this resistor, the resitivity was first calculated using equation (4.1) assuming $n = N_D$, where $N_D = 1 \times 10^{20} cm^{-3}$. In this case a constant mobility of $300 cm^2 V^{-1} s^{-1}$ was used in the simulation which gives a resistivity of $2 \times 10^{-4} \Omega cm$. The resistance was calculated for the appropriate length, width and height using equation (4.2).In the case of a resistor $50 \times 50 \times 50$ nm this would be 41.6Ω and for the case of a resistor $30 \times 20 \times 20$ nm this would be 156.25Ω.

Once the analytical value for the resistance was obtained the continuously doped resistor was simulated using the classical Drift Diffusion approach. A perfect match was found betweeen the analytical value for the resistance and the results of the Drift Diffusion simulation. This confirmed the accuracy of the simulation for the case of the continuously doped resistor which was then used as a benchmark for comparison with the results from the "atomistic" cases.

In the case of the "atomistic" simulations we chose to simulate and average

the results from an ensemble of 200 "atomistic" devices as a statistical sample indicative of the effects caused due to the inclusion of discrete random dopants. The problem of charge localisation is closely related to the use of mesh space dependent charge densities to represent discrete random dopants. One possible solution to the problem is to "smear" out the associated charge over a broader (preferably constant) area and so reduce the amount of the charge density at any given mesh point, so reducing the depth of the associated potential well. In order to do this, various charge assignment schemes can be used. In this section we compare three different charge assignment methods, Nearest Grid Point (NGP), Cloud in Cell (CIC) [99] and a method we refer to as Gaussian Smearing.

4.2.1 Charge Assignment Scheme

As discussed in Section 3.3.3 when introducing discrete random dopants within our simulated device, the silicon crystalline lattice is used within the device as a reference and the discrete random dopants are placed at the individual lattice sites based on a probability created from the volume and the associated doping concentration.

These dopant positions do not, however, match the mesh points generated by our simulation grid and it is therefore necessary to assign the charge density associated with a discrete dopant atom to these mesh points. This can be accomplished by the use of different charge assignment schemes. The main principle of the various charge assignment schemes is that an individual dopant atom will make a contribution to the charge density assigned to the neighbour-

ing mesh points in the simulation. The amount of this contribution generally depends on the distance between the dopant position and the corresponding mesh point and results in a weighting between 0 and 1 determining the fraction of the doping density, $\frac{1}{V}$ (where V is the volume of the mesh cell) assigned to the corresponding node. In three dimensions this can be written as in equation (4.3).

$$\rho(x,y,z) = w_x w_y w_z \frac{1}{V} \tag{4.3}$$

Here the nodal contribution of a single dopant to the charge density ρ at a position (x,y,z) is given by the accumulated weighting,w in the x, y and z directions of the charge density $\frac{1}{V}$. The various charge assignment schemes vary in terms of the weighting factors w_x, w_y and w_z and therefore in the positional dependance of the contribution to a particular mesh point.

The simplest charge assignment scheme is Nearest Grid Point (NGP). In the NGP approach all of the charge associated with a single dopant atom is assigned to the single closest mesh point. In this case the expression for calculating the weighting in one dimension is shown in equation (4.4).

$$w_x = \begin{cases} 1, & |x_i - x_m| \leq \frac{1}{2} \\ 0, & otherwise \end{cases} \tag{4.4}$$

Here a weighting of 1 is given to the closest mesh point while all other weightings are 0. The benefits of this method are that it is quick and simple to implement. However in respect to the effects associated with charge trapping this

is the worst charge assignment scheme, because NGP assigns all of the charge associated with a single dopant atom to a single mesh point it produces very sharp and deep potential well. Along with this, the NGP scheme also means that the placement of dopants is strongly coupled to the mesh, particularly if a very coarse mesh is used, introducing artificial correlation in the doping distribution. A more complex scheme for charge assignment is the Cloud in Cell (CIC) approach. In this approach the charge associated with a single dopant atom in three dimensions is spread across the surrounding eight mesh points. The appropriate formula for calculating the weighting in one dimension is given in equation 4.5 where x_i is the x coordinate of the dopant and x_m is the x coordinate of the mesh node.

$$w_x = \begin{cases} 1 - |x_i - x_m|, & |x_i - x_m| \leq 1 \\ 0, & otherwise \end{cases} \tag{4.5}$$

Here the weighting depends on the position of the dopant within the surrounding box, the closer a dopant is to a mesh point the larger the doping density assigned at that point. Figure 4.4 illustrates this method of charge assignment. The dopant atom represented by the red sphere is assigned to the surrounding eight mesh nodes. If the arrows indicate the distance to a particular mesh point then the back, lower left node would receive the greatest proportion of the associated charge while the back, upper right would receive the least.

This method of charge assignment is an improvement over NGP as the charge associated with a single dopant atom is spread over a region of eight nodes as opposed to one node. For a dopant close to the centre of the cell this gives the

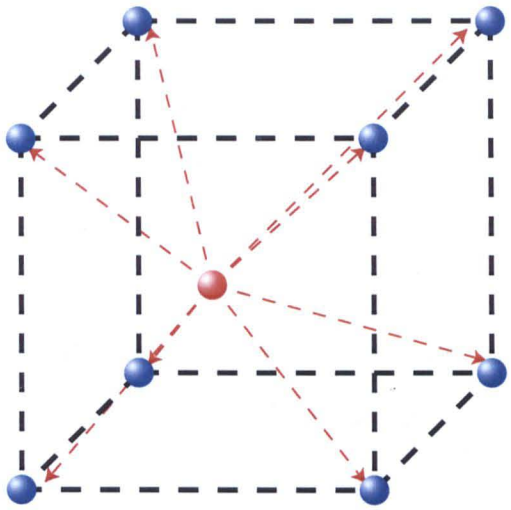

Figure 4.4: Schematic view of the Cloud in Cell charge assignment scheme, the charge associated with a single dopant atom is applied to the eight surrounding mesh cell.

best possible improvement compared to NGP, assigning one eighth of the doping density to each node of the cell. However as a dopant moves closer to a single mesh point this charge assignment scheme becomes more like NGP. CIC exhibits charge trapping which increases with the reduction in mesh size, however the effect is significantly reduced when compared with the NGP scheme.

A wider spreading of the charge can be achieved by using the Triangular Shaped Cloud (TSC) approach to charge assignment. However this approach is time consuming and more complicated to implement than the previous assignment schemes [99]. The expression for calculating the weighting in TSC is shown in equation (4.6). Again x_i is the x component of the dopant position and x_m is the x component of the mesh nodes position in real space.

$$w_x = \begin{cases} \frac{3}{4} - |x_i - x_m|^2, & |x_i - x_m| \leq \frac{1}{2} \\ \frac{1}{2}(\frac{3}{4} - |x_i - x_m|)^2, & \frac{1}{2} < |x_i - x_m| \leq \frac{3}{2} \end{cases} \tag{4.6}$$

In this method the charge assignment extends outside of a single mesh cell to the surrounding cells. However when these cells vary in both size and volume it becomes difficult to calculate the correct weighting, resulting in charge spreading over more than one cell. This is particularly undesirable at the Si/SiO_2 interface in MOSFET simulations. TSC has not been used when comparing effects of charge assignment schemes on charge localisation.

The last method of charge assignment with which we have been experimenting is to use a Gaussian smearing. In this case a dopant is represented as a normalised Gaussian distribution, which has a total associated charge of one. The doping concentration at each of the mesh points is then calculated from this distribution using equation (4.7). Where $\rho(x)$ is the charge density at position x, x_i is the x coordinate of the dopant in real space and x_m is the x coordinate of the mesh node in real space. σ^2 is the variance of the Gaussian distribution and gives the width of the distribution's spread.

$$\rho(x) = exp\left(\frac{(x_i - x_m)^2}{\sigma^2}\right) \tag{4.7}$$

The width of this distribution is set to 2 nm which corresponds to the effective diameter of a dopant using the Bohr model for hydrogen adjusted for silicon. The precise width (and indeed this method in general) is open to discussion. This approach is somewhat similar to the attempt to split the Coulomb potential into long and short range components [50, 119] as the point at which the split

is taken is arbitrary and there is a possibility of double counting the screening effects since screening is also included in the Drift Diffusion system of equations.

The Gaussian spreading approach is purely empirical and may result in a loss of resolution with respect to actual "atomistic" effects as they become washed out as a result of the dopants spreading over such a wide area. This is particularly true in very small devices where only a few discrete dopants determine the device characteristics. However, this method is the most successful in restoring the correct resistivity of the simulated resistors obtained using the continuously doped profile.

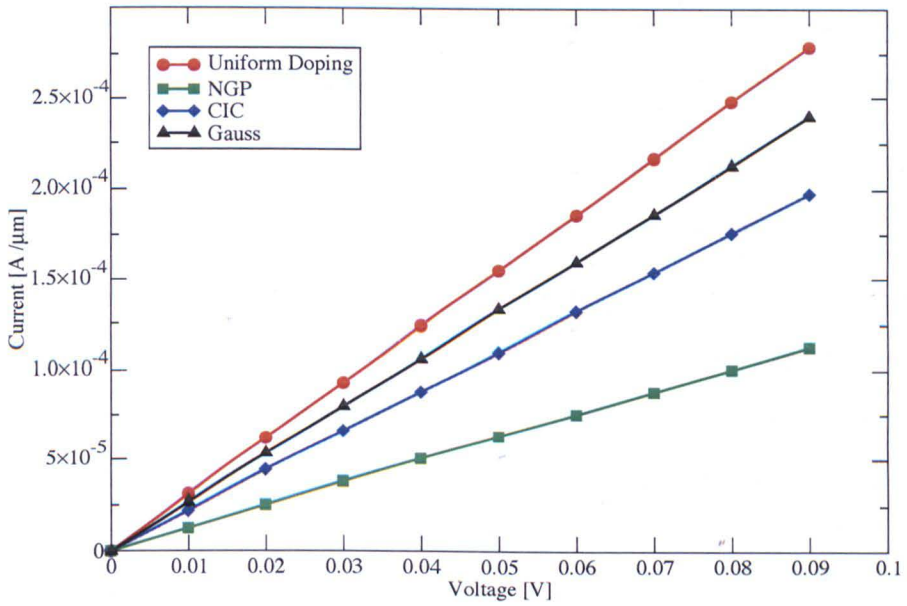

Figure 4.5: The IV characteristics of a 50 × 50 × 50 nm resistor with 1nm mesh spacing. Showing the cases of a continuously doped resistor as well as the atomistic average of 200 devices for three different charge assignment schemes, being Nearest Grid Point, Clound in Cell and Gaussian Smearing.

The capability of the described charge assignment schemes to reduce the charge trapping is illustrated in figure 4.5 which depicts the IV characteristics obtained from the classical simulation of a $50 \times 50 \times 50$ nm resistor with 1nm mesh spacing, for both continuous doping and for the three methods of charge assignment. It can be seen that the resistance of the device increases with the inclusion of discrete random dopants and that when a coarse method of charge assignment such as NGP is used the increase in resistance is greatest. This is due to the sharply resolved potential wells associated with the high doping concentrations representing the discrete random dopants. When using CIC charge assignment the resistance is lowered in comparison to the NGP approach as for an identical mesh cell size the magnitude of the doping concentration at individual mesh points is reduced and so the associated depth of the potential well is also.

However it can also be seen that even using CIC we do not recover the IV characteristics and resistance corresponding to the continuous doping. If the Gaussian Smearing technique is used the results are closer to the continuous case. As we have mentioned however the use of Gaussian Smearing is purely an empirical approach relying on an almost arbitrary choice of the standard deviation and therefore may result in the loss of resolution in respect of "atomic" scale effects.

4.2.2 Quantum Corrections

An alternative approach to the use of charge assignment schemes for reducing charge localisation is to introduce quantum corrections, in particular the use of the Density Gradient (DG) method. As has been discussed previously this

method has been used successfully to model confinement effects at the interface of a MOSFET and the corresponding shape of the electron distribution peaking away from the interface. If these quantum corrections are applied to the region that contains discrete random dopants a marked reduction in the charge localisation at impurity sites can be observed. In the case of a sharply resolved potential well electrons are unable to follow the real electrostatic potential because their concentration is determined by the effective quantum potential which introduces a force pushing the electrons out of the well. The result of this is that when DG is applied in conjunction with the Drift Diffusion system of equations it smoothes out the unnaturally large variations in the electron concentration around the dopants and more carriers are able to contribute to the current flow in the simulation.

The impact of the inclusion of quantum confinement effects in simulations by use of the Density Gradient approximation is illustrated in Figure 4.6. Quantum confinement in the sharp potential wells smoothes the overall electron concentration by reducing the sharper peaks eliminating the charge localisation. It can also be seen that this effect raises the concentration in the surrounding area allowing more charge to participate in the simulation instead of being trapped in the potential wells. Because less of the simulation's mobile charge is trapped within the Coulomb potential wells, the overall resistance of the simulated slab is reduced.

With the inclusion of quantum corrections the simulation time for 200 devices of a resistor of dimensions $50 \times 50 \times 50$ nm became prohibitively expensive and so the size of the simulated slab was reduced to $30 \times 20 \times 20$ nm. The same

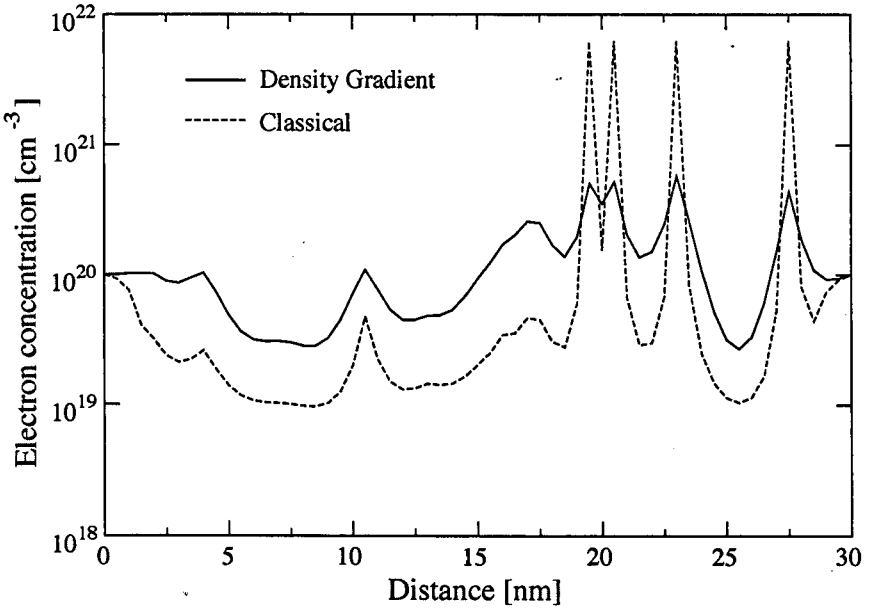

Figure 4.6: The Electron Concentration along the x-direction of a $30 \times 20 \times 20$ nm resistor, showing both the classical electron concentration due to Boltzmann statistics and the quantum electron concentration generated by the inclusion of Density Gradient quantum corrections.

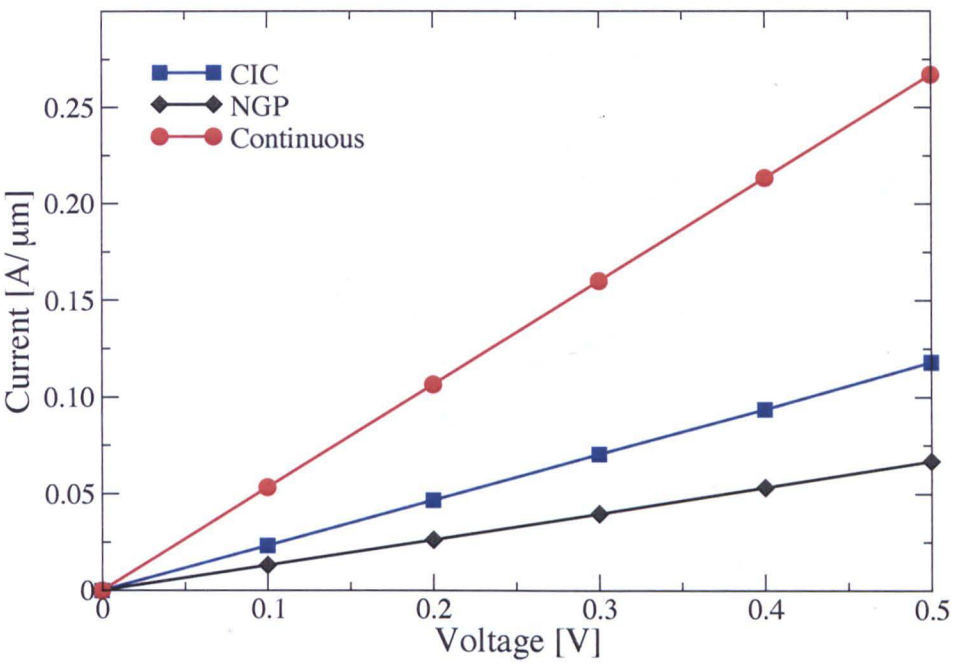

Figure 4.7: The IV characteristics of a 30 × 20 × 20 nm resistor with 1 nm mesh spacing. Showing the cases of a continuously doped resistor as well as the atomistic average of 200 devices for two different charge assignment schemes, being Nearest Grid Point and Clound in Cell.

classical simulations where carried out for this smaller resistor using initially CIC and NGP charge assignment schemes. Figure 4.7 depicts the IV characteristics obtained from classical simulations of the resistor for both continuous doping and for the two methods of charge assignment for a mesh spacing of 1 nm. The same increase of resistance can be seen as before in the case of the "atomistic" simulations.

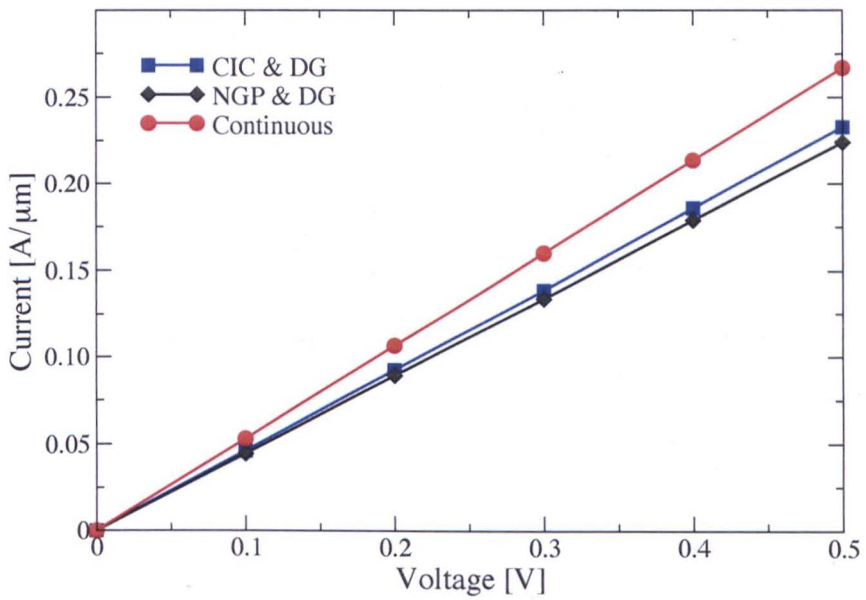

Figure 4.8: The IV characteristics of a 30 × 20 × 20 nm resistor with 1 nm mesh spacing. Showing the cases of a continuously doped resistor as well as the atomistic average of 200 devices for two different charge assignment schemes, being Nearest Grid Point and Clound in Cell with the inclusion of DG quantum corrections.

Figure 4.8 illustrates the impact of the introduction of DG quantum corrections. When DG is applied to the simulation it smoothes the electron concentration. The overall effect of this smoothing process is a reduction in the amount of charge which is trapped around the individual discrete dopants. This

reduction in trapped charge leads to a decrease in the effective resistance of the device being simulated. This effect is clearly visible in Figure 4.8 where a significant improvement in the "atomistically" simulated resistance can be observed with the introduction of the quantum corrections for both CIC and NGP charge assignment schemes.

4.2.3 Mesh Sensitivity

The amount of charge that becomes localised is strongly dependant on the mesh size used in the simulation. If we consider the case of NGP the charge density at a given mesh point is $\frac{1}{V}$ where $V = h_x h_y h_z$. If we the reduce the mesh size a subsequent reduction in volume causes the charge density at that point to increase. The fine mesh in combination with the charge density increase results in a more sharply resolved Coulomb potential well leading to a greater charge localisation.

Figure 4.9 illustrates the mesh size dependence of the electrostatic potential and the effective quantum potential associated with the DG formalism obtained from the mesh-based solution of the Poisson's equation around a single discrete donor. As expected the singularity in the electrostatic potential increases with the reduction of the mesh size. At the same time the effective quantum potential is much less sensitive to the mesh size, showing no differences below 0.5 nm mesh spacing. The effective potential in the middle of the Coulomb well which is approximately 40 mV corresponds roughly to the energy level of the electron ground state and is in good agreement with the ground state of a hydrogenic model of an impurity in Si.

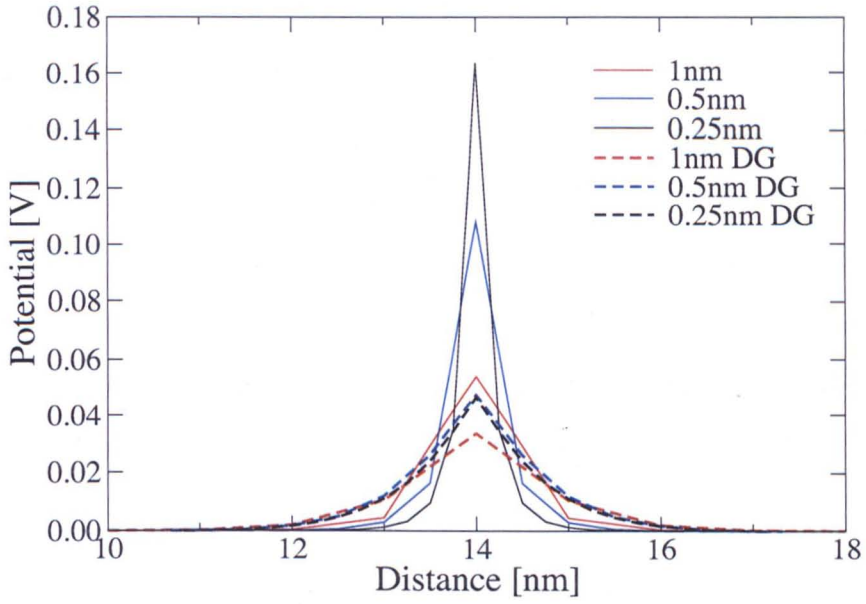

Figure 4.9: The Electrostatic potential and the effective quantum potential around a single point charge for mesh sizes of 1 nm, 0.5 nm and 0.25 nm

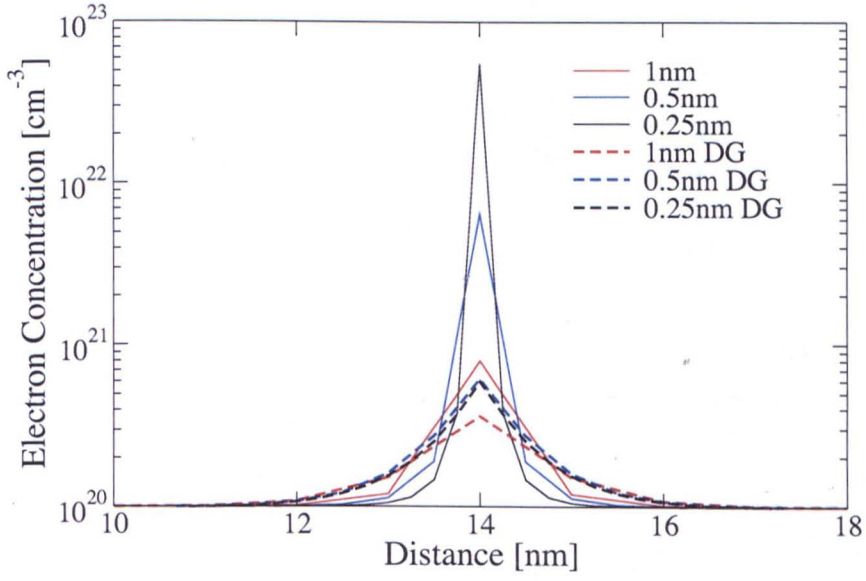

Figure 4.10: The classical and quantum electron concentrations around a single point charge for mesh sizes of 1 nm, 0.5 nm and 0.25 nm

Figure 4.10 depicts the electron concentration associated with a single point charge. In the classical case the electron concentration follows the solution of the Poisson equation which in turn is strongly coupled to the mesh size. Because of this the electron concentration is strongly coupled to the mesh size and as a result the smaller the mesh size the more mobile charge becomes localised by the sharply resolved potential well. We can also see that with the inclusion of quantum corrections the solution not only reduces the amount of charge being localised but for mesh sizes below 0.5 nm becomes mesh independent.

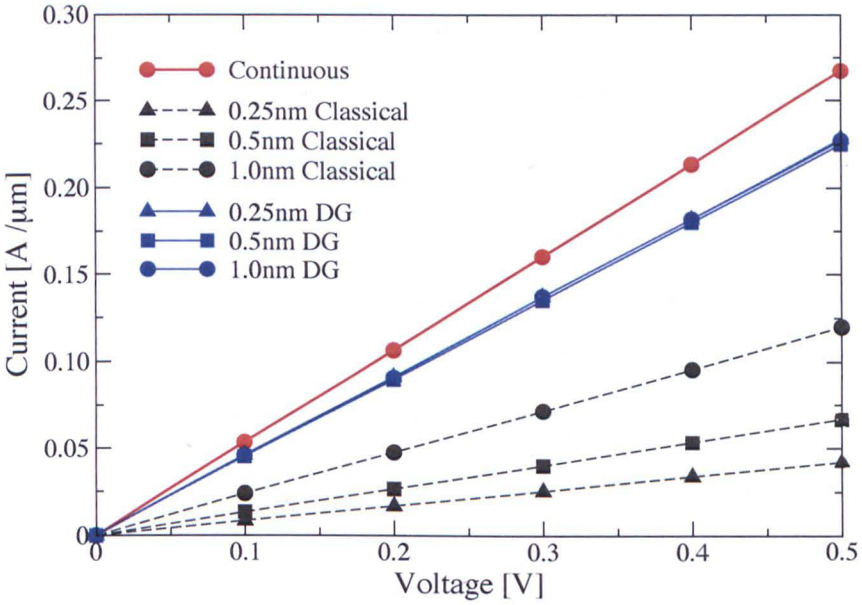

Figure 4.11: The IV characteristics of a 30 × 20 × 20 nm resistor for both continuous and "atomistic" doping using the NGP charge assignment scheme. Showing the average of 200 "atomistically" doped resistors for mesh sizes of 1 nm, 0.5 nm and 0.25 nm in both classical and quantum corrected simulations.

The effects of the mesh size on the overall resistance of the simulated slab is shown in Figure 4.11. For an "atomistically" doped resistor the use of NGP charge assignment scheme results in a progressively large increase in the resis-

tance of the slab as the mesh spacing is reduced. With the inclusion of DG quantum corrections not only is most of the charge localisation removed from the simulation but also the mesh dependance of the solution is greatly reduced below 0.5 nm mesh spacing.

4.3 MOSFET simulation

If DG quantum corrections are now included into an "atomistic" MOSFET simulation the average on current for an ensemble of devices will increase, approaching more closely the current obtained in the continuously doped case. Figure 4.12 shows the linear and logarithmic plots of the $I_D V_G$ characteristics of a 50 × 50 nm gate transistor. Results from both continuously doped device and the average of 50 "atomistically" different devices are shown, using purely classical Drift Diffusion simulations and Drift Diffusion simulation with DG quantum corrections. It can be seen that when DG quantum corrections are applied this results in an increase in the current above threshold when compared to the purely classical case.

4.4 DG corrections for Holes

It has been sufficient to outline the problems associated with charge trapping by only considering the case of electrons being trapped within the source and drain regions of a MOSFET, which leads to a increase in the series resistance of the device. However for the accurate "atomistic" simulation of a MOSFET it

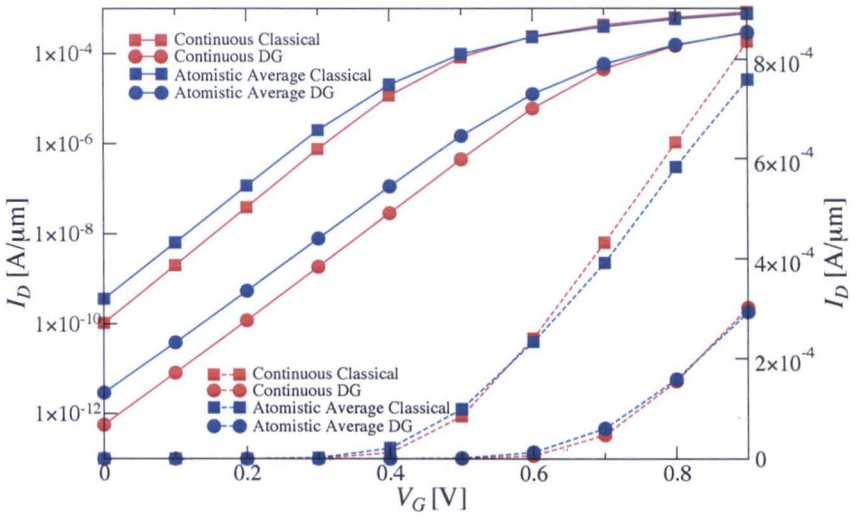

Figure 4.12: I_D/V_G characteristics of a 50×50 nm transistor for both continuous doping and "atomistic" doping, with and without the inclusion of DG quantum corrections.

is also important to consider the trapping of holes in the channel region when "atomistic" acceptors are considered. This trapping leads to a change in both the size and shape of the depletion region. This in turn alters the characteristics of the simulated device by reducing, for example, the threshold voltage.

To overcome this problem DG corrections for holes have to also be considered in the MOSFET. Typically in the DG Drift Diffusion simulation of n-channel MOSFETs it is sufficient to solve, self-consistently, Poisson's equation and the electron current continuity equation adding eventually DG corrections for electrons. However both the trapping of electrons in the source/drain regions and the trapping of holes in the substrate have to be avoided in the "atomistic" simulation and therefore the DG corrections have been introduced for the two types of carriers even when simulating "unipolar" MOSFETs.

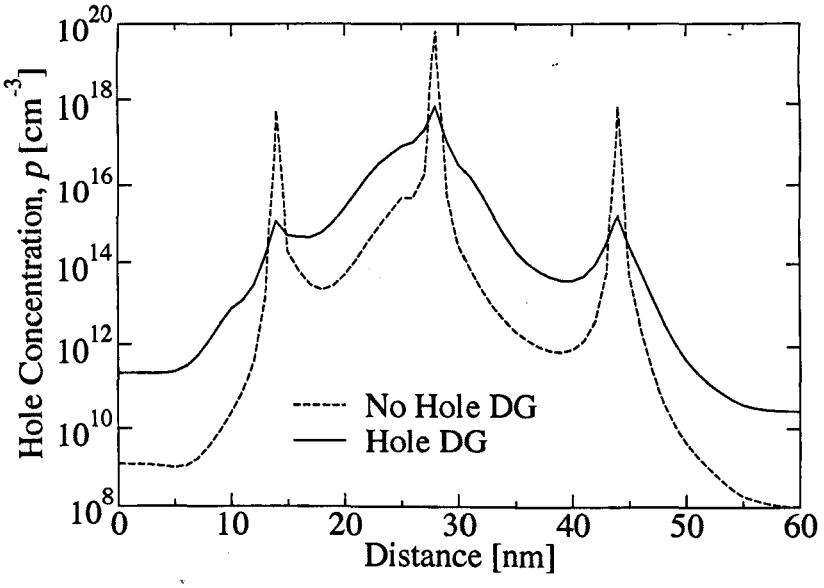

Figure 4.13: The hole concentration along the channel of a 50x50nm transistor, showing both the classical hole concentration due to Boltzmann statistics and the quantum hole concentration generated by the inclusion of Density Gradient quantum corrections.

Figure 4.13 shows the hole concentration underneath the channel of a 50×50 nm MOSFET. It is clear that in the classical case the holes become localised around individual dopant atoms producing high concentration spikes. With the inclusion of DG quantum corrections for holes these high concentration regions are reduced and the resulting concentration in the surrounding regions is increased. This in turn modifies the shape of the depletion region under the gate and alters the threshold voltage.

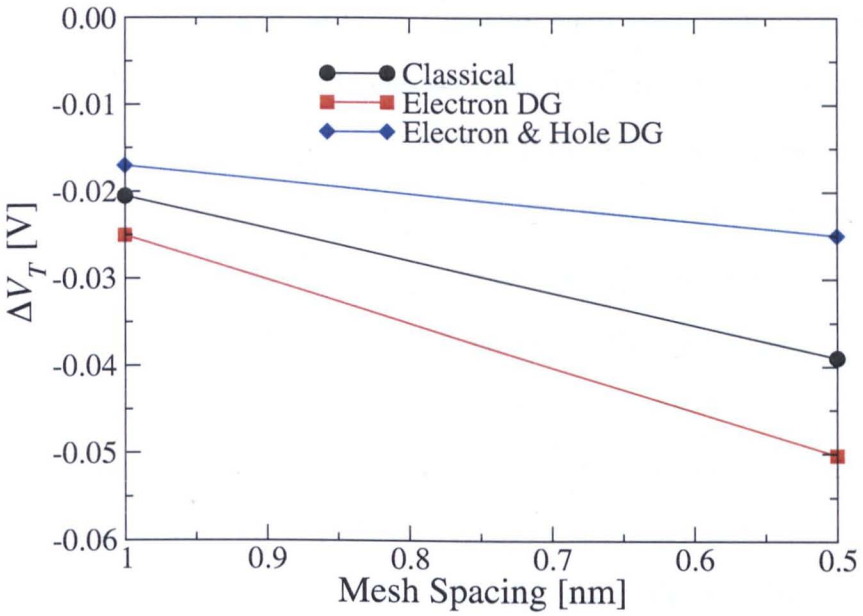

Figure 4.14: Change in threshold voltage for a single atomistic MOSFET device for a mesh spacing of 1 nm and 0.5 nm for three different forms of simulation, classical, electron density gradient and full density gradient

The effects of applying DG quantum corrections to both electrons and holes is illustrated in Figure 4.14. This plot shows the change in threshold voltage for a single "atomistic" configuration of a 50×50 nm MOSFET when a classical, electron only and electron and hole DG quantum corrections are introduced in

the simulations. It can be seen that the mesh size variation in threshold voltage due to artificial charge trapping is not eliminated completely with the inclusion of electron only DG corrections because the charge trapped in the source and drain does not effect the threshold voltage. However with the inclusion of hole density gradient the mesh spacing sensitivity of the threshold voltage due to artificial charge trapping is greatly reduced.

The problem of charge localisation becomes significant if "atomistic" effects are to be studied using standard Drift Diffusion simulation techniques and fine resolution of individual dopants is needed. The corresponding charge localisation alters the characteristics of the modelled MOSFET by increasing the access resistance and reducing the threshold voltage. The removal of this charge localisation can be achieved in two ways. In a purely classical approach this is done by using charge assignment techniques which smear out the charge associated with a single dopant atom. However these techniques do not succeed in removing all of the charge "trapping" associated with a single dopant atom. The higher order schemes which give the best improvement can be arbitrary in their nature and can lead to a loss of resolution for the "atomic" scale effects. Also the various methods of charge assignment do not address the problem of mesh dependance and leave the results strongly coupled to the simulation mesh.

A better approach is to include quantum confinement effects through the use of Quantum Corrections. This approach not only includes the quantum effects into the simulation which is itself beneficial. It also eliminates both the problems associated with charge trapping by reducing the large concentration of carriers in the sharply resolved potential wells, and the problems associated with mesh

dependance removing the strong coupling between the simulation mesh size and the results. It is also possible to expand the scope of these quantum corrections to include not only the electrons in the system but also to include the holes. This reduces the localisation of holes around individual dopant atoms leading to a better model of the depletion region within the channel and more accurate values of the threshold voltage.

However even with the inclusion of quantum corrections it can be seen from both Figure 4.11 and Figure 4.12 that a sizeable discrepancy remains at high voltages. In Figure 4.11 it can be seen that when an voltage of $0.5V$ is applied to the resistor, even with the inclusion of quantum corrections there is an approximately 15% to 20% decrease in current. This is a a sizeable amount of current which would become apparent in the linear regime of a MOSFETs operations. In Figure 4.12 this can be observed in both the classical and quantum corrected case. The reduction is much reduced in the quantum case but is not eliminated. This along with the applicability of the Drift Diffusion formalism implies that we cannot trust the results found above threshold and therefore in the remainder of this work only the sub-threshold and threshold region of the device operation will be studied.

With all of these effects in place and the applicability of our simulations we can now consider the simulation of realistic MOSFETs and can study the effects that intrinsic parameter fluctuations have upon the operation of such devices.

Chapter 5

Real Device Simulation

In the previous chapters, enhancements to the Glasgow "atomistic" device simulator, aiming to allow the simulation of realistic devices been discussed. These improvements include the accurate placement of discrete random dopants based upon the Si crystalline lattice and the implementation of better charge assigment schemes such as Cloud In Cell and Nearest Grid Point.

In addition a more physical and accurate mobility model was implemented taking into account the doping concentration and the field dependence of the mobility. Finally Density Gradient (DG) quantum corrections were introduced for both electrons and holes to avoid trapping of electrons and holes in sharply resolved potential wells associated with the discrete donors and acceptors respectively. The purpose of these improvements was to improve the accuracy and to avoid artefacts associated with the simulation of MOSFETs with realistic structures, obtained for example from a commercial TCAD simulator.

Once reliable models and methodologies were developed and introduced in

the simulator it became possible to focus on the main task of this project which was to study the impact of different sources of intrinsic parameter fluctuations on real devices. In this chapter we describe the benchmark 35 nm MOSFET published by Toshiba [32] and used to study the impact of various sources of intrinsic parameter fluctuations and their combinations. In particular we study the effects of LER, OTF and discrete RD upon the device characteristics in isolation or combined together.

5.1 35 nm Toshiba MOSFET

The purpose of this work is to study the effects of intrinsic parameter fluctuations on realistic devices. This device should be representative of the devices expected to be in production within the following years to give a good indication of the problems associated with the scaling the next generation of device. The International Roadmap for Semiconductors (ITRS) [10] states that by 2006 mass production MOSFET's will have a printed gate length of 40 nm and an actual channel length of 28 nm. Ten years later the gate length will be reduced to 13 nm and 9 nm respectively. At the present time mass production is moving into the 90 nm technology node with a gate length of 37 nm [30, 31]. The detailed structure of such devices will become increasingly important. Near the end of the ITRS they will contain only a few thousand silicon atoms, and intrinsic parameter fluctuations introduced by the discreteness of charge and matter will have an increasingly large impact on the performance of these devices. Research carried out by semiconductor corporations have confirmed that for conventional

MOSFETs nanoscale dimensions are feasible. Nano-CMOS bulk MOSFETs such as the Intel 20 nm MOSFET [24], the Advanced Micro Systems (AMD) 15 nm MOSFET [25] and the NEC 5 nm MOSFET [26] have been published already. Non conventional transistors such as the IBM 6nm gate SOI FET [122] have also been demonstrated.

The device that we have chosen for this work is the 35 nm gate length MOSFET produced by Toshiba [32], this device is of a comparable size with the most advanced devices of the present generation 90nm technology node. This particular transistor has been selected due to the breadth of technological information avaliable from the corresponding publications and through personal contacts. The extraction of the device characteristics, the device structure and the doping profiles as well as the calibration of commercial TCAD tools were performed by Fikru Adamu-Lema and are covered in detail in his PhD dissertation [123]. In this work an outline of the device structure will be presented and the results of the calibration of the device will be shown. More detailed information about the device and the calibration process can be found in [123].

A 3D view of the 35 nm MOSFET produced using Taurus Process^{TM}[113] is shown in Figure 5.1. This general purpose device modelling code was used to deduce with confidence the structure of the 35 nm device and to match its characteristics. The 35 nm MOSFET has shallow As extensions to suppress short channel effects followed by deeper junctions to reduce the access resistance. The As doping concentration in the junctions is shown in red, the poly Si gate is also doped with As. The channel of this 35 nm device has a retrograde indium doping, along with two HALO boron extensions. The HALO doping regions

provide better punch though control without unnecessarily increasing the doping beneath the gate.

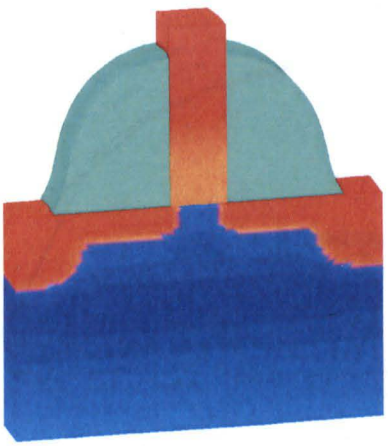

Figure 5.1: 3D potential profile from the Avanti! Taurus Process and Device simulator used to extract the doping and device characteristics of the 35 nm Toshiba MOSFET device.

The structure and the I_D/V_G characteristics of the 35 nm device at both high and low drain voltages were matched using Taurus Process and Device simulator. Figure 5.2 compares the experimental I_D/V_G to the results of a Taurus simulation at low drain voltage of $50mV$. Figure 5.3 compares the experimental I_D/V_G characteristics to the results of a Taurus simulation at high drain voltage of $850mV$. In both cases there is very good agreement between the extracted experimental results and the Taurus simulation as the simulation almost perfectly follows the experimental data. These two results give confidence that the models and corresponding parameters deduced in [123] well represent the structure and the characteristics of the real 35 nm MOSFET. This is a good starting point to our study of the intrinsic parameter fluctuations in such devices with realistic structures.

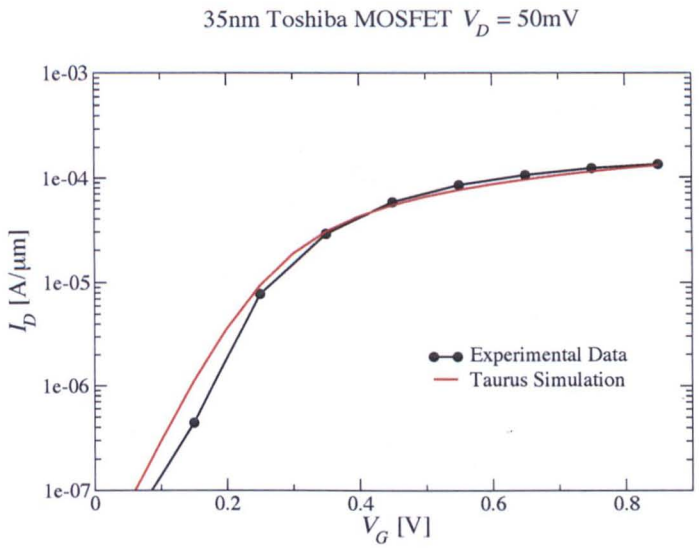

Figure 5.2: I_D/V_G characteristics of the 35 nm Toshiba MOSFET at $V_D = 50mV$, comparing extracted experimental data of the behaviour of the device to that generated by the use of the Avanti! Taurus simulation.

5.1.1 Importing the device structure from Taurus

A complicated doping strategy is used in the 35 nm MOSFET to achieve good device performance characteristics and a reasonable threshold voltage. The inclusion of extensions and HALO doping to improve the device performance make it difficult to use an analytical doping model within the "atomistic" device simulator. Attempts to realistically describe the doping profile using the simple analytical model in the original simulator were unsuccessful and the I_D/V_G characteristics were a poor match to the Taurus results. In Taurus the device simulation is preceded by a comprehensive process simulation which models accurately the key processing steps and produces doping profiles that closely match the experimental measured doping profiles in the real device. Therefore it was de-

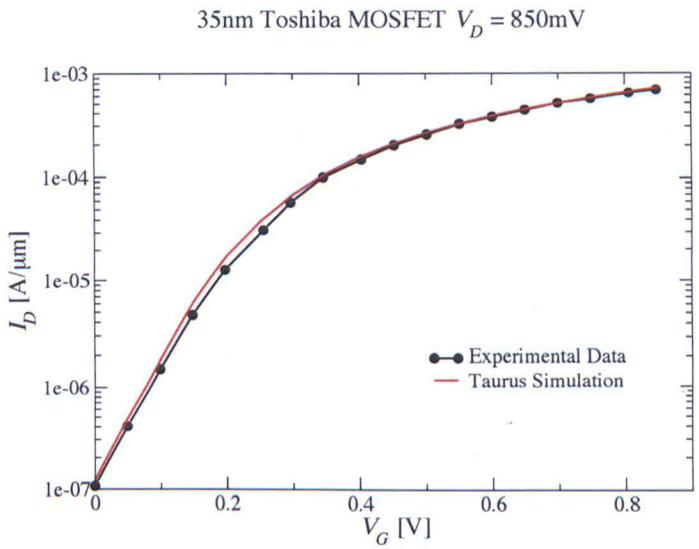

Figure 5.3: I_D/V_G characteristics of the 35 nm Toshiba MOSFET at $V_D = 850mV$, comparing extracted experimental data of the behaviour of the device to that generated by the use of the Avanti! Taurus simulation.

sirable to directly import the doping profile obtained from Taurus Process into the "atomistic" simulator.

There are different methods which can be used to export doping and structural data from Taurus Process into the "atomistic" simulator. One possible solution is to extract the required information from 1D or 2D slices of the device and then to interpolate the doping concentration onto the simulation mesh used by our simulator. The simplest was to use the extract command which provides data for 1D slices as well as single points, in the format:

```
Extract(element, file, points, sample type)
```

Therefore the above command returns the doping at a particular coordinate within the simulation domain. A very simple procedure has been used to pro-

vide a doping profile for each node of the simulation mesh. The first step in the process was the development of a sub-program that returns the list of all the co-ordinates of the nodes within the simulation mesh which require the specification of a doping concentration. A section of sample output is given below.

```
....
point(x= 45.0nm,y=   .0nm),
point(x= 45.0nm,y=   .5nm),
point(x= 45.0nm,y=  1.0nm),
....
```

The point keyword from the Taurus extract command syntax is part of this output to allow the list of points to be more easily merged into the Taurus input file. Once the simulation mesh had been generated and a list of the nodes of that mesh had been created this can be fed into a simple shell script which creates the desired Taurus Input file from an adapted base file. The shell script for an indium extraction would look like:

```
# create new base file for indium extraction
cat toshiba_base.pdm > profile.pdm
echo "Extract(only(indium)," >> profile.pdm
echo 'extractfile="indiumpoints.dat",' >> profile.pdm
echo "points(" >> profile.pdm
cat points.dat >> profile.pdm
echo ")," >> profile.pdm
echo "sample=values)" >> profile.pdm
```

```
echo "Stop () " >> profile.pdm
# runs taurus extraction for indium,boron,arsenic
/apps/avanti taurus profile.pdm
# strips name information from files
sed s/'indium '//g indiumpoints.dat | sed s/' cm-3'//g > indium.dat
```

The list of mesh points is stored in a file called `points.dat` and is entered at the appropriate position in the `extract` command of the Taurus input file. The file to which the extracted points are to be saved is labeled `indiumpoints.dat`. Once the file `profile.pdm` has been constructed to include the `extract` command, the shell script then executes Taurus Process. As part of the process simulation an output file, `indiumpoints.dat`, is produced containing the indium doping at the required mesh points. This file is then passed through the `sed` Unix command to strip off the extraneous information and to leave a list of doping concentrations in a file called `indium.dat`. The same process can be followed to obtain the doping profiles within the device for the arsenic implantation and the boron implantation. Once the data for all doping species has been obtained it is read into the "atomistic" simulator and used as the doping profile instead of an analytical profile. This method of transferring the doping within the 35 nm MOSFET from Taurus Process results in much better agreement with the simulated output from Taurus and hence the realistic device characteristics were obtained.

Parameter	Value
$\mu_{n_{min}}$	$55.24cm^2/Vs$
$\mu_{n_{max}}$	$1429.23cm^2/Vs$
N_{refn}	$1.072 \times 10^{17} atoms/cm^3$
α_n	0.73
Ecn	$4 \times 10^5 V/cm$
v_n^{sat}	$1.1 \times 10^7 cm/s$
βn	2

Table 5.1: List of the parameters used in the combined Analytical and Caughey-Tomas mobility model used in the simulation of the 35 nm MOSFET device, along with the values used in the simulation of the 35 nm device

5.1.2 Calibration

Once the doping profile for the 35 nm MOSFET device has been transferred into the "atomistic" simulator, it can be used to calibrate its results in respect of the Taurus device simulation, which has already been calibrated to the experimental data. The reason for calibrating our simulation to the Taurus simulations instead of directly to the experimental data is the inclusion of external resistances in the Taurus simulations to adjust the above threshold current to match the experimental data. Our "atomistic" device simulator does not have the ability to include external resistances and therefore cannot match well the experimental on current values. After the calibration of the Taurus simulation to the experimental device characteristics the external resistances where removed from the simulation to allow the "atomistic" simulator to be calibrated.

In the Taurus simulation of the 35 nm MOSFET an analytical concentration dependent model coupled with the Caughey-Tomas mobility model [111], outlined in Chapter 3 was used. The parameters for the concentration, per-

pendicular field and lateral field dependence of the mobility are shown in Table 5.1. The same parameters where used in the "atomistic" simulator to match the mobility by Taurus. Another important parameter in the simulation was the permativity, ϵ_0 of the gate oxide. The 35 nm Toshiba MOSFET has as a gate insulator OxyNitride which has a greater permittivity than SiO_2. The exact value of the OxyNitride permittivity is not know but values varying from 3.5 to 7 are accepted in the literature. In the calibration of the Taurus simulation a value of 5.45 was used. For consistency the same value has been adopted in this work.

The results of the calibration of the "atomistic" simulator are shown in Figures 5.4 and 5.5. In these simulations both the "atomistic" simulator and Taurus utilise continuous doping to allow a fair comparison. Figure 5.4 shows the I_D/V_G characteristics of the 35 nm device at $V_D = 100mV$ obtained from the "atomistic" simulator and Taurus without any series resistance. The good agreement between both I_D/V_G curves grants confidence that the "atomistic" simulator is able to replicate the behaviour of the 35 nm device at low drain biases. In Figure 5.5, the I_D/V_G characteristics of the 35 nm device at $V_D = 850mV$ are compared. Despite a slight shift in the subthreshold slope of the "atomistic" simulator the overall agreement between the I_D/V_G characteristics at high drain voltage is also remarkably good.

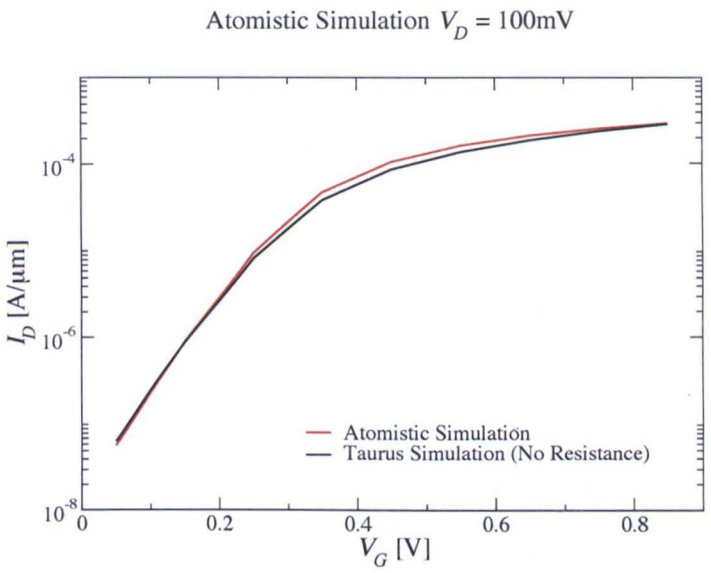

Figure 5.4: I_D/V_G characteristics of the two 35 nm models, comparing the Avanti! Taurus model to the Glasgow "Atomistic" simulator at $V_D = 100mV$.

5.2 Effects of Intrinsic Fluctuations on Device Characteristics

The "atomistic" device simulator can be used to study the effects of various sources of intrinsic parameter fluctuations on the behaviour of MOSFET's. The sources of fluctuations that are included at present in the device simulator are random discrete dopants, line edge roughness and oxide thickness fluctuations. In this section we study the effect of each one of these sources on the average "atomistic" devices characteristics as well as the statistical distribution of the threshold voltage, off-current and on-current.

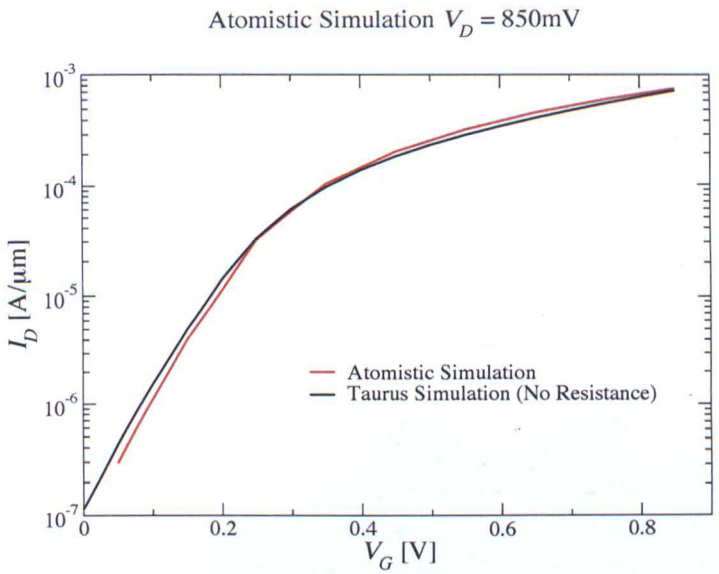

Figure 5.5: I_D/V_G characteristics of the two 35 nm models, comparing the Avanti! Taurus model to the Glasgow "Atomistic" simulator at $V_D = 850mV$.

5.2.1 Random Discrete Dopants

When MOSFETs are fabricated dopant atoms will be implanted into the existing bulk silicon lattice. The microscopic distribution of discrete individual dopant atoms affects the electrostatics and transport within the individual transistor and results in variation in the device characteristics and performance. Figure 5.6 illustrates the 3D potential distribution in the 35 nm Toshiba MOSFET. Random dopants are introduced in the inner region and continuous doping is used in the outer regions to facilitate the handling of boundary conditions. Both the heavily doped source and drain regions as well as the extension regions are clearly visible. The potential fluctuations in the "atomistic" part of the device contrast the smooth potential distribution in the continuously doped

regions. The random dopants also alter the shape of the depletion region as the edges of this region become rough due to various local dopant positions. All of these create a more complicated potential profile than would original exist in a continuously doped transistor.

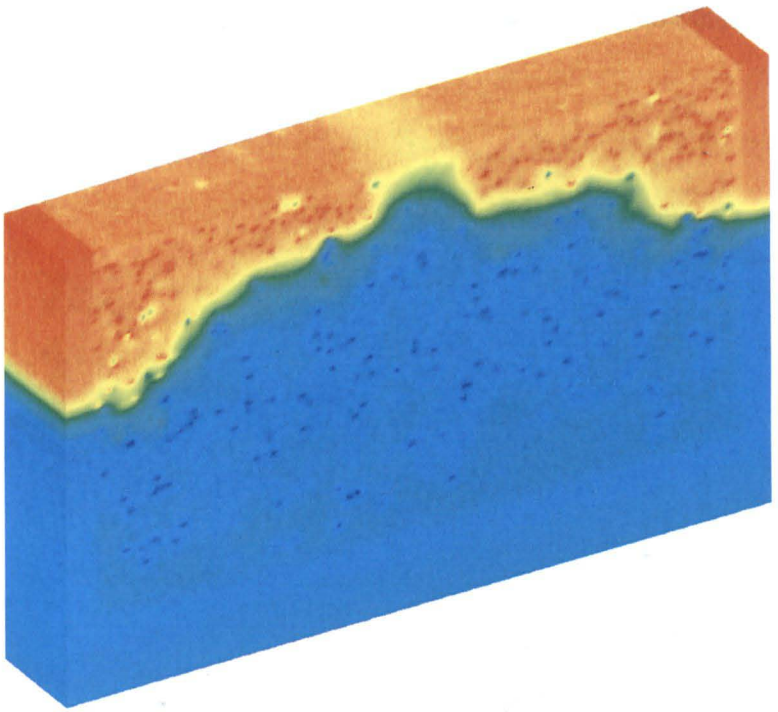

Figure 5.6: 3D "atomistic" potential profile of the Toshiba 35 nm MOSFET, the "atomistic" nature of both the source/drain regions and the channel region can clearly be seen as well as the varying potential across the channel due to random dopants. The simulation domain is $240 \times 35 \times 175$ nm.

Figure 5.7 displays the individual positions of the discrete random dopants within the same transistor. At close inspection the retrograde Gaussian doping profile within the channel region can be seen as an increasing density of individual dopants beneath the depletion region. The contour plot overlayed on the individual dopants represents an electron equi-concentration contour. A con-

ducting path connecting the source and drain is visible. Such random conducting paths alter the performance of a MOSFET. If a path is formed the device will turn on more quickly than in the continuous doping case. If the channel region is uniformly blocked by dopant atoms, a conducting path will form more slowly and at higher gate voltage resulting in an increase in the threshold voltage.

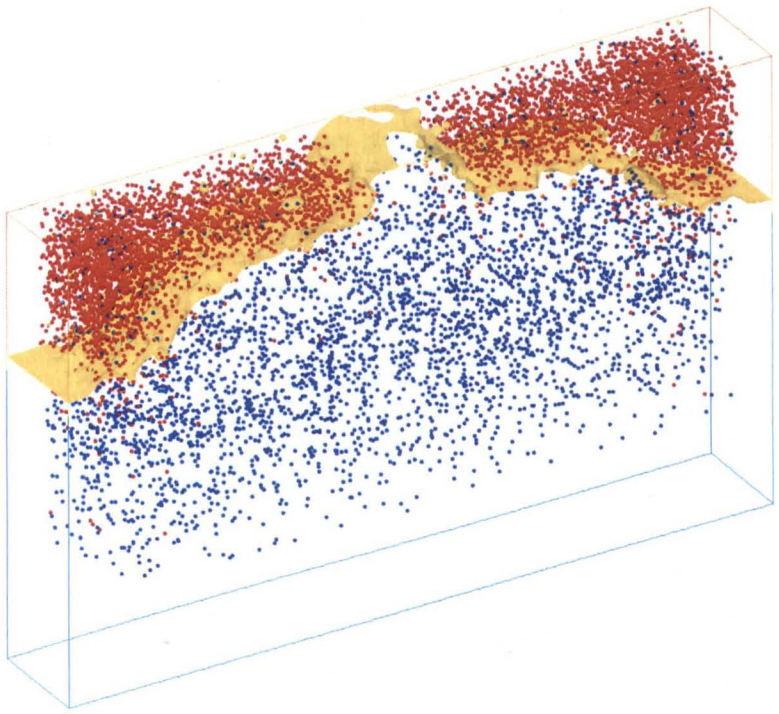

Figure 5.7: 3D dopant profile of the Toshiba 35 nm MOSFET showing the position of the individual dopants within the MOSFET and a concentration of electrons across the channel and within the source/drain regions of this device. The simulation domain is $240 \times 35 \times 175$ nm.

Figure 5.8 shows the I_D/V_G characteristics of 200 microscopically different "atomistic" MOSFET simulations. In each of these simulations the individual dopants are placed at random points in the crystal lattice based on a rejection technique. The uniformly doped device is shown as a reference and the spread

in the characteristics of the various "atomistic" devices can clearly be seen. All simulations are performed at low V_D because Drift Diffusion does not capture well the non-equilibrium transport effects in such short gate length devices. The relative magnitude of the fluctuations is large in the subthreshold region and is reduced with the increase of the gate voltage above threshold. It is also important to note that the "atomistic" average does not perfectly match the simulated characteristics of a continuously doped device. The average subthreshold slope is shifted and the average on-current is reduced.

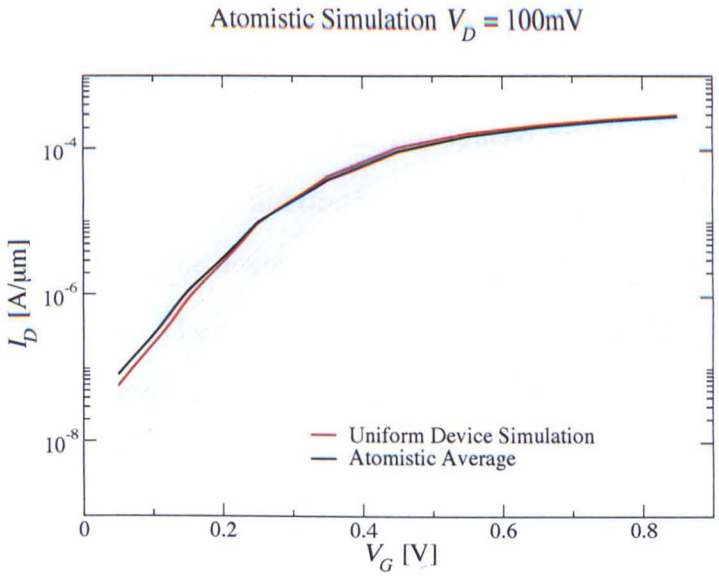

Figure 5.8: The I_D/V_G characteristics of 200 "atomistic" 35 nm devices produced by the "atomistic" simulator. The uniformly doped device is shown as a reference along with the average "atomistic" device.

A more detailed analysis of the differences between the "atomistic" average and the uniformly doped characteristics is shown in Figure 5.9 which depicts the percentage difference between the uniformly doped device and the "atom-

istically" doped device.

This comparison is important because up to now continuously doped devices were the basic of TCAD simulations in device design. It can be seen from Figure 5.9 that below threshold there is a 30% increase in the average "atomistic" current compared to the continuous doping characteristics complemented by a slight increase in the subthreshold slope. This is due to early percolation of current through valleys in the potential landscape and results in an overall increase in the subthreshold leakage. The difference is reduced with the increase of the gate voltage until it crosses over around threshold at which point a reduction in the average "atomistic" current is observed. This behaviour can be understood in terms of potential fluctuations introduced by the individual dopants. Below threshold the potential fluctuations associated with the discrete dopants are unscreened allowing the device to turn on more quickly by creating conducting percolation paths. The effect is reduced above threshold when the mobile carriers in the inversion layer start to screen the potential fluctuations with an increasing efficiency. Also above threshold the series resistance of the "atomistically" doped source and drain exceeds the series resistance of the continuously doped device resulting in a relative reduction in the average current.

Another important parameter in the operation of a MOSFET is the threshold voltage, V_T. In the case of the uniformly doped device the threshold voltage deduced from the simulations was $132mV$. The average threshold voltage obtained from the "atomistic" simulations was $133mV$, which indicates a very small positive shift in threshold voltage. This observation, however depends on the threshold voltage criteria. The lack of threshold voltage lowering in the

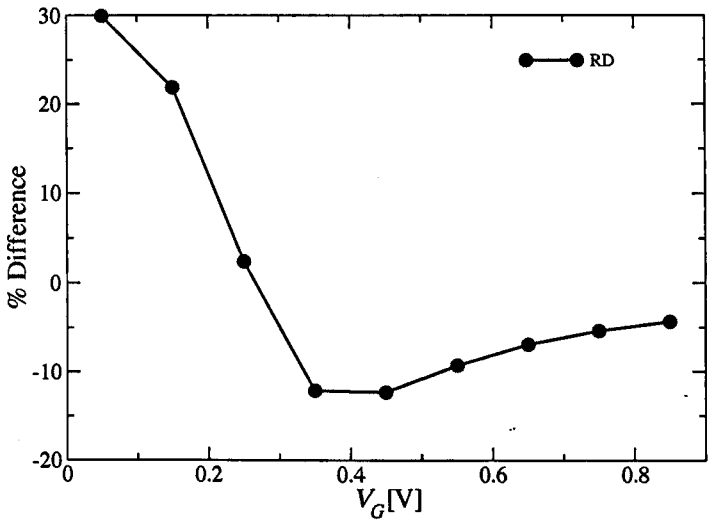

Figure 5.9: The percentage difference between the "atomistic" average I_D/V_G characteristics and the uniformly doped I_D/V_G characteristics.

Toshiba MOSFETs can be attributed to the channel doping design including the retrograde doping and the inclusion of pockets. Figure 5.8 shows that despite the minor shift in the average V_T there is still a significant threshold voltage spread. To better characterise the spread statistically we need to analyse the distribution of the threhold voltages obtained in the "atomistic" simulations.

Figure 5.10 shows a histogram of the "atomistic" threshold voltages, which indicates close to a Gaussian distribution which has also been observed previously [18]. The calculated standard deviation of this distribution is $33.2mV$ which is a 24.8% deviation from the average. For Gaussian distributions two thirds of the samples are found within one standard deviation of the mean and almost all of the samples are found within three standard deviations. Therefore 67% of the devices will have a threshold voltage within the range of 24.8% of

the average. The threshold voltage is directly related to the leakage current via the subthreshold slope and its variation will also impact the amount of leakage current within a sample of devices.

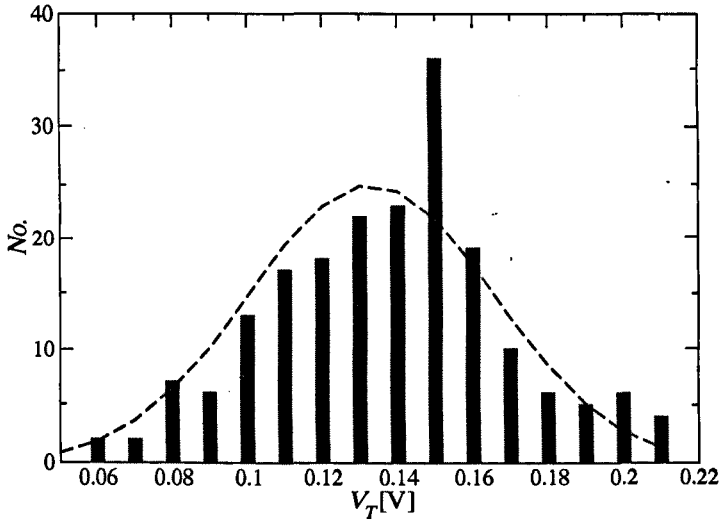

Figure 5.10: Histogram of the distribution of threshold voltages among 200 "atomistic" 35 nm MOSFET's

The off-current which is the leakage current for a gate voltage of 0V is an important parameter because it determines the minimum amount of standby power that a device will consume. This is particularly important in low power and embedded systems where the amount of standby current has to be reduced to allow longer operation. For the 35 nm Toshiba MOSFET the continuously doped simulations produce an off-current of $1.46 \times 10^{-8} A/\mu m$ and the "atomistic" average an off-current of $2.2 \times 10^{-8} A/\mu m$. This gives a 50% increase in leakage current due to the inclusion of random dopants in the simulations.

To understand further the effect of the random dopants upon the leakage

current in the 35 nm MOSFET it is useful to carry out statistical analysis. However we should take into account that the distribution of the "atomistic" leakage currents, shown in Figure 5.11, is not normal. This can be understood if we consider that on a semilogarithmic scale the subthreshold current is close to a straight line having a constant gradient which determines the subthreshold slope. Therefore a normal distribution of the threshold voltage will transfer to a log-normal distribution of the off current.

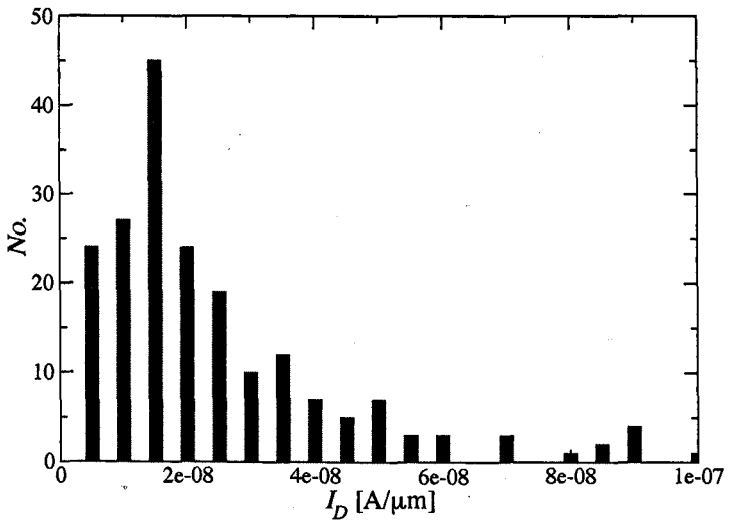

Figure 5.11: Histogram of the distribution of drain current with zero applied bias to the gate, of 200 "atomistically" different devices.

For the log-normal distribution the standard deviation is no longer a sufficient description of the shape of the distribution. To give a better indication for the shape of the distribution we calculate the skew and the kurtosis which are related to the higher order moments of the distribution. The skew of a distribution is a measure of the symmetry of the distribution around a mean value. Positive

values for a skew imply that the distribution is shifted to the right of the mean, while negative values imply that the distribution is shifted to the left. The kurtosis of a distribution is a measure of the shape of a distributions peak. For positive values of kurtosis a distribution would have a sharp peak, while negative values would imply a flat peak. The value of kurtosis for a normal distribution is 3. The kurtosis minus 3 is often used as a measure for the degree of deviation from the normal distribution. Figure 5.12 shows the percentage standard deviation of the current distribution as a function of the gate voltage. Below threshold the percentage standard deviation is large and is reduced with the increase in the gate voltage.

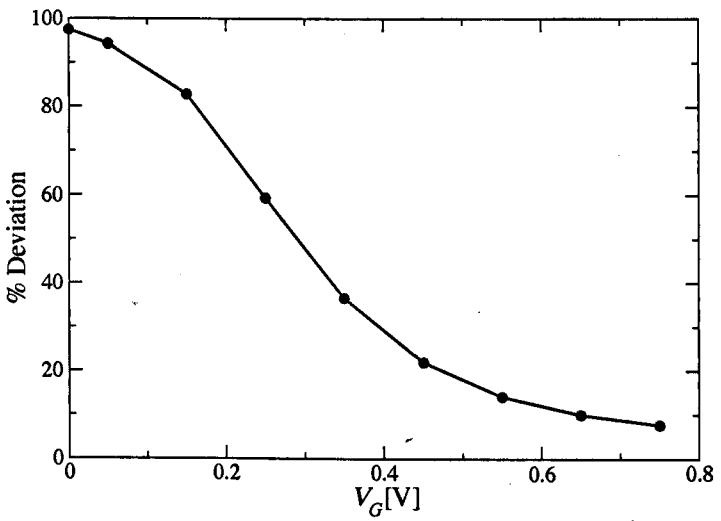

Figure 5.12: The percentage deviation of the I_D/V_G distrubutions of the "atomistic" 35 nm device, shown for all measured gate voltages between 0V and 0.85V

The gate voltage dependance of the skew, illustrated in Figure 5.13, indicates a very slight negative skew above threshold implying that the distribution is close

to normal and is shifted slightly to the left. Below threshold the the distribution is skewed excessively to the right. This would mean that most of the I_D/V_G curves below threshold are distributed below the mean curve.

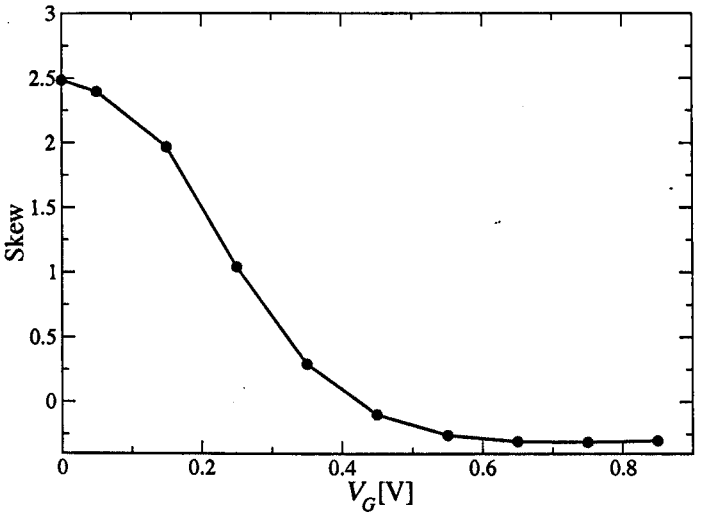

Figure 5.13: The skew of the I_D/V_G distrubutions of the "atomistic" 35 nm device, shown for all measured gate voltages between 0V and 0.85V

The gate voltage dependance of the excessive kurtosis of the I_D/V_G curves is illustrated in figure 5.14. Similar to the gate voltage dependance of the skew above threshold the kurtosis indicates close to a normal distribution with a very small kurtosis value. However below threshold this value increases with the reduction of the gate voltage implying that the distribution at this point is sharply peaked and that a large number of the I_D/V_G curves are clustered near to the mean value.

Although the percentage standard deviation, skew and kurtosis describes the distribution fully they are not very easy to understand particularly for use in

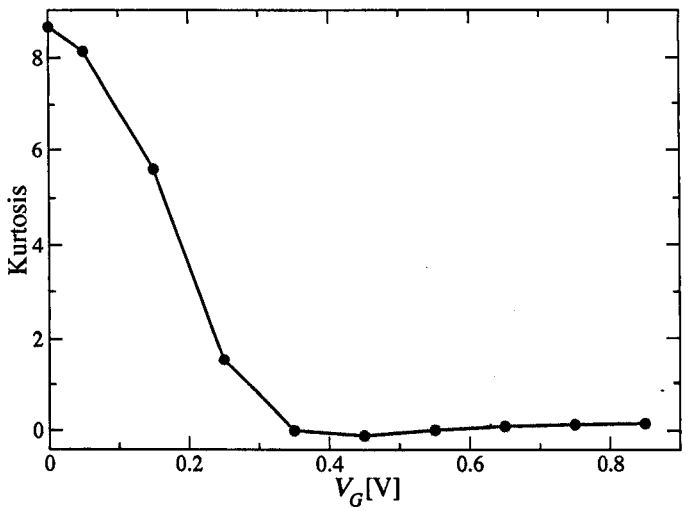

Figure 5.14: The kurtosis of the I_D/V_G distrubutions of the "atomistic" 35 nm device, shown for all measured gate voltages between 0V and 0.85V

higher level design. A better approach for quick extraction of information from the I_D/V_G distributions in the subthreshold region is to use the distribution of $log_{10}(I_D)$ instead of I_D. A histogram of the off-current distribution is shown in Figure 5.15. The distribution is close to a normal distribution confirming the log-normal behaviour of the distribution in the subthreshold region. Such a distribution is sufficiently well described by the standard deviation making the use of skew and kurtosis redundant. For the distribution in Figure 5.15 the standard deviation of the $log_{10}(I_D)$ is 0.39. Since $10^{0.39} \approx 2.5$ the standard deviation of distribution is two and a half times the mean value, or has a value of $8dB$. This figure is a far more meaningful measure of the distribution indicating that nearly all of the devices will fall in the range of approximately 16 times the mean value or $24dB$.

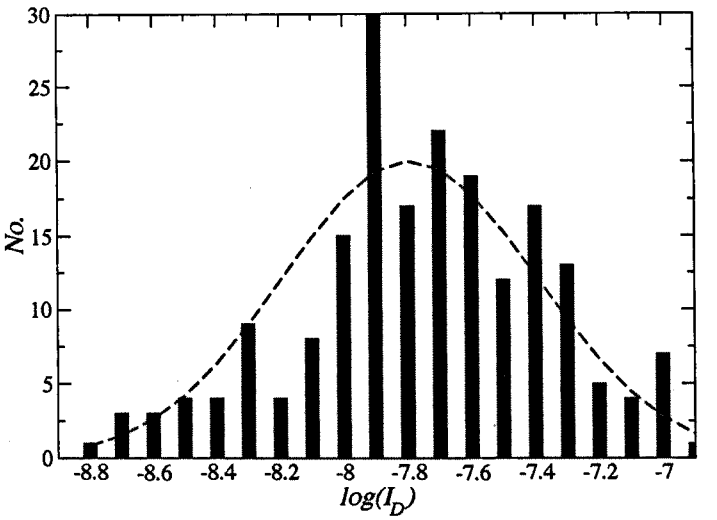

Figure 5.15: Histogram of the distribution of the log of the drain current with zero applied bias to the gate, of 200 "atomistically" different devices.

5.2.2 Line Edge Roughness

Line edge roughness (LER) is caused by the molecular polymer aggregates found within the resist material. As the large polymer aggregates dissolve at different rates than the surrounding material their impression can be left in the sidewalls of the gate edge causing a rough edge. In conventional simulations the gate edge is assumed to be straight however the rough edges caused by this processing step can lead to large fluctuations in the MOSFET performance. Figure 5.16 shows the 3D electron concentration distribution of a square 35 nm MOSFET when LER is included in the simulations. Square or small W/L ratio devices are often used in SRAM cells and because of their small size will be the first to suffer from the effects of intrinsic parameter fluctuations. The small size of the MOSFET used in the simulation makes the effects of LER more prominent.

There will be less parameter variations in a wider device because most of the LER effects would self-average.

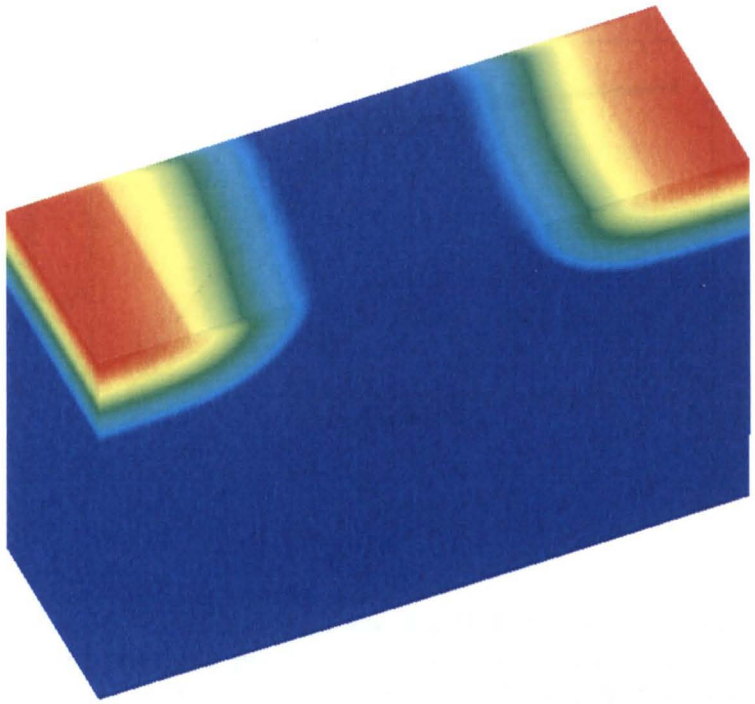

Figure 5.16: 3D electron concentration profile of the Toshiba 35 nm MOSFET with applied gate edge roughness. The roughness can be seen along the edge of the channel region of the device shortening the effective channel length. The simulation domain is $107 \times 35 \times 75$ nm, with the values for electron concentration on a logarithmic scale varying from $1 \times 10^{14} cm^{-3}$ to $2 \times 10^{20} cm^{-3}$.

The simulated I_D/V_G characteristics of 200 different devices in the presence of LER are shown in Figure 5.17. Each of these devices has a randomly generated line for each side of the gate edge to simulate the effects of LER. A correlation length of 30nm and an rms amplitude of 2nm is used in these simulations to describe the LER. The averaged I_D/V_G curve together with the characteristics corresponding to a device with straight gate edges are shown in the same figure. Compared to the similar results associated with random discrete dopants in

Figure 5.8 a larger spread is found above the uniform device characteristics.

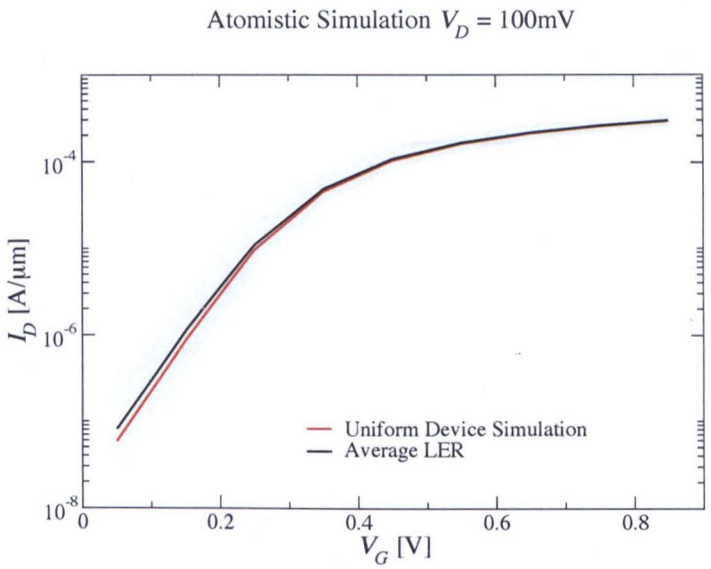

Figure 5.17: The I_D/V_G characteristics of 200 35 nm devices with applied line edge roughness to the gate produced by the "atomistic" simulator. The uniformly doped device is shown as a reference with straight line edges compared to the average of the LER devices.

The percentage difference between the average current in the LER devices and the uniform device is shown in Figure 5.18. Similar to the case of random discrete dopants the average LER curve is not identical to the one from the transistor with straight gate edges. A larger difference reaching 30% is observed in the subthreshold region. The magnitude of this difference is comparable to the effects that have already been observed in the case of random discrete dopants. In the case of LER the sign of the difference does not change and the magnitude increases with the increase of the gate voltage. The effects of LER become less pronounced above threshold. This result indicates that the LER

by locally shortening the physical channel length reduces the average threshold voltage. This is more pronounced in our simulations due to the small width of the transistor which is almost comparable to the correlation length used.

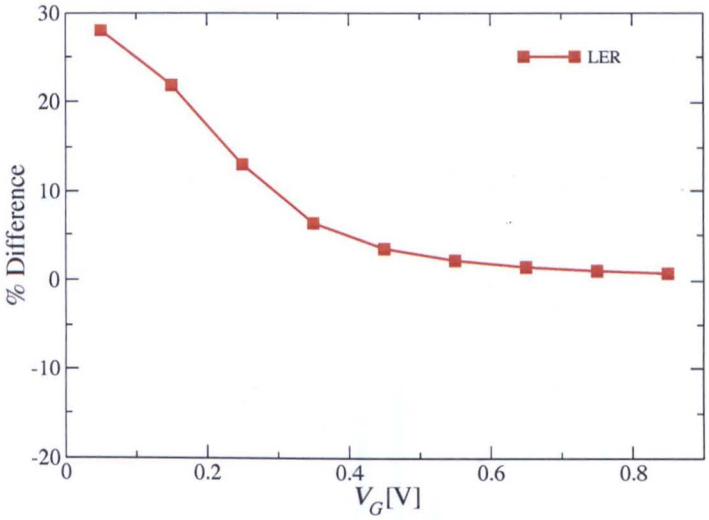

Figure 5.18: The percentage difference between the LER average I_D/V_G characteristics and the uniformly doped I_D/V_G characteristics.

The difference between the average device characteristics and the ideal straight edge device results in a threshold voltage shift. The threshold voltage of the ideal device is $132mV$ and the average threshold voltage of the applied LER devices is $126mV$ which corresponds to a 4.5% shift in the threshold voltage. The distribution of threshold voltages for 200 different 35 nm devices simulated with LER is shown in Figure 5.19, which has close to a normal distribution. The standard deviation of the threshold voltage is $19mV$ which is a 15% deviation from the average. Most of the devices will have a threshold voltage in the region of 3σ with two thirds falling in the region of one standard deviation. The

variation in the threshold voltage together with its average shift to lower gate voltages will have a negative impact on the leakage current within the device. The leakage current of the ideal device is $1.46 \times 10^{-8} A/\mu m$ whereas the average leakage current of the 200 devices is $2.12 \times 10^{-8} A/\mu m$ leading to a 45.5% increase in the leakage current. This increase is comparable with that associated with random discrete dopants. Such a difference when compared to the ideal device advocated the need to consider the sources of intrinsic parameter fluctuations in the TCAD design.

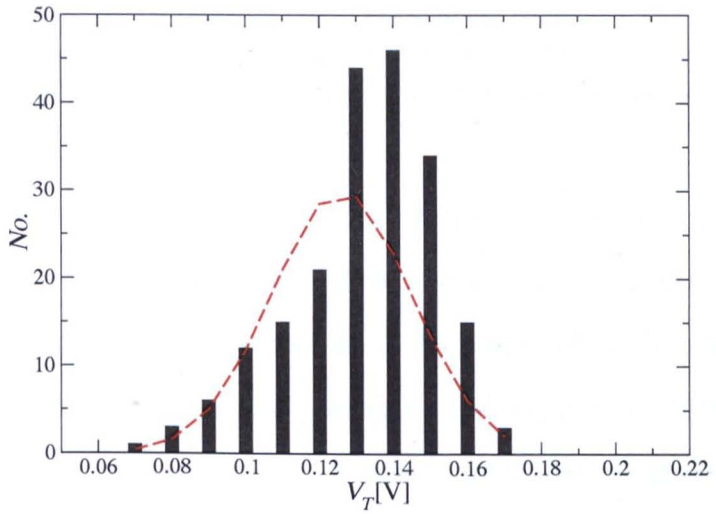

Figure 5.19: Histogram of the distribution of threshold voltages among 200 different 35 nm MOSFET's, with randomly applied line edge roughness.

Figure 5.20 shows the distribution of the off-current for 200 individual different devices simulated with LER. Similar to the case of random discrete dopants the distribution is no longer normal but log-normal. The standard deviation does not completely describe such a distribution and the skew and kurtosis should also

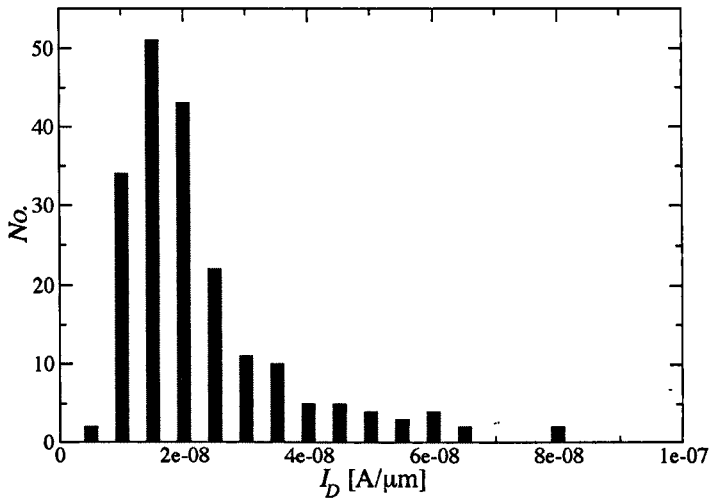

Figure 5.20: Histogram of the distribution of Drain current with zero applied bias to the gate, of 200 devices with different, randomly applied line edge roughness.

be evaluated. Figures 5.21, 5.22 and 5.23 illustrate the gate voltage dependence of the percentage standar deviation, skew and kurtosis. Similar to the case of random discrete dopants below threshold the standard deviation is large and is reduced with the increase in gate voltage. The magnitude of the percentage standard deviation associated with LER is smaller when compared to the case of random discrete dopants. Whereas for random discrete dopants the maximum standard deviation corresponding to $V_G = 0V$ was 90% for line edge roughness it reaches approximately 70%.

The skew for the distributions is shown in Figure 5.22. Again similarly to the case of random discrete dopants below threshold the distribution is shifted to the right of the mean. However unlike the case of random discrete dopants where above threshold there was a slight negative skew, in the case of LER a

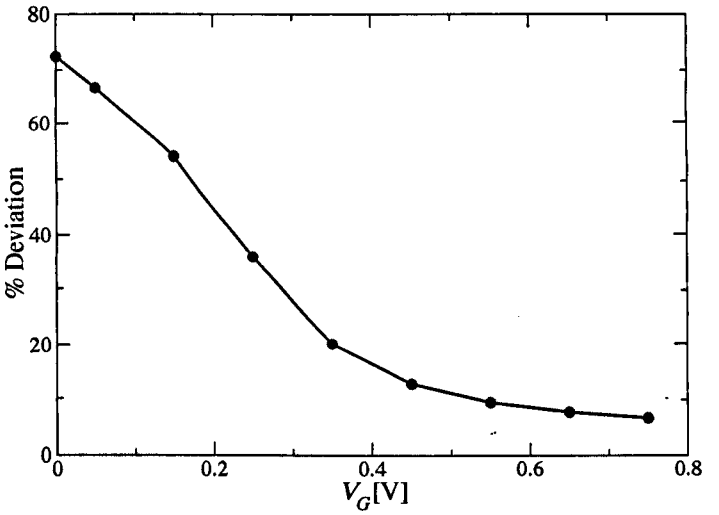

Figure 5.21: The percentage deviation of the I_D/V_G distrubutions of 200 devices with randomly applied line edge roughness, shown for all measured gate voltages between 0V and 0.85V

reasonably large positive skew remains above threshold altering the shape of the Gaussian profile. This is most likely due to the variation in the channel length which cannot be screened by the gate voltage.

The gate voltage dependence of the excessive kurtosis is shown in Figure 5.23. Similar to the case of random discrete dopants above threshold any change in the shape of the peak is negligible and the distribution remains close to normal. However below threshold the peak of the distribution becomes sharper as the distribution begins to cluster around the mean value. .

A better way of statistically describing the distribution of the I_D/V_G curves is to use $log_{10}(I_D)$. A histogram of this distribution is shown in Figure 5.24. The $log_{10}(I_D)$ distribution similarly to the case of random discrete dopants is close to normal, allowing only the standard deviation to be used as a reasonable

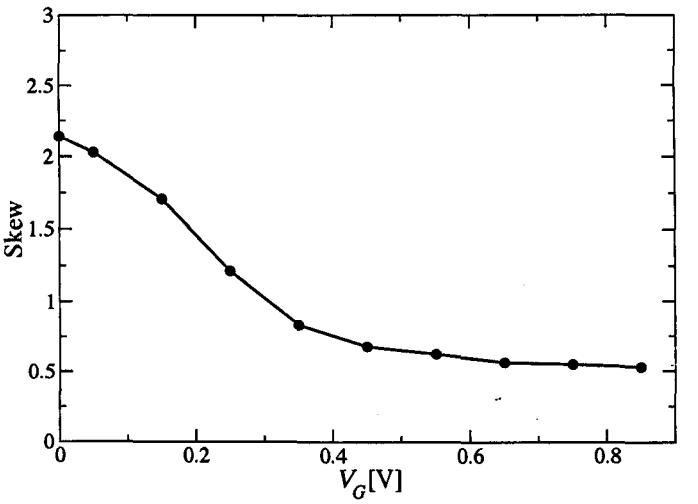

Figure 5.22: The skew of the I_D/V_G distrubutions of 200 devices with randomly applied line edge roughness, shown for all measured gate voltages between 0V and 0.85V

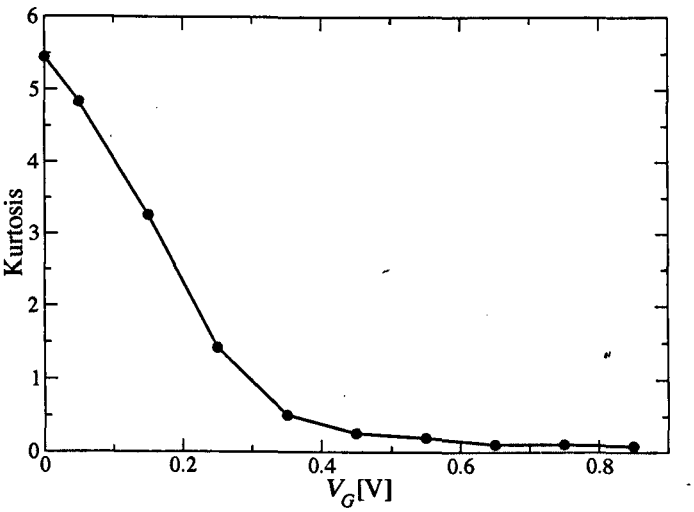

Figure 5.23: The kurtosis of the I_D/V_G distrubutions of 200 devices with randomly applied line edge roughness, shown for all measured gate voltages between 0V and 0.85V

statistical measure. For the log-normal distribution the standard deviation of the $log_{10}(I_D)$ is 0.25 since $10^{0.25} \approx 1.75$ the standard deviation of the distribution is 1.75 times the mean value, or has a value of approximately $5dB$. This figure is a far more meaningful measure of the distribution. Three standard devations implies that nearly all of the devices will fall in the range of approximately 5.4 times the mean value or $14.6dB$.

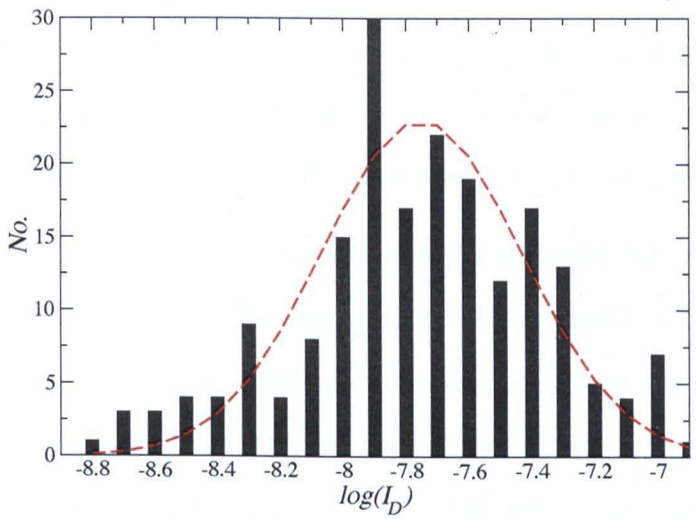

Figure 5.24: Histogram of the distribution of the log of the drain current with zero applied bias to the gate, of 200 devices with different, randomly applied line edge roughness.

5.2.3 Oxide Thickness Fluctuations

Interface roughness and the corresponding oxide thickness fluctuations are introduced during oxidation. They are most probably related to the strain at the interface as the amorphous oxide material pushes aside the silicon lattice. SiO_2

grows by eating away the silicon surface and intruding down into the bulk silicon. In the case of the 35 nm MOSFET, OxyNitride is used as a gate dielectric which has a larger dielectric constant than that of SiO_2. Since the OxyNitride is fabricated by nitrification of the thermally grown SiO_2, to describe interface roughness we use the well documented interface roughness parameters for SiO_2. Figure 5.25 shows the electron concentration distribution in a 35 nm MOSFET with interface roughness included in the simulations. A correlation length of 1.8 nm and an rms amplitude of 0.3 nm is used in these simulations to describe the LER. The interface roughness can be clearly seen, with the blue regions representing the areas where the oxide protrudes into the bulk silicon.

The I_D/V_G characteristics of 100 devices with different interface roughness are shown in Figure 5.26. The average value of the current is shown along with the current obtained from the simulation of a device with a flat interface. It is immediately obvious that the spread in the characteristics is smaller compared to the case of random dopants and line edge roughness. This is due to the small correlation length of the interface roughness compared to the device dimensions resulting in self averaging in the interface roughness induced variations. Also the variations are bounded between the maximum and the minimum thickness of the oxide which differ by 0.3nm,

The percentage difference between the average of the OTF devices and the ideal device is show in Figure 5.27 along with the percentage difference for LER and RD. In the case of OTF we obtain a similar result to that of LER, the effects of the random OTF acts to increase the leakage current. Below threshold there is approximately a 35% increase in the average value of current. This due to local

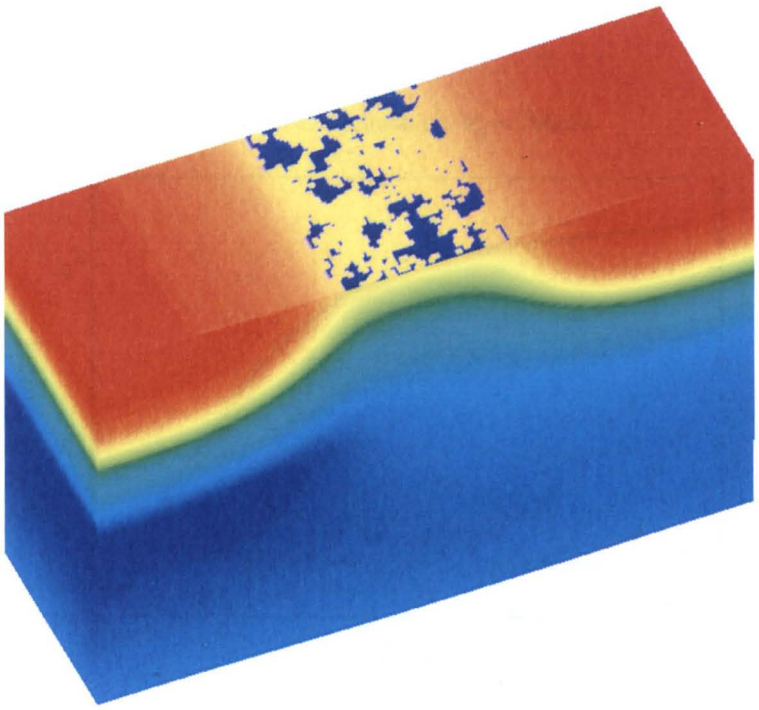

Figure 5.25: 3D electron concentration of a 35 nm MOSFET with applied oxide thickness variations, the blue regions within the channel of the MOSFET are areas where the oxide protrudes into the bulk silicon and removes electrons. The simulation domain is $107 \times 35 \times 75$ nm, with the values for electron concentration on a logarithmic scale varying from $1 \times 10^{14} cm^{-3}$ to $2 \times 10^{20} cm^{-3}$.

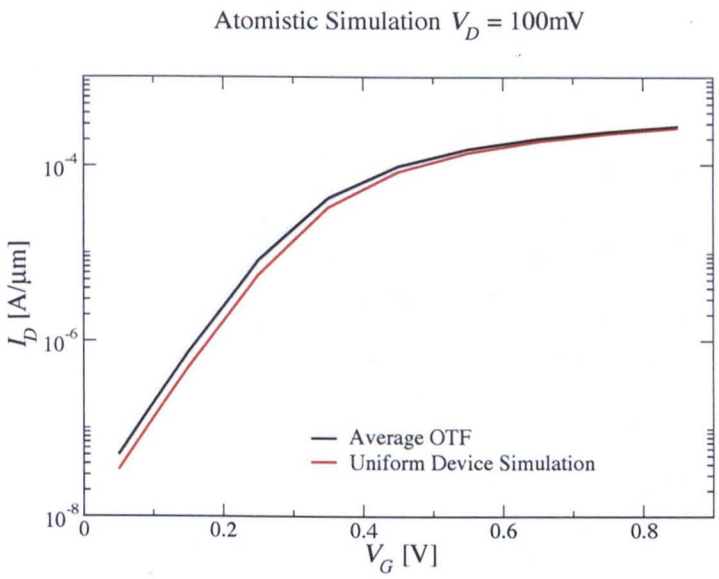

Figure 5.26: The I_D/V_G characteristics of 100 35 nm devices with different random rough interfaces. The uniformly doped device is shown as a reference with no oxide thickness variations.

areas that have a reduced oxide thickness which increases the amount of leakage current due to a increased local capacitance. Above threshold the fluctuations in the oxide thickness has less pronounced effect on the device characteristics due to the well developed inversion layer everywhere.

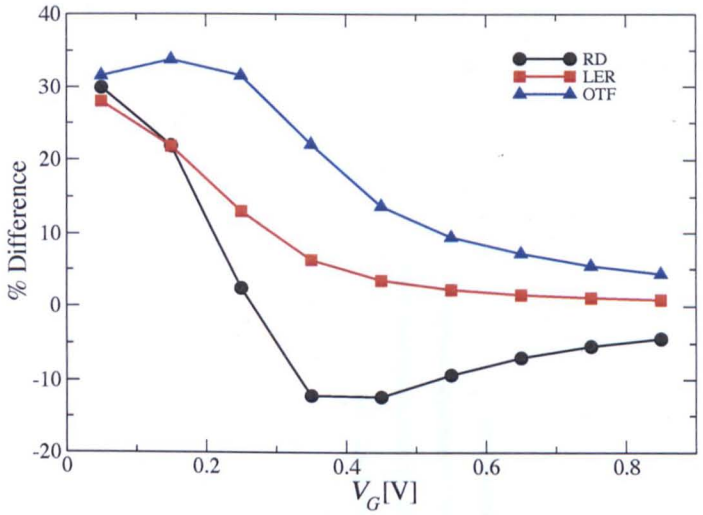

Figure 5.27: The percentage difference between the OTF, LER and RD average I_D/V_G characteristics and the uniformly doped I_D/V_G characteristics.

The threshold voltage of the ideal device is $132mV$ and the average threshold voltage of the collection of devices with applied OTF is $122mV$ which is a 7.5% reduction in the threshold voltage. The distribution of threshold voltages for 100 different 35 nm devices with OTF is shown in Figure 5.28 and is close to a normal distribution. The standard deviation due to the inclusion of OTF is $1.8mV$ which is a 1.48% deviation from the average. Most of the devices will fall in the region of 3σ with two thirds falling in the region of one standard deviation, therefore most of the devices will have a threshold voltage deviating between $0mV$ and $5.4mV$ form the average. This will be added to the 7.5% shift

of the average V_T.

To simulate the effects of OTF a much more finely resolved mesh must be used which alters the numerical behaviour of the simulation. In this case the leakage current of the ideal device is $8.876 \times 10^{-9} A/\mu m$ which is reduced when compared with the previous simulations and the average leakage current of the 100 devices is $1.26 \times 10^{-8} A/\mu m$ this leads to a 29.5% shift in the leakage current from the ideal to the uniform.

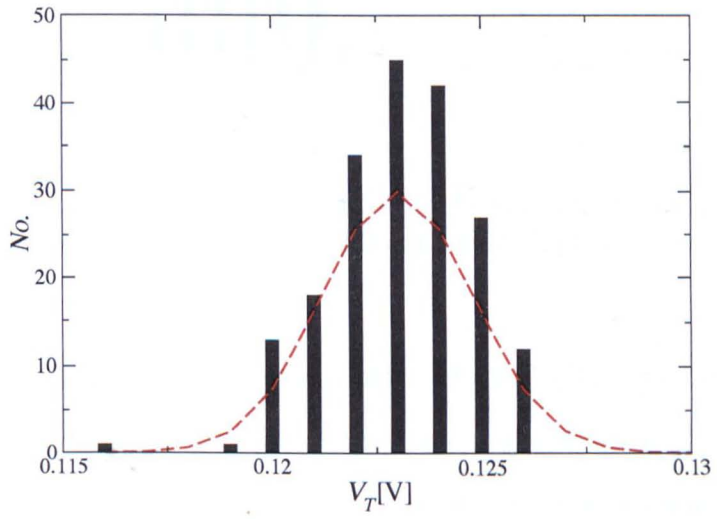

Figure 5.28: Histogram of the distribution of threshold voltages among 100 different 35 nm MOSFET's, with local varying oxide thicknesses.

Figure 5.29 shows the distribution of the off-current or leakage current for 100 individual different devices with applied oxide thickness fluctuations, unlike the prior two cases of intrinsic parameter fluctuation Figure 5.29 appears to be Gaussian in distribution, which is most likely due to the fact that there is really such a small spread. Again we are required to look not only at the standard

deviation but also the skew and kurtosis to understand what is happening.

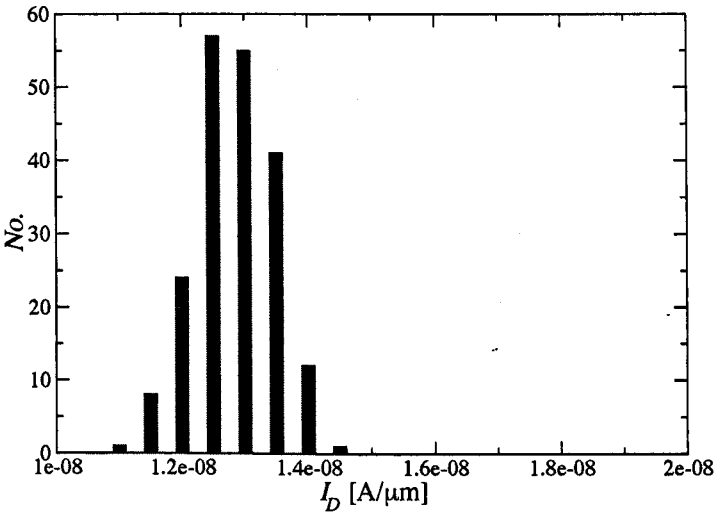

Figure 5.29: Histogram of the distribution of Drain current with zero applied bias to the gate, of 100 devices with different oxide thickness variations.

Figure 5.30 shows the percentage deviation of the I_D/V_G distrubutions of 100 devices with applied oxide thickness variations. We can see from the plot that there is very little deviation in the drain current with a maximum value of only about 6% of the average and that this happens mainly below threshold. It is interesting to note the trend turning down as we get to the value of off-current, the values are so small at this point that this 1 or 2 percent fluctuation is probably error or statistical noise in the results.

The possibility of statistical noise seems more apparent as we view Figure 5.31, it can be seen that the skew fluctuates wildly around a range of very small values, again most likely due to the very small distribution which is found, the small distribution coming from the tight boundary constraints of the thickness

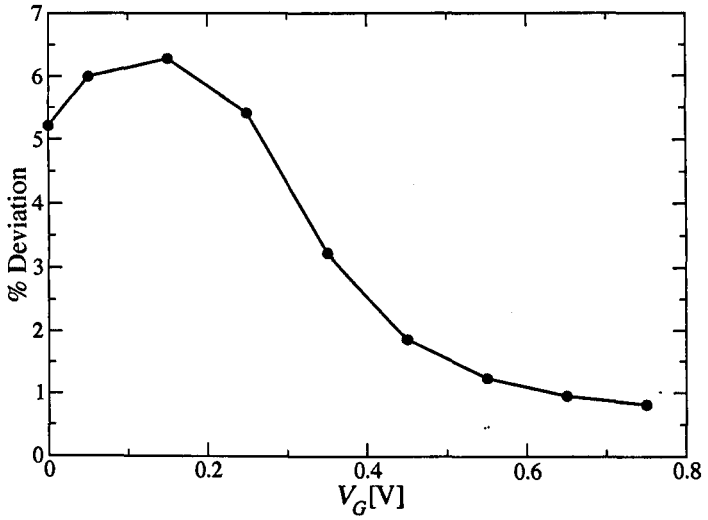

Figure 5.30: The percentage deviation of the I_D/V_G distrubutions of 100 devices with applied oxide thickness variations, shown for all measured gate voltages between 0V and 0.85V

of the oxide.

The graph of the excessive kurtosis shown in Figure 5.32 tell a similar story, the amount by which the distribution moves is very small, particularly when compared to the prior sources of fluctuations, it does appear to apply that the distribution is more sharply peaked than a standard Gaussian distribution however with the uncertainty introduced by Figure 5.29 and 5.31 it would be difficult to say this with any certainty.

As before we can plot a histogram of the $log_{10}(I_D)$, which is shown in Figure 5.33, here we can see that the distribution still exhibits a Gaussian profile. For the distribution above the standard deviation of the $log_{10}(I_D)$ is 0.0225. This can be interpreted as a power such that $10^{0.0225} \approx 1.05$ so we can say that the standard deviation of distribution is 1.05 times the mean value, or has a value

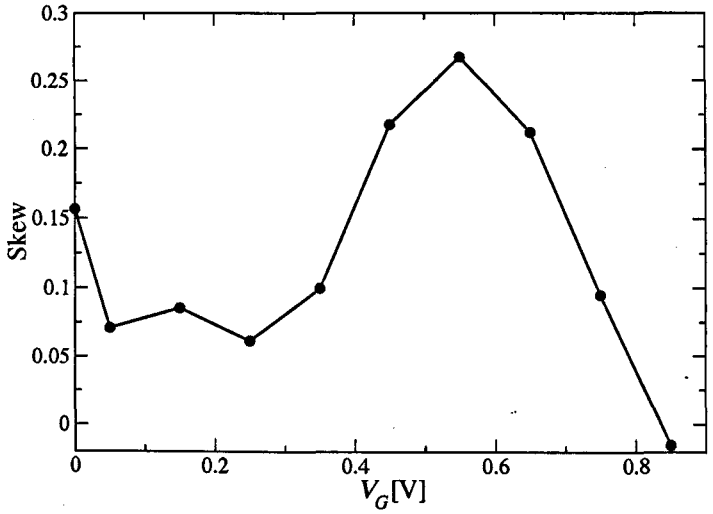

Figure 5.31: The skew of the I_D/V_G distrubutions of 100 devices with applied oxide thickness variations, shown for all measured gate voltages between 0V and 0.85V

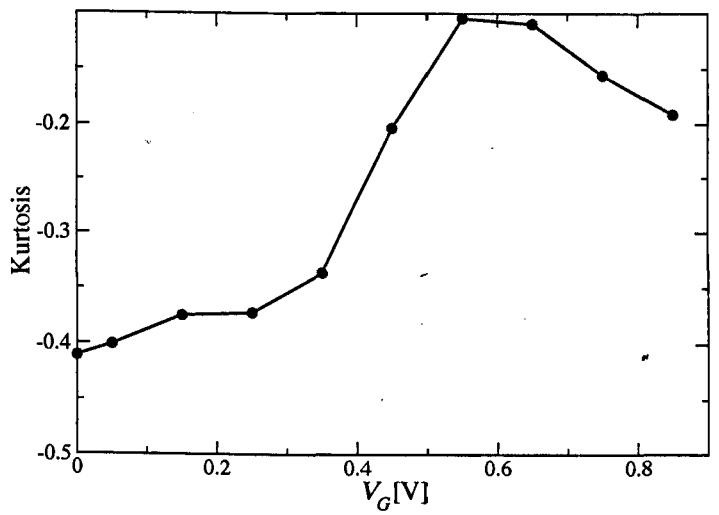

Figure 5.32: The excessive kurtosis of the I_D/V_G distrubutions of 100 devices with applied oxide thickness variations, shown for all measured gate voltages between 0V and 0.85V

of approximately $0.45dB$. As can be seen these values are very small showing that the overall effect of oxide thickness fluctuations are negligible in this device, even three standard devations only give a multiplication factor of 1.17 times the mean value or a $1.35dB$ change. All of the above results imply that the most important effect associated with the inclusion of oxide thickness fluctuations is to lower the average threshold voltage.

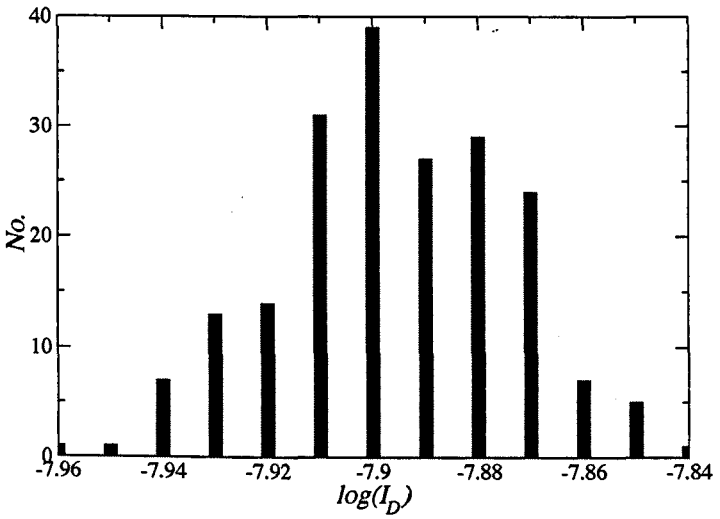

Figure 5.33: Histogram of the distribution of the log of drain current with zero applied bias to the gate, of 100 devices with different oxide thickness variations

5.2.4 Summary

The results of the study of single sources of intrinsic parameter fluctuations are compared in Table 5.2. This table shows the average threshold voltage for the ensemble of devices which contain random discrete dopant, line edge roughness and oxide thickness variations. It also compares the standard deviation and

the percentage standard deviation of these ensembles of threshold voltages. It is clear that random discrete dopants induce the largest fluctuation in threshold voltage, with line edge roughness producing slightly less fluctuations. The amount of fluctuations produced by oxide thickness variations is very small when compared to the other sources of fluctuations. The table also contains the average leakage current produced by each of the sources of fluctuations, and its standard deviation in decibels. It can be seen that random dopant cause the largest shift in leakage current increasing the amount of leakage by 50% while line edge roughness caused the amount of current to rise by 45.5%. An interesting result was that the oxide thickness fluctuations caused a reduction in leakage current by 13.2%. These results show that as device are reduced in size simulations of devices with continuous doping and straight boundaries are not enough to understand the device behaviour and to perform TCAD device design.

Parameter	RD	LER	OTF
Average V_T	133 mV	126 mV	122 mV
σV_T	33.2 mV	19 mV	1.8 mV
$\%\sigma V_T$	24.8%	15%	1.48%
Leakage I_D	$2.2 \times 10^{-8} A/\mu m$	$2.12 \times 10^{-8} A/\mu m$	$1.26 \times 10^{-8} A/\mu m$
% shift I_D	50%	45.5%	-13.2%
σI_D	8dB	5dB	0.5dB

Table 5.2: Summary of the results of various simulation of intrinsic parameter fluctuations, showing average V_T, σV_T, $\%\sigma V_T$, Leakage I_D and σI_D

5.3 Combined Sources of Intrinsic Fluctuations

The sources of intrinsic parameter fluctuations which can be separated in simulations will occur simultaneously within a single MOSFET. To understand how these fluctuations will interact and to what extent they are statistically independent we have carried out simulations with more than one source of fluctuation present. In this section pairs of fluctuations are simulated in order to test their statistical interdependence and possible correlation. The three pairs considered include (i) discrete random dopants with line edge roughness, (ii) discrete random dopants with oxide thickness variations, and (iii) line edge roughness with oxide thickness fluctuations. The simulation of multiple sources of fluctuations requires more computational effort and therefore less devices where simulated with samples ranging between 50 and 100 devices focusing only on the threshold voltage variations.

5.3.1 RD and LER

The electron concentration distribution obtained from the combined simulation of discrete random dopants and line edge roughness in the 35 nm MOSFET is illustrated in Figure 5.34. The discrete random dopants superimposed on the rough gate edges can clearly be seen. It is a more difficult task to see the effects of line edge roughness as the random dopants act to break up the gate edge making the rough line more difficult to observe.

To understand how the two sources of fluctuations interact with one another we can analyse the covariance and correlation of the individual distributions.

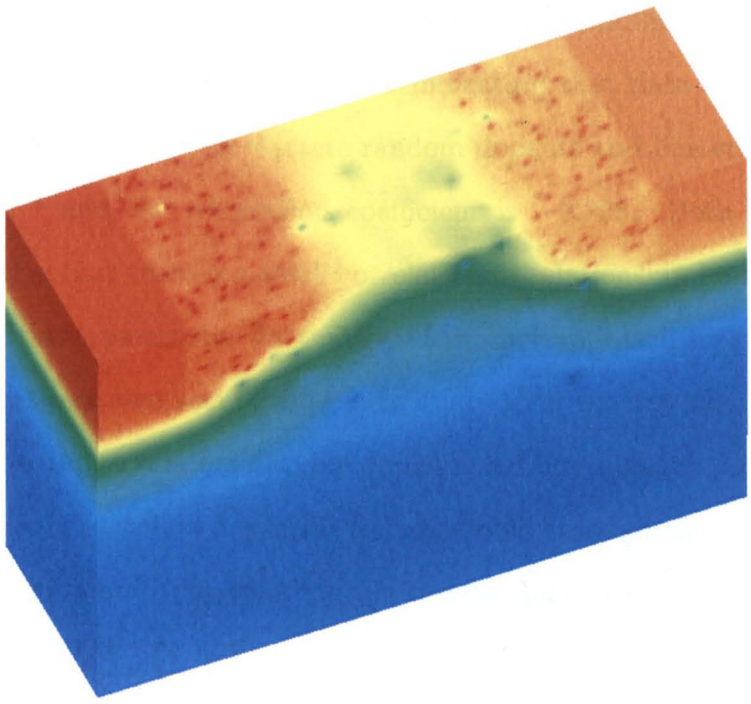

Figure 5.34: 3D electron concentration of a 35 nm MOSFET with the inclusion of discrete random dopants and applied line edge roughness. The simulation domain is $107 \times 35 \times 75$ nm, with the values for electron concentration on a logarithmic scale varying from $1 \times 10^{14} cm^{-3}$ to $2 \times 10^{20} cm^{-3}$.

The covariance and the correlation are strongly related in that they both give a measure of how strongly related one set of data is to another. A large negative covariance suggests that large data values in a set X are strongly related to small values in a set Y while a large positive number suggests that large data values in X are related to large data values in Y. A small covariance close to 0 suggest that the two data sets are statistically independant. Similarly the correlation coefficient varies between -1 to 1 with 0 indicating that data is uncorrelated. The calculated covariance for discrete random dopants and line edge roughness is -6.2×10^{-5} and the correlation coefficient is -0.098. Both of these are small indicating that there is a small correlation or relationship between the two sources of intrinsic parameter fluctuations. Another approach for observing the correlation between two sets of data is to study their scatter plot. The scatter plot corresponding to random discrete dopants and line edge roughness is shown in Figure 5.35. The points are fairly well scattered through the plot indicating a lack of correlation. If they were highly correlated a distinct elongated pattern would be observable.

The average threshold voltage produced by these two sources of intrinsic parameter fluctuations is $126mV$ with a standard deviation of $38.7mV$ which is a 30.6% deviation from the average. It is interesting to note that the average threshold voltage has become lower and is now equal to that corresponding to line edge roughness only. The combined effect of two statistically independent variables on the standard deviation is described by the relationship $\sigma_{1+2} = \sqrt{\sigma_1^2 + \sigma_2^2}$. If we use this expression to add together the standard deviations associated with random discrete dopants and line edge roughness the

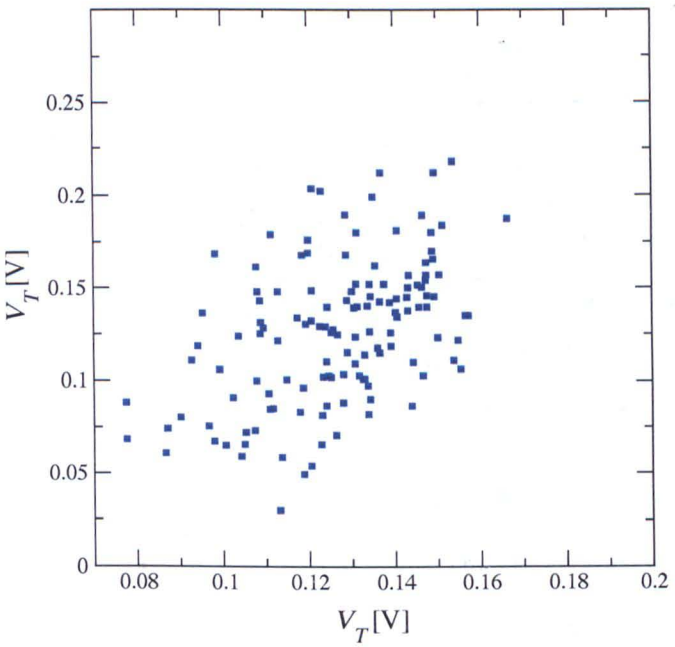

Figure 5.35: Scatter plot of the threshold voltages produced by discrete random dopants and line edge roughness induced fluctuations.

combined effect is $38.25mV$ which is very close to the value obtained from the simultaneous simulation of the two fluctuation sources. This provides strong additional evidence that these two sources of fluctuation are uncorrelated.

5.3.2 RD and OTF

The next combination of intrinsic parameter fluctuations considered simultaneously was discrete random dopants and oxide thickness fluctuations. Figure 5.36 shows the 3D electron concentration distribution in a 35 nm MOSFET with both discrete random dopants and oxide thickness fluctuations present in the simulation. The random dopants are visible in the source and drain region and the oxide thickness fluctuations are visible in the channel region of the device. The blue regions are the areas where the oxide protrudes down into the substrate and so reduces the associated electron concentration.

In order to understand the interaction between the individual random dopants and the oxide thickness variations we calculate the covariance and the correlation coefficient for the individual threshold voltage distributions. In this case the covariance is 6.3×10^{-7} and the correlation coefficient is 0.0104. With both of these values small there is little correlation between the above sources. The corresponding scatter plot of the threshold voltages is shown in Figure 5.37, indicating also very little correlation.

The average threshold voltage produced by the simultaneous presence of random discrete dopants and oxide thickness fluctutations is $123mV$ with a standard deviation of $33.9mV$ which is a 27.53% deviation from the average. It is also interesting to note that the average threshold voltage is almost the same

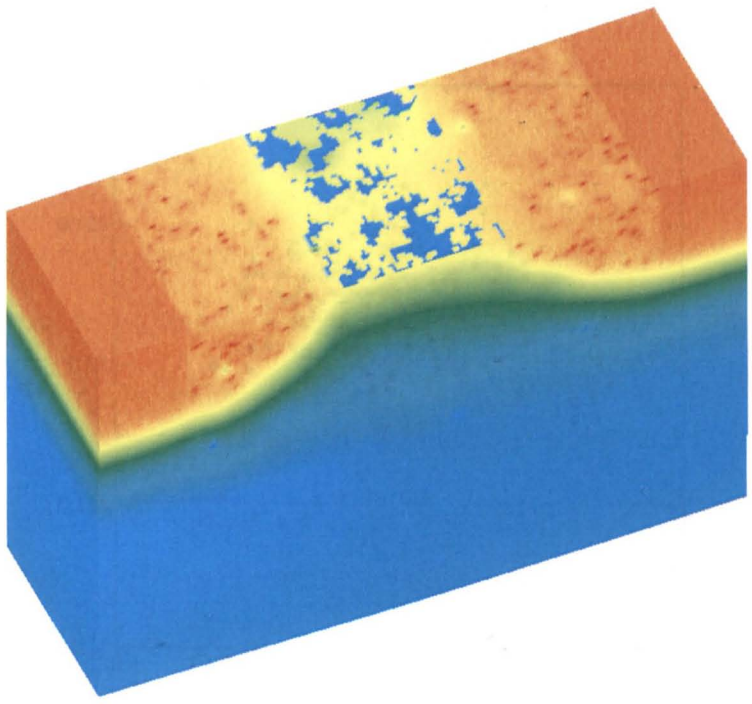

Figure 5.36: 3D electron concentration of a 35 nm MOSFET contain both discrete random dopants and applied oxide thickness fluctuations. The simulation domain is $107 \times 35 \times 75$ nm, with the values for electron concentration on a logarithmic scale varying from $1 \times 10^{14} cm^{-3}$ to $2 \times 10^{20} cm^{-3}$.

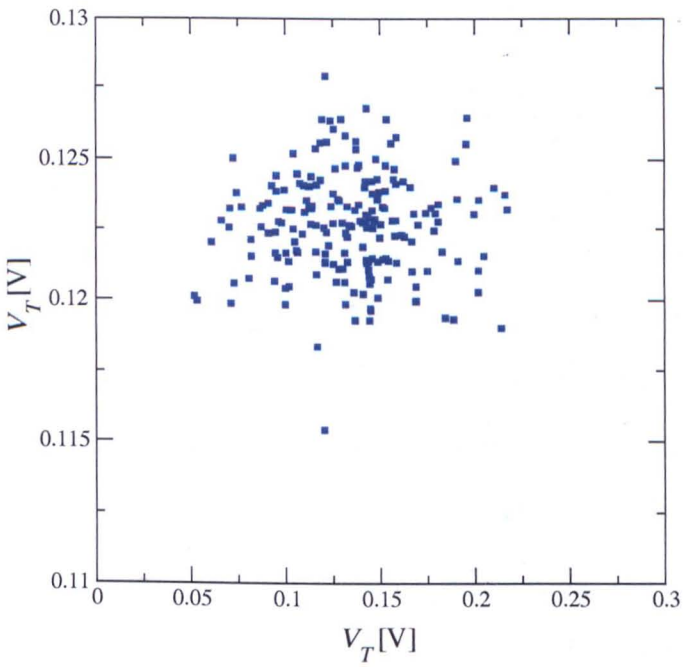

Figure 5.37: Scatter plot of the individual threshold voltages for both discrete random dopants and oxide thickness fluctuations.

as that produced by the stand alone oxide thickness fluctuations and that this reduction still exists when combined with the discrete random dopants. The statistically combined standard deviation from the independent simulation of the two sources is $33.3mV$ which is very close to the value obtained from the combined simulations. This is an additional indicator that these two sources of fluctuations are uncorrelated.

5.3.3 LER and OTF

The last combination of intrinsic parameter fluctuations which will be considered include line edge roughness and oxide thickness fluctuations. Figure 5.38 shows the 3D electron concentration distribution in the 35 nm device with both oxide thickness fluctuations and line edge roughness present in the simulation. The oxide thickness fluctuations are visible in the channel of the MOSFET however this form of fluctuation obscures the line edge roughness between the gate.

For line edge roughness and oxide thickness fluctuations the calculated covariance is 1.06×10^{-6} and the calculated correlation coefficient is 0.031. This values of covariance and correlation in combination with the inspection of the scatter plot shown in Figure 5.39 indicates little correlation between line edge roughness and oxide thickness fluctuations.

The average threshold voltage produced by the inclusion of both line edge roughness and oxide thickness fluctuations is $113mV$ with a standard deviation of $22.8mV$ which is a 20.48% deviation from the average. The average threshold voltage is lower compared to the values obtained from the independent simulation of each of the above sources. It is interesting to note that by adding

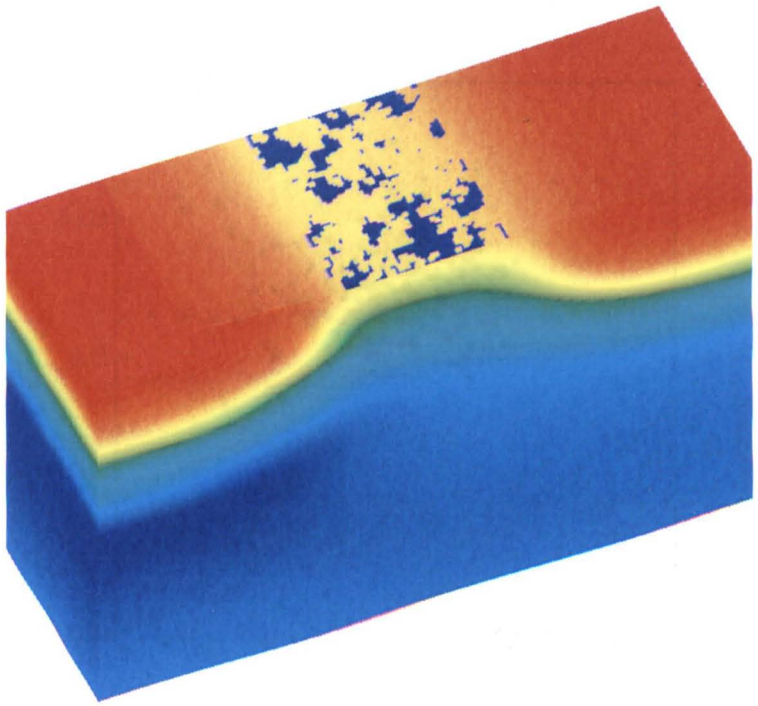

Figure 5.38: 3D electron concentration of a 35 nm MOSFET device with applied line edge roughness and oxide thickness fluctuations. The simulation domain is $107 \times 35 \times 75$ nm, with the values for electron concentration on a logarithmic scale varying from $1 \times 10^{14} cm^{-3}$ to $2 \times 10^{20} cm^{-3}$.

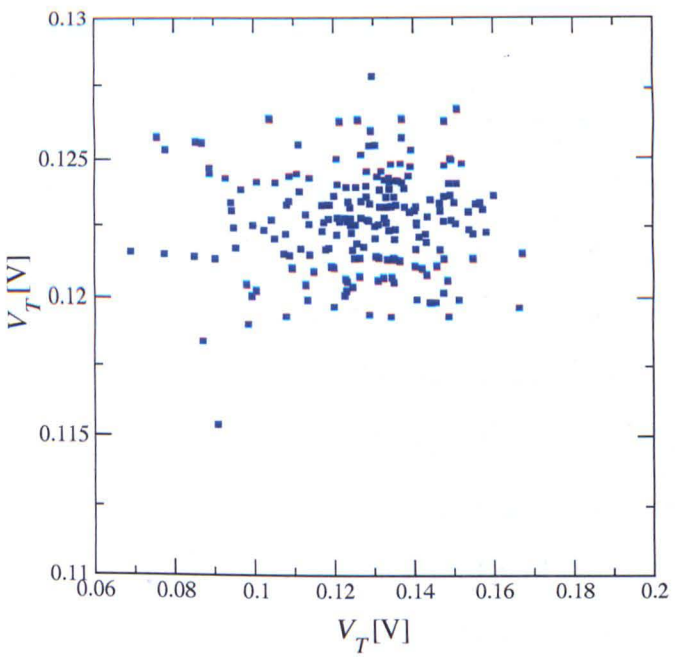

Figure 5.39: Scatter plot of the individual threshold voltages for both line edge roughness and oxide thickness fluctuations.

the shift in threshold voltage for both line edge roughness and oxide thickness fluctuations we arrive at a value of approximately $115mV$ which is very close to the value obtained from the combined simulations. The statistically combined standard deviation of the two sources acting independently is $19.1mV$ which is close to the value obtained from the combined simulations.

5.3.4 Summary and Conclusion

Table 5.3 summarises the data that has been collected throughout this chapter in respect of threshold voltage fluctuations. This is one of the most important parameter in MOSFET design as it has an impact on all other parameters and characteristics of the device. It is clearly seen that the most important source of fluctuations in the 35 nm MOSFET are the random discrete dopants. Although they do not appear to affect the average threshold voltage, most likely due to the inclusion of a retro-grade channel doping. The next main source of fluctuation is line edge roughness which also results in lowering of the threshold voltage. At the bottom of the scale of influence are oxide thickness variations which do not cause large amounts of fluctuations but do affect the average threshold voltage.

It is also clear from Table 5.3 that the three sources of intrinsic parameter fluctuations appear to be statistically independent which allows us to use standard statistical theory to add the effects of the corresponding parameter fluctuations together in order to estimate their combined effect.

The results shown in Table 5.3 allow us to predict the magnitude of fluctuations that would be found in a 35 nm MOSFET which included all the sources of intrinsic parameter fluctuations. A 3D electron concentration distribution ob-

Fluctuation	Average V_T	σV_T	$\%\sigma V_T$	calculated σV_T
RD	133 mV	33.2 mV	24.8%	-
LER	126 mV	19 mV	15%	-
OTF	122 mV	1.8 mV	1.48%	-
RD and LER	126 mV	38.7 mV	30.6%	38.2 mV
RD and OTF	123 mV	33.9 mV	27.53%	33.3 mV
OTF and LER	113 mV	22.8 mV	20.48%	19.1 mV

Table 5.3: Summary of the parameters of both individual intrinsic fluctuations and combined intrinsic parameter fluctuations.

tained from the simulation of the 35 nm MOSFET that included all three sources of intrinsic parameter fluctuations is shown in Figure 5.40. This MOSFET would most likely have an average threshold voltage of $113mV$ which combines the effect of LER and OTF, bearing in mind that random dopants have little effect on the average threshold voltage. The standard deviation of this device would be approximately $39mV$ which would be 34.5% of the average threshold value and the 3σ value would then be approximately $117mV$ which would imply the presence of fluctuations larger than the threshold voltage itself.

The calculated value for the standard deviation of threshold voltage, σV_T, compares favourably with experimental data found in literature. In [124] the σV_T due to intrinsic parameter fluctuations for a 90 nm technology node device with a 40 nm physical gate length [125] and a 65 nm technology node device [126] is given. The values presented are scaled to a gate length of 35 nm assuming that σV_T is approximately equal to $\frac{1}{\sqrt{LW}}$. The scaled value of σV_T for the 90 nm process device $49mV$, this is very close to the calculated value of $39mV$. It is expected that the measured value would be larger than our predicted value because only three sources of fluctuations are presented here while more (such

as grain boundary effects in the poly-silicon gate) would be found within an actual device. It does however give confidence that the results presented here are believable and within the correct range. The calculated σV_T for the 65 nm process device is $63mV$, this again is larger than the value which we calculated, however this device is from a different technology node which will use different implantation methodologies and densities, this is likely to increase the potential for intrinsic parameter fluctuations. Both of these figures grants confidence in the analysis and results presented in this chapter showing that the figures generated are believable when compared to experiment.

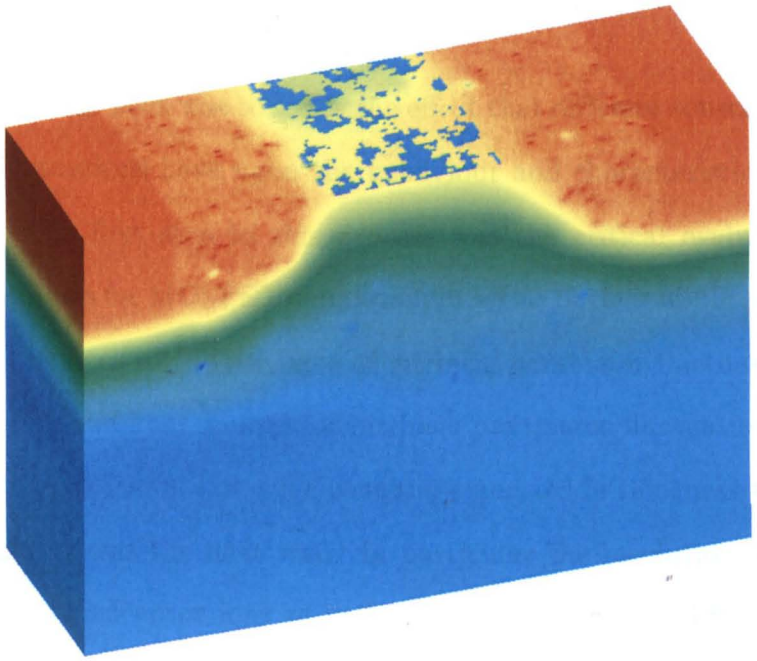

Figure 5.40: 3D electron concentration plot of a 35 nm MOSFET device which includes all sources of intrinsic parameter fluctuations, random dopants, line edge roughness and oxide thickness variations. The simulation domain is 107 $\times$ 35 $\times$ 75 nm, with the values for electron concentration on a logarithmic scale varying from $1 \times 10^{14} cm^{-3}$ to $2 \times 10^{20} cm^{-3}$.

Chapter 6

Conclusion

The purpose of this work was to study the effect that different sources of intrinsic parameter fluctuations have upon the behaviour and characteristics of realistic MOSFETs. The first step in this process was the critical evaluation of the existing techniques for MOSFET simulation in terms of their applicability for 3D simulation of devices in the presence of intrinsic parameter fluctuation sources. The inclusion of different sources of intrinsic parameter fluctuations including random discrete dopants, line edge roughness and oxide thickness variations in the 3D simulator was the next step. In particular the resolution of individual discrete donor or acceptor ions in drift-diffusion simulations creates problems which have been successfully resolved. The developed simulator has been used to study the individual and combined effect of the above sources of intrinsic parameter fluctuations in a 35 nm MOSFET fabricated and characterised by Toshiba [32].

In Chapter 2 the various sources of intrinsic parameter fluctuations found in

the literature have been reviewed. Three key sources of intrinsic parameter fluctuations which are critical for the operation of conventional MOSFETs, and are amenable to numerical simulation and analysis have been identified. These sources include individual random discrete dopants, line edge roughness of the gate edge and oxide thickness fluctuations. The specific requirements associated with the numerical simulation of intrinsic parameter fluctuations were also identified. Full 3D simulations are mandatory due to the inherently 3D nature of the intrinsic parameter fluctuation sources. Computational efficiency is paramount because statistical samples containing 100-200 devices were to be simulated to allow statistical analysis of their characteristics with sufficient accuracy. The simulation technique should correctly capture the physics in the area of interest which in our case includes the subthreshold slope region of operation and threshold voltage. The only simulation technique that satisfies the above three requirements was the Drift Diffusion approach. The other simulation techniques like the hydrodynamic and ensemble Monte Carlo approaches that capture more accurately the transport in deep submicron MOSFETs were too computationally expensive. The ability to include, in an inexpensive way, quantum corrections in the Drift Diffusion framework became very important when solving the problems associated with the inclusion of individual discrete dopants in the Drift Diffusion simulations.

In Chapter 3 the Drift Diffusion simulator that was developed in the Device Modelling group at Glasgow University was described along with the various physical models included in it. The methodology used to include line edge roughness and oxide thickness variations in the simulations was also described.

Due to a lack of accuracy of the currently used approach an original method has been developed for the accurate inclusion of discrete random dopants. The new method evaluates the probability for a dopant placement at each individual site of the silicon crystalline lattice in the active region of the device. The modelling of the transport within the Drift Diffusion framework was also enhanced, compared to the previously used constant mobility, by the inclusion of the more realistic Caughey-Thomas field dependant mobility model coupled to an analytical concentration dependant mobility model. After a careful calibration the mobility models provided good agreement with experimental results.

In Chapter 4 we examine problems associated with the inclusion of random discrete dopants in earlier drift-diffusion simulations. It was found that when discrete dopants were modelled by introducing localised charge density within classical Drift Diffusion a significant portion of the mobile carrier charge became localised ("trapped") around the donor or acceptor position. This charge trapping causes an increase in the resistivity of the correspondingly heavily doped regions. In the case of a MOSFET this means it artificially increases the resistance of the source / drain and reduces the on state current in the simulations. Since the magnitude of this problem increases with the finer resolution of the Coulomb potential of the individual dopants the drift diffusion simulations also become mesh size sensitive. Two possible solutions to these problems were implemented and discussed. The first was to use a charge assignment schemes that "smear" out the associated charge over more than one mesh node. Although successful in reducing the associated charge trapping such schemes remain mesh size sensitive or somewhat arbitrary in terms of choosing the charge smearing

radius. The second solution was to introduce Density Gradient quantum corrections for both the electrons and holes in the Drift Diffusion simulations. This approach was more successful in both reducing the amount of charge trapped and in removing the mesh sensitivity of the solution. This instils more confidence in the results obtained using our simulator.

In Chapter 5 we present the results of the simulation of intrinsic parameter fluctuations in a realistic MOSFET. For our study we have chosen a 35 nm Toshiba MOSFET because it was well documented and the internal doping profiles were readily available. The transistor corresponds in size to the late generation of the 90 nm technology node and has well optimised I_D/V_G characteristics. The "atomistic" simulation was carefully calibrated in respect of the characteristics of the 35 nm transistor. Methodology has been developed to transfer the doping profile of the device from the TCAD simulator Taurus into our simulator. The individual effects of random discrete dopants, line edge roughness and oxide thickness variations were then investigated. It was found that the discrete random dopants have the greatest impact on both the threshold voltage and the leakage current fluctuations. The corresponding standard deviation of the threshold voltage was $33.2mV$ and a percentage increase in the average leakage current was 50%. The line edge roughness had the second greatest impact on the intrinsic parameter fluctuations resulting in a standard deviation of the threshold voltage of $19mV$ and percentage increase in the average leakage current of 45.5%. The smallest impact was caused by the oxide thickness variations resulting in a standard deviation of the threshold voltage of $1.8mV$ and a 13% increase in the average leakage current.

The intrinsic parameter variations associated with the individual sources were found to be statistically independent when combined in simulations. The combination of discrete random dopants and line edge roughness has the largest impact producing a standard deviation in the threshold voltage of $38.7mV$, followed by the combination of random dopants and oxide thickness fluctuations producing a standard deviation of the threshold voltage of $33.9mV$. The smallest impact was the combination of line edge roughness and oxide thickness fluctuations which resulted in a standard deviation of the threshold voltage of $22.8mV$. It is clear that in the conventional 35 nm MOSFET the random discrete dopants are the greatest source of intrinsic parameter fluctuations which cannot be avoided unless the device architecture is changed to adopt ultrathin undoped channels. It is also interesting to note that the effects of line edge roughness are almost as large as the effect of random discrete dopants. If lithographic techniques are not improved LER could become the dominant source of intrinsic parameter fluctuations in the next generation MOSFET. The oxide thickness fluctuations produce the smallest intrinsic parameter fluctuations, however our Drift Diffusion simulations do not include the effects of surface roughness on the channel mobility which will increase the current fluctuations in the simulated device. Finally the inclusion of all sources of fluctuations results in a standard deviation of the threshold voltage of $39mV$ which is 34.5% of the average threshold voltage.

6.1 Future Work

There are several areas in which this work could be extended. The first is the inclusion of other sources of fluctuation. The scaling of the conventional MOSFET beyond the 45 nm technology node requires the introduction of high-κ gate dielectrics. The deposition and annealing of the most promising high-κ candidate $HfSiO$ results for example in crystalline HfO_2 islands in amorphous SiO_x material which results in variations in the local dielectric value. The local fluctuations in the dielectric property will cause macroscopic variations in performance between devices and it becomes important to study these effects. The study of strain induced fluctuations is a another possible extension to the work. The processing steps involved in the fabrication of MOSFETs introduce strain and strain variations in the Si crystalline lattice which locally alters the mobility. This strain has been used intentionally to increase the performance of MOSFETs but there is the possibility that unintentional strain variations will cause unwanted fluctuations in device performance.

Introduction of roughness at the $Poly - Si/SiO_2$ interface is a natural extension of this work. It will also be important to study the impact of roughness between the high-κ dielectric and corresponding metal gate, and the roughness between the high-κ dielectric and the suboxide layer. At a simulator development level it may also be useful as devices are reaching nanometer dimensions to introduce a discretisation of the simulation domain based on the nodes of the silicon crystalline lattice. This too may increase the accuracy when studying random dopant effects by reproducing more accurately the dopant configuration that would appear in a real device, and possibly correlations in the dopant

placement. Crystal lattice discretisations will also benefit the study of interface roughness related effects.

Finally the impact of random discrete dopant induced parameter fluctuations on circuit performance have been studied in ring oscillators and SRAM cells. In this work we have shown that other sources of intrinsic parameter fluctuations like LER will play an important role in the future and their impact on circuits and systems should also be studied.

Bibliography

[1] C. Canali, C. Jacaboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta. Electron Drift Velocity in Silicon. *Phys. Rev. B*, 12:2265–2284, 1975.

[2] G. E. Moore. Cramming More Components onto Integrated Circuits. *Electronics*, 38:114–117, 1965.

[3] G. E. Moore. Progress in Digital Integrated Electronics. *IEDM Tech. Digest*, pages 11–13, 1975.

[4] H. S. P. Wong. Beyond the conventional transistor. *IBM J. Res. and Dev.*, 46:133–165, 2002.

[5] R. Dennard, F. H. Gaensslen, H. N. Yu, L. Rideout, E. Bassous, and A. R. LeBlanc. Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions. *IEEE J. Solid State Circuits*, SC-9:256, 1974.

[6] T. Skotnicki. Heading for decananometer CMOS – Is navigation among icebergs still a viable strategy? *ESSDERC*, pages 19–33, 2000.

[7] Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S. H. Lo, G. Sai-Halasz, R. Viswanathan, H. J. C. Wann, S. Wind, and H. S. Wong. CMOS Scaling into the Nanometer Regime. *Proc. IEEE*, 85:486, 1997.

[8] H. S. P. Wong, D. Frank, P. M. Solomon, H. J. Wann, and J. Welser. Nanoscale CMOS. *Proc. IEEE*, 87:537, 1999.

[9] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H. S. Wong. Device Scaling Limits of Si MOSFETs and Their Application Dependancies. *Proc. IEEE*, 89:259–288, 2001.

[10] ITRS. International Technology Roadmap for Semiconductor Devices. 2001.

[11] B. Hoeneisen and C. A. Mead. Fundemental Limitations in Microelectronics – I MOS Technology. *Solid State Electron*, 15:819, 1972.

[12] R. W. Keyes. The Effect of randomness in the distribution of impurity atoms on FET thresholds. *Appl. Phys.*, 8:251–259, 1975.

[13] R. W. Keyes. Fundamental Limits of Silicon Technology. *Proc IEEE*, 89:259–288, 2001.

[14] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivack. Surface roughness at the Si(100)-SiO2 Interface. *Phys. Rev. B.*, 32:8171–8186, 1985.

[15] T. Mizuno, M. Iwase, H. Niiyama, T. Shibata, K. Fujisaki, T. Nakasugi, A. Toriumi, and Y. Ushiku. Performance Fluctuations in 0.1 um MOS-

FETs - Limitation of 0.1 um ULSIs. *Symposium on VLSI Digest of Technical Papers*, pages 13–14, 1994.

[16] T. Mizuno, J. Okamura, and A Toriumi. Experimental Study of Threshold Voltage Fluctuation Due to Statistical Variation of Channel Dopant Number in MOSFETs. *IEEE Trans. Elec. Dev.*, 41:2216–2221, 1994.

[17] T. Yamaguchi, H. Namatsu, M. Nagase, K. Yamazaki, and K. Kurihara. Nanometer-scale linewidth fluctuations caused by polymer aggregates in resist films. *Appl. Phys. Lett.*, 71, Oct 1997.

[18] A. Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub 50 nm MOSFETs: A 3D "atomistic" simulation study. *Nanotechnology*, 10:153–158, 1999.

[19] H. S. Wong, D. Frank, and P. Solomon. Device Design Considerations for Double-Gate, Ground-Plane and Single Gated Ultra-Thin SOI MOSFETs at the 25nm Channel Length Generation. *IEDM Tech. Digest*, page 407, 1998.

[20] T. Sekigawa and Y. Hayashi. Calculated Threshold Voltage Characteristics of a XMOS Transistor Having an Additional Bottom Gate. *Solid State Electron.*, 27:827, 1984.

[21] H. S. P. Wong. Novel Device Options for Sub-100nm CMOS. *IEDM Short Course: Sub-100nm CMOS*, 1999.

[22] X. Huang, W. C. Lee, C. Ku, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. K. Choi, K. Asano, V. Subramanian, T. J.

King, J. Bokor, and C. Hu. Sub 50-nm FinFET: PMOS. *IEDM Tech Digest*, page 67, 1999.

[23] D. Fried, A. Johnson, E. Nowak, J. Rankin, and C. Willets. A sub-40 nm Body Thickness N-Type FinFET. *Proceedings of Device Research Conference*, page 24, 2001.

[24] R. Chau. 30nm and 20nm Physical Gate Length CMOS Transistors. *Silicon Nanoelectronics Workshop*, 2001.

[25] B. Yu H. Wang A Joshi Q. Xiang et al. 15nm Gate Length Planar CMOS Transistor. *IEDM Tech. Dig.*, 2001.

[26] H. Wakabayashi S. Yamagami N. Ikezawa et al. Sub-10-nm Planar-Bulk-CMOS Devices using Lateral Junction Control. *IEDM Tech. Dig.*, pages 989–991, December 2003.

[27] S. Selberherr. *Analysis and Simulation of Semiconductor Devices.* Springer-Verlag, 1984.

[28] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva. Simulation of Intrinsic Parameter Fluctuations in Decananometer and Nanometer-Scal MOSFETs. *IEEE Tran. Elec. Dev.*, 50:1837–1852, 2003.

[29] J. H. Davies A. Asenov, A. R. Brown and S. Saini. Hierarchical approach to "atomistic" 3D MOSFET simulation. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 18:1558–1565, 1999.

[30] T. Schafbauer et al. Integration of high-performance, low leakage and mixed signal features into a 100nm CMOS technology. *Symp. VLSI Tech. Dig. Tech. Papers*, page 62, 2002.

[31] K. Fukasaku et al. UX6-100 nm generations CMOS integration technology with Cu/low-k interconnect. *Symp. VLSI Tech. Dig. Tech. Papers*, page 64, 2002.

[32] S. Inaba, K. Okano, S. Matsuda, M. Fujiwara, and et al. High Performance 35nm Gate Length CMOS with NO Oxynitride Gate Dielectric and Ni SALICIDE. *IEDM Tech. Dig.*, page 641, 2001.

[33] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Sato, and H. Ivai. 1.5nm direct tunneling gate oxide Si MOSFETs. *IEEE Trans. Elec. Dev.*, 43:1233–1241, 1996.

[34] G. F. Cardinale and et al. Demonstration of pattern transfer into sub-100nm polysilicon line/space features patterned with extreme ultraviolet lithography. *J. Vac. Sci. Tech. B.*, 17:2970–2974, 1999.

[35] K.R. Lakshikumar, R. A. Hadaway, and M. A. Copeland. Characterization and modeling of mismatch in MOS transistors for precision analogue design . *IEEE Journal of Solid State Circuits*, SC–21, 1986.

[36] M. Steyaert, J. Bastos, R. Roovers, P. Kinget, W. Sansen, B. Graindourse, A. Pergot, and E. Janssens. Threshold mismatch in short-channel MOS transistors. *Electron Lett.*, 30:1546–1548, 1994.

[37] R. Difrenza, P. Llinares, and G. Gibaundo. Impact of short channel and quantum effects on MOS transistor mismatch. *Proc. 3rd Eur. Workshop Ultimate Integration Silicon*, pages 127–130, 2002.

[38] O. R. dit Buisson and G. Morin. MOSFET mactching in deep submicron technology. *Proc. ESSDERC*, pages 731–734, 1996.

[39] J. T. Horstmann, U. Hilleringmann, and K. F. Goser. Matching analysis of deposition defined 50-nm MOSFETs. *IEEE Trans. Elec. Dev.*, 45:299–306, 1998.

[40] K. Takeuchi, T. Tatsumi, and A. Furukawa. Channel engineering for the reduction of random-dopant-placement-induced threshold voltage fluctuations. *IEDM Technical Digest*, 1996.

[41] T. Hagivaga, K. Yamaguchi, and S. Asai. Threshold voltage variation in very small MOS transistors due to local dopant fluctuations. *Proc. Symp. VLSI Technol. Dig.*, pages 46–47, 1982.

[42] K. Nishiohara, N. Shiguo, and T. Wada. Effects of mesoscopic fluctuations in dopant distributions on MOSFETs threshold voltage. *IEEE Trans. Elec. Dev.*, 39:634–639, 1992.

[43] P. A. Stolk and D. B. M. Klaasen. The effect of statistical dopant fluctuations on MOS device performance. *IEDM Tech. Dig.*, 1996.

[44] A. Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1um MOSFETs: A 3D "atomistic" simulation study. *IEEE Trans. Elec. Devices*, 45:2505–2513, 1998.

[45] P. A. Stolk, F. P. Widdershoven, and D. B. M. Klaasen. Device modelling of statistical dopant fluctuation in MOS transistors. *Proc. SISPAD*, pages 153–156, 1997.

[46] H. S. Wong and Y. Taur. Three dimensional "atomistic" simulation of discrete random dopant distribution effects in sub-0.1um MOSFETs. *IEDM Tech Dig*, pages 705–708, 1993.

[47] D. Vasileska, W. J. Gross, and D. K. Ferry. Modelling of deep-submicrometer MOSFETs: random impurity effects, threshold voltage shifts and gate capacitance attenuation. *IWCE*, pages 259–262, 1998.

[48] J. R. Zhou and D. K. Ferry. Three dimensional simulation of the effect of random dopant distribution on conductance for deep submicron devices. *Proc. 3rd Int. Workshop. Comput. Elec.*, pages 74–77, 1994.

[49] A. Asenov, M. Jaraiz, S. Roy, G. Roy, F. Adamu-Lema, A. R. Brown, V. Moroz, and R. Gafiteanu. Integrated Process and Device Simulation of decananometre MOSFETs. *Proc. SISPAD*, pages 87–90, 2002.

[50] T. Ezaki, T. Ikezawa, A. Notsu, K. Tanaka, and M. Hane. 3D MOSFET Simulation Considering Long-Range Coulomb Potential Effects for Analyzing Statistical Dopant Induced Fluctuations Associated with Atomistic Process Simulator. *Proc. SISPAD*, 91, 2002.

[51] G. Slavcheva, J. H. Davies, A. R. Brown, and A. Asenov. Potential Fluctuations in MOSFETs generated by randomly distributed impurites in the depletion layer. *J. Appl. Phys*, pages 4326–4334, 2002.

[52] A. Asenov and S. Saini. Suppression of Random-Induce Threshold Voltage Fluctuations in Sub-0.1um MOSFETs with Epitaxial and delta-doped Channels. *IEEE Transactions on Electron Devices*, 46, Aug 1999.

[53] A.Asenov G. Slavcheva A. R. Brown J. H. Davies and S. Saini. Quantum enhancement of random dopant induced threshold voltage fluctuations in sub 100nm MOSFET's: A 3-D density-graient simulation study. *IEEE Trans. Electron Dev.*, 48:722–729, 2001.

[54] A. Asenov and S. Saini. Polysilicon Gate Enhancement of the Random Dopant Induced Threshold Voltage Fluctuations in Sub-100nm MOSFETs with Ultrathin Gate Oxide. *IEEE Transactions on Electron Devices*, 47, Apr 2000.

[55] M. Nagase, H. Namatsu, K. Kurihara, K. Murase, and T. Makino. Nanoscal fluctuations in electron beam resist pattern evaluated by atomic force microscopy. *Microelectronic Engineering*, 30:419–422, 1996.

[56] H. Namatsu, M. Nagase, T. Yamaguchi, K. Yamazaki, and K. Kurihara. Influence of edge roughness in resist patterns on etched patterns. *J. Vac. Sci. Technol. B*, 16:3315–3321, 1998.

[57] S. Kaya, A.R. Brown, A. Asenov, D. Magot, and T. Linton. Analysis of Statistical Fluctuations due to Line Edge Roughness in sub-0.1um MOSFETs. *SISPAD*, 2001.

[58] S. Mori and et al. Reduction of line edge roughness in the top surface imaging process. *J. Vac. Sci. Tech. B.*, 46:3739–3743, 1998.

[59] P. Oldigies, L. Qimghuamg, K. Petrillo, M. Sanchez, M. Ieong, and M. Hargrove. Modeling line edge roughness effects in sub 100 nanometer gate length devices. *Proceeding of SISPAD*, 2000.

[60] S. Winkelmeier, M. Sarstedt, M. Ereken, M. Goethals, and K. Ronse. Metrology method for the correlation of line edge roughness for different resists before and after etch. *Microelec. Eng.*, 66:57–58, 2001.

[61] M. Yoshizawa and S Moriya. Edge roughness evaluation method for quantifying at-size beam blur in electron-beam lithography. *Electr. Lett.*, 36:90, 2000.

[62] C. H. Diaz, H. J. Tao, Y. C. Ku, A. Yen, and K. Young. An experimentally validated analytical model for gate line edge roughness(LER) effects on technology scaling. *IEEE Electron Dev. Lett.*, 22:287, 2001.

[63] J. Wu, J. Chen, and K. Liu. Transistor width dependance of LER degradation to CMOS device characteristics. *Proc. SISPAD*, pages 95–98, 2002.

[64] S. D. Kim, S. Hong, J. K. Park, and J. C. S. Woo. Modeling and analysis of gate line edge roughness effects on CMOS scaling toward deep nanoscale gate length. *Extende Abstracts Int.Conf. Sold-State Device Materials*, pages 20–21, 2002.

[65] T. Linton, M. Giles, and P. Packan. The impact of line edge roughness on 100nm device performance. *Proc. Silicon Nanoelectronics Workshop*, page 82, 1998.

[66] T. D. Linton, S. Yu, and R. Shaheed. 3D modeling of fluctuation effects in highly scaled VLSI devices. *VLSI design*, 13:103–109, 2002.

[67] A. Asenov, S. Kaya, and A. R. Brown. Intrinsic Parameter Fluctuations in Decananometer MOSFETs Introduced by Gate Line Edge Roughness. *IEEE Transactions on Electron Devices*, 50:1254–1260, May 2003.

[68] S. Thompson et al. An enhanced 130 nm generation logic technology featuring 60 nm transistors optimized for high performance and low power 0.7-0.4V. *IEDM Tech. Dig.*, pages 257–260, 2001.

[69] B. Daoyle et al. Transistor elements for 30nm physical gate length and beyond. *Intel Technol. J.*, 6:42, 2002.

[70] M. Niva, T. Kouzaki, K. Okada, M. Udagawa, and R. Sinclair. Atomic order planarization of ultrathin SiO2/Si(001) interface. *Jpn. J. Appl. Phys.*, 33:388–394, 1994.

[71] D. Z. Y. Ting, E. S. Daniel, and T. C . McGill. Interface roughness effects in ultrathin gate oxides. *VLSI Des*, 8:47–51, 1998.

[72] E. Cassan, P. Dolfus, S. Galadin, and P. Hesto. Calculation of direct tunneling gate current though ultrathin oxide and oxide/nitride stacks in MOSFETs and H-MOSFETs. *Microelec. Reliability*, 40:585–588, 2000.

[73] A. Asenov and S. Kaya. Effect of oxide roughness on the threshold voltage fluctuations in decanano MOSFETs with ultrathin gate oxide. *Proc. SISPAD*, pages 135–138, 2000.

[74] M. Koh, W. Mizubayashi, K. Ivamoto, H. Murakami, T. Ono, M. Tsuno, T. Mihara, K. Shibahara, S. Miyazaki, and M. Hirose. Limit of gate oxide thickness scaling in MOSFETs due to apparent threshold voltage fluctuations introduced by tunneling leakage current. *IEEE Trans. Electron. Dev.*, 48:259–264, 2001.

[75] M. Gotoh, K. Sudoh, H. Itoh, K. Kawamoto, and H. Iwasaki. Analysis of SiO2 Si(001) interface roughness for thin gate oxides by scanning tunneling microscopy. *Appl. Phys. Lett.*, 81:430–432, 2002.

[76] Y. Yoshinobu, A. Iwamoto, K. Sudoh, and H. Iwasaki. Scaling of Si/SiO2 interface roughness. *J. Vac. Sci. Tech. B*, 13:1630–1634, 1995.

[77] T. Yamanaka, S. J. Fang, H-C. Lin, J. P. Snyder, and C. R. Helms. Correlation Between Inversion Layer Mobility and Surface Roughness Measured by AFM. *IEEE Elec. Dev. Lett.*, 17:178–180, 1996.

[78] A. Pirovano, A. L. Lacaita, G. Ghidini, and G. Tallarida. On the Correlation Between Surface Roughness and Inversion Layer Mobility in Si-MOSFETs. *IEEE Elec. Dev. Lett*, 21:34–36, 2000.

[79] M. V. Fischetti, F. Gamiz, and W. Hänsch. On the enhanced electron mobility in strained-silicon inversion layers. *J. of Appl. Phys.*, 92:7320–7324, 2002.

[80] S. Jallepelli, J. Bude, W. K. Shih, M. R. Pinto, C. M. Maziar, and A. F. Tasch Jr. Electron and hole quantization and their impact on deep sub-micron silicon p- an n-MOSFET characteristics. *IEEE Trans. Electron. Dev.*, 44:297–303, 1997.

[81] C. S. Rafferty, B. Biegel, Z. Yu, M. G. Ancona, J. Bude, and R. W. Dutton. Multi-dimensional quantum effects simulation using a density-gradient model and script level programming technique . *Proc. SISPAD*, pages 137–140, 1998.

[82] A. Asenov S. Kaya J. H. Davies and S. Saini. Oxide Thickness Variation Induced Threshold Voltage Fluctuations in Decanano MOSFET's: A 3D Density Gradient Simulation Study. *Superlattices and Microstructures*, 28:507–515, 2000.

[83] G. Bersuker, J. H. Sim, C. D. Young, R. Choi, B. H. Lee, P. Lysaght, G. A. Brown, P. M. Zeitoff, M. Gardner, R. W. Murto, and H. R. Huff. Effects of Structural Properties of Hf-Based Gate Stacks on Transistor Performance. *Mat. Res. Soc. Symp. Proc.*, 811:31–35, 2004.

[84] S. Saito, D. Hisamoto, S. Kimura, and M. Hiratani. Unified Mobility Model for High-κ Gate Stacks. *IEDM Tech, Dig.*, pages 797–800, 2003.

[85] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies. RTS amplitudes in decanano MOSFETs: A 3D simulation study. *IEEE Trans. Electron Dev.*, 50:839–845, 2003.

[86] X. Tang, V. K. De, and J. D. Meindl. Intrinsic Parameter Fluctuations Due to Random Dopant Placement. *IEEE Trans. on VLSI Systems*, 5:369–376, 1997.

[87] The Impact of Intrinsic Device Fluctuations on CMOS SRAM Cell Stability. A. Bhavnagarwala and X. Tang and J. Meindl. *IEEE Journal of Solid-State Circuits*, 36:658–665, 2001.

[88] K. A. Bowman, X. Tang, J. C. Eble, and J. D. Meindl. Impact of Extrinsic and Intrinsic Parameter Fluctuations on CMOS Circuit Performance. *IEEE Journal of Solid-State Circuits*, 35:1186–1193, 2000.

[89] S. Roy, A. Lee, A. R. Brown, and A. Asenov. Applicability of Quasi-3D and 3D MOSFET Simulations in the "Atomistic" regime. *Journal of Computational Electronics*, 2:423–426, 2003.

[90] S. Roy, B. Cheng, G. Roy, and A . Asenov. A methodology for quantitatively introducing "atomistic" fluctuations into compact device models for circuit analysis. *Journal of Computational Electronics*, 2:427–431, 2003.

[91] B. Cheng, S. Roy, G. Roy, and A. Asenov. Integrating "atomistic", intrinsic paramter fluctuations into compact model circuit analysis. *Proc. ESSDERC 2003*, pages 437–440, 2003.

[92] B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, and A.Asenov. The Impact of Random Doping Effects on Decanano CMOS SRAM Cell Stability. *Proc. ULIS 2004*, 2004.

[93] U. Ravaioli. Hierarchy of simulation approaches for hot carrier transport in deep submicron devices. *Semicond. Sci. Technol.*, 13:1–10, 1998.

[94] T-W. Tang H. Kosnar T. Grasser and S. Selberherr. A Review of Hydrodynamic Energy-Transport Models for Semiconductor Device Simulation. *Proc. IEEE*, 91:251–271, 2003.

[95] M. G. Ancona and G. J. Iafrate. Quantum correction to the equation of state of an electron gas in a semiconductor. *Physical Review B*, 39:9536–9540, 1989.

[96] D. K. Ferry, R. Akis, and D. Vasileska. Quantum effects in MOSFETs: Use of an Effective Potential in 3D Monte Carlo Simulationof Ultra Short Channel Devices. *IEDM*, 2000.

[97] K. Blotekjaer. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Elec. Dev.*, 17:38–47, 1970.

[98] C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation.* Vienna: Springer, 1989.

[99] R. W. Hockney and J. W. Eastwood. *Computer Simulation using Particles.* New York: McGraw-Hill, 1981.

[100] K. Hess ed. *Monte Carlo Device Simulation: Full Band and Beyond.* Norwell MA: Kluwer, 1991.

[101] M. V. Fischetti. Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures I. Homogenous transport. *IEEE Trans. Elect. Dev.*, 38:634–649, 1991.

[102] M. V. Fischetti and S. E. Laux. Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures II. Submicrometer MOSFET's. *IEEE Trans. Elect. Dev.*, 38:650–660, 1991.

[103] A. Asenov R. Balasubramaniam A. R. Brown and J. H. Davies. Effect of Single electron trapping in decanano MOSFET's: a 3D "Atomistic" simulation study. *Superlattices and Microstructures*, 27:411–416, 2000.

[104] A.Asenov G. Slavcheva A. R. Brown R. Balasubramaniam and J. H. Davies. Statistical, 3D "atomistic" simulation of Decanano MOSFET's. *Superlattices and Microstructures*, 27:215–227, 2000.

[105] A. Asenov. Quantum correction to the "atomistic" MOSFET simulation. *VLSI Design*, 13:15–21, 2001.

[106] A. Asenov S. Kaya and J. H. Davies. Intrinsic Threshold Voltage Fluctuations in Decanano MOSFET's due to Local Oxide Thickness Variations. *IEEE Trans. Electron Dev.*, 49:112–119, 2002.

[107] H. K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Elec. Dev.*, 11:455–465, 1964.

[108] R. M. Feenstra, M. A. Lutz, F. Stern, K. Ismail, P. Mooney, F. K. LeGoues, C. Stanis, J. O. Chu, and B. S. Meyerson. Roughness analysis of Si/SiGe heterostructures. *J. Vac. Sci. Technol. B. Microelectron. Process. Phenom.*, 13:1608–1612, 1995.

[109] D. J. Frank, Y. Taur, M. Ieong, and H-S. P. Wong. Monte Carlo Modeling of Threshold Variation due to Dopant Fluctuation. *Symposioum on VLSI Circuits Digest of Technical Papers*, pages 171–172, 1999.

[110] N. C. Barford. *Experimental Measurements: Precision, Error and Truth.* Addison and Wesley, 1967.

[111] D. M. Caughey and R. E. Tomas. Carrier Mobilities in Silicon Empirically Related to Doping and Field. *Proceedings of the IEEE*, pages 2192–2193, December 1967.

[112] W.R. Thurber, R. L. Mattis, and Y. M. Liu. Resistivity-Dopant density Relationship for Phosphorus Doped Silicon. *J. Electrochem. Soc.: Solid State Sci and Technology*, 127:1807–1812, 1980.

[113] Synopsis Inc. *Taurus User Manual Vol. 1.* Synopsis, 2001.

[114] G. Roy, A. R. Brown, A. Asenov, and S. Roy. Quantum Aspects of Resolving Discrete Charges in "Atomistic" Device Simulations. *J. of Comp. Elec.*, 2:323–327, 2003.

[115] G. Roy, A. R. Brown, A. Asenov, and S. Roy. Bipolar quantum corrections in resolving individual dopants in "atomistic" device simulation. *Superlattices and Microstructures*, 34:327–334, 2003.

[116] T. R. Sandin. *Essentials of Modern Physics.* Addsion-Wesley, 1989.

[117] P. Y. Yu and M Cardona. *Fundamentals of Semiconductors.* Springer: Germany, 1996.

[118] R. A. Smith. *Semiconductors.* Cambridge University Press, 1959.

[119] N. Sano, K. Matsuzawa, M. Mukai, and N Nakayama. Role of Long-Range and Short-Range Coulomb Potentials in Threshold Characteristics under

Discrete Dopants in Sub-0.1um Si-MOSFETSs. *IEDM Tech. Digest.*, 275, 2000.

[120] Z. Qin and S. T. Dunham. Modeling Fermi Level in Atomistic Simulation. *Proc. Mater. Res. Soc. Symp.*, 717:C3.8, 2002.

[121] Z. Qin and S. T. Dunham. Atomistic simulations of the effect of Coulombic interactions on carrier fluctuations in doped silicon. *Physical Review B*, 68, 2003.

[122] B. Doris, M. Ieong, T. Kanarsky, and et al. Extreme Scaling with Ultra-Thin Si Channel MOSFETs. *IEDM Tech. Dig.*, 2002.

[123] F. Adamu-Lema. *Scaling and Intrinsic Parameter Fluctuation of Nano-CMOS Devices* . PhD thesis, University of Glasgow, Electronics and Electrical Eng. Dept., 2005.

[124] G. Roy, A. Brown, F. Adamu-Lema, S. Roy, and A. Asenov. Intrinsic Parameter Fluctuations in Conventional MOSFETs of the Next Four Generations Technology Nodes. *In Preparation for Submission to the IEEE Trans. Electron Dev.*, 2005.

[125] H. Fukutome, Y. Momiyama, Y. Tagawa, T. Kubo, T. Aoyama, H. Arimoto, and Y. Nara. Direct measurement of effects of shallow-trench isolation on carrier profiles in sub-50nm n-MOSFETs. *Symp. on VLSI Tech. Dig.*, pages 140–141, 2005.

[126] F. Arnaud, B. Duriez, B. Tavel, and et al. Low cost 65 nm CMOS platform for low power and general purpose applications. *Symp. on VLSI Tech. Dig.*, pages 10–11, 2004.