University of Glasgow

Arokianathan, Clinton Rudra (1998) Methods for the atomistic simulation of ultrasmall semiconductor devices. PhD thesis

http://theses.gla.ac.uk/6275/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

Methods for the atomistic simulation of ultrasmall semiconductor devices

Clinton Rudra Arokianathan

March, 1998

A Thesis presented to the University of Glasgow
Department of Electronics and Electrical Engineering
in partial fulfilment of the requirements for
the degree of Doctor of Philosophy.

© *Clinton Rudra Arokianathan, 1998*

ABSTRACT

As the feature sizes in VLSI technology shrink to less than 100 nm the effects due to the quantisation of electronic charge begin emerge. There are a small number of carriers and impurities and the statistical variation in their number have significant effects on the threshold characteristics of the devices that hampers their large scale integration into future ULSI.

These considerations imply that a full three dimensional simulation need to be carried out where the charges are treated as discrete objects and all the electrostatic interactions between them must be modelled. This procedure provides a more realistic description of ionised impurity and carrier-carrier scattering than conventional methods. Particle based simulations are very computationally intensive and so this work presents a new method based on the theory of Brownian motion that is particularly useful for studying systems near equilibrium conditions where drift-diffusion models are applicable. It is considerably cheaper to implement than conventional Monte Carlo methods (being up to 70 times faster) and does not suffer from the large statistical fluctuations due to infrequent scattering processes. A practical demonstration of the Brownian method is demonstrated through the study of a short diode where the simulation gave the correct value for the built in potential and a good agreement in the I-V characteristics to the Shockley model and a commercial drift-diffusion simulator.

The complex potential landscape arising from the Coulomb force, with its sharp localised peaks and troughs, faces problems due to band limiting in meshes and places heavy burdens on the integration techniques. A computationally efficient solution to the problem of band-limiting is presented and is shown to provide an accurate description of the electrostatics. This work also introduces a highly efficient and numerically stable multigrid solver, for Poisson's equation, that can cope with the complex potential distributions on

large meshes.

The study of ionised impurity scattering is used to validate these molecular dynamics simulations. Results have shown that the Brownian method — despite precluding the use of adaptive integration schemes — gives a good approximation to the standard results and has the advantage of smoothing away errors that can build up during the integration of motion and drives the system towards thermal equilibrium.

The greatest hurdle to be cleared before these three-dimensional simulations can be practicable is the sheer computational effort that is required. The implementation of the problem on parallel architectures has been explored and discussed.

The methods developed in this work are demonstrated through the simulation of an 80 nm dual-gate MESFET. The results were verified by comparing them with those from a commercial drift-diffusion simulator.

The threshold behaviour of devices has been investigated through the study of the formation of conduction channels in blocks. The percolation threshold gives the point when conductive paths form across the gate barrier. The results from the FET simulation was found to be in agreement with the earlier studies on the blocks.

PREFACE

The work carried out for this PhD had been completed by October 1996 and the first draft had been largely completed by December 1996. No additional material has been added to the work since then and so I hope that this fact is borne in mind when reading this work.

I have worked as a Research assistant for the early part of 1997 in the Electronics & Electrical Engineering Department before taking up the post of Research Engineer for Roke Manor Research Limited in Romsey in May.

I would like to convey my thanks and appreciation to my PhD supervisors, Dr Asen Asenov and Dr John Davies, for their help, advice and support. I am also grateful for the support given by my fellow device modellers — Dr Scott Roy, Dr Sharief Babiker and Andrew Brown. Sam Kilgour deserves special commendation for his valiant efforts to keep the Parsytec systems in working order and Stephen *et al* for managing the computer systems.

CONTENTS

Publications xi

1. *Introduction* 1
   - 1.1 Trends in miniaturisation 1
   - 1.2 Scaling submicron devices and possible limits to miniaturisation 2
   - 1.3 The need to consider the discrete nature of dopants and carriers 6
   - 1.4 A brief prospectus 12

2. *A simple method for discrete electrons* 16
   - 2.1 The Langevin Equation and its Discrete Time Approximation . 20
   - 2.2 Practical Implementation 24
   - 2.3 Verification 27

3. *Calculation of the electric field from discrete random charges* 37
   - 3.1 The solution of Poisson's equation 39
     - 3.1.1 Charge assignment 40
     - 3.1.2 The solution of Poisson's equation 43
     - 3.1.3 Obtaining the Electric Field 53
   - 3.2 The short range forces 55

4. *The dynamics of carriers in media with discrete impurities* 64
   - 4.1 Scattering models 67
     - 4.1.1 The Conwell-Weisskopf model 67
     - 4.1.2 The Brooks-Herring model 68
     - 4.1.3 Statistical Screening 70
   - 4.2 The dynamics of carriers in a Coulomb potential 73
   - 4.3 The unscreened impurity scattering 79
     - 4.3.1 The initial distribution 80

4.3.2 The numerical results 81
4.3.3 Screened carrier-impurity scattering 86
4.4 The threshold equipotentials in samples with discrete impurities 87
4.5 summary 91

5. *Atomistic device simulation* 92
5.1 The types of supercomputers 92
5.2 The parallel implementation of our models 95
5.2.1 Partitioning the problem on a network 95
5.2.2 Topologies and Message passing 99
5.2.3 Partitioning and particles 104
5.3 The atomistic simulation of a dual-gate MESFET 106

6. *Conclusions* 114
6.1 Scope for further study 117

Appendix 121

A. *The Langevin Equation* 122
A.1 The relation to the Drift-diffusion equation 125

B. *The Biconjugate gradient method.* 127

C. *Material parameters* 130

LIST OF FIGURES

2.1 The relaxation of velocity towards equilibrium. 26
2.2 Equilibrium velocity distributions for constant fields. 28
2.3 The simulation domain of the short diode. 30
2.4 The development of the built in potential. 32
2.5 V_T studied over a 0.4 ns. 33
2.6 The carrier distribution in the diode. 34
2.7 The temperature distribution across the diode. 35
2.8 I-V characteristics of the short diode 36

3.1 Aliasing in meshes. 42
3.2 BiCGSTAB(2) vs SOR 47
3.3 An illustration of the multigrid method. 50
3.4 MESFET test structure 51
3.5 Convergence of the multigrid method with γ and number of pre-smoothing steps 51
3.6 Comparison of the three solvers for a typical problem. 52
3.7 Extrapolation at the boundaries. 54
3.8 The force between two charges along the [100] direction. 55
3.9 Interlacing and the [100] mesh force 56
3.10 Interlacing and the [110] mesh force 57
3.11 Interlacing and the [111] mesh force 58
3.12 Mesh forces along the directions of symmetry for a cubic mesh . 59
3.13 The mesh force at a range of 1.0625 mesh spacings as a function of $\cos^2\theta$ & $\sin^2\phi$. 60
3.14 The interpolation weights as a function of (a) $\cos^2\theta$ and (b) $\sin^2\phi$. 62
3.15 The test for the approximation to the mesh force along the [411] direction. 63

4.1 The trajectory of an electron around an ionised donor. 65
4.2 The theoretical electron mobility for a neutral sample of bulk GaAs. 72
4.3 The geometry of the orbit of an electron around a donor. 75
4.4 Rutherford scattering by attractive field 77
4.5 Rutherford scattering by repulsive field 78
4.6 The mobility of non-interacting carriers at room temperature vs doping. 83
4.7 The mobility of non-interacting carriers in a bulk sample from ionised impurity scattering. 84
4.8 Electron dynamics in GaAs at 300 K using Brownian dynamics and Ewald summation. 87
4.9 A 2D example of a sample with periodic boundaries. 88
4.10 The average threshold potentials in $(50\,nm)^3$ samples containing randomly placed 125 donors. 90

5.1 The SMP architecture. 94
5.2 The MPP architecture. 94
5.3 The distribution of a 3D array on a 2D network is equivalent to mapping a 2D array of vectors on the network. The partition of a 10×10 array of vectors on a 3×4 mesh. 98
5.4 The partition of the 10×10 array of vectors on a 3×4 mesh that achieves a better load distribution. 98
5.5 The 2D mesh and torus 99
5.6 The data exchanged every partial update for the simple problem. 100
5.7 The simple binary tree. 100
5.8 The 'H'-shaped data path that is suited for rectangular topologies.. 101
5.9 The fylfot offers more data paths to the centre than an 'H'-shaped topology. 102
5.10 The wholesale shifting of data across the network. 102
5.11 The multigrid method on a rectangular topology 103
5.12 The use of buffers to exchange particles. 105
5.13 The problem of high order differencing and assignment schemes on MPP nodes. 106

5.14 Alleviating communications bottlenecks. 107
5.15 The double gate MESFET. 108
5.16 The I-V characteristics of the 80 nm double gated MESFET with a smeared doping. 109
5.17 The I-V characteristics of the 80 nm double gated MESFET with atomistic donors. 110
5.18 A cross section of the centre of the channel through the xy-plane at threshold. 112
5.19 A cross section of the centre of the channel through the yz-plane at threshold. 113

LIST OF TABLES

1.1 The constant field scaling of a typical MOSFET for a scaling factor of κ. . . . 3
1.2 The use of two scaling factors — κ for the potentials and λ for the lengths. . . . 3

2.1 The influence of the time-step on the velocity distribution. . . . 25

3.1 The discretisation of the MESFET 53

4.1 The effect of mesh spacing on ionised impurity scattering for a doping of $10^{24}\,\mathrm{m}^{-3}$ 85
4.2 The threshold potentials in $(50\,nm)^3$ blocks. 89

5.1 A $30 \times 30 \times 20$ mesh on a 2D rectangular topology. 104

PUBLICATIONS

The following publications have resulted during the course of this work:

1. C. R. Arokianathan, A. Asenov, and J. H. Davies, J. Appl. Phys **80**, 226 (1996).

2. C. R. Arokianathan, A. Asenov and J.H. Davies, Brownian approach to the simulation of small semiconductor devices,Presented at 4th International Workshop on Computational Electronics, Tempe (1995), VLSI Design, in press.

3. C. R. Arokianathan, J. H. Davies and A. Asenov, Ab-initio Coulomb scattering in atomistic device simulation, Presented at 5th International Workshop on Computational Electronics, Notre Dame (1997), VLSI Design, in press.

4. C. R. Arokianathan, A. Asenov and J.H. Davies, Mesh-based particle simulation of sub-0.1 micron FETs, SIMD'97 Maui, accepted for Semiconductor Science and Technology.

5. A. Asenov, A. Brown, S. Roy, C. Arokianathan, J. Davies and J.R. Barker, Parallel 3D Simulation of Semiconductor Devices, in Proc. Second NASA Device Modeling Workshop, Ames, pp. 85-99 (1997). Invited

6. J.R. Barker, A. Asenov, A.R. Brown, J. Cluckie, S. Babiker and C.R. Arokianathan, Parallel simulation of semiconductor devices, in Massively Parallel Processing, Applications and Development, Dds. L. Dekker, W. Smith, J.C. Zuidervaart, Elsevier, pp. 683 - 690 (1994)

An oral presentation of "Brownian Device Simulation" was given at the 4th International Workshop on Computational Electronics, Tempe (1995) by myself. A poster on "Ab-initio Coulomb scattering in atomistic device simulation" was presented, by Asen Asenov, at the 5th International Workshop on Computational Electronics, Notre Dame (1997).

1. INTRODUCTION

The main aim of this work is to develop the simulation methods that can be used to model ultra-submicron devices. It will be shown that the discrete nature of carriers and impurities need to be considered in order to study structures with feature sizes below $0.1\,\mu$m. The main thrust of the work is to examine the methods of dealing with particulate carriers in the complex and sharply varying potential landscapes due to the discrete charges.

This section aims to introduce the problem, while highlighting the major issues, and set out an outline for the body of the work. Firstly there will be an overview of the current trends in miniaturisation, followed by an outline of the issues when scaling devices down to the order of tens of nanometres. The possible limits to this scaling will be discussed while paying specific attention to the underlying device physics. The following section proposes the need to consider the particulate nature of carriers and impurities. And finally a summary of the remainder of this work will be given.

1.1 Trends in miniaturisation

The demand for faster microprocessors and larger memories has driven the semiconductor industry towards the fabrication of devices with smaller feature sizes. As a rough rule of thumb, Moore's law predicts a shrinkage of an order of magnitude in the feature sizes in VLSI technology every decade. The gate length is the most important feature size, as far as FET performance is concerned. Current state of the art integrated circuits using gate lengths of $0.35\,\mu$m have enabled processors, such as the PowerPC 604e [1] and the Pentium Pro [2], to be clocked at 200 MHz and have made 64 Mbit DRAM ICs possible. Next year, $0.25\,\mu$m technology will allow clock speeds over 800 MHz and push memory capacities to 256 Mbit. By around the year 2000, ICs should appear with gates that are shorter than $0.1\,\mu$m [3] and so pushing clock speeds

into the Gigahertz range and memory capacities into the order of Gigabits.

MOSFETs with gate lengths of 40 nm have been fabricated [4] using novel lithographic techniques such as resist thinning [5]. They have been found to display transistor action at room temperature and so raising the possibility of 128 Gbit DRAMs. This translates into a time-scale of a further 16.5 years of miniaturisation, by Moore's law.

This process of continuous miniaturisation begs the question whether there is a fundamental limit to this exercise. The following section discusses the behaviour of these small devices and looks at any pitfalls that may arise in the fabrication of integrated circuits using these devices.

1.2 Scaling submicron devices and possible limits to miniaturisation

There has been much speculation as to whether there is a fundamental limit to this trend in miniaturisation. Many times in the past people have claimed a lower limit on the size of a FET, and many times this limit has been exceeded. For the case of the MOS transistor, Hoeneisen and Mead [6] estimated the minimum oxide thickness to be 14 nm in order to prevent tunnelling; that the channel lengths should exceed 400 nm to avoid drain-source punch-through; and a maximum substrate doping level of around $2 \times 10^{27}\,\mathrm{m}^{-3}$ due to field emission in the source and drain junctions. There may well be other factors which may impose limitations on the extent of this miniaturisation process such as the fluctuations in the threshold characteristics. These limits have clearly been breached by today's mass production technology.

The first concern when shrinking FETs is how to scale the lengths, doping and potentials so that the proper transistor action is observed under normal operating conditions. The first problem is due to the short channel effects: as FETs get smaller their channel lengths become comparable to the widths of the depletion regions at the source and drain, and thereby reducing the influence of the gate and increasing the likelihood of punch-through. Short channel effects can cause severe degradation of subthreshold characteristics and an unacceptable dependence of threshold voltage on channel length. The simplest approach that minimises this effect is constant field scaling [7], that uses a

Tab. 1.1: The constant field scaling of a typical MOSFET for a scaling factor of κ.

Quantity	Scaled by	Limitation
Threshold Voltage	$1/\kappa$	Off leakage
Bias Voltages	$1/\kappa$	V_T, speed
Capacitances	$1/\kappa$	Packing density
Doping	κ	Junction leakage, mobility
Saturation Currents	$1/\kappa$	power dissipation
Current densities	κ	Electromigration
Maximum operating frequency	κ	
Switching power	$1/\kappa^2$	
Power dissipation	$1/\kappa^2$	

scaling factor ($\kappa > 1$) that shrinks the lengths and potentials by $1/\kappa$ to preserve the electric field distribution in the scaled device. The doping has to be scaled by κ. Table 1.1 shows how this scaling affects the physical characteristics of a typical FET. It shows that miniaturisation is highly advantageous because it gives higher transconductances and higher operating frequencies (which are further enhanced by velocity overshoot in submicron devices) with no increase in power dissipation per unit area on a wafer where the packing density has been increased by κ^2.

However not all characteristics scale linearly with this approach. An important exception is diffusion — being dependent on the temperature and mobility of carriers — which controls the subthreshold characteristics of the device. The built in junction potentials do not scale and so become larger

Tab. 1.2: The use of two scaling factors — κ for the potentials and λ for the lengths.

Quantity	Scaled by
Threshold Voltage	$1/\kappa$
Bias Voltages	$1/\kappa$
Capacitances	$1/\lambda$
Doping	λ^2/κ
Saturation Currents	λ/κ^2
Current densities	λ^3/κ^2
Maximum operating frequency	λ^2/κ
Power dissipation	λ/κ^3

in comparison to the smaller applied biases, as a consequence of this scaling law, and so give rise to higher fields and larger depletion regions and so a more elaborate scaling law (see table 1.2 using separate constants (κ for the potentials and λ for the lengths) has been proposed [8].

There are limitations to this rather simplistic approach. Electromigration places an upper limit on the maximum acceptable current density (*e.g.* $10^9\,\mathrm{Am}^{-2}$ for Al) [9]. The magnitude of the operating levels must be large in comparison to the thermal energy ($k_BT \approx 25\,m\mathrm{eV}$ at room temperature) for the sake of noise immunity. Increasing the channel doping — in the face of high transverse fields— leads to a reduction in channel mobility and so degrading the transconductance and maximum operating frequency.

An empirical formula has been found for the minimum gate length (L_{min}) that still exhibits long channel behaviour [10].

$$L_{\mathrm{min}} = A\sqrt[3]{dt_{\mathrm{ox}}(W_{\mathrm{S}} + W_{\mathrm{D}})^2}, \tag{1.1}$$

where d is the junction depth, t_{ox} is the oxide thickness and the widths depletion region are W_{S} and W_{D} respectively. A is a fitting parameter. The results from equation 1.1 seem to give intuitively obvious results since it predicts a larger L_{min} when either the junction depth or oxide thickness is increased where the influence of the gate is reduced. However the minimum gate length cannot shrink to zero as this equation may imply since there has to be some limiting value.

These empirical rules must only been seen as a rough guide and are no substitute for the proper modelling of these devices, especially under conditions where diffusion is likely to be important. The models themselves need to be fast and have to be validated against experimental results, in order for them to gain widespread acceptability. A new method based on Brownian motion, which meets these criteria, will be presented in chapter 2.

The prevailing view is that tunnelling through the oxide places a lower limit of $3\,n\mathrm{m}$ [11, 12], but devices with an oxide thickness of $1.5\,n\mathrm{m}$ have been fabricated [13] that have demonstrated transistor action at room temperature [4]. The reason for this may be that the tunnelling current — being proportional to the gate area — would become smaller in comparison to the source-drain current as the gate-length is reduced (to a greater degree than the width).

Tunnelling can also be assisted by traps and the charge transported through the oxide accumulates a 'charge to breakdown' [14, 15, 16]. Gate induced drain leakage (GIDL) is another major problem [17] that occurs when there is some overlap between the gate and highly doped drain. The thinner the oxide, the higher the capacitance and so a large charge is induced in the drain. The minority carriers in this non-equilibrium depletion region can be drained via the substrate. The high electric fields across the oxide promote tunnelling that can be facilitated by traps near the surface. Traps resulting from hot electron damage [18] and Fowler-Nordheim leakage [19] exacerbate gate induced drain leakage by introducing additional tunnelling. This problem requires the careful fabrication of these devices, so as to minimise the formation of traps, and the use of lower drain voltages as the oxide thickness is reduced.

The high fields give rise to hot electrons that increases the likelihood of injection into the oxide — inducing threshold voltage shifts and degradation in transconductance — and so shortening the lifetimes of the device. This effect is most marked in the region around the drain where the electric fields are very high. Great attempts have been made to overcome this very serious problem by rendering the oxide more immune to hot carrier damage [20, 21, 22] or by attempting to reduce the fields. The largest fields occur between the gate and the drain and the use of a lightly doped drain (LDD), with a spacer layer between the gate and drain, is becoming common practice [23, 24, 25, 26, 27, 28]. The reduction in the large lateral electric fields is more effective when the LDD region extends under the gate, but there is a penalty [29, 30] of increased channel resistance and an increased parasitic gate-drain capacitance. Extending the doping under the gate also increases the susceptibility to GIDL. High electric fields can lead to Zener-Bloch oscillations (of electrons between the top and bottom of the conduction band) and band to band tunnelling (of electrons from the valence to the conduction band) [31, 32].

The problems due series resistances and parasitic capacitances are exacerbated as a result of miniaturisation and higher operating frequencies. Self aligned contacts ameliorate the problem of series resistances by facilitating the formation of large area contacts and placing the metallised contacts over the field oxide greatly reduces the junction-substrate capacitances while also allowing looser tolerances on mask alignment. The use of self aligned silicide

has been shown to reduce contact resistances as well as parasitic capacitances. However the resistivity of the silicide is large for small patterns (less than 200 nm and the silicide process is difficult to control for shallow junctions less than 50 nm [33]. Thicker silicide layers produce larger contact resistances. One of the limitations to miniaturisation may come from interconnects [34]. Shrinkage results in increased resistance and the closer separation leads to increased capacitance that lengthens the $[RC]$ propagation delay and worsens the problem of cross-talk. The interconnects start to look like tall pillars at these small scales since the resistance is determined by the cross sectional area and so the scope for reducing the heights is restricted as the lines become narrower. The use of copper and its alloys would help because of its low series resistance and immunity to electromigration [35, 36, 37].

Isolation becomes a serious problem as packing densities are increased as a result of miniaturisation. The influence of the gate is reduced near the edges of the channel as the field lines spread apart. The narrow width effect is said to occur in this region where the threshold voltage is higher than in the middle. This region becomes a lager fraction of the total channel width as the device is shrunk and so can cause an increase in threshold voltage.

The impact of processing errors are expected to increase as devices are shrunk [38, 39, 40, 41, 42]. Fluctuations in the gate length and oxide thickness produce significant changes in the threshold voltage [43, 44, 45, 46]. The small size of current and future semiconductor devices introduces further complications because of the discrete nature of electronic charge. The next section maps out these problems and gives the motivation for this present work.

1.3 The need to consider the discrete nature of dopants and carriers

The previous section reviewed the possible pitfalls in the process of miniaturisation resulting from the physics of small devices. This section serves as in introduction to the main thrust of this work which argues that the small numbers of carriers and impurities in these mesoscopic devices will have a significant impact on their incorporation into future integrated circuits.

When devices are scaled down to less than 100 nm, there are only a few tens

or hundreds of impurities in the active region and the statistical variation in their number n (which varies as $\sqrt{n}$) becomes comparable to the numbers that are actually present. This has the effect of producing significant fluctuations in the $I - V$ characteristics between devices that are fabricated even on the same piece of substrate and poses some difficulty in the fabrication of integrated circuits using such small devices.

The first attempt to study this problem was made by Keyes in 1975 [47], who used a statistical approach and percolation theory to investigate the formation of conductive paths between two electrodes that represented a simple FET. He divided his devices into cubes — with dimensions that were the thickness of the depletion layer [48] — whose state of conduction was determined by their impurity concentration. As the threshold voltage is approached, the cubes with the lowest number of impurities turned on first, and as more and more of these turn on a conducting channel is formed form the source to the drain. The number of impurities in a cube follows a Poisson distribution, and from this the probability of this region being conductive can be calculated. And so, by estimating the probability of there being a conductive path between the source and drain, the random fluctuation in threshold voltage can be studied. In order to keep the algebra manageable, Keys only considered straight line paths between the contacts and calculated the fluctuations in threshold potentials from the variation in space charge due to the impurity numbers in the device. He extended his idea to to include more complex paths [49] through the use of percolation theory [50]. One limitation with this approach is that the cells have been assumed to have no influence from their neighbours. This is clearly not the case from the solution of Poisson's equation. For the case of a n-MOSFET, a cell with a smaller than average number of ionised acceptors represents a hill in the potential map and increases the chances that a neighbouring cell would conduct by increasing the local mean potential, and vice versa. In other words the charges also have a more global influence on the potential distribution.

Early simulations of MOSFETs using a two-dimensional drift-diffusion simulation, by Hagiwara *et al.* [51] demonstrated that the significance of the fluctuations in threshold voltage, due to variations in local doping, increased as the gate length was reduced. He deduced that the variance of the threshold voltage

fluctuations was inversely proportional to the volume of the depletion layer. After applying constant field scaling, he estimated the standard deviation in the threshold voltage (ΔV_{th}) as a fraction of drain bias (V_{DD}) as,

$$\frac{\Delta V_{\mathrm{th}}}{V_{\mathrm{DD}}} \approx L^{-\frac{3}{2}} \quad (1.2)$$

Later Nishinohara used a combination of two- and three-dimensional simulations to study the effects of shrinking on the threshold voltage [52]. A Poisson distribution was used to obtain an integer multiple of electronic charge for the charge density ascribed to each mesh point. They found both a reduction in threshold potential and an increase in its fluctuation as the MOSFET was shrunk.

Variations in the subthreshold characteristics and threshold voltages in sub-micron MOSFETs caused by random fluctuations in the local impurity distributions have been clearly demonstrated in simple 3D drift-diffusion simulations. H. Wong and Y. Taur, at I.B.M [53], used a 3-dimensional drift-diffusion simulator to investigate the effects of different random dopant distributions on silicon MOSFET's. The ionised dopants were represented as discrete entities on the nodes of a uniform mesh. Closer attention was paid to three device characteristics - the threshold current, the subthreshold gate voltage that produced a current of $2nA/\mu m$ and the subthreshold slope.

They found a spread of I-V curves along the gate voltage axis and a general reduction in threshold voltage, along with a slight degradation and fluctuation of the subthreshold slope. These effects appeared to increase with decreasing gate-controlled channel volume. The fluctuation of the number of dopants in the depletion region could not fully account for the variations in the threshold voltage and the average threshold voltage shift, in the simulation, was much larger than that expected due to the fluctuation in the number of dopants. This led them to the conclusion that these effects were caused by the discrete nature of the dopant charge, producing an inhomogeneous channel potential, which allowed for early turn-on in parts of the channel and resulted in the threshold voltage shift and sub threshold slope degradation. The asymmetry of the I-V characteristics, seen after interchanging the source and drain terminals, was taken as further evidence of the importance of this discrete microscopic dopant distribution.

Studies of MOS capacitors [54] have shown significant contributions to the the spreading of the capacitance-voltage curves due to fluctuations in the dopant distributions in the space charge region. This effect is most marked at higher doping concentrations, but at lower doping levels the random interface charges [55] play a more important role along with fluctuations in oxide thickness.

Measurements on 8192 0.1 μm n-MOSFETs by Mizuno *et al.* [56, 57] have demonstrated these fluctuations in threshold voltage. The devices had been patterned on a small piece of the wafer (0.7 mm) to minimise the other sources of fluctuation, such as variations in oxide thickness or doping levels. By studying the variation of the threshold voltage with position and its autocorrelation, they showed that the magnitude fluctuations were completely random and not related to the location of the device. Moreover the fluctuations were shown to follow a Gaussian distribution whose width was larger for devices with a shorter gate.

A rough calculation showed that the fluctuation in the number of dopants in the channel contributed to 60 % of the threshold fluctuations. A later analysis of the experimental results [57] showed that the standard deviation of fluctuation in dopant number (n_a) was actually $\sqrt{n_a}/2$. They considered the variation in dopant distribution in the channel as a combination of a lateral and vertical component as seen by the relationship between the threshold voltage fluctuations and the drain and substrate biases respectively. The experimental results indicated that including the fluctuations in the distribution of impurities as well as the variation in their number in the analytic model could give a good approximation to the magnitude of the threshold voltage fluctuations.

J.-R. Zhou and D. K. Ferry looked at $sub-0.1\mu$m MESFETs [58], of total length 360 nm and depth 100 nm, with a 24 nm gate. They investigated the dependence of the I-V characteristics on the device widths for widths in the range of 42 nm to 162 nm, using a hydrodynamic model. The doping in the channel was $1.5 \times 10^{24}\,\mathrm{m}^{-3}$ and so, for a width of 42 nm, they found a mean of 36 impurities under the gate with a standard deviation of 5.99.

Not surprisingly, they found significant fluctuations in the I-V characteristics which depended on both the number of impurities under the gate and their positions. These fluctuations were more pronounced whenever a large nega-

tive gate bias was applied since even fewer impurities were present in narrowed conducting channel.

J. H. Davies *et al.* [59] have showed the random distribution of donors in the δ-doped layer above a two-dimensional electron gas can give rise to noticeable fluctuations in threshold voltage. However, the length scales of these fluctuations were too large for them to be caused by individual dopants. This reflects the physical separation of the confined electrons from the donors by the spacer layer.

A two dimensional analogy (for a flow of electrons) can be a dried up river bed where the ionised acceptors can be viewed as peaks and donors as pits. The landscape is also not completely flat since the height of the ground is a little elevated in regions containing more than an average density of donors (or where there are fewer acceptors) and depressed where the donors are scarce or acceptors are clustered. As the water level (*ie* the Fermi level) rises the the hollows and pits are the first to fill and they begin to merge. Little trickles (or conducting paths) form as the percolation threshold is approached. As the level rises these paths become larger and merge. They are also augmented by additional pathways that appear as a result of the rising level. The total rate of flow then rapidly increases above the conduction threshold.

In a real device (as threshold is approached) 'bubbles' of conducting regions form in the device which expand, following their local equipotential lines, and coalesce as threshold is approached to form these conducting paths. The shorter the device the fewer these regions that are needed to form a conducting path. A small cross sectional area reduces the statistical averaging and so the subthreshold level is smeared to a greater extent. As a result this effect is expected to increase in smaller devices.

The particulate nature of carriers is also expected to have a significant influence in the behaviour of mesoscopic devices. There are a small number of carriers in the channel and so the trapping and de-trapping of individual carriers can produce an appreciable random telegraph noise [60, 61, 62, 63]. The oxide traps affect the charge transport by altering the number of carriers and by introducing additional Coulomb scattering centres (whose effect are exacerbated by their small numbers). The distribution and number of these traps and defects vary from one device to another and so this effect varies

considerably for different devices. Individual carriers can induce modulations in the potential landscape, of very small devices, that can significantly affect the blocking characteristics of the barrier — especially between the source and channel [64, 65, 66, 67]. This effect is small as it appears only at liquid helium temperatures for a 100 nm MOSFET and in a 1 nm MOSFET at room temperature [64].

The small number of carriers raises other considerations. The separation between carriers is often larger than the Debye screening length (for example, in GaAs where the electron density is 10^{24} m^3 the screening length $L_D = 4.3\,nm$ while the average electron separation is 10 nm). Popular models that are used to represent screening consider the screening carriers as a continuous distribution of charge. This is often not true when there is less than one intervening electron, on average, that can contribute to screening and when transit times are short. Carriers have to be treated as individual entities and all electrostatic interactions must be considered.

Methods of studying systems using particulate carriers have been proposed by some authors to study deep submicron MOSFETs. The early work used 2D simulations [68] since the major problem was the computational expense. Monte Carlo simulations — where all Coulomb interactions are considered as interactions between discrete point-like particles — have been performed by Ferry *et al* [65]. They used the method of Ewald summation to calculate the electrostatics in their samples with periodic boundaries. Molecular dynamics simulations of actual MOSFETs have been performed by Tarnay *et al* [69] who solved Laplace's equation on a mesh to resolve the slowly varying fields from the boundaries and electrodes, while analytically calculating all forces between the point charges. This method adequately copes with the boundaries while preserving the discreteness of electronic charge. The major concern is that the electric field, resulting from the discrete charges, must be included for the von Neumann boundaries and the corresponding potentials added to the Dirichlet boundaries in order to treat the electrostatics properly. This method suffices for 2D or small 3D meshes with relatively few particles, but becomes very expensive for larger meshes. It will be shown (in chapter 3) that the mesh derived forces are accurate for particle separations that are greater than three or four mesh spacings and so much of this additional effort is unnecessary.

1.4 A brief prospectus

The aim of this work is to develop methods that can be employed to investigate the effects caused by these discrete impurities and to look at the results from using discrete carriers. It will be shown that the discrete impurities can have significant effects on the potential map in their immediate vicinity. I will also address the problem of ionised impurity scattering. This is a complex many-body problem that is not well understood.

This is a new field and the aim of this work is to study the various approaches to modelling these small devices and highlighting their pitfalls. The simulation of carrier dynamics has three main components:

1. implementing scattering,
2. resolving the forces acting on the carriers,
3. integrating these sharply varying forces.

Modelling the scattering processes lies at the heart of all semiconductor modelling. The standard approach is based on the solutions to the Boltzmann transport equation and the next chapter will deal with this matter. It will be argued that the main problem with this equation is that it only applies to continuous charge distributions and does not represent the stochastic nature of the collision events. Models based of the solution of partial differential equations (arising from the first few moments of the Boltzmann equation) will be reviewed briefly. Particulate models (such as the Monte Carlo method and Cellular automata) will then be outlined briefly. Monte Carlo methods have a long history and provide an accurate semiclassical representation of carrier movement [70, 71]. However they are costly to implement, despite the efforts to improve their speed using variance reducing techniques [72] and reducing 'wasted' effort from self-scattering [73]. Cellular Automata have been used for device simulations only fairly recently [74, 75, 76, 77, 78]. Their proponents claim that this method is much faster than traditional Monte Carlo techniques [79], but however the large memory requirements of these methods make them less favourable on distributed computing architectures and on systems with limited caches.

In the next chapter, I will describe a new approach to device simulation — based on Brownian dynamics and the Langevin equation — that does view scattering as a continuous and random process. It is very useful when a particulate representation of carriers is required but where the full-blown Monte Carlo method is not necessary or too expensive. Apart from its simplicity our method may be particularly useful for isolating and studying some specific aspects of the carrier dynamics related to the 'atomistic' nature of carriers and impurities. It may be more transparent than the Monte Carlo technique in distinguishing the effects due to the discrete nature of particles and their interaction with impurities, because these effects are not obscured by structure arising from the other scattering rates. I shall start with a formal description of the Langevin theory and its discrete time approximation. The practical implementation of the Brownian dynamics approach is outlined and tested in a simulation of a p-n junction diode. The results of this test are in excellent agreement both with the ideal diode equation and with the drift-diffusion results obtained from the commercial simulator MEDICI.

The following chapter concentrates on the important question of resolving the forces that drive the carriers and highlights the problems caused by the electrostatic forces between the discrete charges. This problem is the major challenge because the forces have to be calculated in a simulation domain that can contain a large number of particles and have complicated boundaries. It would be impractical to calculate all inter-particle forces directly because of their large numbers and, worse still, there can be an infinite number of image charges resulting from the boundary conditions. Attempting to save on computational effort by solving Poisson's equation on a mesh poses more problems. The sharp spikes resulting from the discrete nature of the charges impede the convergence of most solvers and as the charges move, a fresh solution is needed every time-step. Traditional simulation methods evolve towards a constant steady state distribution and so less effort is needed to as convergence is reached, but this does not happen here. I have developed a very efficient multigrid method for the solution of Poisson's equation that can cope with large three-dimensional meshes containing discrete charges which can also be implemented on massively parallel computers.

It will be shown that the solution of Poisson's equation on a mesh band

limits the force and underestimates the force between particles when they approach within around three mesh spacings. This can be a problem when describing the scattering between discrete charges and so a new heuristic method has been devised to correct the erroneous force that has been derived from the mesh. The cubic symmetry of the mesh has been exploited to create a set of interpolating functions to estimate the inaccurate component of the mesh force between the particles. This wrong result can be corrected with an analytic determination of the forces between the particles in close proximity. Therefore analytic calculations of the inter-particle force has been restricted to only those charges that are within four grid spacings.

Chapter 4 will concentrate on the influence of the discrete ionised impurities on the dynamics of carriers. Three standard models will be reviewed (which are the Conwell-Weisskopf, the Brooks-Herring and Ridley's third body exclusion technique) and the effects of considering the asymptotes between the incoming and outgoing trajectories — despite the fact that the separation between the ions is finite — will be discussed. Firstly the ability of the simple techniques, for obtaining the forces from a mesh, that are used in standard Monte Carlo simulations to model ionised impurity scattering will be investigated. The short range forces are the most important in ionised impurity scattering and so the method of Ewald summation was used to obtain analytic estimated of the forces between charges so that the effects due the errors introduced by the finite mesh spacing can be studied.

The difficulties of integrating the r^{-2} forces will also be discussed. It is important to integrate the sharply varying forces accurately and that the energy of a particle is conserved when using free dynamics. An adaptive Runge-Kutta method has been implemented to study the motion of free carriers. This is not such a great problem in Brownian dynamics since the combination of the frictional and stochastic terms tend to erode any errors in integration and drive the system toward thermal equilibrium.

The last part of chapter 4 has great relevance in describing the characteristics of devices at threshold. The development of conductive pathways across the channel, as threshold is approached, can be treated as a percolation problem. The random distributions of charges in a material produce variations in the threshold potential throughout the sample. This means that some regions

are turned on before others and so, as threshold is approached, these regions would expand and fuse with one another forming tortuous pathways that snake across the gate barrier. The standard view was to divide the sample into arrays of independent cells — each of which has a turn on potential that is determined by its doping level. This discrete percolation model will be shown to be incorrect since there is a degree of interaction between different regions of a material because local fluctuations in doping can also influence the long range potential profile. The reduction in the percolation threshold in the samples, due to the presence of random arrangements of particulate impurities, will be compared with the results from FET simulations.

The penultimate chapter is concerned with the practical implementation of a device simulator with particulate charges. The major obstacle to the molecular dynamics simulation of these ultrasmall FETs is the heavy computational overhead inherent in this problem. Therefore, the implementation of these simulations on supercomputers is outlined with particular emphasis placed on the massively parallel processing (MPP) architectures of the department's Parsytec systems. Efficient methods of partitioning the problem onto a network of computing nodes will be explored and their influence on the implementation of the solution will be discussed. Finally, the study of an $80\,nm$ dual-gate MESFET is given as a practical demonstration of the general concepts and methods developed in this work. The results will be validated by comparing them with those from a commercial drift-diffusion simulator. The influence of the atomistic impurities will also be considered.

Finally the conclusions reached during the course of this PhD will be presented and suggestions will be made into how this work could be extended.

2. A SIMPLE METHOD FOR DISCRETE ELECTRONS

In this chapter a new method computationally efficient method based on the theory of Brownian motion is presented. Firstly the current approaches will be discussed in order to elucidate the need for this approach.

A broad range of methods are used for simulating semiconductor devices, varying widely in their accuracy, predictive power and computational needs. The simplest approach is to construct an equivalent circuit model and to obtain the required parameters by fitting the model to the measured characteristics of actual devices. Such models are very fast, requiring little computational effort and are commonly employed in packages such as SPICE. They are essential for circuit design but their physical predictive capability is very restricted.

In order to enhance the predictive capabilities of the simulations, where new physical phenomena start to influence the device characteristics, one must pay attention to the carrier dynamics in response to the potential distribution across the device. The foundation of the semi-classical approach for the simulation of devices is the Boltzmann transport equation (BTE) which states that the carrier distribution function $f(\mathbf{r}, \mathbf{k}; t)$, in phase space and time, obeys the following differential equation,

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \dot{\mathbf{k}} \cdot \nabla_k f = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} . \tag{2.1}$$

The right hand side of this equation represents the overall rate at which the carriers enter the state $(\mathbf{r}, \mathbf{k})$ from the rest of phase space, at a given time, due to scattering processes to which they are subjected.

$$\left(\frac{\partial f}{\partial t} \right)_{\text{coll}} = \frac{V}{(2\pi)^3} \int \left[f(\mathbf{k}')(1 - f(\mathbf{k}))W(\mathbf{k}', \mathbf{k}) - f(\mathbf{k})(1 - f(\mathbf{k}'))W(\mathbf{k}, \mathbf{k}') \right] d\mathbf{k}' \tag{2.2}$$

where V is the total volume of the sample and $W(\mathbf{k}_1, \mathbf{k}_2)$ is the total transition rate from state $(\mathbf{r}, \mathbf{k}_1)$ to $(\mathbf{r}, \mathbf{k}_2)$ from all scattering processes. The first set of

terms account for the scattering rate into the state $(\mathbf{r}, \mathbf{k})$ while the second indicate the total rate out of this state. The Pauli exclusion principle requires the $(1 - f)$ factor but in non-degenerate semiconductors where $f \ll 1$ this is often neglected. This model, therefore, neglects the stochastic nature of the scattering and replaces them with flow rates between the different points in phase space.

The quantities of interest, such as the current density across the device, can be extracted from this distribution function. The BTE, however, is complicated and difficult to solve by conventional numerical techniques. (Iterative methods have been proposed by some authors [80, 81, 82].) It also requires considerable input in the form of band structure and scattering rates for accurate results. Indeed these input parameters are often doctored by many device modelers so as to obtain good agreement between simulation results and experiment.

In large devices the carriers can be represented as a continuous distribution of charge whose flow is described by partial differential equations. In this case the BTE within the relaxation time approximation is often simplified by taking its first few moments.

The simplest of these approximations to the BTE is the drift-diffusion (DD) approach [83]. It arises from the zero and first moments of the BTE, which are basically statements of particle and momentum conservation.

$$\frac{\partial n}{\partial t} - \nabla \cdot (n\mathbf{v}) = G - R, \tag{2.3}$$

where G and R are the generation and recombination rates respectively.

$$\mathbf{J}(\mathbf{r}, t) = qn\mu\mathbf{E}(\mathbf{r}, t) - qD\nabla n(\mathbf{r}, t). \tag{2.4}$$

As the name implies, the current arises from a diffusive component driven by the local density gradients and a drift component driven by the local electric fields. It is assumed that the carriers are close to thermal equilibrium, and that their velocity depends only on the local electric field (i.e. they have the same drift velocity as in a large sample with a constant electric field of that value). Although it is unsophisticated, the DD approach rarely gives qualitatively wrong results since it is derived from conservation laws. The exceptions are when the fine details of the scattering rates are important such

as in the Gunn effect. Because of the local relationship between the mobility and the electric field, even in extensions with field-dependent mobility, the drift diffusion approach also cannot describe overshoot effects with great accuracy in small devices.

A more sophisticated approximation to the BTE yields the hydrodynamic approach [84]. It can be derived from the BTE by retaining the full momentum conservation equation and by adding the equation for energy conservation, derived from the second moment of the BTE. This gives small sets of partial differential equations which are amenable to numerical solution.

$$\frac{\partial \mathbf{p}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{p} + q\mathbf{E} + \frac{1}{n}\nabla(nk_{\mathrm{B}}T) = -\frac{\mathbf{p}}{\tau_p}, \quad (2.5)$$

$$\frac{\partial \mathcal{E}}{\partial t} - \mathbf{v}\cdot\nabla\mathcal{E} + q\mathbf{E}\cdot\mathbf{v} + \frac{1}{n}\nabla\cdot(k_{\mathrm{B}}Tn\mathbf{v}) - \frac{1}{n}\nabla\cdot(\mathcal{K}\nabla T) = -\frac{\mathcal{E}-\mathcal{E}_0}{\tau_{\mathcal{E}}} \quad (2.6)$$

The carrier distribution function has been replaced with the carrier concentration and $\mathbf{p}$ is the momentum and $\mathcal{E}$ is the total energy of the carriers. T is the carrier temperature and τ_p and $\tau_{\mathcal{E}}$ are the momentum and energy relaxation times respectively. These relaxation times need not be constants but can be functions of momentum or energy that can be physically measured. They can also be extracted from Monte Carlo simulations. $\mathcal{K}$ is the thermal conductivity.

In the hydrodynamic (HD) approach the momentum relaxation time and consequently the mobility are related to the average energy of carriers rather than to the local electric field. As a result, non-equilibrium effects like spatial and dynamic overshoot are included in the HD simulations. Phenomenological energy-dependent relaxation times — often extracted from steady-state Monte Carlo simulations or from experiment — are used.

These PDE methods, and especially the more elaborate modifications of the HD approach, can result in complicated sets of equations which are difficult to solve and whose boundary conditions are difficult to define. Also, we are concerned with devices containing a small number of carriers whose discrete nature is expected to be important and so a particulate representation of carriers is required.

The ensemble Monte Carlo method [72, 85] provides an exact solution to the Boltzmann transport equation [86] by following the phase-space trajectories of an ensemble of fictitious particles through the simulated device. The scattering

processes are treated as discrete events separated by 'times-of-flight' which are chosen at random, using the total scattering rate. Many such flights for all fictitious particles in the device are used to build up a statistical picture of the device. This is a computationally intensive process because the statistical uncertainty after N measurements varies as $\sqrt{N}$. Thus a high density of data is required throughout phase space although variance-reducing techniques allow some effort to be saved. As well as its application to strongly non-equilibrium transport effects, MC is the only approach to provide *ab initio* information on the intrinsic noise properties of a device.

Phase space can be subdivided into cells and the BTE can be integrated over these to yield transition rates between them. This is the essence of the cellular automaton approach [75] to direct solution of the BTE. By choosing finite time steps, phase space can be restricted to a finite set of cells with a finite number of allowed transitions between them. This enables the use of look-up tables to determine a possible final state for each initial state after every time step, and so can result in the saving of much computational effort. The problem is that transitions for a state with energy $\mathcal{E}$ are not localised in phase space since the final state can lie anywhere within a shell of bounded by $\mathcal{E} - \Delta\mathcal{E}$ and $\mathcal{E} + \Delta\mathcal{E}$. The sheer volume of data associated with this approach requires huge computer memories.

Our interest in ultra-small structures necessitates an 'atomistic' approach that follows the dynamics of individual particles in three dimensions. Clearly this is computationally expensive even before any description of scattering has been considered. Fortunately there are several situations where a simple model for scattering is adequate, we need not have recourse to the detailed treatment of typical MC programs, and a much more economical approach can be used. For example, the electric field is small and carriers are only weakly perturbed from equilibrium in large regions of conventional devices. Moreover, the particular focus is on new features due to the discrete nature of carriers, and these should not depend qualitatively on the precise form of scattering. I therefore propose the use of a Langevin equation, which satisfies most of these requirements for accurate dynamics while treating scattering in an extremely straightforward way.

The Brownian simulation approach described here is essentially an ensem-

ble Monte Carlo method with a greatly simplified scattering term. This makes it much cheaper to implement, while retaining a good description of the dynamics of individual carriers and their scattering from randomly distributed individual charges. The Brownian approach also retains to some extent the ballistic description of particle dynamics, which is essential as length scales approach the mean free path. The Langevin equation on which this method is based shall now be described and two approximations in discrete time, one in velocity space and the other in real space.

2.1 The Langevin Equation and its Discrete Time Approximation

The motion of carriers in the semiconductor device can be modelled as that of a collection of particles undergoing Brownian motion (as described physically back in 1828 by R. Brown [87] and mathematically in 1905 by A. Einstein [88]). This approach is based on Langevin theory [89, 90, 91] can then be used to describe the motion of carriers in the semiconductor device. It assumes that the acceleration of a carrier, of mass m and charge q, is due to two forces: the electric field $\mathbf{E}$ within the device and a force $\mathbf{F}_{sc}(t)$ which represents the scattering. The latter force may itself be resolved into two parts, an average frictional part $-m\mathbf{v}/\tau$ (which can instead be written as $-e\mathbf{v}/\mu$ where μ is the mobility of the carriers), and a rapidly fluctuating random force $\mathbf{F}(t)$ which averages to zero over a long time. The Langevin equation is then

$$m\frac{dv(t)}{dt} = q\mathbf{E}(\mathbf{r},t) - \frac{m\mathbf{v}(t)}{\tau} + \mathbf{F}(t) \ . \tag{2.7}$$

An important feature of this approach is that the impurity and the carrier-carrier scattering can be treated ‘atomistically’ by assuming that $\mathbf{F}(t)$ represents only the phonon scattering; the impurities and other carriers exert their influence on the motion of the individual particle through modulations of the electric field $\mathbf{E}(\mathbf{r},t)$.

Using the Langevin approach the complicated scattering mechanisms of the Monte Carlo have been reduced to two terms. The frictional term that is used to model the dissipative effects of the scattering process employs the momentum relaxation time τ_m, that can be obtained from the mobility μ.

The random fluctuating force $\mathbf{F}(t)$ represents the stochastic kicks received by the particle from the phonon bath. This random acceleration is usually represented as a white noise having a mean of zero and being delta-correlated (*i.e.* its autocorrelation function is some constant times a delta function). Its magnitude is set by the equipartition theorem.

The balance between the frictional and the fluctuating forces leads to a Boltzmann distribution of velocity in thermal equilibrium. There is already some averaging inherent in this method, associated with the frictional force and incorporated in the momentum relaxation time, and so fewer flights in the simulations may be required to get statistically acceptable measurements.

The ensemble average of Equation 2.7 can be solved in a constant electric field to give the time dependence of the expected carrier velocity given that $\mathbf{v} = \mathbf{v}_0$ at $t = 0$,

$$\langle \mathbf{v}(t) \rangle = \mu \mathbf{E} + (\mathbf{v}_0 - \mu \mathbf{E}) \exp\left(-\frac{t}{\tau}\right) . \tag{2.8}$$

After a sufficiently long time the average velocity relaxes to $\mu\mathbf{E}$ as expected from Ohm's law.

In order to derive a discrete-time approximation to the Langevin equation, which could be employed in practical simulations, equation 2.7 is integrated over a short time interval Δt. The simplest approach is Euler integration which gives

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \left[\frac{q\mathbf{E}(t)}{m} - \frac{\mathbf{v}(t)}{\tau} \right] \Delta t + \Delta \mathbf{V}(t) \tag{2.9}$$

where

$$\Delta \mathbf{V}(t) = \frac{1}{m} \int_t^{t+\Delta t} \mathbf{F}(t') dt'.$$

The random force $\mathbf{F}(t)$ requires special treatment and has to be integrated separately because it is rapidly varying on *any* time scale. It averages to zero ($\langle \Delta \mathbf{V}(t) \rangle = 0$) and is δ-correlated

$$\langle \Delta \mathbf{V}(t) \Delta \mathbf{V}(t + s) \rangle = \sigma^2 \delta(s). \tag{2.10}$$

This auto-correlation function indicates the degree of influence that a value, at any instant in time, has on the future and so it can be viewed as a measure of the duration of the 'memory' of the fluctuating force. The δ-function clearly

indicates that the value of $\Delta\mathbf{V}$ at any instance is totally unrelated to its value at any other point in time. This is the property of 'white noise'.

The statistical properties of the integral $\Delta\mathbf{V}(t)$, called a Wiener process, can be obtained through the central limit theorem. This states that the sum of an infinite number of uncorrelated random events follows a Gaussian distribution. The random force induces a random walk in velocity space, and the associated diffusion coefficient D_v (in velocity space) must be related to τ through the fluctuation-dissipation theorem. The variance of the distribution of $\Delta\mathbf{V}(t)$ can therefore be written in the two forms

$$\langle(\Delta\mathbf{V}(t))^2\rangle = 2D_v\Delta t = 2\frac{k_\mathrm{B}T}{m}\frac{\Delta t}{\tau} \tag{2.11}$$

where k_B is Boltzmann's constant and T is the equilibrium temperature of the particles. This equation also relates the diffusion coefficient in real space D_x to D_v via the Einstein relation

$$D_x = D_v\tau^2. \tag{2.12}$$

The time integrated version of the Langevin equation can be written as a finite difference equation . Introducing a discrete time $t_n = n\Delta t$ and the notation $\mathbf{v}_n = \mathbf{v}(n\Delta t)$,

$$\mathbf{v}_{n+1} = \gamma\mathbf{v}_n + \frac{q}{m}\mathbf{E}_n\Delta t + \Delta\mathbf{V}_n \tag{2.13}$$

where $\mathbf{E}_n = \mathbf{E}(n\Delta t)$ and $\gamma = 1 - \Delta t/\tau$. It can be shown that this discrete-time approximation to the Langevin equation can be solved in a constant field for the expected velocity at the nth time step to give

$$\langle\mathbf{v}_n\rangle = \mu\mathbf{E} + (\mathbf{v}_0 - \mu\mathbf{E})\gamma^n . \tag{2.14}$$

The analogy with equation 2.8 in continuous time is clear. For this system to relax towards the true solution, it is required that $\gamma < 1$ (*i.e.* $\Delta t < \tau$). For $\Delta t \ll \tau$ the discrete-time result approximates very well the transient in the continuous-time solution since

$$\gamma^n = \left(1 - \frac{\Delta t}{\tau}\right)^n = \left(1 - \frac{t}{n\tau}\right)^n \longrightarrow \exp\left(-\frac{t}{\tau}\right) \tag{2.15}$$

as $n \longrightarrow \infty$. This expression verifies that $\langle \mathbf{v}_n \rangle \longrightarrow \mu \mathbf{E}$ when $n \longrightarrow \infty$ in the correct manner.

The treatment above used the simplest approach to integrating the equation of motion. A better discrete time approximation would use the average velocity during the time-step, $(\mathbf{v}_n + \mathbf{v}_{n+1})/2$, in the frictional term of equation 2.9. This gives

$$\beta \mathbf{v}_{n+1} = \gamma \mathbf{v}_n + \frac{q}{m} \mathbf{E}_n \Delta t + \Delta \mathbf{V}_n \tag{2.16}$$

where $\beta = 1 + \Delta t/(2\tau)$. This improved approximation to the Langevin equation has a similar solution for the expected velocity exactly as equation 2.14 but with a different factor

$$\gamma' = \frac{\gamma}{\beta} = \frac{\left(1 - \frac{\Delta t}{2\tau}\right)}{\left(1 + \frac{\Delta t}{2\tau}\right)}.$$

This gives a better approximation for the exponent in the continuous time solution with half the error due to discretisation.

The discrete time formalism described above, based in velocity space, is only valid for time steps smaller than the momentum relaxation time τ_m. This physically obvious result is confirmed by equation 2.15. Over time-scales longer than the momentum relaxation time, the velocity would have largely relaxed to its equilibrium distribution and so $\langle \dot{\mathbf{v}} \rangle \approx 0$. The motion may then be viewed as a random walk in *real* space and the Brownian motion of the carriers can then be described by a stochastic equation in position

$$\dot{\mathbf{x}}(t) = \mu \mathbf{E}(\mathbf{x}, t) + \mathbf{V}(t) \tag{2.17}$$

which can be derived from the Langevin equation 2.7 by letting $\langle \dot{\mathbf{v}} \rangle = 0$. Now the particle's velocity has only two components. The drift $\mu \mathbf{E}$ contains both the acceleration due to the electric field and the average retarding force, while $\mathbf{V}(t)$ provides stochastic diffusion. This is close to the drift-diffusion picture; the random walk induced by $\mathbf{V}(t)$ leads to diffusion, and the balance between this and $\mu \mathbf{E}$ ensures that the particles acquire a Boltzmann distribution of velocities in equilibrium.

Integration of equation 2.17 over a time interval Δt leads to

$$\Delta \mathbf{x} = \mu \mathbf{E} \Delta t + \Delta \mathbf{X} \tag{2.18}$$

where $\Delta\mathbf{X}$ is another Wiener process, a random displacement which averages to zero and has a Gaussian distribution with a variance given by

$$\langle(\Delta\mathbf{X}(t))^2\rangle \;=\; 2\,D_x\,\Delta t \;=\; 2\frac{k_B T\,\tau\,\Delta t}{m}\,. \tag{2.19}$$

Here D_x is the diffusion coefficient in real space. Introducing a discrete time $t = n\Delta t$ as before, equation 2.18 can be written in terms of finite differences,

$$\mathbf{x}_{n+1} \;=\; \mathbf{x}_n \;+\; \mu\,\mathbf{E}_n\,\Delta t \;+\; \Delta\mathbf{X}_n \tag{2.20}$$

This is a discrete analogue of the drift-diffusion equation and can be used in device simulation, but all description of ballistic processes has been lost. The time step Δt is no longer limited by the momentum relaxation time but by the scale on which $\mathbf{E}(\mathbf{x}, t)$ varies: a 'jump' $\mathbf{x}_{n+1} - \mathbf{x}_n$ must be smaller than the spatial scale of $\mathbf{E}$, and Δt should itself be small on the temporal scale of $\mathbf{E}$. Note that the jump contains contributions from both $\mathbf{E}$ and $\Delta\mathbf{X}_n$, which itself grows as Δt is increased.

The practical implementation of the Brownian Dynamics method shall now be considered.

2.2 Practical Implementation

The implementation of the Brownian simulation approach based on the discrete time approximation of the Langevin equation is straightforward. The only input parameters required are the mobilities and effective masses of the carriers and the lattice temperature, as in the drift-diffusion method. The procedures involved in the simulation are very similar to those used in the ensemble Monte Carlo approach. The Brownian simulation is an ensemble technique where the particles are moved at each time step in response to local electric fields. The simplest way of obtaining the fields is to subdivide space into a mesh and solve Poisson's equation, after assigning the charges to the appropriate nodes, using some weight function. The resulting potential map can be differenced to get the local electric fields and so, while the motion is continuous, the fields are calculated from discrete values. In very small devices, each of those particles may represent an individual carrier. Otherwise

the concept of 'superparticles' should be used, where each superparticle behaves dynamically like a single carrier but represents multiple charges in the solution of the Poisson's equation. After some transition time required for equilibrium to be established, the quantities of interest are extracted and the whole process is repeated until all measurements have been made.

For the time step $\Delta t < \tau$, the formalism based in velocity space can be used. The velocity $\mathbf{v}_{n+1}$ of a particle at the end of the nth time step can be calculated from the velocity $\mathbf{v}_n$ at the beginning of the step using the finite difference equation 2.16. The trapezoidal integration in time (where the average velocity is used to determine the drag) gives a much improved result for very little computational expense (as shown in figure 2.1). The components of the random velocity $\Delta\mathbf{V}_n$ are randomly chosen from a Gaussian distribution of variance $2D_v\Delta t$. This procedure gives the correct statistical distribution of $\Delta\mathbf{V}_n$ immediately and so it is beneficial in the simulation of very small devices where particles take a small number of time steps to traverse the active region.

Tab. 2.1: The shape of the velocity distribution with time-step for (a) fixed magnitude ($\pm\sqrt{2D_v\Delta t}$) and (b) Gaussian Weiner process.

Gaussian

Abs. dev.	Mean	σ	Skew	Kurtosis (-3)
$\sqrt{\frac{2}{\pi}} \approx 0.7979$	0	1	0	0

(a)

$\Delta t/\tau$	Abs. dev.	Mean $\times 10^{-12}$	σ	Skew $\times 10^{-3}$	Kurtosis (-3)
1	0.9458	1.438	1.006	2.403	-1.604
$\frac{1}{2}$	0.8360	3.016	0.9966	-12.68	-0.9409
$\frac{1}{3}$	0.8266	1.422	1.004	-5.406	-0.6615
$\frac{1}{20}$	0.8022	2.289	0.9994	11.21	-0.1293
$\frac{1}{324}$	0.7973	-0.1144	1.000	1.961	-0.01149

(b)

$\Delta t/\tau$	Abs. dev.	Mean $\times 10^{-12}$	σ	Skew $\times 10^{-3}$	Kurtosis (-3)
1	0.7984	-0.4105	0.9997	5.355	-0.01063
$\frac{1}{2}$	0.7983	-2.323	1.002	-22.74	0.009409
$\frac{1}{10}$	0.7946	2.127	0.9927	-5.047	0.008555

The last term in equation 2.16 is a random walk in velocity space and can

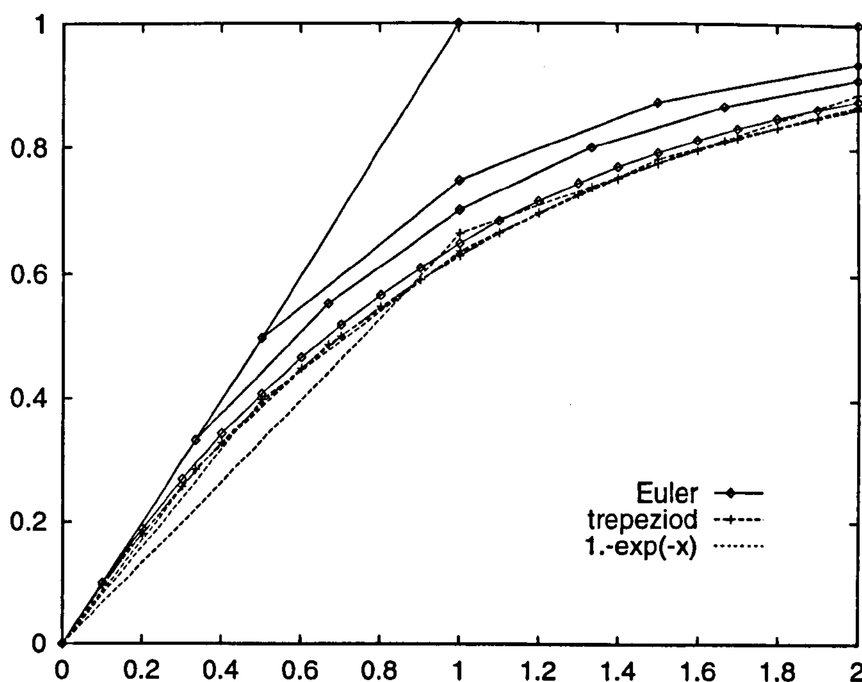


Fig. 2.1: The relaxation of velocity towards equilibrium as determined by the two integration methods. τ_p has been chosen as the unit of time and $\Delta t = \tau$, $\Delta t = \tau/2$, $\Delta t = \tau/3$ and $\Delta t = \tau/10$ have been graphed.

be represented by any distribution with the correct variance, such as steps of constant magnitude. The central limit theorem ensures that the sum of velocities will approximate a Gaussian distribution after a large number of steps and so the components of $\Delta \mathbf{V}_n$ can be chosen at random from the set $\{\pm\sqrt{2D_v\Delta t}\}$. The frictional term is helpful here since it smoothes these stochastic jumps. The simplicity of this approach pays great dividends when $\Delta t/\tau$ becomes small. In fact the smallest time scale is the energy relaxation time ($\tau_\mathcal{E} = \tau_p/2$) and in practice any time step $\Delta t \leq \tau_p/2$ suffices. Table 2.1 shows the advantageous of choosing the random term from a normal distribution.

The new positions of the particle at the end of the nth time step can be obtained from the average velocity,

$$\mathbf{x}_{n+1} = \mathbf{x}_{n+1} + \frac{1}{2}(\mathbf{v}_n + \mathbf{v}_{n+1})\Delta t.$$

This is consistent with equation 2.16. In very small devices, where there are at most only a few particles at a node on the Poisson mesh, the movement of a

particle out of the cell can result in a significant change in the potential map. Therefore the time steps must be smaller than the time taken for a particle to cross the width of a cell.

Alternatively the description in real space can be used which allows a time step larger than τ. Its implementation is simpler, requiring only the position of the particle. The position of an individual particle at the end of the nth time step $\mathbf{x}_{n+1}$ can be calculated from the position at the beginning of the step $\mathbf{x}_n$ using equation 2.20. Again $\Delta\mathbf{X}_n$ should be chosen from a Gaussian distribution whose variance is given by equation 2.19.

To model small devices, including the fluctuating potential of randomly distributed impurities and interacting carriers, it is advisable to use the velocity representation of the Brownian simulation rather than the spatial one. The scattering from randomly distributed ionised impurities can then be incorporated into the electric field $\mathbf{E}$ through their potential obtained from the solution of the Poisson's equation. This allows such scattering to be studied in detail, rather than averaged as in the usual approach where it would be lumped into an overall mobility. In this case the 'mobility' or relaxation time in the simulation should include only the phonon scattering and can be deduced from the phenomenological mobility at low doping concentration.

A source of concern is that the relaxation time approximation is not strictly applicable to inelastic scattering events, which dominate at room temperature in moderately doped semiconductors. However it provides a reasonable approximation where the system is not driven far from equilibrium. Methods based on partial differential equations also rely on relaxation time approximations and obtain remarkably accurate results. The Brownian method, like the drift-diffusion model, is applicable where a relaxation time can be defined. In practice this covers a wide range of operating conditions.

2.3 Verification

A series of tests was performed to confirm the validity and accuracy of the Brownian simulation method.

It is obviously essential that the method gives the correct behaviour at equilibrium. The central limit theorem is expected to produce a Gaussian

distribution of velocities. This Figure 2.2 shows the distribution of velocities for the model based in velocity space (equation 2.16) after it has come to a steady state.

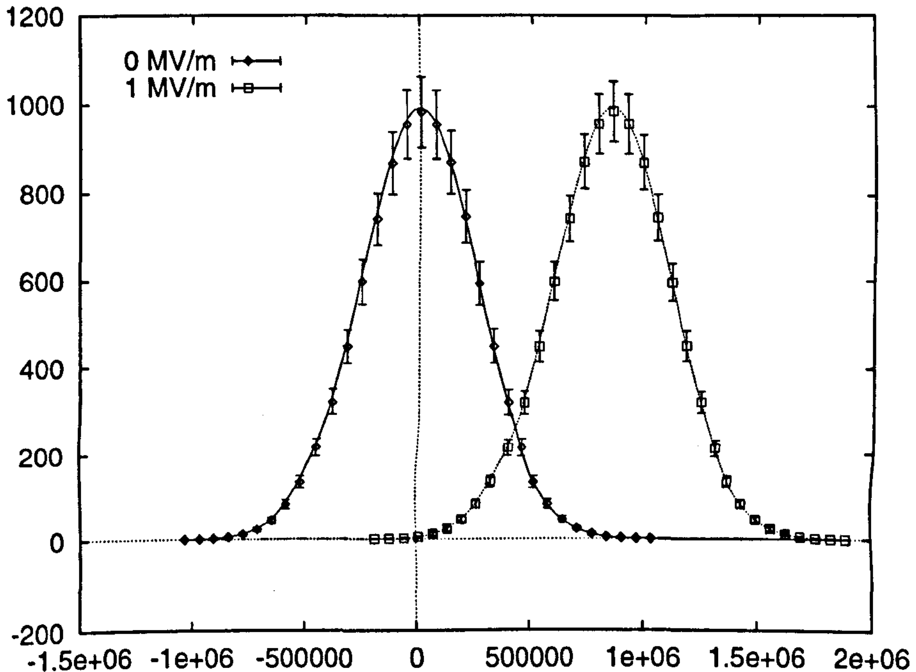


Fig. 2.2: Distribution of velocities in the simulation at equilibrium (◇) and under a constant electric field of $1\,\mathrm{MV\,m^{-1}}$ (−). For comparison, the full curve is a Maxwell-Boltzmann distribution at 300 K, while the broken curve is the same distribution shifted to the expected drift velocity of $850\,\mathrm{km\,s^{-1}}$. Note that an electric field displaces the distribution but does not broaden it within this model. The error bars denote the $\sqrt{N}$ statistical error

It is clearly in excellent agreement with a Maxwell distribution at the expected temperature of 300 K. The particles' temperature represents the equilibrium point where the power delivered by the fluctuating force matches the losses due to the frictional term. The diffusion coefficient in real space was also checked to have had the correct value (equation 2.19).

Constant electric fields were applied to verify that drift was predicted correctly and Ohm's law was clearly obeyed with the correct mobility. This is in keeping with the analytic calculations. Figure 2.2 also shows the distribution of velocities in a field of $1\,\mathrm{MV\,m^{-1}}$ for the model based in velocity space.

The distribution is centred on the drift velocity but has the same width as before. Note that the temperature of the electrons does not increase with electric field in this model. This is to be expected from its close relation to the drift-diffusion approximation, and contrasts with other approaches such as the relaxation-time or drifted-Maxwellian approximation. The calculations in appendix A show that the distribution is unaffected by the electric field.

The simulation therefore confirms the correct choice for the magnitude of the Wiener process. The spatial diffusion coefficient was evaluated from the variance of the position vector and the correlation between the velocity and position vectors. The simulations predict the correct diffusion coefficient in real space, properly related through the Einstein relation to the mobility which is an input parameter in the simulations. This reconfirms the correct choice of $\Delta\mathbf{V}$ as diffusion is entirely related to the fluctuating force. Thus we have confidence that this simulation gives the correct results in equilibrium.

The simulation of a *pn*-junction diode provides an excellent test for the Brownian approach because both drift and diffusion are critical. At equilibrium, the width of the depletion region and the built in potential are established by the balance between drift and diffusion. In forward bias at low current level the diode characteristic is dominated by the diffusion of the minority carriers away from the depletion layer edges. This leads to an exponential current voltage characteristic (the Shockley equation). At high current level the series resistance and the drift current in the undepleted regions of the diode lead to a linear (Ohmic) behaviour. Hence the diode operation can be described satisfactorily using drift-diffusion simulations. The comparison of standard drift-diffusion results with the Brownian simulations for a simple diode will establish further the credibility of the new simulation approach.

A $1\,\mu$m long, symmetrical Si diode, with doping concentrations of $10^{22}\,\mathrm{m}^{-3}$ on both sides of the metallurgical junction, was studied as a further test of the Brownian method. The 1D simulation domain was uniformly subdivided into 100 blocks with width $\Delta x = 10\,nm$ over which Poisson's equation was discretized and the charge assignment carried out. This ensures that there are around 20 mesh points over each half of the depletion region. A cross sectional area was selected so that there would be a manageable number of particles (arbitrarily chosen to be 100) in each block. The time step must satisfy two

constraints: it must be small compared with the momentum relaxation time, and should ensure that the particles are only likely to travel between adjacent blocks; this is determined by the thermal velocity rather than the drift velocity. $\Delta t = 0.01\,ps$ was used, compared with $\tau = 0.2\,ps$, which allows the exponential relaxation of the average velocity to be adequately characterised (equation 2.15). A typical time to travel between adjacent blocks at the thermal velocity is $0.1\,ps$ so the second condition is also satisfied. The solution domain and the ohmic contact regions are illustrated in figure 2.3. A width of 4 blocks was chosen for the contacts on the both sides. The tridiagonal system of equations arising from the discretization of Poisson's equation was solved directly.

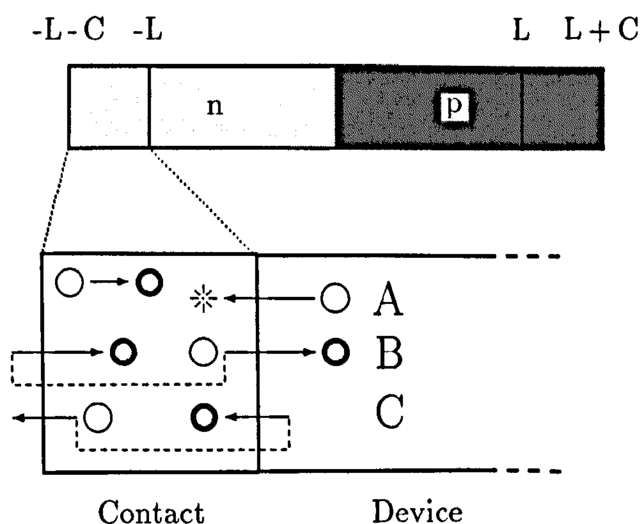


Fig. 2.3: (a) Domain of the simulation, showing a symmetric diode between $\pm L$ with additional regions of thickness C for the contacts. (b) Expanded view of the left contact region, showing the processes that occur when a particle meets the boundary. (A) A particle entering the contact from the device is counted and annihilated. (B) When a particle from the contact enters the device, a replica is created and returned to the far edge of the contact using periodic boundary conditions. (C) Particles hitting the far edge of the contact are returned using periodic boundary conditions. These processes ensure that the contact retains a constant number of particles and remains in equilibrium.

Special care must be taken when the modelling of the contacts. In a good model of an ohmic contact the carriers should have an infinite recombination

rate and should remain in thermal equilibrium during the whole simulation. If, however, finite ohmic contact regions are introduced in the simulation, energetic carriers are more likely to leave the contact than less energetic ones and there would be a contact cooling effect. To overcome this problem, the following procedure, illustrated in figure 2.3(b), is used. The ohmic contact regions were considered separately from the domain in which Poisson's equation is solved. They contain no electric field and have periodic boundaries. If a particle enters the contact from the device it is immediately removed from the simulation and counted (A). When a particle leaves the contact region into the device, a duplicate is created in the contact using periodic boundary conditions (B). Procedure (A) ensures that there is no heating from incoming particles and along with the periodic boundary conditions (B and C) ensures that the contacts are always in thermal equilibrium while conserving the number of particles. This contact also gives the correct projection profile into the device with particle velocity: particles of higher velocity can penetrate more deeply than those that are slower.

The widths of these contact blocks must be greater than the mean free paths of the carriers in order to avoid introducing any correlations in the motions of the particles injected into the device. This procedure allows equilibrium conditions in the device to be established.

The balance between the carriers which are removed from the simulation domain through the contact boundary and the carriers injected from the contact region through the same boundary constitute the total contact current. A sufficient number of time steps must elapse after any change in bias before any such current can be reliably measured.

First the development of the depletion layer from charge neutrality was simulated, with one boundary of the solution domain (at $x = -L$) fixed to 0V and the other (at $x = L$) allowed to float free. The von Neumann boundary should be treated very carefully since there is now less control in the injection rates or in the numbers of carriers in the device. The presence of both positive and negative mobile carriers can cause problems if recombination is ignored. A slight excess of carriers of a given charge in any region will set up an electric field that will attract carriers of opposite charge because the solution of Poisson's equation depends only on the net charge in each block. This can lead to

an incessant accumulation of carriers in the device. The contact at the floating boundary must not cause any unwanted injection or absorption effects, and must be at thermal equilibrium at all times. The procedure for ohmic contacts described on page 30 satisfies these requirements. Recombination is necessary for a depletion region to form, but does not affect its profile and so it is not necessary to model it very accurately. The depletion region is formed such that the thermal diffusion currents are countered by the drift currents due to the proper magnitude of the built in potential (0.69 V for this doping). However this is a slow process with a exponential relaxation to the equilibrium condition and prone to large fluctuations (see figure 2.4). This was because the number of carriers that can cross the depletion forms decreases markedly as it develops. The simulation was repeated with initial condition where the carriers had been removed in the cells lying within the depletion region and the results for a 400 *ps* simulation are shown in figure 2.5 and clearly indicate that the appropriate built in potential is maintained. This shows the advantages of the Brownian method (for the modelling of steady state conditions)

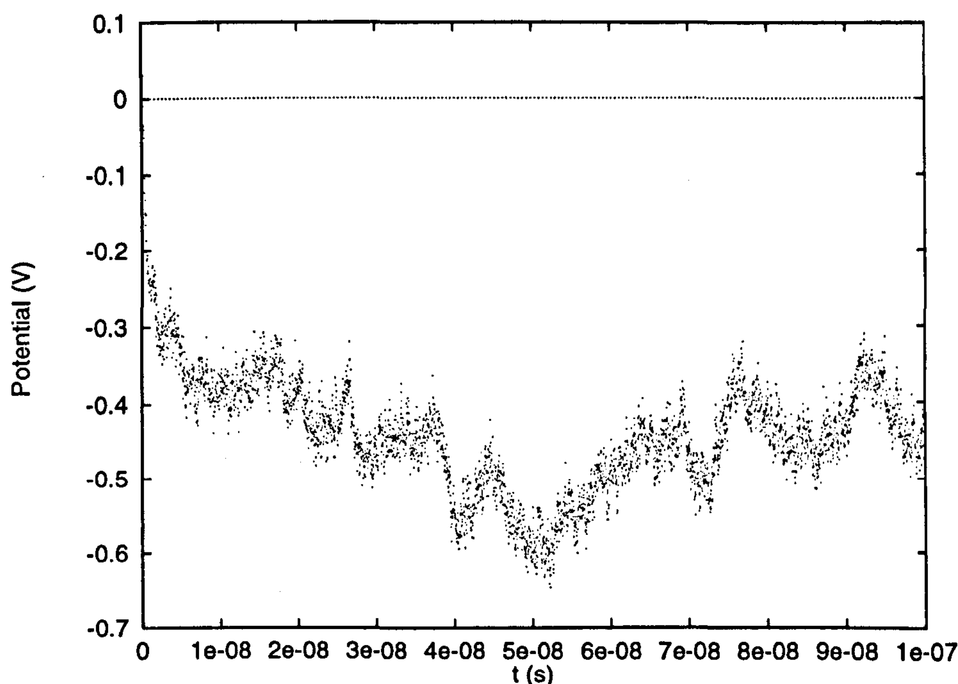


Fig. 2.4: The development of the built in potential.

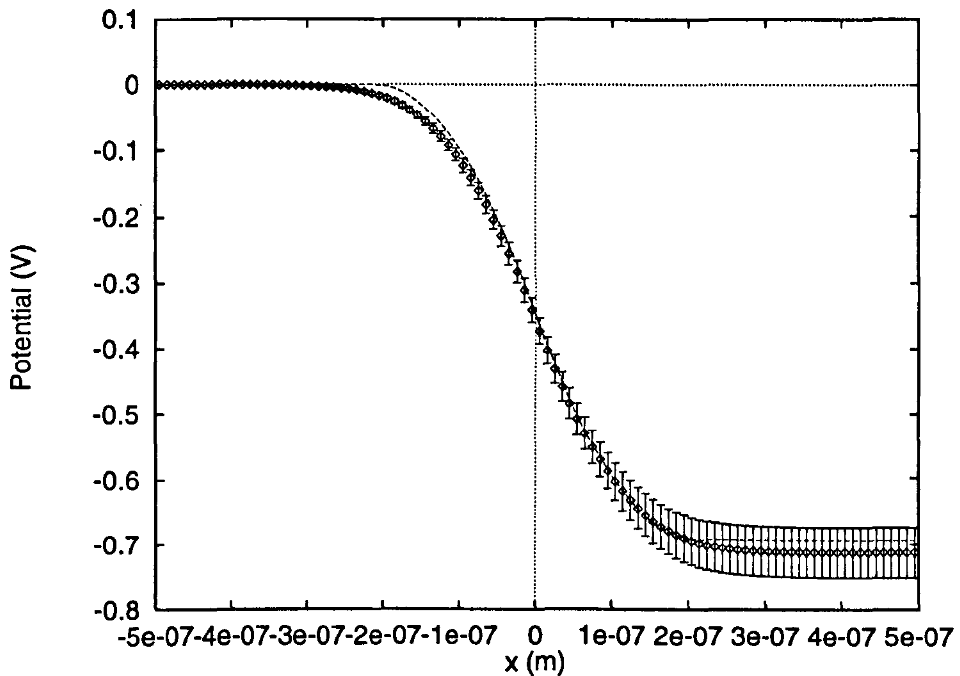


Fig. 2.5: V_T studied over a 0.4 ns. (The line represents the initial condition.) Note that the value stays around the correct value. This is a very good test of the Brownian method and the efficacy of the method of dealing with the contacts.

where the large infrequent changes in momenta arising from the Monte Carlo scattering are replaced by numerous and much smaller terms. This result also demonstrates that the procedure for the treatment of the contacts meets all these required criteria. The error bars show the magnitude of the standard deviation in the potential at each cell. Figure 2.6 shows the the equilibrium distribution over this long interval. The width of the depletion region arising from the Brownian model is in good agreement with the simple abrupt view. The differences are due to the spreading of the charge distribution caused by diffusion.

The temperature distribution across the diode provides further verification of the Brownian method but it gives an idea of the range of kinetic energies across the device — a result that cannot be obtained from drift-diffusion. The large fluctuations in the depletion region are due to the low numbers of carriers and the fluctuations in high fields due to noise. Much of the potential difference

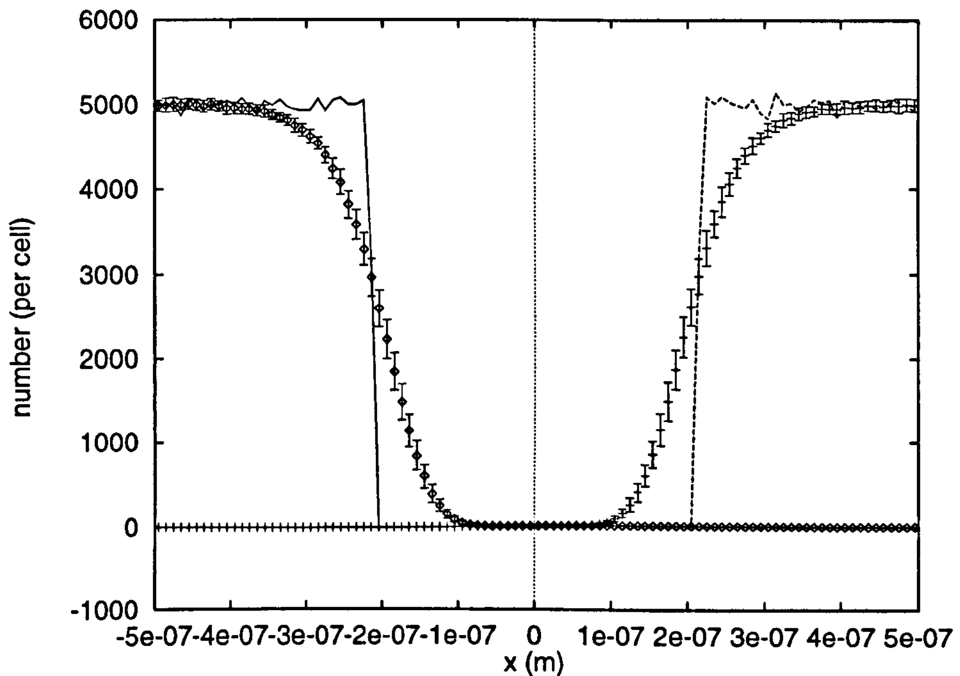


Fig. 2.6: The electron and hole numbers /(10 nm). (The lines represent the initial condition.) Note the large fluctuations in the depletion regions.

is dropped across the depletion region and so when V_T is more positive more carriers can diffuse further into the depletion region but when V_T drops they experience a larger electric field and so can be accelerated to higher energies. The flustuations in V_T can therefore act as another random fluctuating force that heat the carriers. Electron energies within the depletion region can be as high as 1700 K.

The $I(V)$ characteristics in forward bias were simulated using Dirichlet boundary conditions on both sides of the solution domain. Recombination was neglected since the length of the diode was much shorter than the minority carrier diffusion lengths on both sides of the depletion region. The results (see figure 2.8) were compared with the Shockley equation for an ideal short diode and found to be in good agreement for small forward currents. The commercial device simulator MEDICI was also used in drift-diffusion mode to simulate the same diode at higher current levels where the series resistance limit applies. The agreement between the Brownian approach and the MEDICI

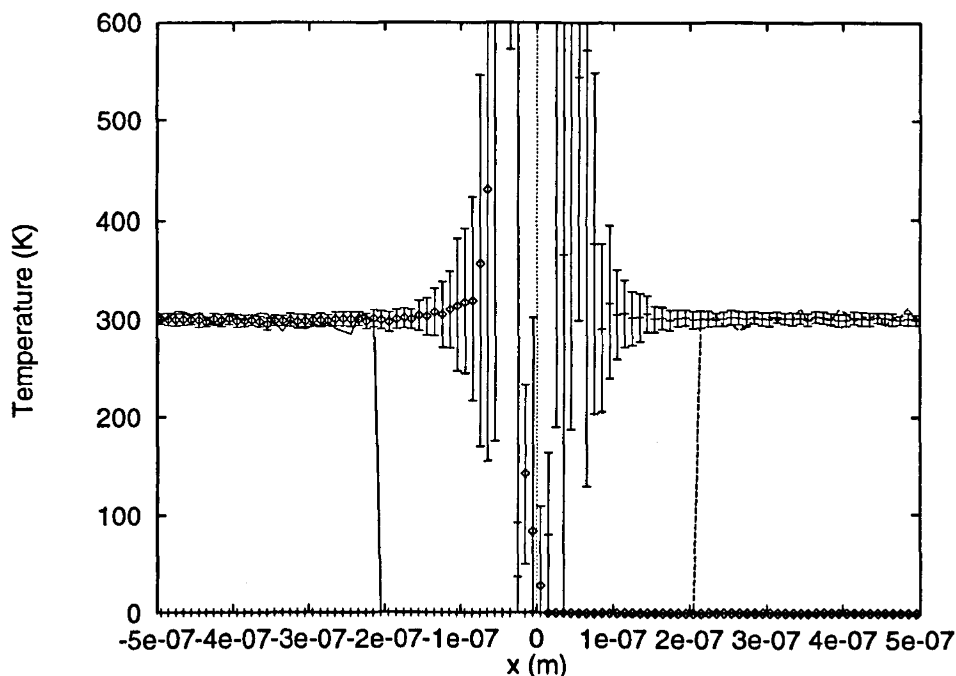


Fig. 2.7: The temperature distribution across the diode. (The lines represent the initial condition.) Note the large fluctuations in the depletion regions.

drift-diffusion simulation is excellent both at low and high current levels and shows that the new method properly describes the effects due to diffusion at low bias and those due to drift at higher applied voltages. The statistical noise due to the discreteness of carriers is apparent during the small injection rates that are present when a small forward bias is applied. The very small discrepancy at large forward bias may be due to insufficient injection rates at the contacts, which were modelled as reservoirs at thermal equilibrium with no external fields.

The Brownian method has therefore been shown to provide a good description of both drift and diffusion throughout its region of applicability.

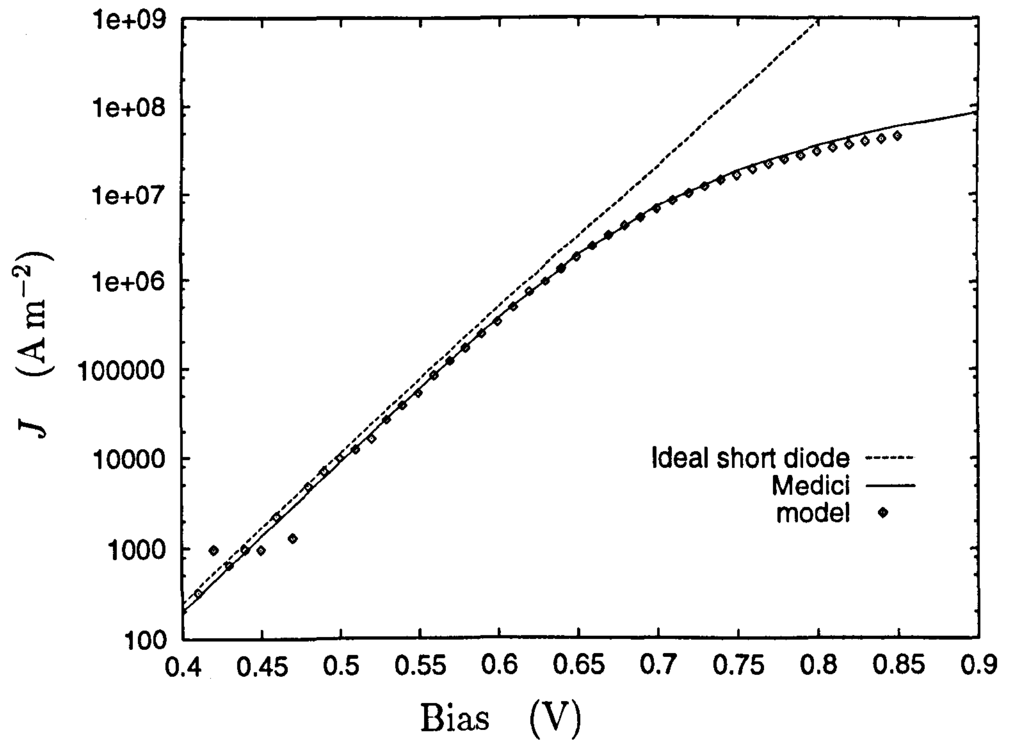


Fig. 2.8: Current-voltage characteristics of a short diode, comparing the Brownian simulation (circles) with results from MEDICI (full line) and the ideal diode equation (broken line). Agreement is excellent both at low bias where diffusion dominates and at high bias when drift is important.

3. CALCULATION OF THE ELECTRIC FIELD FROM DISCRETE RANDOM CHARGES

Now that we have a computationally efficient method of dealing with phonon scattering, the next important step is to integrate the equations of motion resulting from the complicated electrostatic forces arising from these discrete charges. The crux of the problem is that the presence of discrete point-like charges result in sharply varying electrostatic forces that are numerically difficult to calculate and integrate through space and time. These forces can be split into an internal component, due to the electrostatic force between the particles, and an external component, resulting from the applied potentials and boundaries. The internal forces show the rapid spatial variation (i.e. they have high spectral components), while the external forces provide a slowly varying background.

The solution of Poisson's equation on a mesh is the mainstay of device modelling. It is an efficient method that can deal with complex boundaries and large numbers of carriers. It obviates the need to sum what are often infinite sets of image charges at boundaries. However, the act of sampling the charge density and the potential distribution on a finite sized mesh places an upper limit on the spacial frequency of these quantities. The finite size of each mesh point means that it cannot truly represent the pointlike nature of individual ions and carriers or the spikes in the potential from these discrete charges.

The direct analytic evaluation of the electrostatic forces does not suffer from the problem of band limiting but is not practical for real devices. The Coulomb force obeys the inverse square law and so the high spatial frequency components are only significant around the vicinity of each pointlike charge. It then becomes only necessary to calculate the forces directly when the particles are close enough for the mesh to have a significant effect on the measurements.

Now since the more slowly varying long range forces are extracted from the mesh, a suitable and efficient method is required to estimate the erroneous contributions to the force from nearby charges. Estimating the force that is obtained from the mesh is a tricky problem because it depends on both the positions of the charges within the mesh cells and the orientation of the vector that separates them. (A mesh cell is the volume enclosed by eight adjacent abscissae for the case of a rectangular mesh.) Therefore the main objective is to use a scheme that gives forces that are simple to predict. This problem has been studied by other authors [92, 93] who used 'optimised' Green's functions to make the force obtained from the mesh approximate that which occurs between spherical distributions of charge. Their method gives forces that are not affected by the coordinate of the particles with respect to the mesh nor their relative orientations. However, this procedure places limits on the choice of Poisson solver since the use of Green's functions require the use of methods based on Fourier transforms. They used an efficient FACR algorithm [94] in their work [93, 95]. Section 3.2 introduces a novel and computationally inexpensive method, that has been devised as part of this work, which adequately determines the mesh-derived force for point charges in close proximity. It uses the symmetry of the problem to estimate the mesh force and so divorces problem of solving Poisson's equation from estimation of the mesh derived force. This makes the Poisson solver simpler to set up and allows the most suitable method to be used.

This chapter will firstly review the standard technique for the resolution of forces — the solution of Poisson's equation on a mesh. Attention will be paid to the problem of band-limiting because this determines the volume over which the interparticle interactions have to be directly evaluated, for a given mesh. Now since the number of interactions varies as the square of the number of charges which in turn varies as the cube of the length of the cell, this is an important consideration when designing a method to follow the equations of motion.

The use of weighting functions to assign the discrete charges to the mesh and to obtain the resulting forces will be discussed. The important point here is that the same weights must be used for both these processes to avoid any unphysical or 'fictitious' forces.

The solution of Poisson's equation will then be discussed. Three methods have been developed and applied to typical structure. The first is the simple SOR scheme; the second is the BiCGSTAB(2) which is a combination of a two-step bi-conjugate gradient method and a GMRES method; and finally a highly efficient multigrid method with several enhancements is presented. It will be shown that the nature of these large problems with discrete charges is subtly different from those found in standard device modelling. A comparison of the SOR, BiCGSTAB(2) and multigrid methods for a typical case show that relaxation methods are best suited for these very large problems containing discrete charges because they can tolerate rounding errors better. The multigrid method will be shown to be the most efficient for large problems containing discrete charges and so making it the method of choice. The BiCGSTAB(2) and SOR methods have been further enhanced to allow them to run on a rectangular array of transputers of arbitrary size and configuration. Chapter 5 discusses the implementation of these methods on parallel architectures.

The following section deals with the calculation of forces between discrete pointlike charges. The new method of estimating the mesh derived force is central to this work.

3.1 The solution of Poisson's equation

The most common way to resolve the forces acting on the carriers is through the estimation of the local electric field following the solution of Poisson's equation. The use of a mesh allows does not suffer from the the problems caused by the need to sum infinite sets of image charges boundaries and is generally faster for systems with a large number of particles.

There are three steps to this procedure:

1. The assignment of charges to the mesh points using some weight function.

2. Solve Poisson's equation to get the potentials at the mesh points.

3. Difference the potentials and interpolate the fields for each particle.

The operations count for this process can be written as $\alpha n + \beta(N_p)$ for a system on n particles on a mesh with N_p abscissae. α is determined by the

total number of operations required for charge assignment and for differencing and interpolation of the forces. β gives the count for the solution of Poisson's equation. One can compare this to the $n(n-1)$ operations required to calculate the forces between n particles — assuming that one division and one squaring operation is required for each evaluation.

Unfortunately, the use of a mesh introduces inaccuracies in the calculated force due to band-limitation. For a mesh of spacing h, the sampling theorem limits the minimum wavelength of any fluctuations in the mesh values to $2h$. This has the effect of introducing large errors in the short-range force between point-like particles with separations that are smaller than h. In fact the particles behave as though they are clouds of charge with a radius of $h/2$. It is probably more accurate to represent them as cubes of dimensions $h \times h \times h$. The band limitations imposed by the mesh can be unacceptable for studying correlated systems, requiring meshes that are fine enough to resolve the close encounters that can occur between the particles, due to the exorbitant computational loads. However, this can be a boon in uncorrelated systems where this cut-off introduces the properties required to destroy the unphysical correlations between a limited number of superparticles in the simulation. The usual solution to this problem is to use a varying mesh spacing across the simulation domain with the finest mesh in areas where there are sharp variations in electric field. In standard FET simulations, this region is under the gate near the drain. The presence of discrete and mobile charges add to this problem because they induce spikes in the potential map whose positions vary with time.

Screening causes potentials to decay more rapidly and so the scale of the mesh is determined by the Debye length

$$l_{\mathrm{D}} = \sqrt{\frac{\epsilon k_{\mathrm{B}} T}{e^2 n}} \tag{3.1}$$

The processes involved in the solution of Poisson's equation will now be discussed.

3.1.1 *Charge assignment*

The first step in the solution of Poisson's equation is to place the discrete charges on a mesh. The simplest solution is to place each charge on the

nearest grid point which preserves its pointlike nature as far as possible on the mesh. This crude method gives a discontinuous electric field and potential during the motion of the carriers and so it is of little practical use. The need for a smoothly varying potential and electric field requires the use of an interpolation scheme that divides up the charge and assigns the components to nearby mesh points. A couple of problems can arise if care is not taken during this procedure. An important consideration is that the mesh should not give rise to fictitious forces. The resulting equation of motion must conserve both energy and momentum. This problem arises when the discrete charges are ascribed to several mesh points through some weighting function. This condition requires that the same weight functions be used when assigning the charges to the abscissae and when interpolating the particle forces [95].

A simple approach would be to use a piecewise linear scheme to interpolate the values to the nearest 8 abscissae (in three dimensions). This cloud-in cell (CIC) method is very widely used. A triangle shaped cloud (TSC) using a quadratic weighting function, used to ascribe the particulate charge to the nearest 27 abscissae, also has a continuous first derivative. One can continue onwards using more abscissae to give an n-point scheme that is continuous on the first $(n-1)$-derivatives. The penalty for this is that the weights must be ascribed - on a three dimensional grid - to n^3 abscissae, resulting in a rapid increase in computational effort for fewer gains. In practical situations the TSC is found to be adequate.

Spreading the charges on to a small number of mesh points can result in large tails in their Fourier spectrum that leads to the phenomenon of aliasing. To illustrate this problem, first consider 1D problem where a band limited function that is sampled by a train of δ-functions and its Fourier transform

$$\begin{aligned} f_0(x) &= \sum_{n=-\infty}^{\infty} \delta(x-nh)f(x) \\ \Longleftrightarrow \quad F_0(k) &= h \sum_{n=-\infty}^{\infty} \delta(k-nk_m) \otimes F(k) = \sum_{n=-\infty}^{\infty} F(k-nk_m). \end{aligned} \tag{3.2}$$

($\Longleftrightarrow$ represents the Fourier transform operation and $\otimes$ the convolution integral).

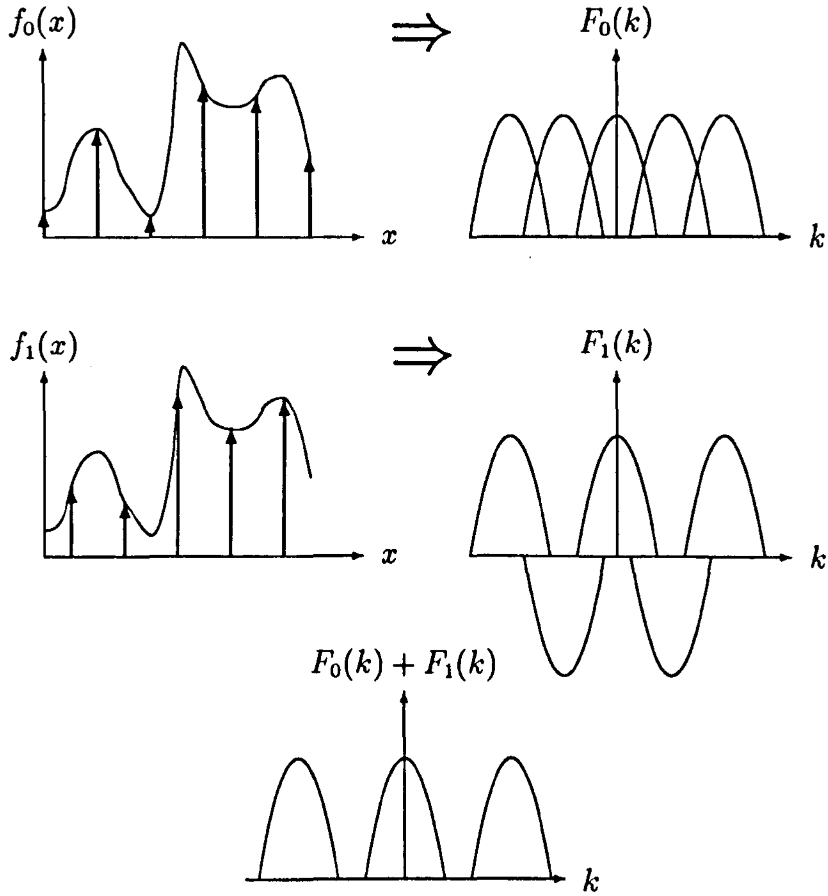


Fig. 3.1: Aliasing in meshes.

The shift theorem can be used to remove the first harmonics

$$\begin{aligned} f_1(x) &= \sum_{n=-\infty}^{\infty} \delta(x-(n+\frac{1}{2})h)f(x) \\ \Longleftrightarrow \quad F_1(k) &= h\sum_{n=-\infty}^{\infty} \delta(k-(n+\frac{1}{2})k_m) \otimes F(k) \\ &= \sum_{n=-\infty}^{\infty} (-1)^n F(k-nk_m)\,. \end{aligned} \tag{3.3}$$

Averaging the 2 together causes all the odd images to cancel and so largely removes the problem of aliasing, as shown in figure 3.1.

The 3D case can be treated as a product of 3 spatial functions

$$f(x,y,z) = f_x(x)f_y(y)f_z(z) \tag{3.4}$$

and so requires eight interlaced meshes. However, the magnitude of the error is most significant along the [100] directions and so one can use two interlaced meshes that are displaced by $(h/2, h/2, h/2)$ (as seen from figures 3.9–3.11). Obviously these arguments are an oversimplification: the quantities represented on the mesh are not truly band limited and so the second and higher harmonics are also present, but interlacing only removes the contributions from the first.

The use of interlacing is not the same as sampling the potential or charge distribution using a finer mesh. It is a device for reducing the influence of the particle's position, relative to the mesh points, on the mesh derived force. This will be discussed in section 3.2.

3.1.2 The solution of Poisson's equation

The solution of Poisson's equation is often the most time consuming part of a three dimensional simulation. In our simulations, over 80% of the cpu time for a time step is spent solving Poisson's equation since there are a relatively few number of particles on very large meshes. Optimising this stage would, therefore, pay back great dividends. Three different iterative methods were considered: the successive over-relaxation method, the bi-conjugate gradient method and the multigrid method.

Firstly, the three methods will be outlined [96] and their relative merits will be discussed.

Successive over relaxation methods [97, 98, 99] are one of the most widely used techniques for solving boundary value problems in science and engineering due to their simplicity and reasonable speed. Essentially, they are Gauss-Seidel methods with overcorrection. Consider the system of linear equations

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} \tag{3.5}$$

The SOR method can be written down as

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \omega(\mathbf{L} + \mathbf{D})^{-1}\mathbf{r}_i\,, \tag{3.6}$$

where **L** and **D** are the lower triangle and main diagonal of **A** respectively, and the residual $\mathbf{r}_i = \mathbf{A}\mathbf{x}_i - \mathbf{b}$.

For a problem of size N, an optimal choice for the overrelaxation parameter $(1 < \omega < 2)$ can reduce the number of iterations required to obtain a solution from the order of N^2, of the Gauss-Seidel, to that of N. This optimal value is given by

$$\omega = \frac{2}{1 + \sqrt{1 - \rho_{GS}}}, \tag{3.7}$$

where the spectral radius ρ_{GS} is the magnitude of the largest eigenvalue of the Gauss-Seidel iteration matrix $(-(\mathbf{L}+\mathbf{D})^{-1} \cdot \mathbf{U})$. (**U** is the upper triangle of **A**.) Its value is just the square of the Jacobi iteration matrix $-\mathbf{D}^{-1} \cdot (\mathbf{L} + \mathbf{U})$. The spectral radius of the Jacobi iteration, for a $N_x \times N_y \times N_z$ mesh with spacing $(\delta x, \delta y, \delta z)$ and either Dirichlet or von Neumann homogeneous boundary conditions, is given by

$$\rho_J = \frac{(\delta y^2 \delta z^2)\cos\frac{\pi}{N_x} + (\delta x^2 \delta z^2)\cos\frac{\pi}{N_y} + (\delta x^2 \delta y^2)\cos\frac{\pi}{N_z}}{\delta x^2 \delta y^2 + \delta x^2 \delta z^2 + \delta y^2 \delta z^2}. \tag{3.8}$$

Overrelaxation often results in the increase in the error by more than an order of magnitude in the early iterations. However this problem can be easily alleviated by the use of Chebyshev acceleration. Here the mesh is subdivided into black and white nodes, as in a chessboard pattern, which are updated alternatively using data from the opposite colour at each half-sweep. The updating of points in this manner also facilitates the the implementation of the SOR method on massively parallel computers because the data required to update each point does not change during each half sweep. This will be discussed further in chaper 5.

The overrelaxation parameter is initially set to unity (*i.e.* starting from the Gauss-Seidel iteration) and gradually refined toward the optimal value over each half-iteration step.

$$\begin{aligned}
\omega_0 &= 1 \\
\omega_{1/2} &= \frac{1}{1 - \frac{1}{2}\rho_J^2} \\
\omega_{n+1/2} &= \frac{1}{1 - \frac{1}{4}\rho_J^2 \omega_n}, \quad n > 1/2 \\
\omega_n &\xrightarrow{n \to \infty} \omega_{optimal}
\end{aligned} \tag{3.9}$$

Chebyshev acceleration always results in the decrease of the norm of the error and this method converges to the correct result eventually, even when the meshes are irregular. This fact, coupled with its simplicity and ease of parallel implementation, makes it the mainstay of many device simulators.

Conjugate gradient methods [100] are widely used for the solution of differential and sets of linear equations $\mathbf{Ax} = \mathbf{b}$. Their essence lies in the minimisation of an error function

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b} + c \tag{3.10}$$

to which this problem provides the gradient. Now $f(\mathbf{x})$ is a quadratic function if $\mathbf{A}$ is symmetric and so if there exists a stationary point (*i.e.* a solution) at $\mathbf{x}_0$ then

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) \tag{3.11}$$

If $\mathbf{A}$ is positive definite then

$$f(\mathbf{x}) > f(\mathbf{x}_0) \quad \forall \mathbf{x} \neq \mathbf{x}_0, \tag{3.12}$$

in other words there is a unique stationary point which is at the minimum.

Now take a set of parallel lines along a direction $\mathbf{p}$ in an n-dimensional vector space, then the minimum of $f(\mathbf{x})$ along these lines at intersection of these lines with a $(n-1)$-dimensional hyperplane π_{n-1}. This plane is normal to $\mathbf{Ap}$ and so any vector $\mathbf{q}$ in π_{n-1} is conjugate to $\mathbf{p}$ ($\mathbf{q}^T\mathbf{Ap} = 0$). The obvious choice for $\mathbf{p}$ is along the direction of steepest descent. Furthermore, the solution $\mathbf{x}_0$ must lie on this hyperplane and so the dimensionality of the problem has been reduced by one. The next step will be along the vector of steepest descent in this hyperplane and so forth giving the solution in at most n steps — assuming exact arithmetic of course.

The important question that has to be resolved now is the size of these steps. Since $\mathbf{A}$ is symmetric then equation 3.10 becomes

$$\begin{aligned} f(\mathbf{x}+\mathbf{p}) &= f(\mathbf{x}) + \mathbf{p}^T(\mathbf{Ax} - \mathbf{b}) + \frac{1}{2}\mathbf{p}^T\mathbf{Ap} \\ &= f(\mathbf{x}) + \mathbf{p}^T f'(\mathbf{x}) + \frac{1}{2}\mathbf{p}^T f''(\mathbf{x})\mathbf{p} \end{aligned} \tag{3.13}$$

and a residual ($\mathbf{r} = -f'(\mathbf{x})$) can be defined.

Let $\mathbf{x}_2$ be the position of the minimum along the line in the direction of $\mathbf{p}$ from an initial point $\mathbf{x}_1$ so that

$$\mathbf{x}_2 = \mathbf{x}_1 + d\mathbf{p}. \tag{3.14}$$

If

$$d = \frac{g}{h} \tag{3.15}$$

$$\begin{aligned} g &= \mathbf{p}^T\mathbf{r} \\ h &= \mathbf{p}^T\mathbf{A}\mathbf{p} \end{aligned}$$

then it is evident that

$$f(\mathbf{x}_1 + 2\mathbf{p}) = f(\mathbf{x}_1) - 2d(g - dh) = f(\mathbf{x}_1)$$

and so this prescription does indeed give the position of the minimum since $f(\mathbf{x})$ is a quadratic function.

Unfortunately $\mathbf{A}$ is neither symmetric nor positive definite in many cases. There are several solutions to this problem. Preconditioning can be used in these cases where a matrix $\mathbf{K} = \mathbf{K}_1 \cdot \mathbf{K}_2$ can be defined which approximates $\mathbf{A}$. The original problem can now be transformed into a more symmetric one after pre-multiplication by $\mathbf{K}_1^{-1}$

$$\mathbf{K}_1^{-1} \cdot \mathbf{A} \cdot \mathbf{K}_2^{-1} (\mathbf{K}_2 \cdot \mathbf{x}) = \mathbf{K}_1^{-1} \cdot \mathbf{b}. \tag{3.16}$$

The problem here is to choose $\mathbf{K}_1$ and $\mathbf{K}_2$ to provide a rapid convergence. There are many choices (such as $\mathbf{K}_1 = \mathbf{L}$ and $\mathbf{K}_2 = \mathbf{U}$ [101]) and there is no universal choice for all types of problems. For large sparse three dimensional systems some authors have proposed the use of incomplete Cholesky decomposition as a preconditioner [102, 103, 104]. This method used $\mathbf{K}_1 = \mathbf{LU}$ and $\mathbf{K}_2 = \mathbf{I}$.

The topic of preconditioning is very large and there is a fine art to choosing the appropriate method, and the small number of discrete charges mean that the solutions to Poisson's equation can vary greatly during the simulation. Therefore, the variants of the conjugate gradient scheme that can be applied to non-symmetric problems have been considered. The bi-conjugate gradient

method [105, 101] in conjunction with a generalised minimal residual method (GMRES) [106, 107], (BiCGSTAB(2)) [108] has been chosen. This method, while still being able to cope with complex unsymmetric problems, is very fast and is amenable to parallel implementation. (See appendix B for a fuller description.)

Figure 3.2 shows the results for a set of simulations carried out on a three-dimensional cubic mesh and with von Neuman boundary conditions. The values of **b** were set to zero but the central point set to a value of 1. One of the corner points was fixed (*i.e.* having a Dirichlet boundary condition) in order to provide a unique solution. It is clear that the bi-conjugate gradient method is about eight times faster than the SOR method with Chebyshev acceleration.

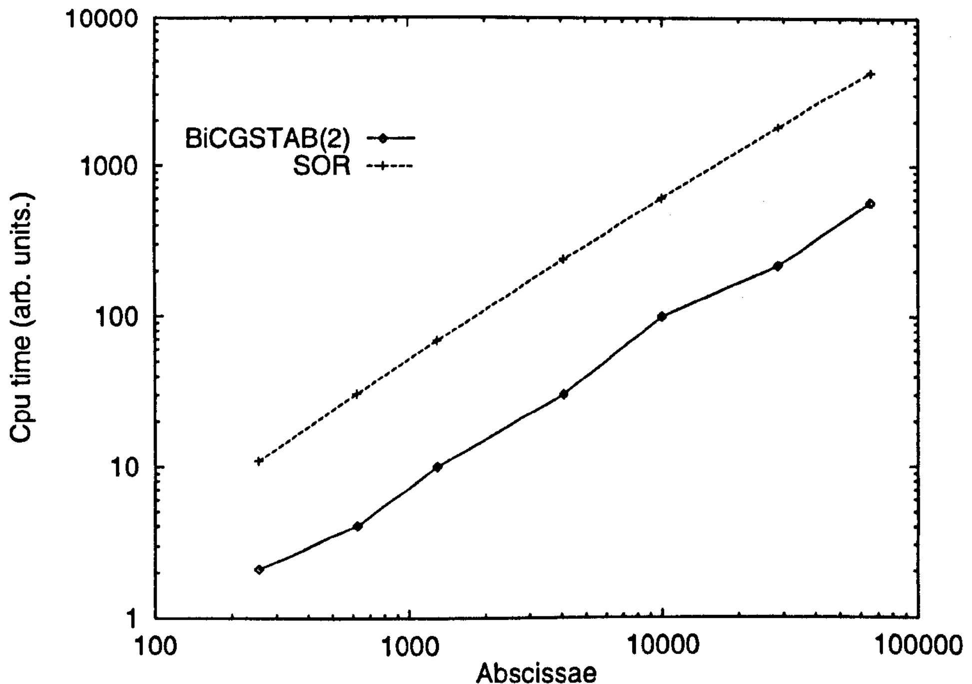


Fig. 3.2: The BiCGSTAB(2) method vs the SOR method for a 3D cube with von Neumann boundaries. One corner was fixed to 0V to prevent the solution from drifting.

Multigrid methods [109, 110, 111, 112] are among the quickest of Poisson solvers, requiring only of $\mathcal{O}(N)$ iterations for the solution of a mesh with N

abscissae. They are ofter faster than the so called 'rapid methods' that use fast Fourier transforms and cyclic reductions (FACR) [113, 114, 94]. Poisson's equation can be written as a set of linear equations after expressing the Laplacian as a finite difference operator and discretising the potential and charge density on a grid. Considering a linear differential operator $\mathcal{L}^l$ applied to a mesh labelled l, if the exact solution is given by

$$\mathcal{L}^l x^l = b^l \,, \tag{3.17}$$

then the correction that is required for the nth iterate for this solution, as denoted by x_n^l, is simply

$$c_n^l = x^l - x_n^l \tag{3.18}$$

The residual, or defect, is then given by

$$\begin{aligned} d_n^l &= \mathcal{L}x_n^l - b^l \\ &= \mathcal{L}(x^l - c_n^l) - b^l \\ &= -\mathcal{L}c_n^l \end{aligned} \tag{3.19}$$

One can proceed by obtaining the defect and solving this for the correction to the next iterative solution. The use of a relaxation method for solving this system of equations leads to a much greater efficiency in reducing the components in the upper half of the spectral band of the mesh than those in the lower half. Sampling the defects in the mesh using coarser grids will be advantageous in eliminating the low frequency components. This is called restriction and there is a corresponding prolongation of the correction to the finer grid. The aforementioned label l can now be used to denote the level of the current mesh. This can be written as the following algorithm :

1. Apply smoothing operations to the problem (pre-smoothing).

2. Compute the defect.

3. Restrict the defect onto a coarser mesh.

4. Solve defect for correction on coarser mesh.

5. Prolong correction to the finer mesh.

6. Update solution.

7. Apply more smoothing (post-smoothing).

A multigrid method can be used to solve for the correction on the coarser mesh and so this results in a recursive algorithm. This process (see figure 3.3), of smoothing and restriction going down to the coarsest mesh where an exact solution to the correction is found and the resulting prolongation and smoothing back to the finest mesh, is called a V-cycle. More effort is required to reduce the low frequency components and so in order to reduce some effort, one may descend γ times before ascending to a higher level. This results in the W-cycle. There is also a variant of the multigrid method that starts with the coarsest mesh and cycles through to the finer meshes. This is very useful the potential within the domain is not known *a priori*, but in our simulations the potential map is very similar, as far as the low spatial frequency components are concerned. It is convenient to use the Gauss-Seidel scheme as a smoothing operation, but over-relaxation must never be used because it destroys the smoothing of the high frequency components and may introduce oscillations. (A recent publication [115] questions this belief and claims a 34 % increase in speed when using a modified V-cycle method for very little computational overhead.) Many authors use the Jacobi method and this is the method of choice when multigrid methods are implemented on massively parallel architectures.

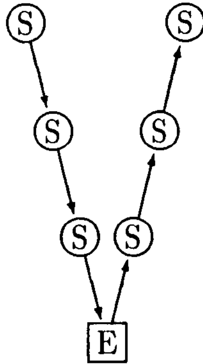
In the case of a regular mesh, linear interpolation functions may be used to derive the restriction ($\mathcal{R}$) and prolongation ($\mathcal{P}$) operators. Therefore the defect on the coarser grid of double the mesh spacing, on a three dimensional mesh, is given by

$$d_n^{l-1}(i,j,k) = \frac{1}{8} \sum_{i'=-1}^{1} \sum_{j'=-1}^{1} \sum_{k'=-1}^{1} 2^{-(|i'|+|j'|+|k'|)}\, d_n^{l}(2i+i', 2j+j', 2k+k') \tag{3.20}$$

and the corresponding prolongation of the correction on to the finer mesh is

$$c_n^{l}(2i+\alpha, 2j+\beta, 2k+\zeta) = 2^{-(|\alpha|+|\beta|+|\zeta|)} \sum_{i'=0}^{\alpha} \sum_{j'=0}^{\beta} \sum_{k'=0}^{\zeta} c_n^{l-1}(i+i', j+j', k+k') \tag{3.21}$$

$$\alpha, \beta, \zeta \in \{-1, 0, 1\}$$



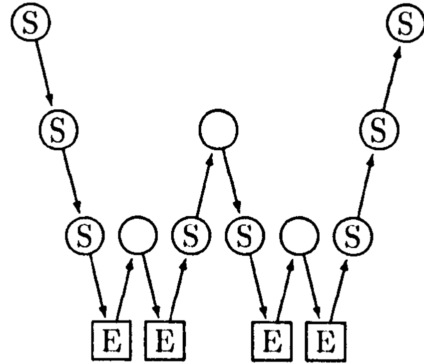


Fig. 3.3: An illustration of the multigrid method. The circles represent the levels of the multigrid iterations and [E] the exact solution at the coarsest mesh. The smoothing is applied at (S). $\searrow$ represents the calculation and subsequent restriction of the defect. $\nearrow$ denotes the prolongation of the correction followed by its application.

It is much clearer to write these equations in the form of a stencil as follows

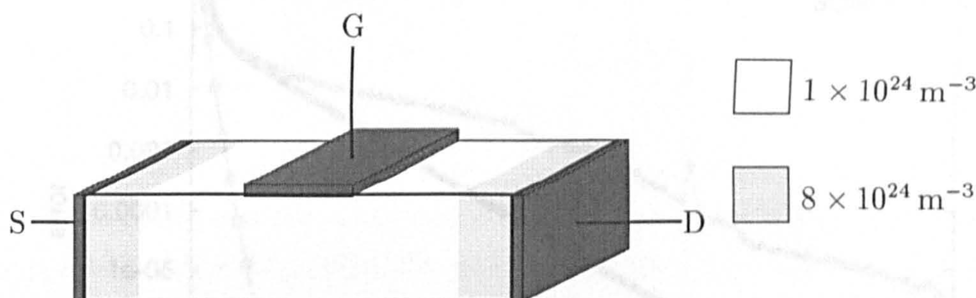
$$\mathcal{R} = \frac{1}{64}\left[\begin{array}{ccc|ccc|ccc} 1 & 2 & 1 & 2 & 4 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 4 & 8 & 4 & 2 & 4 & 2 \\ 1 & 2 & 1 & 2 & 4 & 2 & 1 & 2 & 1 \end{array}\right] \tag{3.22}$$

$$\mathcal{P} = \frac{1}{8}\left[\begin{array}{ccc|ccc|ccc} 1 & 2 & 1 & 2 & 4 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 4 & 8 & 4 & 2 & 4 & 2 \\ 1 & 2 & 1 & 2 & 4 & 2 & 1 & 2 & 1 \end{array}\right] \tag{3.23}$$

The main problem with this method is that one must be always able to define the differential operator for all meshes. This can lead to problems when inhomogeneous boundary conditions are used or if the mesh is irregular, or equivalently if there are materials with different dielectric constants, are present. Fixed (Dirichlet) points must be treated with some care since any correction terms to these points must be zero. Therefore any abscissae on the coarser mesh that map to a fixed point are forced to have a zero defect.

The multigrid method was validated using the FET structure in figure 3.4. The solver was written so that the number number of levels and the num-

Fig. 3.4: The MESFET structure that was used to test and compare the solvers.



ber of points along each direction in the coarsest mesh could be controlled arbitrarily. The value of γ and number of pre- and post-smoothing can also be controlled through the input parameters. The geometry, determining the boundary conditions, of the FET structure was fixed.

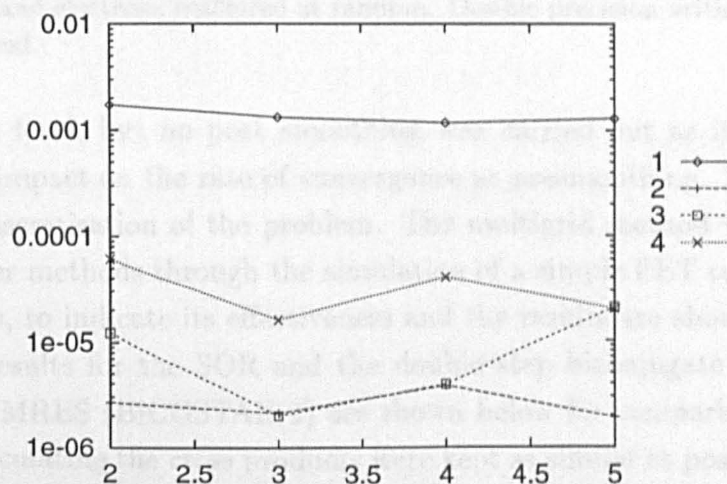


Fig. 3.5: The magnitude of the residual, as a fraction of its initial value, at 12 s cpu time for the multigrid method, with the number of pre-smoothing steps. Graphs are shown for $\gamma = 1, 2, 3$ and 4. There was no post-smoothing.

Figure 3.5 shows that $\gamma = 2$ was best suited to the problem at hand, since it gave the most rapid convergence. Two pre-smoothing iterations were per-

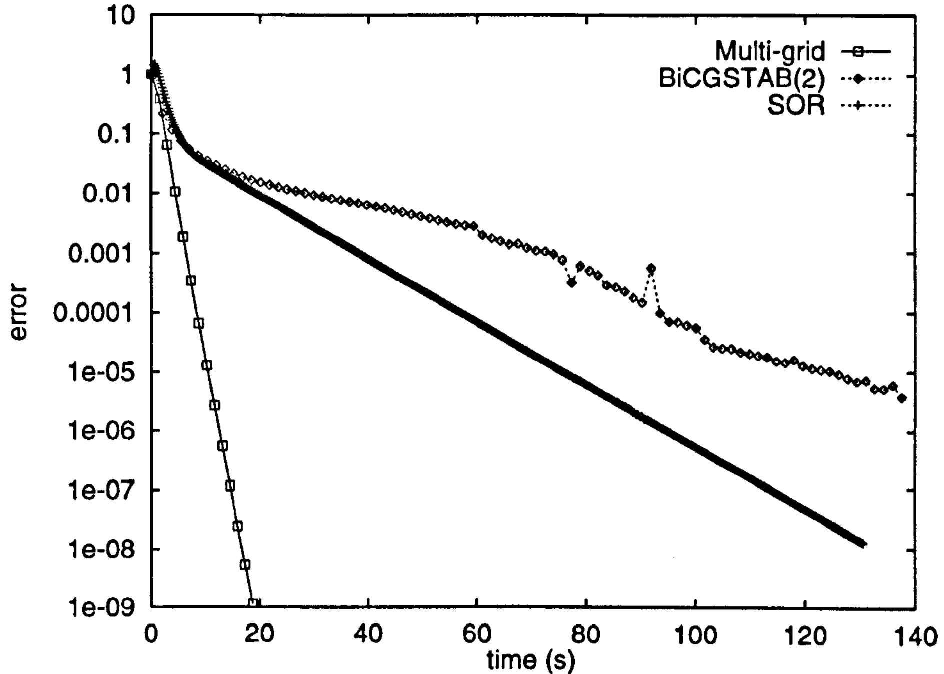


Fig. 3.6: A comparison of the Poisson solvers for a FET structure with particulate donors and electrons scattered at random. Double precision arithmetic has been used.

formed at each level, but no post smoothing was carried out as it did not have the same impact on the rate of convergence as presmoothing. Table 3.1 indicates the discretisation of the problem. The multigrid method was compared with other methods through the simulation of a simple FET containing discrete charges, to indicate its effectiveness and the results are shown in figure 3.6. The results for the SOR and the double step biconjugate gradient method with GMRES (BiCGSTAB(2) are shown below for comparison. The methods for calculating the cross products were kept as similar as possible and only the free abscissae were updated in order to ensure fairness.

Note that the conjugate gradient method (BiCGSTAB(2)) stalls for the complex potential landscape resulting from the point-like charges, despite being initially faster than the SOR scheme. Reducing the highly localised errors seems to be the major problem since only a small fraction of points are affected and so there are small changes in the norms of the vectors. This problem is

Tab. 3.1: The discretisation of the MESFET

number of levels	5
Coarse mesh	$5 \times 5 \times 3$
fine mesh	$65 \times 65 \times 33$
mesh spacing	1.25 nm

still apparent despite the use of double precision arithmetic because of the very large problem size (of 274625 abscissae), although the initial error falls faster than that for the SOR method. In the earlier comparison (see figure 3.2), the potential fell away to zero away from the discrete impurity and all the other nodes had no charge assigned to them. Therefore a small number of nodes around the discrete charge contributed significantly to the dot products. The multigrid method is unaffected and performs excellently. Since only smoothing operations are required it (along with the SOR method) is less reliant on arithmetical precision when the problem becomes large. In other words, relaxation methods seem to be more reliable for studying these more realistic problems containing (both positive and negative) discrete charges and biased contacts.

3.1.3 Obtaining the Electric Field

The discussion outlined below refers to a rectangular grid of constant spacing. The simplest method of getting the forces is to use a centred two-point finite differencing scheme.

$$\mathbf{E} = -\frac{1}{2h}\begin{pmatrix} V(i+1,j,k) - V(i-1,j,k) \\ V(i,j+1,k) - V(i,j-1,k) \\ V(i,j,k+1) - V(i,j,k-1) \end{pmatrix} \tag{3.24}$$

This is accurate to second order and, using Taylor expansion for the difference operator, it gives the leading term in its truncation error to be:

$$\varepsilon = \frac{h^2}{6}\left(\frac{\partial^3 V}{\partial x^3}, \frac{\partial^3 V}{\partial y^3}, \frac{\partial^3 V}{\partial z^3}\right) \tag{3.25}$$

More complicated schemes of higher order can be obtained by extending the differencing scheme to include more distant abscissae. A $2n$-point scheme,

using the nearest neighbours up to a distance of nh is equivalent to fitting a polynomial of degree $2n$ to this neighbourhood and calculating its gradient.

$$\mathbf{E}_{2n} = -\sum_{\varsigma=1}^{n} \frac{\alpha_i}{2\varsigma h} \begin{pmatrix} V(i+\varsigma,j,k) - V(i-\varsigma,j,k) \\ V(i,j+\varsigma,k) - V(i,j-\varsigma,k) \\ V(i,j,k+\varsigma) - V(i,j,k-\varsigma) \end{pmatrix}. \tag{3.26}$$

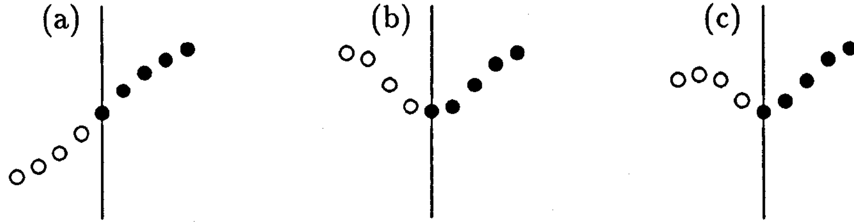


Fig. 3.7: Extrapolation at the boundary ($x = 0$).(a) Dirichlet, (b) von Neumann and (c) Periodic.

The high order difference schemes can still be applied at or near the boundaries by using the appropriate extrapolation, as shown in figure 3.7. For Dirichlet boundaries, the image charges and so the potentials have the opposite sign and thereby giving $V(-x) = -V(x)$. The images have the same sign for von Neumann boundaries where there is no electric field and so $V(-x) = V(x)$. Cells of length L with periodic boundaries just have $V(-x) = V(L-x)$.

There is a balance between the need for accuracy and that of expediency. The 4-point scheme is found to be adequate for the purpose of this work, although most authors use the 2-point scheme. The 4-point difference scheme results in an error of

$$\varepsilon = \frac{h^4}{30}\left(\frac{\partial^5 V}{\partial x^5}, \frac{\partial^5 V}{\partial y^5}, \frac{\partial^5 V}{\partial z^5}\right). \tag{3.27}$$

Figure 3.8 shows the two differencing schemes, in comparison with the Coulomb force. The 100 direction has been chosen as it represents the worst case. The 2-point scheme differs significantly from the true result even when the charges are separated by four mesh spacings.

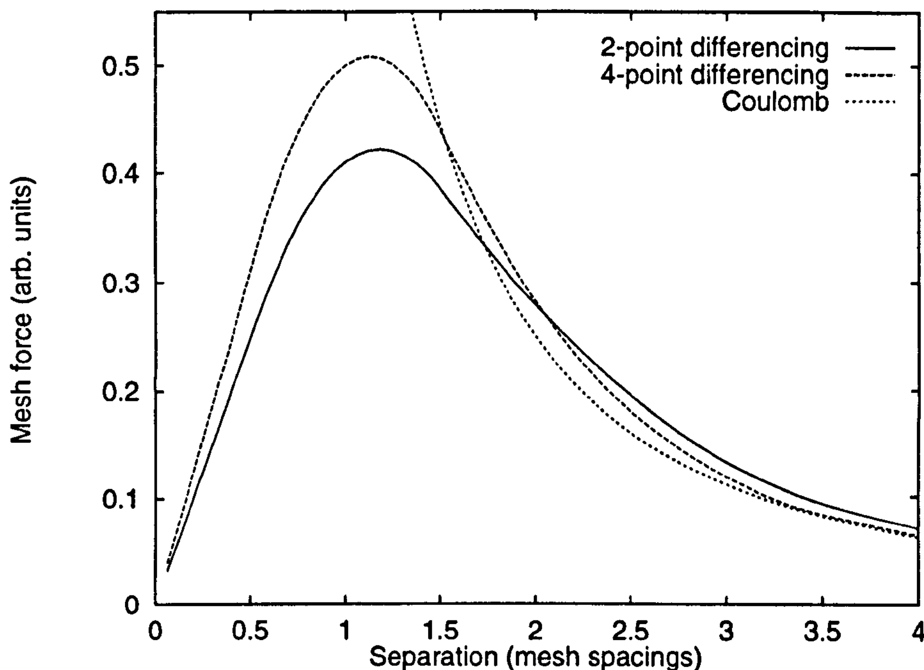


Fig. 3.8: The force between two charges, along the [100] direction, on a mesh using a 2-point and 4-point finite differencing scheme. The r^{-2} Coulomb force has been shown for comparison. Distances are in mesh spacings.

3.2 The short range forces

The previous section dealt with the traditional method of obtaining the electrostatic forces on charged particles. I shall now introduce a new method for estimating the mesh force that is both simple to implement and relatively accurate.

It is clear from figure 3.8 that the mesh gives accurate results when the separation of the particles is greater than around three mesh spacings. The results deviate significantly from the inverse square law when the charges approach to within three mesh spacings, and so it would be better to evaluate these forces analytically rather than use a mesh for these distances. The mesh would still be used to obtain the forces resulting from charges that are more than this distance and the boundary conditions, so the main task is to subtract the erroneous component from the mesh determined force. The problem is that this short range force varies with the relative position of the charges

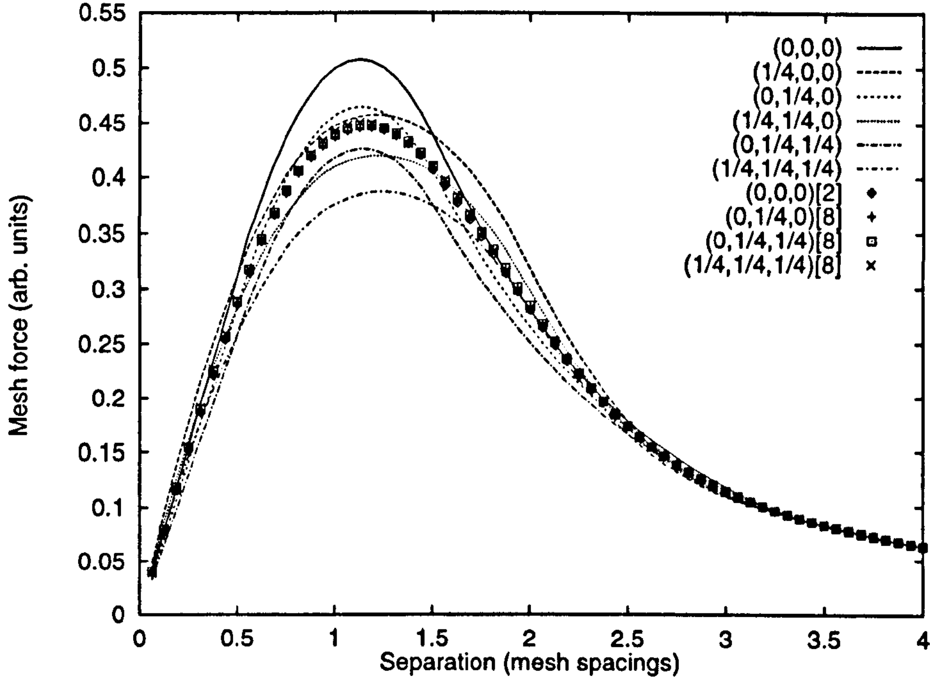


Fig. 3.9: The effects of using a mesh, and two [2] and eight [8] interlaced meshes on the mesh derived force along the [100] direction. The coordinate of the first charge is given.

with respect to the grid points and is determined by the range and relative orientations of the charges in space with respect to each other. This is then a problem of six variables to which an nth-order polynomial approximation can be made as represented below :

$$p(\mathbf{r}_1, \mathbf{r}_2) = \sum_{a=0}^{n} \sum_{b=0}^{n} \sum_{c=0}^{n} \sum_{d=0}^{n} \sum_{e=0}^{n} \sum_{f=0}^{n} \chi_{abcdef} s_1^a s_2^b s_3^c s_4^d s_5^e s_6^f \tag{3.28}$$

where the labelling of the position vectors are chosen so that $|\mathbf{r}_1| \leq |\mathbf{r}_2|$ and

$$\begin{aligned} s_1 &= x_1 \bmod h, & s_4 &= x_2 - x_1, \\ s_2 &= y_1 \bmod h, & s_5 &= y_2 - y_1, \\ s_3 &= z_1 \bmod h, & s_6 &= z_2 - z_1. \end{aligned}$$

Equation 3.28 has $(n + 1)^6$ terms and since a sixth order polynomial is required for a good fit there turns out to be 117649 terms. This scheme is obviously impractical for our purposes.

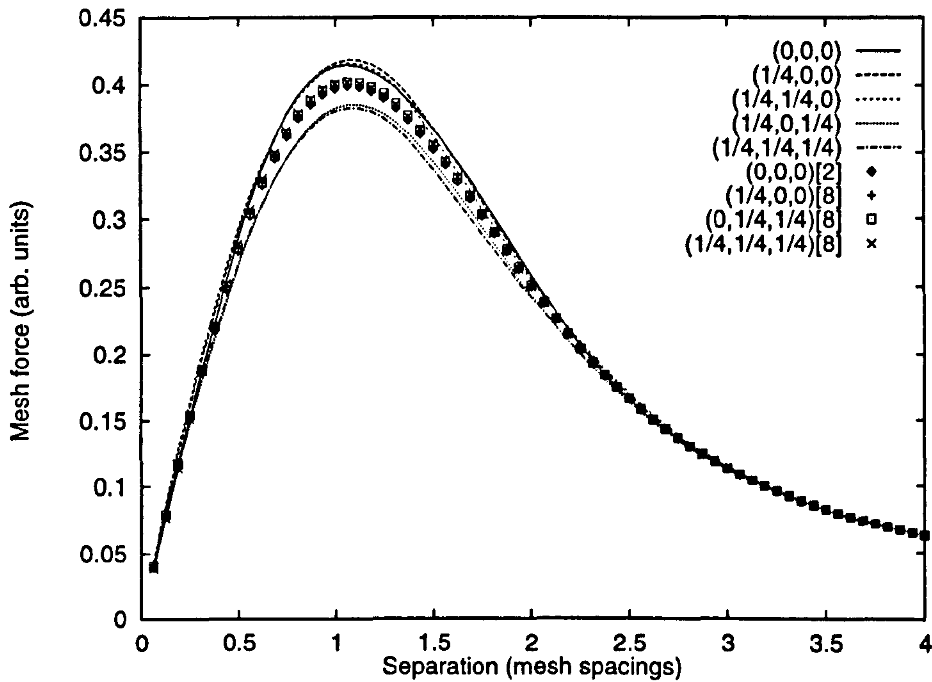


Fig. 3.10: The effects of using a mesh, and two [2] and eight [8] interlaced meshes on the mesh derived force along the [110] direction. The coordinate of the first charge is given.

The symmetry of the cubic lattice may be used to devise a simpler solution. There are three main axes of symmetry, which are the [100], [110] and the [111] directions of this lattice. The main aim is to reduce the dimensionality of the problem and to devise a computationally inexpensive solution to the mesh force.

The first three variables (s_1, s_2, s_3) are the coordinates of the first charge in the mesh cell, bounded by the nearest eight grid points. The first step is to see how the mesh derived force is affected by the coordinate vector (within a cell) and the separation vector between the charges. Figures 3.9–3.11 show the magnitude of force with the charge separation when they are aligned along each of the axes of symmetry. It is evident that the largest variation in the curves occurs when the charges are separated along the [100] direction. The variation is very much reduced for the other two directions with the separation along [111] showing the least dependance on the cell coordinates. A closer inspection

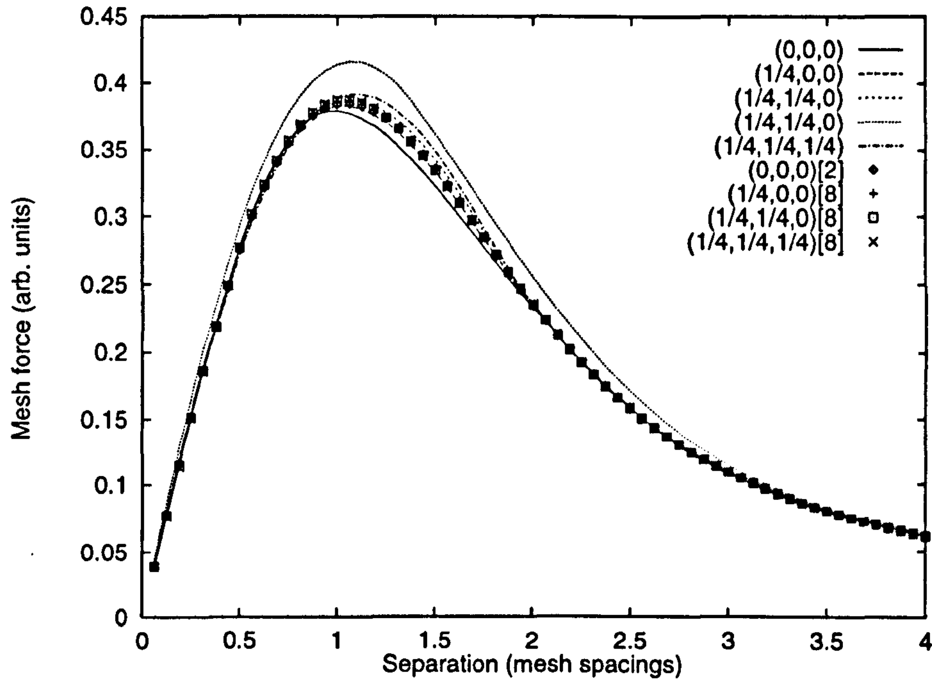


Fig. 3.11: The effects of using a mesh, and two [2] and eight [8] interlaced meshes on the mesh derived force along the [111] direction. The coordinate of the first charge is given.

reveals that the effects of aliasing are more pronounced for displacements in the cell that are in line with the separation. This manifests itself as a reduction in the magnitude of the peak and a shift in the position of the maximum force towards larger separation as the test charge moves further from the mesh point.

The largest contribution to the variation in the mesh force with this position is expected to be due to aliasing effects as described in page 41. Recall that it is the first harmonic that contributes to a large fraction of the problem. The use of an interlaced set of meshes reduces these effects and thereby reducing the problem to one of three variables. Note that the use of eight interlaced meshes gives little improvement in the estimates of force despite the quadrupled effort involved. This is in keeping with the simplified description where only the first harmonics contribute significantly to the potentials and that they are separable functions in x,y and z.

The electric field |E| is tabulated as a function of the relative displace-

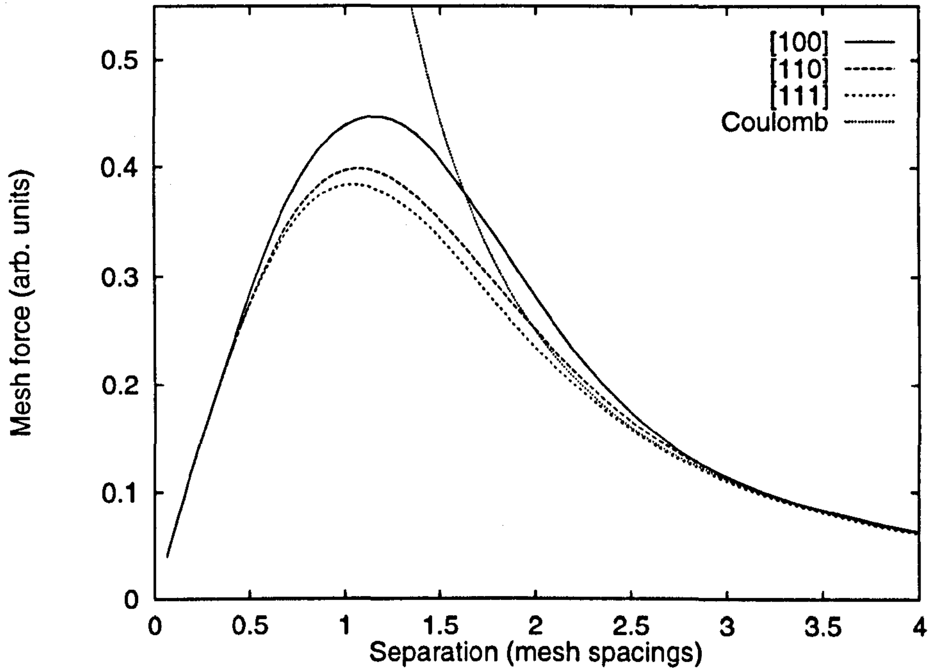


Fig. 3.12: The force as derived from the mesh when the line joining the two charges lie on the [100], [110] and [111] directions.

ment between the two charges, r_{ij}, along each of these directions of symmetry (figure 3.12) and simple polynomial functions are used to interpolate between them to get the mesh force. The assumption that has been made here is that the mesh force is in the direction of the separation vector of the particles (r_{ij}). Although this assumption is not absolutely correct, the method does give results that are reasonably accurate for the 4-point differencing scheme. The first step is to sort the components of $\mathbf{r}_{ij}$ by decreasing order of magnitude and label them l_1, l_2 and l_3. Since we are dealing with magnitude it would be advantageous to deal with squares of numbers as this minimises the operations count.

The quantity

$$\cos^2\theta = \frac{l_1^2}{r_{ij}^2} \tag{3.29}$$

can be used both to interpolate between the forces along the [100] and [111] directions, and also for the interpolation between the [100] and [110] directions.

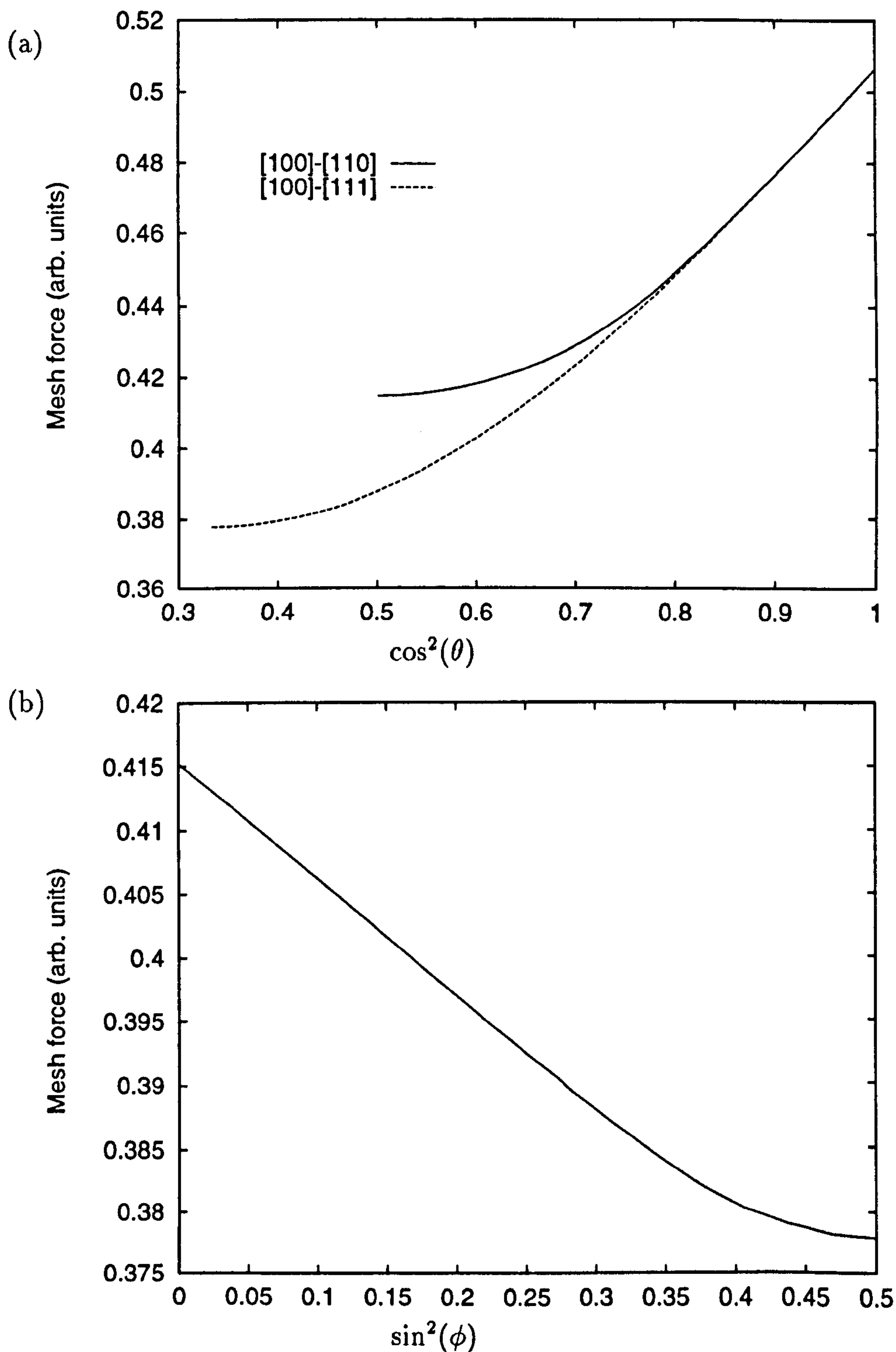


Fig. 3.13: The mesh force at a range of 1.0625 mesh spacings as a function of (a) $\cos^2(\theta)$ and (b) $\sin^2(\phi)$.

Finally,

$$\sin^2\phi = \frac{l_3^2}{l_2^2 + l_3^2} \tag{3.30}$$

is used to interpolate the net force from these two quantities. The magnitude of the mesh force, for when the charges are separated by 1.0625 mesh spacings, is shown in figure 3.13 as a function of $\cos^2\theta$ and $\sin^2\phi$. (This is the range where the mesh forces differ the most along the directions of symmetry.)

It turns out that there is less than a 2.5% difference in the force between $\cos^2\theta = 1/2$ and $\cos^2\theta = 1/3$ and so in order to simplify the calculations and gain some increase in speed, a lower limit of 1/2 was chosen. Figure 3.14 shows the quadratic approximations to the weights for $\cos^2\theta$ that is used for the [100] − [111] and [100] − [110] interpolations and the linear approximation is used for $\sin^2\phi$.

The ideal test is to compute the forces along the [411] direction as this is furthest away from the three directions of symmetry and so this result is shown in figure 3.15. It is clearly evident that my method approximates the actual mesh force to within 2%.

It is only necessary to correct the mesh derived force for separations less than around three mesh spacings and so for these short range interactions, the simulation domain is subdivided into a series of cells, and the appropriate correction was applied to the mesh force between any two particles (within the same cell or in two neighbouring cells) that are within this distance. This procedure is very similar to the use of chaining mesh by Hockney in his particle-particle-mesh method [93, 95] except that an array of pointers, to the index of the particles, are used here instead of a linked list. The choice for the size of these cells is a fine art. If they are too large then too much of the cpu time is spent on rejecting particles that are sufficiently far away. Reducing the size of the cells increase the memory requirement for the storage of pointers and also begins to waste effort in seeking particles from non-neighbouring cells (this is especially true when the problem is placed on a distributed network as discussed in section 5.2.3).

This chapter considered the intricacies of studying the electrostatics in systems with discrete charges. The solution of Poisson's equation on a mesh, although being very efficient, suffers from the problem of band limiting due the

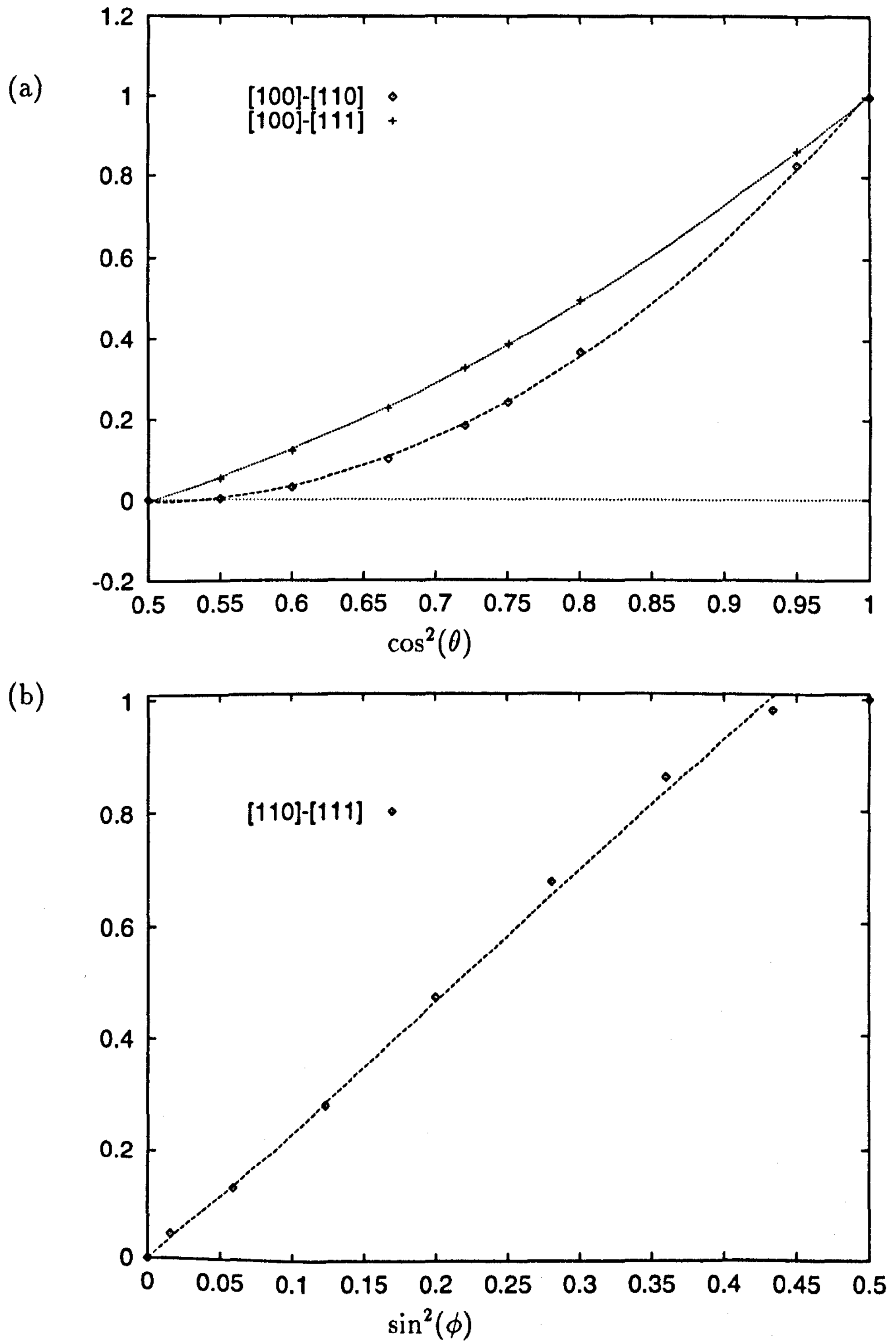


Fig. 3.14: The interpolation weights as a function of (a) $\cos^2\theta$ and (b) $\sin^2\phi$.

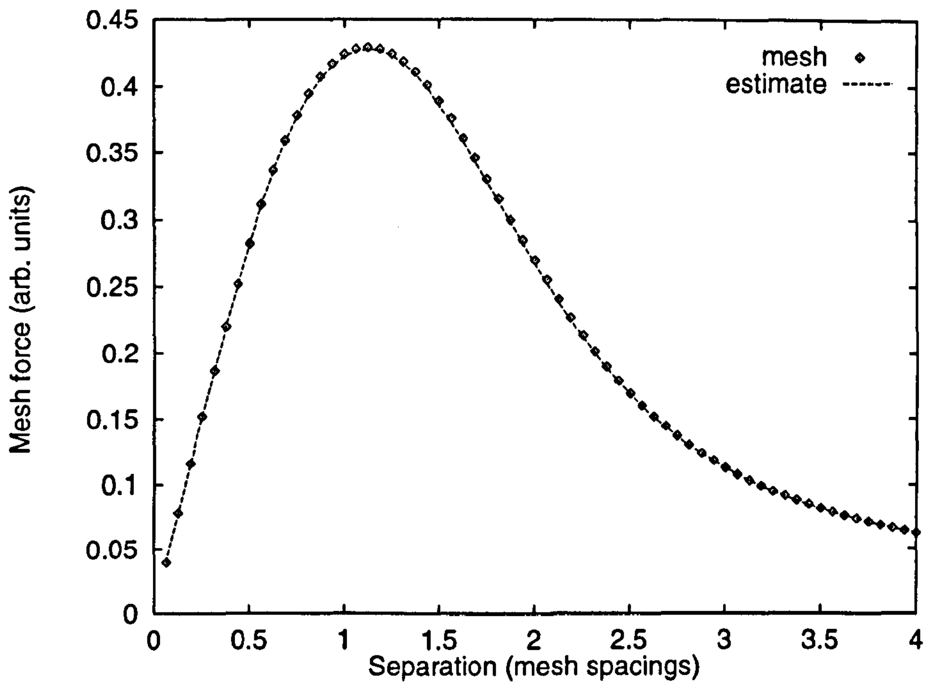


Fig. 3.15: The test for the approximation to the mesh force along the [411] direction.

the sampling of the potential at discrete intervals. A novel method has been proposed to overcome this limitation using the symmetry of the rectangular mesh. The next logical step is to show that the use of the Brownian approach with a mesh can reasonably reproduce the scattering by ionised impurities. The results will provide a qualitative estimate of the errors introduced by the band limiting arising from the use of a mesh.

4. THE DYNAMICS OF CARRIERS IN MEDIA WITH DISCRETE IMPURITIES: A TEST OF THE EQUATIONS OF MOTION

The simulation of sub-0.1 μm devices requires the integration of the Coulomb force between individual particles. In very small devices, both the number of dopants and their distribution in space vary significantly. This chapter focuses on the study of scattering by ionised impurities and considers efficacy of using simple mesh based schemes, that is the mainstay of conventional Monte Carlo models. The '*ab initio*' calculations presented here will demonstrate the effects of including all the ions in the electrostatics, instead of viewing these interactions simply as just another random scattering process. One of the aims is to answer the question whether the precise distribution of ions affects the overall scattering of carriers.

In semiconductor physics, the electrons are represented by Bloch wave functions that occupy the whole of the crystal, but this picture runs into trouble when describing scattering due to ionised impurities. The first issue is the fact that these interactions are relatively well localised in space, especially for particles of high energy. This fact can be clearly seen in the differential cross-section, given by the Rutherford formula

$$\begin{aligned} \sigma(\theta) &= \frac{b}{\sin\theta}\left|\frac{db}{d\theta}\right| \\ &= \left(\frac{Ze^2}{16\pi\epsilon\mathcal{E}}\right)^2 \frac{1}{\sin^4\left(\frac{\theta}{2}\right)} \end{aligned} \tag{4.1}$$

where θ is the angle between the asymptotes of the incident and the scattered trajectories of a particle with energy $\mathcal{E}$. The impact parameter b is shortest distance between each asymptote and the impurity of charge Ze (see figure 4.1). The number of particles that are scattered through an angle between θ and

$\theta + d\theta$, per unit time, can be calculated using the differential cross-section.

$$dn = nv\, N\sigma(\theta)d\Omega \tag{4.2}$$

for incident particles, of density n and velocity v passing through a medium with N scattering centres per unit volume.

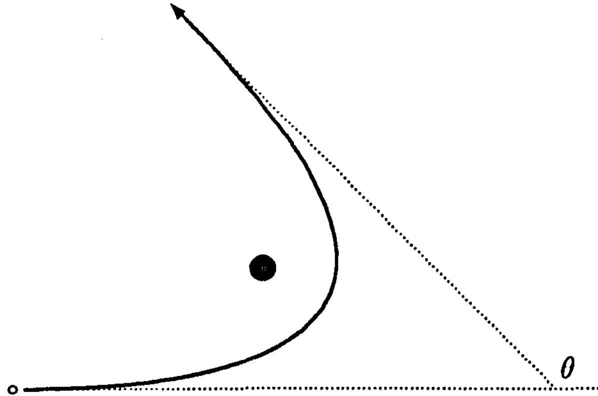


Fig. 4.1: The trajectory of an electron around an ionised donor.

A full quantum mechanical calculation (through the solution of the Schrödinger equation, using the boundary conditions of incident plane waves and spherical scattered waves) gives the same result [116]. The $1/r$ potential influences the wavefunction throughout the whole of space and the cross-section tends to infinity as the carrier's energy tends to zero. What is being said here is that the motion of the carriers is continuously influenced by the discrete charges in the sample and this fact is often neglected in conventional Monte Carlo simulations where ionised impurity scattering is represented as discrete collision events. This potential can be viewed as a combination of sharp localised peaks due to the ion cores and an undulating background due to the more distant charges. The background potential is also significantly influenced by the precise distribution of the ions and other carriers in space. The random scattering events through space and time that is used in conventional Monte Carlo simulations are simply inappropriate when there are few charges that affect the motion of carriers. The fact that shapes of the wavefunctions and their position in space influences the potential distribution which in turn has

repercussions on the wavefunctions indicates that a full quantum transport model may be required to study this problem. The simulations that were undertaken all used classical dynamics and agreed well with experimental results, which indicates this effort may not be necessary.

In this chapter I will firstly review the standard models of scattering by ionised impurities [117, 118]. They provide useful approximations to the scattering rate under various conditions: the Conwell-Weisskopf model best describes the limit of weak screening and light doping while the Brooks-Herring is most appropriate in the limit of strong screening. Ridley's third body exclusion, that attempts to reconcile these approximations, will also be included.

This will be followed by the practical considerations in obtaining the electrostatic forces acting on the carriers and ways of integrating these rapidly varying forces in space. The latter will be dealt with in the study of the trajectory of a carrier around a single impurity.

The dynamics of the carriers in bulk samples will then be investigated by using samples with periodic boundary conditions. Following single carriers in these samples would give the results for unscreened scattering and would serve as a baseline to study the effects of screening. The comparison between these results and the Conwell-Weisskopf model [119] would indicate the significance of the more distant impurities on the dynamics, since this model only considers the interactions between carriers and the nearest impurity. The inclusion of all the carriers in the simulation will introduce screening and the accuracy of the Brooks-Herring model [120, 121] and Ridley's statistical screening [122] can be studied.

Careful attention will be paid to the effects of the constraints imposed by the physical model on the results such as the size of the sample, the size and distribution of the impurities and the calculation of forces and method of their integration. A comparison between the use of Brownian dynamics and free flights will be made to ensure that the introduction of the frictional and stochastic forces do not interfere with the integration of the electrostatic forces.

Finally the effects of the discrete ions on the percolation level will be investigated. This will be of great importance in practical simulations of small devices where the precise distribution of ions can have a significant effect on

the shape of the gate barrier potential.

4.1 Scattering models

The dynamics of carriers in the a doped semiconductor is a complex many body problem but the physical models simplify the dynamics to that of two-body or three-body problems. Three basic types of model that have been formulated to look at the scattering of carriers by ionised impurities will now be outlined. These are the Conwell-Weisskopf, Brooks-herring and Ridley's 'third body exclusion' approach.

4.1.1 The Conwell-Weisskopf model

The Conwell-Weisskopf model [119] is the simplest model and treats electron impurity interactions as a two-body problem (*i.e.* between the electron and the *nearest* impurity). It treats the carrier as a point charge that is scattered by the ion core as in the classical Rutherford model (figure 4.1). The angle through which it is deflected can be estimated from the asymptotes of the trajectory.

$$\frac{\theta}{2} = \arctan\left(\frac{Ze^2}{8\pi\epsilon b\mathcal{E}}\right) \tag{4.3}$$

where b is the impact parameter for a carrier deflected by in impurity of charge Ze. (This formula is valid for both attractive and repulsive interactions.)

This procedure gets around the problem of infinite cross-section by limiting the maximum impact parameter to 1/2 the average impurity separation

$$b_{\max} = \frac{1}{2\sqrt[3]{N_I}}, \tag{4.4}$$

and thereby placing a lower limit on the scattering angle

$$\frac{\theta_{\min}}{2} = \arctan\left(\frac{Ze^2\sqrt[3]{N_I}}{4\pi\epsilon\mathcal{E}}\right) \tag{4.5}$$

The momentum relaxation cross-section for the Conwell-Weisskopf model is

$$\begin{aligned}\sigma &= \int_{\theta_{\min}}^{\pi} \sigma(\theta)\, 2\pi\, (1-\cos\theta)\, \sin\theta\, d\theta \\ &= \frac{Z^2e^4}{32\pi\epsilon^2\mathcal{E}^2} \log\left(1 + \frac{16\pi^2\epsilon^2\mathcal{E}^2}{Z^2e^4 N_I^{\frac{2}{3}}}\right).\end{aligned} \tag{4.6}$$

The scattering rate (as defined by τ_p^{-1} where τ_p is the momentum relaxation time) is just $v\sigma N_I$ and so for parabolic bands this rate is

$$\Gamma = \frac{Z^2 e^4 N_I}{16\pi\epsilon^2\sqrt{2m\mathcal{E}^3}} \log\left[1+\left(\frac{4\pi\epsilon\mathcal{E}}{Ze^2\sqrt[3]{N_I}}\right)^2\right] \tag{4.7}$$

4.1.2 The Brooks-Herring model

Electron-impurity interactions are not just a merely two-body problem. The incoming carrier can transfer some momentum to other mobile carriers and so, in effect, be partially screened from the impurity's electrostatic field. The resulting electrostatic potential around the impurity decays much more rapidly than $1/r$

$$V(r) = \frac{Ze}{4\pi\epsilon r}\exp(-\beta r) \tag{4.8}$$

and β is the inverse screening length

$$\beta = \sqrt{\frac{e^2 n}{\epsilon k_B T}}. \tag{4.9}$$

This Debye-Hücknel potential best describes non-degenerate semiconductors, that are not heavily doped, at low temperatures. The exponential factor resolves the problem of the infinite cross-section of the bare Coulomb potential. The Born approximation is used to calculate the scattering rate and so for a crystal of volume $\mathcal{V}$,

$$\Gamma = \frac{\mathcal{V}}{8\pi^3}\int\frac{2\pi}{\hbar}\left|\langle\mathbf{k}'|eV(\mathbf{r})|\mathbf{k}\rangle\right|^2 \delta(\mathcal{E}'-\mathcal{E})\,d\mathbf{k}'. \tag{4.10}$$

Since the carriers are represented as plane waves :

$$\begin{aligned}\langle\mathbf{k}'|eV(\mathbf{r})|\mathbf{k}\rangle &= \int\exp(-i\mathbf{k}'\cdot\mathbf{r})\frac{Ze^2\exp(-\beta|\mathbf{r}|)}{4\pi\epsilon|\mathbf{r}|}\exp(i\mathbf{k}\cdot\mathbf{r})d\mathbf{r} &(4.11)\\ &= \frac{Ze^2}{\mathcal{V}\epsilon}\frac{1}{|\mathbf{k}-\mathbf{k}|^2+\beta^2}. &(4.12)\end{aligned}$$

The transfer of energy to the ion is negligible since it is so much heavier than the carrier and so $|\mathbf{k}'| = |\mathbf{k}| = k$ and so some simple trigonometry yields

$$\Gamma = \frac{Z^2e^4}{4\pi^2\hbar\epsilon^2\mathcal{V}}\iiint\frac{\delta(\mathcal{E}'-\mathcal{E})}{\left[2k^2(1-\cos\theta)+\beta^2\right]^2}k'^2\,dk'\ \sin\theta\,d\theta\,d\phi. \tag{4.13}$$

This eventually gives the scattering rate to be

$$\Gamma = \frac{Z^2 e^4 N_I}{16\pi\epsilon^2\sqrt{2m\mathcal{E}^3}}\left[\log\left(1+\frac{8m\mathcal{E}}{\hbar^2\beta^2}\right) - \frac{1}{1+\frac{\hbar^2\beta^2}{8m\mathcal{E}}}\right] \tag{4.14}$$

The Born approximation assumes that the screened electrostatic potential has negligible effects on the carrier wavefunction and so can be viewed as a small perturbation. This is only valid if

$$\left|-\frac{m}{2\pi\hbar^2}\int\frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{r}V(\mathbf{r})d\mathbf{r}\right| \ll 1 \tag{4.15}$$

So for the Debye-Hücknel potential

$$\begin{aligned}
1 \gg \quad & \left|-\frac{Ze^2m}{\hbar^2k}\int_0^\infty(e^{-2ikr}-1)\frac{e^{-\beta r}}{r}dr\right| \\
= \quad & \left|-\frac{Ze^2m}{\hbar^2k}\int_0^\infty(e^{-2ikr}-1)\int_\beta^\infty e^{-sr}dsdr\right| \\
= \quad & \frac{Ze^2m}{\hbar^2k}\left|\int_\beta^\infty\left[\frac{1}{s-2ik}-\frac{1}{s}\right]ds\right| \\
= \quad & \frac{Ze^2m}{\hbar^2k}\left|\log(1-2i\frac{k}{\beta})\right| \\
= \quad & \frac{Ze^2m}{\hbar^2k}\sqrt{\left(\frac{1}{2}\log\left(1+\frac{4k^2}{\beta^2}\right)\right)^2+\left(\tan^{-1}\frac{2k}{\beta}\right)^2}
\end{aligned} \tag{4.16}$$

This argument is in keeping with the intuitive notion that the validity of the Born approximation is enhanced when the momentum is very large and where there is strong screening. The right hand side of equation 4.16 has been evaluated for GaAs and Si to determine the justification for the use of the Born approximation in typical cases. (A sample at room temperature and doped to $10^{24}\,\mathrm{m}^{-3}$ was considered for each case.)

	GaAs	Si
k	$2.13\times10^8\,\mathrm{m}^{-1}$	$4.72\times10^8\,\mathrm{m}^{-1}$
β	$2.30\times10^8\,\mathrm{m}^{-1}$	$2.43\times10^8\,\mathrm{m}^{-1}$
Eq. 4.16	0.6	2.1

Eq. 4.16 indicates that the Born approximation is not very safe for GaAs at room temperature and not valid for Si. However, the calculations approximate the observed values and so they are used in many device simulations.

The screening carriers have finite masses and cannot react instantly to the test particle. The dielectric relaxation time, $\tau_e = \epsilon\rho$, gives a measure of how fast the screening charges can respond in a sample with permitivity ϵ and resistivity ρ. If the scattering rate (Γ) is so low that $[(\Gamma\tau_e)^2 << 1]$ then a many-body treatment is required to include correlation effects. The inverse screening length should then be modified and so for non-degenerate semiconductors the scattering rate is given by Takimoto [123] as

$$\Gamma = \frac{Z^2 e^4 N_I}{16\pi\epsilon^2\sqrt{2m\mathcal{E}^3}}\left[\log\left(1+\frac{8m\mathcal{E}}{\hbar^2\beta^2\zeta}\right) - \frac{1}{1+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}}\right] \tag{4.17}$$

where

$$\zeta = \frac{1}{y\sqrt{\pi}}\int_0^\infty x\exp(-x^2)\log\left|\frac{x+y}{x-y}\right|dx \tag{4.18}$$

and

$$x = \frac{\hbar|\mathbf{k}'-\mathbf{k}|}{2\sqrt{2mk_BT}} = \frac{\hbar k\sqrt{1-\cos\theta}}{2\sqrt{mk_BT}} \tag{4.19}$$

4.1.3 Statistical Screening

Statistical Screening, by Ridley, attempts to reconcile the Conwell-Weisskopf and Brooks-Herring models by taking a probabilistic view of screening. This model still treats impurity scattering as a two-body process (between the carrier and its nearest impurity), but rather than using the mean impurity separation as in the Conwell-Weisskopf model, it uses probability of finding a third scattering centre to weight the differential cross-section for each impact parameter. The probability of finding another impurity for an impact parameter lying between b and $b + db$ is $2\pi N_I ab\,db$, where $a \approx (2\pi N_I)^{1/3}$ [124] is the average separation of the ions. If P(b) is the probability that there was no other impurity within a radius of b then

$$P(b+db) = (1 - 2\pi N_I ab\,db)P(b), \tag{4.20}$$

i.e.

$$\frac{dP}{db} + 2\pi N_I abP = 0. \tag{4.21}$$

Since $P \leq 1$, this has the solution

$$P(b) = \exp(-\pi N_I ab^2). \tag{4.22}$$

This exponential decay in probability effectively imposes an upper limit on the impact parameter. Screening is introduced through the use of the differential cross-section of the Brooks-Herring model

$$\sigma(\theta) = \left[\frac{Z^2e^2}{16\pi\epsilon\mathcal{E}}\frac{1}{\sin^2\left(\frac{\theta}{2}\right)+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}}\right]^2 \tag{4.23}$$

and so after noting that

$$b^2 = \frac{Z^2e^4}{16\pi\epsilon^2\mathcal{E}}\frac{\cos^2\left(\frac{\theta}{2}\right)}{\left(\sin^2\left(\frac{\theta}{2}\right)+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}\right)\left(1+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}\right)} \tag{4.24}$$

the new weighted differential cross-section becomes

$$\begin{aligned}\sigma(\theta) &= \left[\frac{Z^2e^4}{16\pi\epsilon^2\mathcal{E}}\frac{1}{\sin^2\left(\frac{\theta}{2}\right)+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}}\right]^2 \\ &\quad \exp\left(-\pi N_I a\frac{Z^2e^4}{16\pi\epsilon^2\mathcal{E}}\frac{\cos^2\left(\frac{\theta}{2}\right)}{\left(\sin^2\left(\frac{\theta}{2}\right)+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}\right)\left(1+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}\right)}\right).\end{aligned} \tag{4.25}$$

This expression gives the scattering rate to be

$$\begin{aligned}\Gamma &= \frac{Z^2e^4N_I}{16\pi\epsilon^2\sqrt{2m\mathcal{E}^3}} \\ &\quad \left\{\exp(f)\left(E_1(f)-E_1(g)\right)-\frac{1}{g}\left[1-\exp\left(-\frac{g}{1+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}}\right)\right]\right\}\end{aligned} \tag{4.26}$$

where

$$\begin{aligned} f &= \pi N_I a\left(\frac{Z^2e^2}{8\pi\epsilon\mathcal{E}}\right)^2\frac{1}{1+\frac{\hbar^2\beta^2\zeta}{8m\mathcal{E}}} \\ g &= 4\pi N_I a\left(\frac{Z^2e^2m}{4\pi\epsilon\hbar^2\beta^2}\right)^2\end{aligned}$$

and the exponential integral

$$E_1(x) = \int_x^\infty \frac{\exp(-t)}{t}dt.$$

It approaches the Conwell-Weisskopf result for weak screening (*i.e.* $\beta \to 0$) and the Brooks-Herring result for strong screening, specifically when $N_I \to 0$.

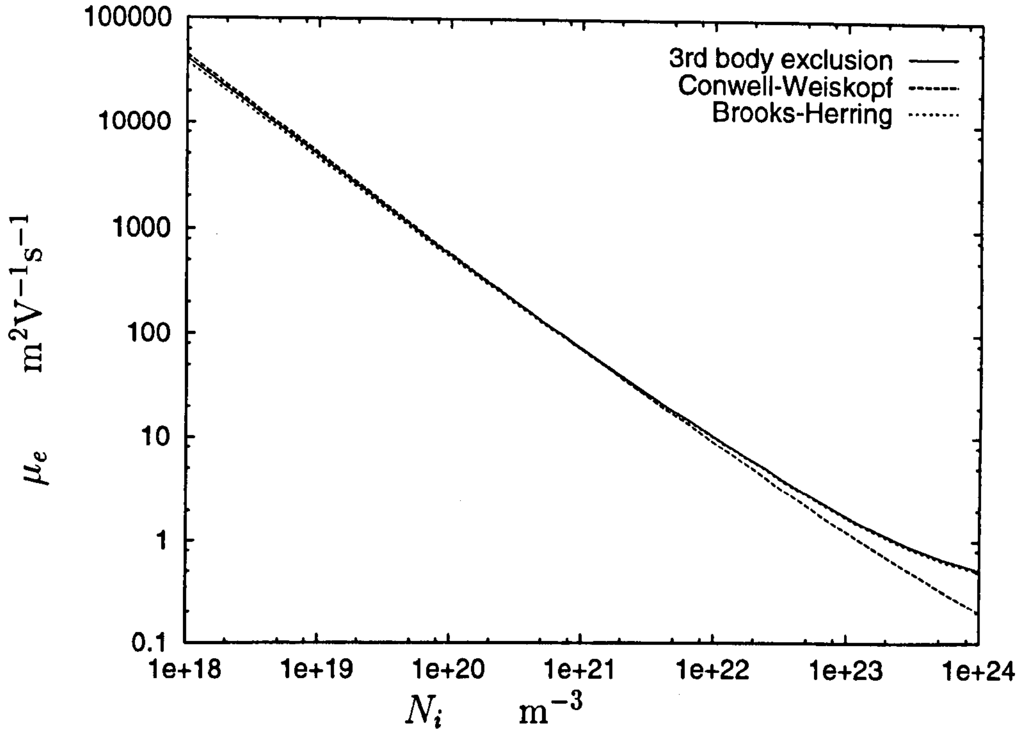


Fig. 4.2: The electron mobility for a neutral sample of bulk GaAs as predicted my the Brooks-Herring, Conwell-Weisskopf and third body exclusion models. Note that the third body exclusion formula produces very similar results to the the Brooks-Herring approach. At low impurity concentrations the results start to move closer to the Conwell-Weisskopf model.

For non-degenerate semiconductors a Maxwell-Boltzmann distribution of particles can be a reasonable approximation. Then the density of states is given by

$$dn = \frac{2\pi}{(\pi k_{\mathrm{B}}T)^{\frac{3}{2}}}\mathcal{E}^{\frac{1}{2}} \exp\left(-\frac{\mathcal{E}}{k_{\mathrm{B}}T}\right) d\mathcal{E} \tag{4.27}$$

and the expected value of an observable is given by

$$\langle A \rangle = \frac{2\pi}{(\pi k_{\mathrm{B}}T)^{\frac{3}{2}}} \int_0^\infty A(\mathcal{E})\mathcal{E}^{\frac{1}{2}} \exp\left(-\frac{\mathcal{E}}{k_{\mathrm{B}}T}\right) d\mathcal{E} \tag{4.28}$$

Figure 4.2 gives the momentum relaxation time as predicted by the Conwell-Weisskopf, Brooks-Herring and third-body exclusion models for an electrically neutral sample of GaAs. Note that Ridley's third-body exclusion model lies

between the Brooks-Herring and Conwell-Weisskopf results and closely approximates the Brooks-Herring as the doping is increased. Ridley defines a dimensionless constant which indicates whether third body exclusion approximates the Brooks-herring or Conwell-Weisskopf

$$\eta = 4Z\frac{N_{\mathrm{I}}^{\frac{1}{3}}}{\beta}\frac{R_H}{E} \qquad \begin{cases} \ll 1 & \text{Brooks-Herring} \\ \gg 1 & \text{Conwell-Weisskopf} \end{cases} . \tag{4.29}$$

For GaAs at room temperature,

	N_I (m^{-3})		
	10^{18}	10^{21}	10^{24}
η (μm)	2.3	0.74	0.0024
a (μm)	0.54	0.054	0.0054
β^{-1} (μm)	4.3	0.14	0.0043

and so (neglecting the influence of the electron's energy) it can be seen that two length scales influence the behaviour of Ridley's third body exclusion method. Not surprisingly these are the lengths that cut off the influence of the ion's electrostatic potential. The mean impurity separation becomes much smaller than the Debye screening length when the screening is weak and so this dominates and makes the result appear similar to the Conwell-Weisskopf model and vice versa.

The value of η is smaller than unity across the doping range where ionised impurity scattering becomes significant for GaAs and so the third body elimination method approximates the Brooks-Herring model. The third body exclusion method approaches the Conwell-Weisskopf model as the doping is reduced, but does not converge towards exactly the same value because the Conwell-Weisskopf model uses a different value for the mean separation than Ridley's model. This is not very important at low doping since the scattering rates are small and all three models give quite similar results.

4.2 The dynamics of carriers in a Coulomb potential

The main objective of this part of the work is to investigate the validity of using the simple methods that are employed in many conventional simulations and

to determine whether they can give a reasonable approximation to scattering by ionised impurities. These methods, using simple linear interpolation and differencing, serve to accentuate the problems associated with a mesh.

The standard models only consider the angle between the asymptotes of the incoming and outgoing trajectories. The use of asymptotes to describe scattering is only valid when the initial and final ranges of the scattered carrier is much larger than the impact parameter. An analytic formula for the path of a carrier around a solitary ion can be obtained by considering the conservation of its angular momentum and total energy under a conservative central force. An angular momentum per unit mass $\mathbf{h}$ can be defined as

$$\mathbf{h} = \frac{\mathbf{L}}{m} = \mathbf{r} \times \mathbf{v} = rv \sin \phi \tag{4.30}$$

where ϕ is the angle between the electron's relative position and velocity vectors and $h = |\mathbf{h}|$ is a constant of motion.

The energy of a particle under the influence of this force is

$$\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + V(r) = E \tag{4.31}$$

Upon making a change of variable $u = 1/r$ and noting that equation 4.30 yields $\dot{\theta} = hu^2$, the energy equation for the orbit becomes

$$\frac{1}{2}m\left[\left(\frac{du}{d\theta}\right)^2 + u^2\right] - ku = E \tag{4.32}$$

when $V(r) = -k/r = -ku$.

The range can be computed, as a function of angle, after separating the the variables and integrating,

$$r = \frac{mh^2k^{-1}}{1 + \sqrt{(1 + 2Emh^2k^{-2})}\cos(\theta - \theta_0)}. \tag{4.33}$$

Therefore, assuming that the initial direction of travel was in the positive x-direction, then

$$\theta_0 = \cos^{-1}\left[\frac{mv_0^2b^2k^{-1}r_0^{-1} - 1}{\sqrt{1 + 2Emv_0^2b^2k^{-2}}}\right] \tag{4.34}$$

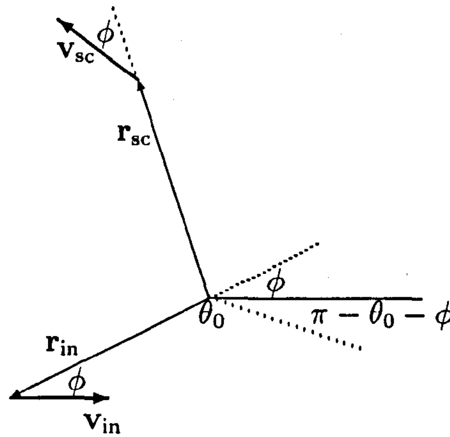


Fig. 4.3: The geometry of the orbit of an electron around a donor. The angle between incident and final scattered positions (where $|\mathbf{r}_{in}| = |\mathbf{r}_{sc}|$) is $2\theta_0 + \phi - \pi$.

The geometry of the problem (figure 4.3) can be used to determine the angle of scatter (between the incident and scattered velocity vectors) to be

$$\theta_v = 2(\theta_0 + \phi_0) - \pi \tag{4.35}$$

when the final range reaches the initial value. The corresponding angle between the initial and final position vectors is

$$\theta_r = 2\theta_0 + \phi_0 - \pi. \tag{4.36}$$

These analytic formulae can be used to verify the procedures for the resolution of the electrostatic forces and the integration of the equations of motion. The inclusion of all electrostatic interactions between the charged impurities and carriers lies at the heart of this work. The complex geometries and boundary conditions necessitate the use of a mesh for the electrostatics, but this results in band-limiting of the forces. In order to cut down on the computational effort, the electric fields were obtained using the simple first order differencing of the mesh

$$\begin{aligned} E_x &= -\frac{1}{h} \Big((1-\beta)\left(V(i+1,j) - V(i,j)\right) \\ &\quad + \beta \left(V(i+1,j+1) - V(i,j+1)\right)\Big) \end{aligned} \tag{4.37}$$

$$E_y = -\frac{1}{h} \Big((1-\alpha)\,(V(i,j+1,k) - V(i,j,k)) + \alpha \;\; (V(i+1,j+1) - V(i+1,j))\Big) \quad (4.38)$$

where

$$\alpha = x \bmod h_x$$
$$\beta = y \bmod h_y$$

are the coordinates of the particle within a cell bounded by eight adjacent grid points in a regular rectangular mesh with spacings h_x, h_y and h_z. This method of force estimation is cruder that the 4th-order finite difference scheme that is used in the actual device simulations and so gives an upper bound on the errors incurred when using a mesh. These first set of experiments had the aim of investigating the errors due to the differencing scheme and so the values at each mesh point was just the analytic evaluation of the Coulomb force between point charges. The impurity ion was placed on the centre mesh point. The Euler method is used to integrate the acceleration due to the electrostatic forces as in the Brownian simulations. The comparison with simulations using direct analytic estimates for the Coulomb force are used to investigate the errors introduced by the differencing scheme for both attractive (figures 4.4) and repulsive fields (figure 4.5). The analytic results (for finite ranges) have not been shown for the repulsive field because there were no significant differences between them, as expected. It is clear that the results differ significantly from the Rutherford formula as the carrier energy is reduced for both attractive and repulsive interactions. This is because there is a more significant change in momentum at lager distances away from the ion then the carrier is travelling more slowly.

The attractive force proves to be difficult to integrate, using the simple Euler method, because both the velocity and the magnitude, and rate of change, of the Coulomb force increase much more rapidly as the carrier nears the impurity. The strength of the force is underestimated as the carrier approaches the impurity and is overestimated as it speeds away and so energy is not conserved during the collision. The oscillations that are evident in the attractive

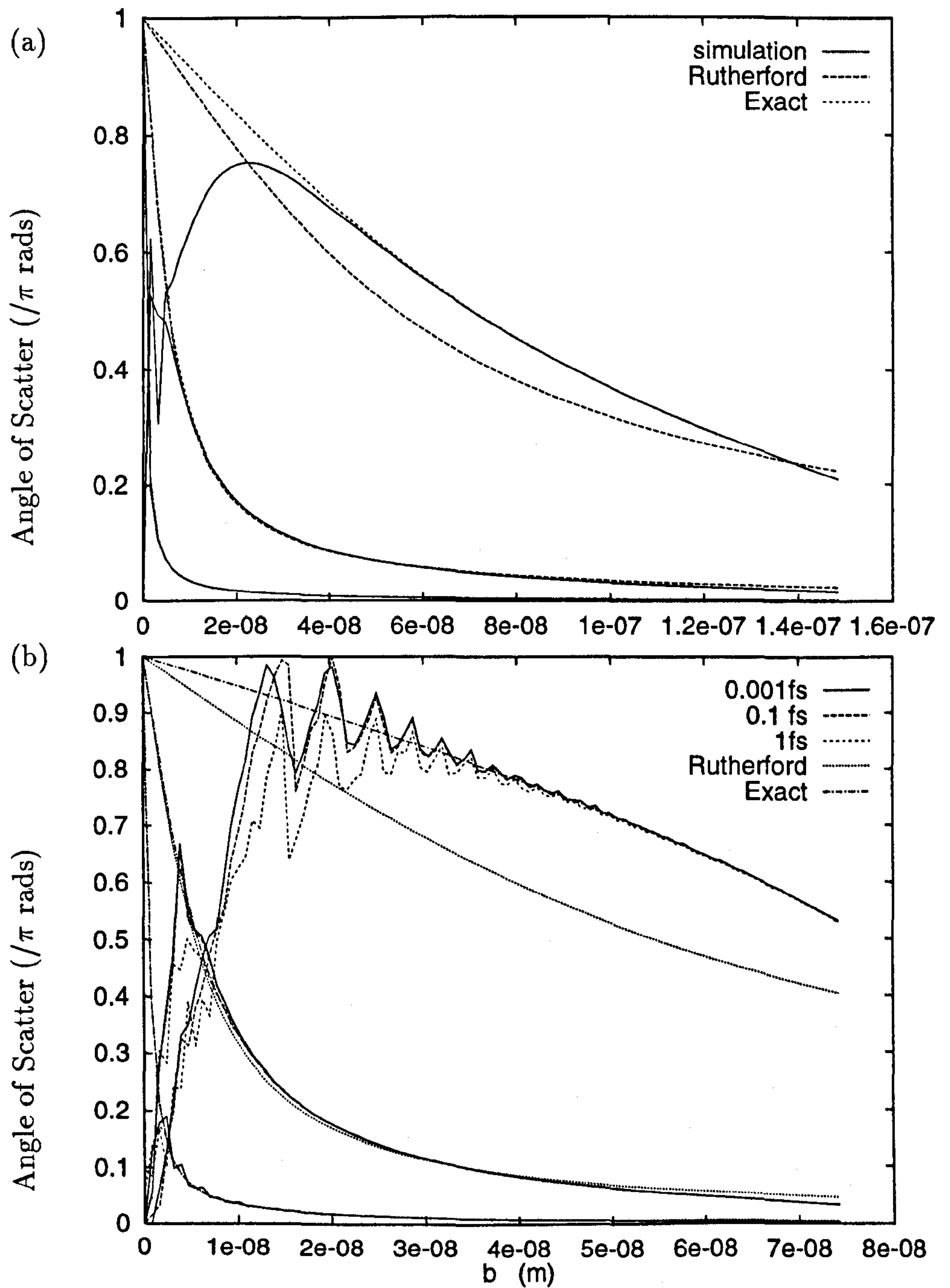


Fig. 4.4: The angle of deflection of a particle, as a function of impact parameter, using Euler integration in an attractive field using (a) an analytic estimate, and (b) a 2 nm mesh. The results are for total energies of 1 meV (top), 10 meV (middle) and 100 meV (bottom).

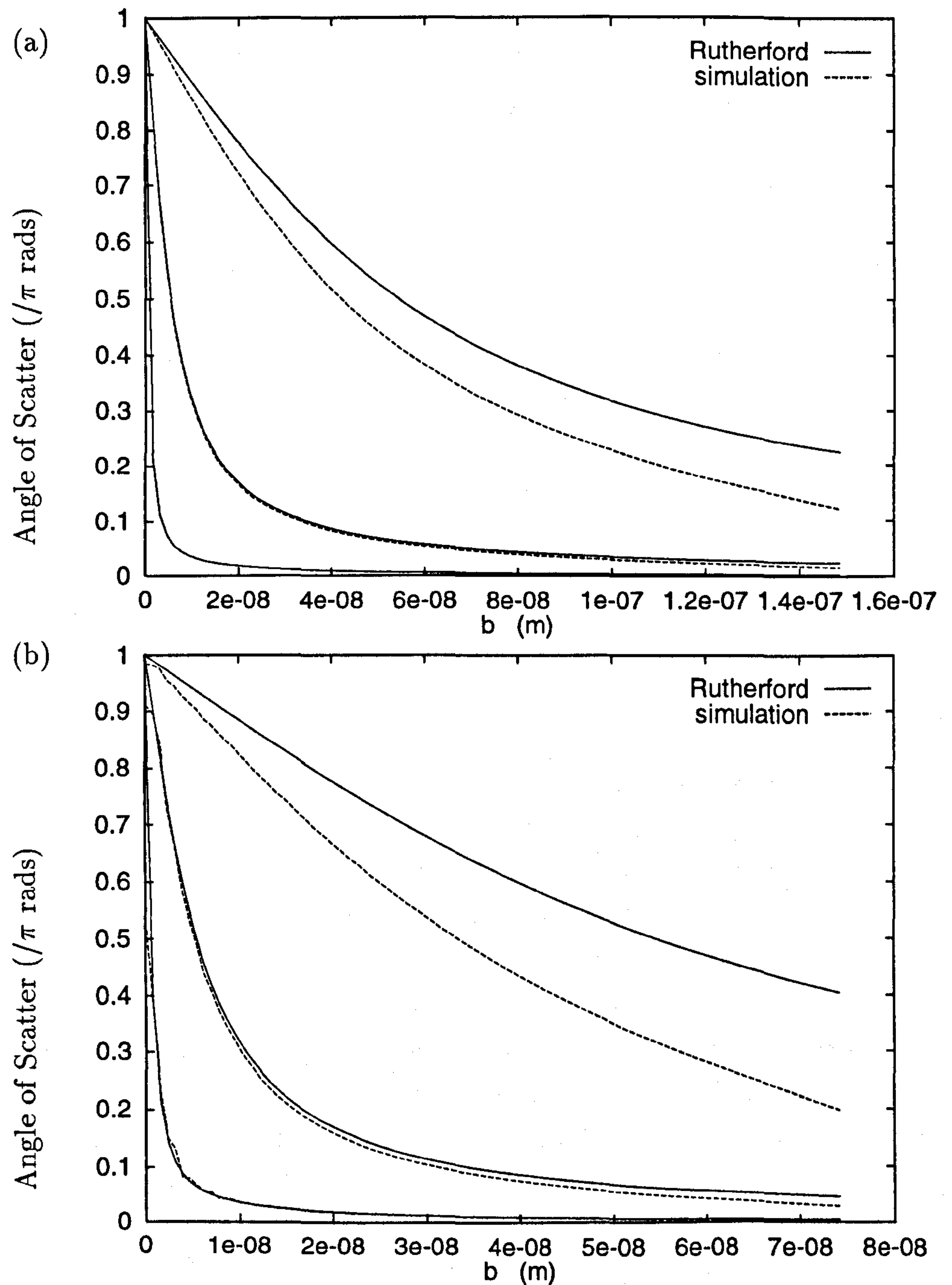


Fig. 4.5: The angle of deflection of a particle, as a function of impact parameter, using Euler integration in an attractive field using (a) an analytic estimate, and (b) a 2 nm mesh. The results are for total energies of 1 meV (top), 10 meV (middle) and 100 meV (bottom).

potential when a mesh is used results from the electron approaching to within a few mesh spacing. These oscillations are also present (at the same impact parameters) at very small time-steps and so are not due to to errors in the integration of the equations of motion over time. The gradient of a linear interpolation function has been used and so this results in discontinuities in the electric field between adjacent pairs of mesh-points. The magnitude of these discontinuities increase as you get closer to an impurity and the lower the kinetic energy the closer the approach for a given impact parameter and so these effects are most marked for low energy electrons.

4.3 The unscreened impurity scattering

The main purpose of this work is to study the dynamics of carriers through the complex potential distribution due to the discrete charges. The major computational burden is in the determination of the forces acting on a carrier and the integration of the resultant force through time. The infinite crystal can be modelled by a cell with periodic boundary conditions. Solving Poisson's equation can prove to be an efficient way of obtaining the force on each carrier, but the components of the force with shorter wavelengths than twice the mesh spacing cannot be represented. However it is these components that result in scatterings with large angular deflections and so play a significant role in ionised impurity scattering. Chapter 3 dealt with the task of calculating forces using a mesh and a possible method of calculating the short range force. The method of Ewald summation [125] proves to be a useful alternative for studying the electrostatics of this problem. It does not use a mesh and so does not cut off the strong high frequency components of the inter-particle force.

Carriers don't just feel the electrostatic forces from the nearest impurity ion, as envisaged in the standard models that treat these scattering processes as a two-body problem. The carriers are continuously experiencing the electrostatic forces due to all the ions and so the simple descriptions involving free flights that separate discrete scattering events may not valid. Therefore, this work also includes the contributions from the other impurities by following the paths of individual carriers through the complex potential distributions due to all the impurities in bulk samples. This work will investigate the rela-

tionship between mobility and the impurity concentration for non-degenerate systems and the results will be compared to those of the Conwell-Weisskopf, Brooks-Herring and third-body exclusion models.

The trajectories of the carriers will be studied using the Brownian dynamics model and also by integrating the free flights using an adaptive Runge-Kutta method [96]. The former model includes the phonon scattering and would simulate the measurements that can made from actual substrates. Matthiessen's rule can be used to extract the ionised impurity scattering rate. The validity of Matthiessen's rule can be checked by studying the free dynamics and since the phonon scattering and impurity scattering rates are independent in this model, then there should be good agreement between the two results.

The first step is to get the initial distribution of carriers in the test sample and this will now be discussed.

4.3.1 The initial distribution

This work will focus on non-degenerate semiconductors and so it is adequate to describe the carrier distribution using the Maxwell-Boltzmann distribution (that results from the use of Brownian dynamics). The probability distribution function can be written as

$$f(\mathbf{v},\mathbf{r}) = A\exp\left(-\frac{\frac{1}{2}m|\mathbf{v}|^2 + V(\mathbf{r})}{k_{\mathrm{B}}T}\right) \tag{4.39}$$

where A is the normalisation constant. Equation 4.39 seems to provide simple technique for the initial state since $f(\mathbf{v},\mathbf{r})$ is a separable function in $\mathbf{v}$ and $\mathbf{r}$. The particles can be placed within the sample using an exponential weighting, in accordance to the potential distribution, and their velocities can be chosen at random from a Maxwell-Boltzmann distribution. The problem with this approach is that the spatial component cannot be normalised since

$$\begin{aligned}\int V(\mathbf{r})f(V(\mathbf{r}))d^3\mathbf{r} &= \frac{Ze^2}{\epsilon}\int_0^\infty -r\exp\left(\frac{Ze^2}{4\pi\epsilon\, k_{\mathrm{B}}T\, r}\right)dr \\ &= 2\pi\, k_{\mathrm{B}}T\left(\frac{Ze^2}{4\pi\epsilon\, k_{\mathrm{B}}T}\right)^3\left\{\left[\frac{e^{-u}}{u^2}\right]_\infty^0 - \left[\frac{e^{-u}}{u}\right]_\infty^0 + \int_\infty^0 \frac{e^{-u}}{u}du\right\}\end{aligned} \tag{4.40}$$

after making a change of variable to the dimensionless quantity,

$$u = \frac{Ze^2}{4\pi\epsilon\, k_{\mathrm{B}}T\, r}.$$

This cannot be evaluated at $u = 0$ and so after considering a small $\delta \to 0$ the terms in the curly brackets approximate

$$\frac{e^{-\delta}}{\delta^2} - \mathrm{E}_1(\delta) \longrightarrow \infty \text{ as } \delta \to 0$$

The solution to this quandary lies in the fact that the vast majority on the electrons are trapped within the potential well and only the finite number of free electrons in the conduction band are of any interest. Furthermore the conduction band energy is the average value throughout space and this can be set as the zero level in the sample.

Unfortunately, excluding the trapped carriers is not a simple matter since the potential spikes due to the individual donors in a $1\,nm$ mesh are very much bigger than $k_{\mathrm{B}}T$ and so there would be a vast number of rejections with this simple procedure. In order to cut down on the computational effort, no attempts were made to place the electrons on points where the potential exceeded some threshold value - which would result in an excessive number of rejections. This threshold has to be as low as possible to obtain a closer approximation to the Maxwell-Boltzmann distribution but not unduly increasing the cost of the initialisation procedure.

The use of the Brownian dynamics model to follow the particle's trajectory has the advantage that it ensures that the carriers are in thermal equilibrium with the crystal lattice. Therefore it is sufficient to place the carriers at random in the sample but chose the velocity with the appropriate thermal distribution. The correct spatial distribution will be attained through the balance between the drift due to the potential fluctuations and diffusion.

4.3.2 The numerical results

The first point that will be addressed is whether the overall mobility of a carrier is affected by the precise distribution on impurities in a bulk sample. In a physical sample the random placement of the ionised impurity charges results in local density fluctuations that result in varying scattering rates throughout the medium. The effect of this increases as the sizes are scaled down. The impurity distributions can have a more subtle influence on the carrier dynamics by creating channels in the equipotential profile that facilitate travel along

certain orientations. The figure 4.10 shows these patterns for some typical distributions in a $(50\,nm)^3$ sample doped to $10^{24}\,\mathrm{m}^{-3}$. (Note that the total number of donor impurities have been kept constant as it is the particular distributions that are of interest at the moment.) The significance of these channels diminish rapidly with increasing energy.

Ideally the objective is to simulate the infinite crystal, but practical limitations forces the use of a small sample. The momentum relaxation time indicates the the time scale over which the particle 'forgets' its initial state (as demonstrated from the velocity auto-correlation function) and the mean free path gives the expected distance travelled over this time. For doping levels of $10^{24}\,\mathrm{m}^{-3}$ and electron mobility $\mu_e \approx 0.4\mathrm{m}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ the mean free path is given by

$$L_D = v_{th}\tau_p \approx 40n\mathrm{m}.$$

The number of charges scales as the cube of the length of the simulation domain and so a cubic sample of $(80n\mathrm{m})^3$ was chosen.

An estimate of the mobility is calculated from the diffusion coefficient resulting from the trajectories of the carriers. Let the displacement of a particle (i) from its initial position, at time t be $\mathbf{r}_i(t) = \mathbf{x}_i(t) - \mathbf{x}_i(0)$, then the spatial diffusion coefficient is

$$D_x = \frac{1}{2}\frac{\partial}{\partial t}\left(\langle r(t)^2\rangle - \langle r(t)\rangle^2\right). \tag{4.41}$$

Therefore, a good method of calculating the diffusion coefficient is to perform a least squares fit to the variance in this displacement as it varies with with time and to extract it from the gradient. The Einstein relation can be used to obtain the mobility

$$\mu = \frac{e}{k_{\mathrm{B}}T}D_x.$$

The results from the simulations following the dynamics of free electrons, using a $1\,nm$ mesh-spacing, all agreed very well (giving $0.417 \pm 0.008\,\mathrm{m}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ for the mobility) indicating that the size of the sample is not really an issue. What this indicates is that the overall mobility is not dependent on the precise distribution of impurities in the sample. These results were as expected since these are time averaged results (over tens of picoseconds) and most particles

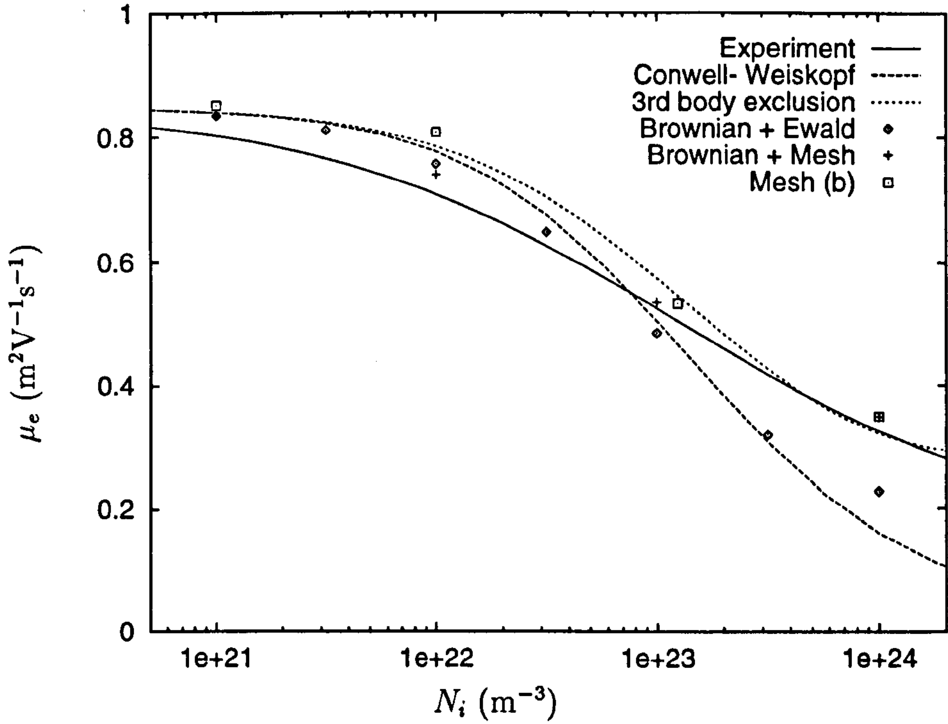


Fig. 4.6: The relationship between the mobility of non-interacting carriers in a bulk sample, with 125 randomly placed donors in each periodic cell, and the doping level. The results for a 50^3 mesh and the Ewald sum, using Brownian dynamics with $\mu_e = 0.85m^2V^{-1}s^{-1}$ are given. Mesh (b) refers to a simulation that was carried out for 200 ps on a 64^3 mesh.

would have visited much of the sample and so all the results would be similar. Samples ranging from $(10\,nm)^3$ to $(200nm)^3$ gave very similar results. The $(10\,nm)^3$ sample represented an ordered periodic square lattice and the larger the sample the more disordered the lattice and so this result reinforces the fact that the overall mobility of a bulk sample is not affected by any microscopic fluctuations in the density of ionised impurities.

The next step is to verify the dependence of mobility on the ionised impurity concentration. In order to save on computational effort the ionised impurities are placed on randomly chosen points of a mesh and Poisson's equation is solved. During the simulation the electric field is computed by taking the gradient of a first-order interpolation of the potential. 125 donors were scattered

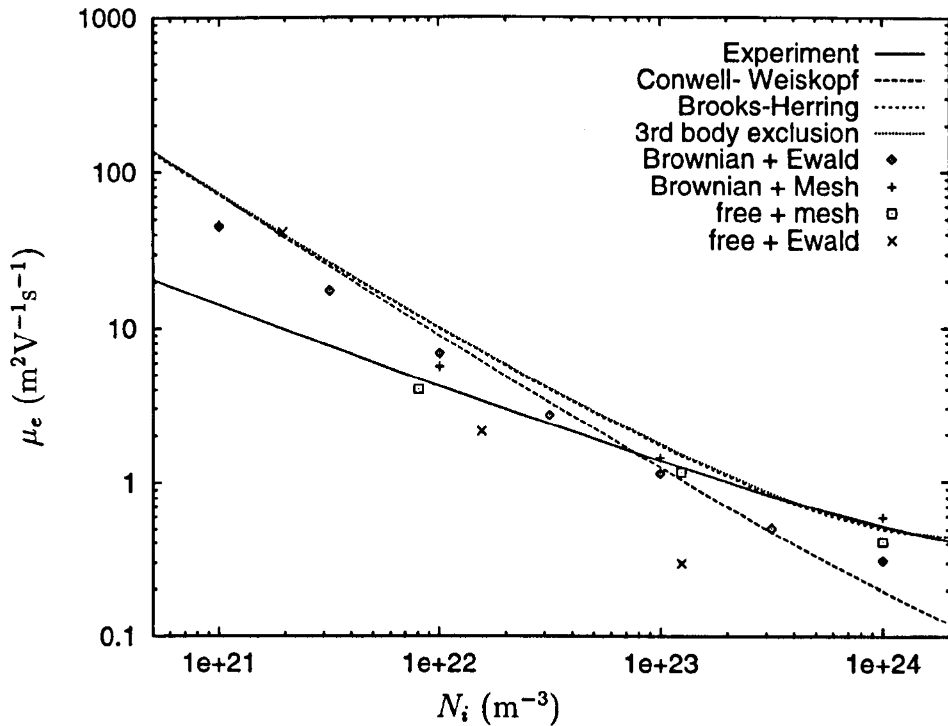


Fig. 4.7: The mobility of non-interacting carriers in a bulk sample (due to ionised impurity scattering only), with 125 randomly placed donors in each periodic cell, and the doping level. Mattheissen's rule has been used to remove phonon scattering from the Brownian results and experimental data.

at random in a 50^3 mesh and the doping level was determined by the mesh spacing. Brownian dynamics were used to follow the trajectories of 10,000 non-interacting electrons in these meshes. (Non-interacting means that the simulation results were the ensemble average of 10,000 one-electron systems.) The results, shown in figure 4.6, are in good agreement with experimental results for doping levels in the range 10^{21}–10^{24} m^{-3}.

The sizes of the potential spikes at the impurity sites are dependent on the mesh size. The diameter of the ions is of the order of the mesh-spacing in these simulations and so care must be taken to ensure that these results are not merely due to the decrease in the size of the potential spikes. Therefore, the simulation was repeated using Ewald summation (using the 8th-order polynomial approximation [126]) to obtain the forces on the electrons. This provides

Mesh spacing	Mobility ($m^2V^{-1}s^{-1}$)
0.1 *nm*	0.337 ± 0.004
1.0 *nm*	0.417 ± 0.008
Ewald sum	0.321 ± 0.001

Tab. 4.1: The effect of mesh spacing on ionised impurity scattering for a doping of $10^{24}\,m^{-3}$

a very accurate estimate of the short-range forces (between the electron and nearby donors) but is much slower and slightly less accurate for the long-range components. The error is under 2 % for much of the sample. The results are also shown for comparison in figure 4.6. These results agreed with the mesh at low concentrations there the probability of a small impact parameter was very small and small angle deflections accounted for most most scattering. The mesh results gave an overestimate for the mobility at high doping levels. This was most likely due to the band-limiting in the mesh. Table 4.1 shows that reducing the mesh size results in a better agreement between the Ewald sum and the mesh and so confirming effects due to artifacts introduced by the mesh.

The ionised impurity scattering rate has been extracted from the Brownian results and plotted in figure 4.7 for comparison with the standard models. The use of Mattheissen's rule to extract the extract the ionised-impurity scattering rate, from the overall mobility, may seem questionable since relaxation time for this process is not a constant but varies with energy. It must be ascertained whether the two scattering processes (phonon and impurity scattering) are truly independent and whether the inclusion of frictional and stochastic terms in the Brownian model has any significant affect on the scattering by ionised impurities. The answer to these questions lies in the study of free carriers in the same potential distribution, which are also included in figure 4.7.

The study of free carriers proved to be problematic, despite the use of adaptive integration methods. Checks were made to ensure that the total energy of each particle remained constant in order to verify the integration scheme. Now that there is no longer any mechanism that drives the system towards thermal equilibrium, any small errors in the initial distribution can have a significant influence on the results. The results from the Ewald summation

were increasingly underestimated for higher impurity concentrations and this is probably due to trapping of carriers in the deep potential wells at the sites of the donors.

It was expected that mobilities should best approximate the Conwell-Weisskopf model and there should be good agreement at low doping levels but at higher doping the Conwell-Weisskopf model should give too low a mobility. This is because the impact parameter can get large before the contribution to the electric field from the next nearest impurity becomes significant in lightly doped media and the scattering angle diminishes rapidly with impact parameter. In heavily doped materials the electric field experienced by the electron falls more rapidly than envisaged in the Conwell-Weisskopf model, due to the influence of the neighbouring ions, and so the actual scattering rate is lowered and the magnitude of this discrepancy is related to the minimum scattering angle of this model. The simulations seem to support this hypothesis. At these doping levels screening dominates in neutral semiconductors and so the Brooks-Herring model would have been used in any case.

4.3.3 Screened carrier-impurity scattering

The central premise of this work has been to take a particulate view of both carriers and dopants. Most simplified scattering models represent the screening charges by continuous charge distributions (from the solutions to Poisson's equation in the case of the models described above). Screening from discrete mobile charges is a much more complex process that has to be modelled by following all these charges, paying close attention to all their interactions. All these charges have to be independent and so the periodic images from the finite simulation domain do pose problems. The size of the cell has to be chosen so that the electrostatic force from the periodic image of any charge is negligible. For a cubic cell of length L, the electric field scales as L^{-2} while the number of charges varies as L^3 and so the number of interactions increases as L^6. This fact leads to a very large computational burden. As a reasonable compromise, the results presented here are for cells containing 512 electrons and 512 impurities. The force due to the next nearest image is $1/64^{\text{th}}$ of the nearest charge.

A simulation was carried out using Brownian Dynamics and the mobility

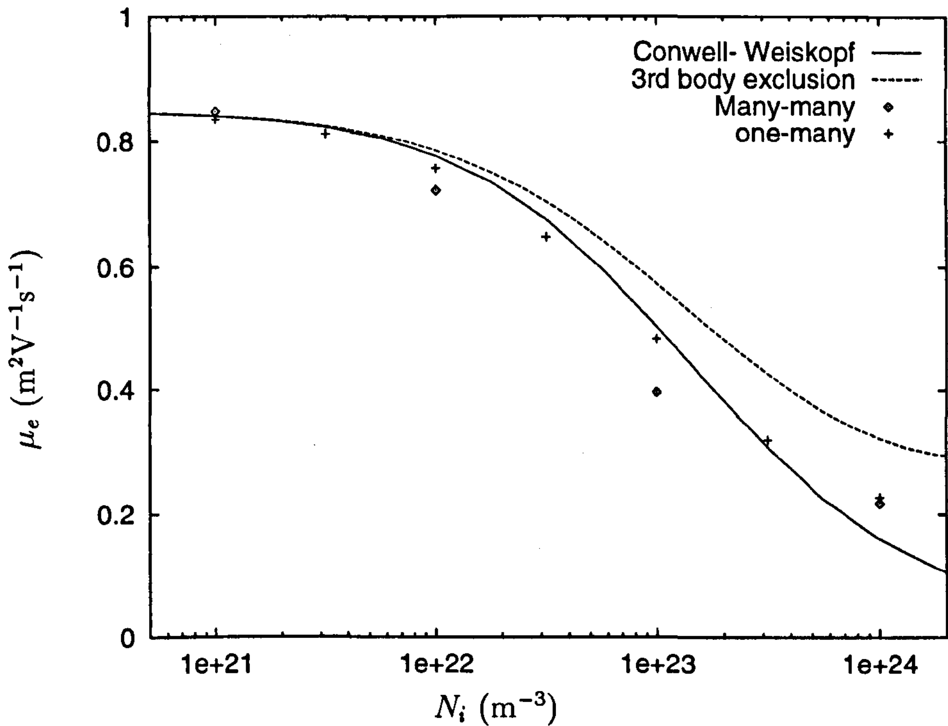


Fig. 4.8: The Molecular Dynamics simulation of electron dynamics in GaAs at 300 K using Brownian dynamics and Ewald summation. The periodic cell contains 512 donors and 512 electrons.

was obtained from the diffusion constant extracted by using a least squares fit on the variance of the particle distribution in space. The results shown in graph 4.8 broadly agree with the unscreened results but do show a much lower than expected set of mobilities. One probable cause of this was the small size of the periodic cell, but unfortunately there was insufficient time to perform a larger simulation. This experiment does validate the methods used and given enough time and computational resources a full set of results can be obtained.

4.4 The threshold equipotentials in samples with discrete impurities

Now that the treatment of ionised impurity scattering has been validated, the next step is to study the ways that these discrete impurities influence

the percolation of carriers in very small samples at threshold. This is a very important topic as far as the study of threshold characteristics is concerned since the movement of carriers across the gate region, as the channels open, can be viewed as a percolation problem. The distribution of the particulate charges make different regions of the channel turn on at different potentials. Therefore little islands of conducting regions form in the channel, according to the local charge distribution, as threshold is approached. These islands increase in size and begin to merge. Complete paths begin to form — that traverse the gate barrier — as the percolation threshold is reached. The channel should now begin to conduct. Percolation describes the device just as it turns on and its effects should show up in the subthreshold current as turn-on is approached.

Simple geometries were considered so as to exclude other extraneous effects resulting from the complex fields that can occur in a real FET. No potential gradients due to external biasing have been included and cubic samples were used — so as to model three instances of ion distributions by exploiting the symmetry of the problem. (The source and the drain could be between any of the three pairs of opposing faces.) The precise distribution of ionised impurities in a small finite sample would affect the potential distribution in the medium so that the percolation threshold varies from sample to sample. This effect was studied for a $(50\,nm)^3$ sample containing 125 donors (corresponding to a doping of $10^{24}\,\mathrm{m}^{-3}$.

The first problem is to find the percolation threshold for free electrons. The total energy of the electrons is constant during their travels, and so they

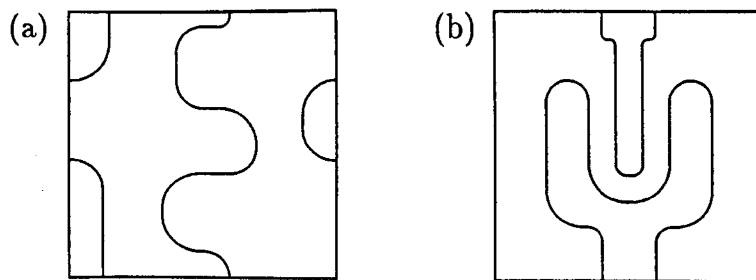


Fig. 4.9: A 2D example of a sample with periodic boundaries. (a) shows allowable free paths and (b) demonstrates a closed loop that must not be classed as open.

follow their own geodesic, on their equipotential surface, through the periodic boundaries. For the electron to be classed as free then this geodesic (that lies on the energy surface that corresponds to the electron's energy) must connect with itself through the periodic boundary in such a way as to form an open loop. There is a simple criterion for a cubic sample which is that the path must connect two faces on the opposite sides. It is permissible to make use of the periodic boundaries when defining the path, with the two conditions that any planar slice (that is parallel to the faces) through the sample must contain a segment of this path (as shown in figure 4.9) and the path must cross it an odd number of times. Obviously the percolation threshold must be greater or equal to the lowest potential energy that is found on one of these faces and on a plane cut half way between the two faces. (Note that the periodic boundary conditions give a very potential distribution on these opposing faces and so any two planes can only be separated by half the length of the sample at the most.)

Tab. 4.2: The threshold potentials in a $(50\,nm)^3$ block. The values for each pairs of faces, labelled 1,2 & 3, are given along with the average of the three. The fraction of points in the conducting state are also shown.

Threshold potential				Fraction conducting			
1	2	3	mean	1	2	3	mean
-0.02224	-0.02977	-0.00958	-0.02053	0.1603	0.1056	0.2972	0.1752
-0.02992	-0.01754	-0.02494	-0.02413	0.1576	0.2537	0.1921	0.1982
-0.04246	-0.03463	-0.02293	-0.03334	0.1000	0.1423	0.2321	0.1514
-0.00291	-0.05722	-0.01280	-0.02431	0.3849	0.0607	0.2763	0.1840

Table 4.2 shows the threshold values for the four samples used in the illustration in figure 4.10. These were four randomly selected cases from the thirty used in the study. The mean potential in the sample was taken as the zero and it turns out that this is the threshold value for a periodic rectangular lattice. It is apparent that (for samples of this size) the threshold potential is lowered by the same order of magnitude as the electron's thermal energy at room temperature. This should give a noticeable fluctuation in the threshold level in a real device of this size. The percolation threshold is reached when the fraction of abscissae in the conducting state reaches 0.19 ± 0.03. The site percolation threshold for a cubic lattice is 0.31 and the simulation results are

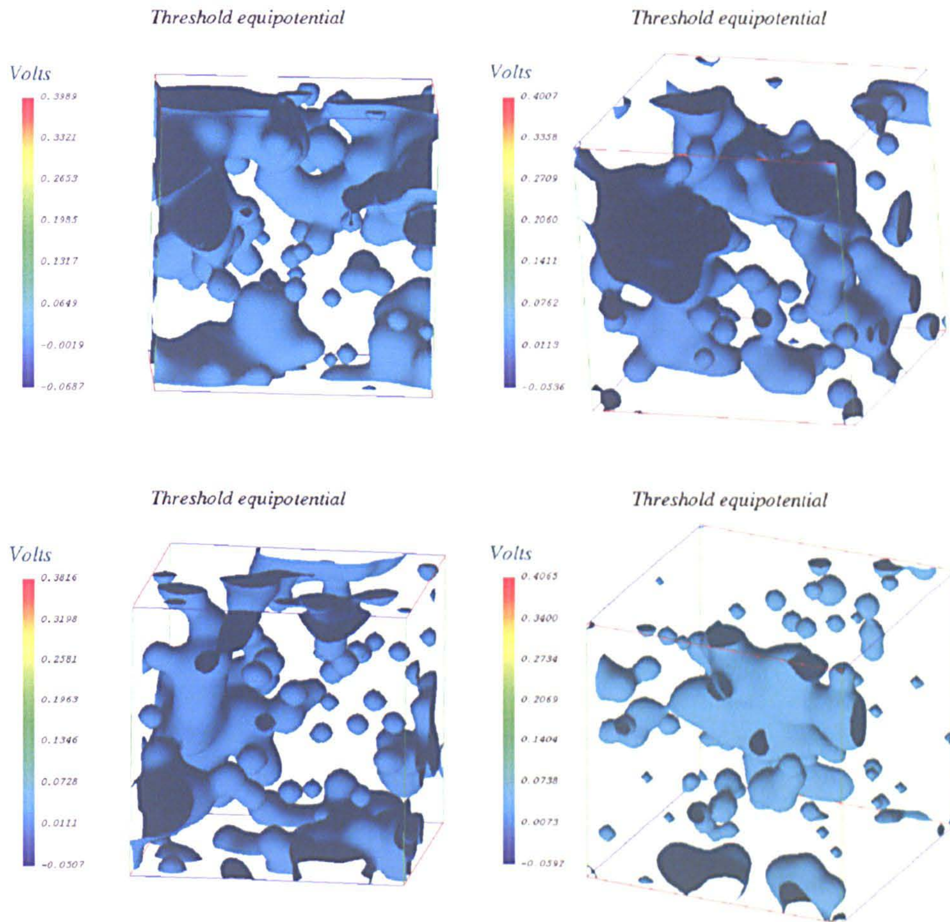


Fig. 4.10: The average threshold potentials in $(50\,nm)^3$ samples containing randomly placed 125 donors.

much lower than this value. This percolation threshold value assumes that the states of the points are independent of each other but this is not true for the potential map because of the longer range fluctuations induced by each ion. An ion sitting at a site lowers the threshold potential in neighbouring cells and make them more likely to conduct and so a lower fraction of sites need to be in the conducting state for a conductive path to form.

4.5 summary

This section has demonstrated that the Brownian method adequately describes the electrostatic interactions between discrete charges. It has the great advantage of driving the system towards equilibrium and so dampens down the errors in the initial distribution and those that are accrued during the integration of motion. The standard methods of differencing and interpolation, used in conventional Monte Carlo simulations seem to be adequate enough to give a reasonable approximation to ionised impurity scattering. The methods that were developed in chapter 3 should do much better since higher order interpolation and differencing schemes have been used that produce a much smoother and continuous force. The new method for estimating the mesh force between closely separated particles can be used to calculate and thus correct the erroneous force derived from the mesh.

The numerical experiments have shown that the precise distribution of impurity ions do not have a significant affect on the carrier mobility but the discreteness of ions and carriers do have an appreciable affect the potential distribution in a small sample. This indicates that the discrete impurities exert their influence mainly through their influence on the potential distribution in the channel by producing different patterns of conductive channels near threshold.

The next chapter deals with the main obstacle to carrying out numerical studies of actual devices. This is the large computational burden of these three dimensional molecular dynamics simulations.

5. ATOMISTIC DEVICE SIMULATION

So far the, the main components of device simulation have been discussed. An efficient solution for the description of carrier dynamics incorporating lattice scattering and electrostatic scattering have been presented in chapter 2. the solution of Poisson's equation was explored in chapter 3 and it is clear that this is the major computational burden that must be borne.

The main obstacle to the molecular dynamics simulation of small semiconductor devices is due to the limitations in computational resources: a great number of operations need to be carried out on very large sets of data. This necessitates the use of "supercomputers" to model realistic three-dimensional structures. The computer architecture has a great bearing on the implementation of the simulation and so this will be the starting point of the discussion. This will be followed by a discussion of the practical points of device simulation using the architectures available in the department.

5.1 The types of supercomputers

There are three broad categories that describe typical computer systems, as first described by Flynn. The first is the single instruction stream, single data stream (SISD) where the cpu can issue only one instruction at any one time on a given data input. This simple architecture was used in most desktop computers until very recently. The second is the single instruction stream, multiple data stream (SIMD) that can apply their instructions to many data values at a given point in time. Older systems had many arithmetic logic units that can perform the same instruction on many data elements at once such as the DAP or Thinking Machines' Connection Machine. The third type is the multiple instruction stream, multiple data stream (MIMD) architecture that could execute many independent instructions simultaneously on many [different] data inputs. The simplest way to implement this is through the use

of multiple processors and these parallel architectures will be discussed later.

The earliest machines relied on vector processors. These machines were optimised to handle data arranged in arrays and streamed their elements through a pipeline architecture to obtain their impressive performances. The same operation (or sets of operations) was carried out on each element of the input streams. They worked most efficiently on contiguous sets of data. Anything that disrupts the smooth operation of the pipeline, such as conditional statements or branches, could severely affect the overall performance.

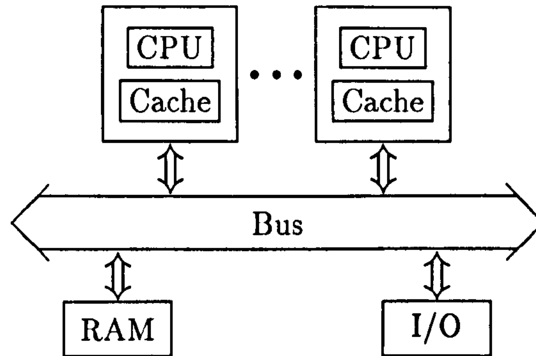
The greatest limitation on the speed of current machines is the speed at which data can be fetched and stored from the CPU. The limiting factor is the bandwidth of the bus and access times of the memory. Current thinking favours the use to MIMD architectures to divide the problem among many computing nodes and this will be discussed in the following section.

The performance of CPUs has risen exponentially over the past decades but so has the scale of the computational problems. Fortunately most applications involve the application of independent instructions on large arrays of data and so great benefits can be gained through the use of parallel computing.

Most workstations use symmetric multi-processing (SMP) where there are more than one CPUs that share all the system resources using the same bus. The operating system shares the computational load through all the processors and so this system is transparent to the end user (*i.e.* the code does not have to be especially written for these systems). Unfortunately the bus bottleneck limits the number of processors to around eight and performance is severely degraded if large amounts of data need to be transferred across the bus. (The latter can be ameliorated through the use of caches.)

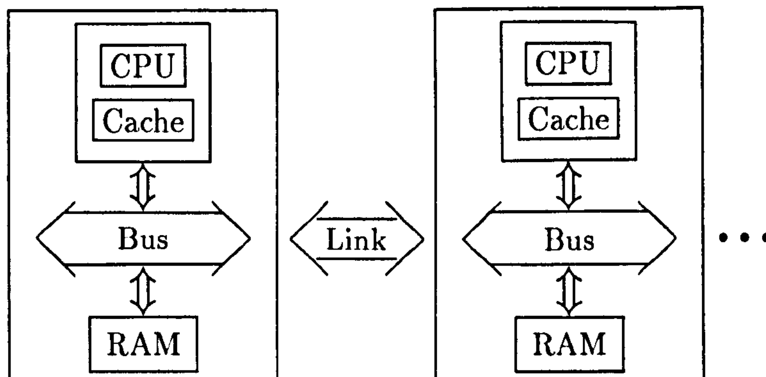
In order to get large numbers of processors, the system resources can be divided onto many independent computational nodes (as in the massively parallel processing (MPP) architecture). Systems can be built with thousands of nodes, each having its own memory and communications hardware. The use RISC processors can offer good performance while keeping the cost of the whole system reasonable. The main problems in MPP systems are due to data synchronisation and communications bottlenecks when tasks are spread over the whole network. The problems occur when data is required from other nodes for the computation to proceed and since all nodes must reach the same

Fig. 5.1: The SMP architecture.



stage in the computation the overall performance of the system is limited by the slowest node. There is a fine art in balancing the load on each node. Code has to be specially written or modified in order to take full advantage of these parallel systems. Some compilers (such as High Performance Fortran (HPF)) offer some help in this undertaking by handling the distribution, of the nominated arrays, and synchronisation between the nodes. Of course, each node can have have a SPM architecture and the fastest machine at this moment is the 4536 node TFLOPS system [127] comprising of 9072 Pentium Pro CPUs and 283 Gbytes of RAM.

Fig. 5.2: The MPP architecture.



It would be advantageous to have the simplicity of the SMP systems and

the scalability of the MPP architecture and so enter the scalable parallel processing (SPP) machines such as the Convex Exemplar. This uses a two-tier memory architecture with each node having a SMP system with multiple processors sharing the same memory, as before, but the second tier makes the memory that is distributed on all the nodes appear as a single global shared memory space. The data flow between the nodes is kept to a minimum by only exchanging updates to keep the memory coherent among the nodes.

The Parsytec systems used in the Electronics and Electrical Engineering Department at Glasgow University are of the MPP architecture. The 64-node Supercluster and 32-node Multicluster networks have configurable links but they can only be initialised as a two-dimensional rectangular network of arbitrary size. However they can support a wide range of virtual topologies, such as the tree network. Each node comprises of a T805 transputer, 4 megabytes of memory and four serial links to the four nearest neighbours. More distant connections are routed through intermediate nodes. All communications to the outside world is carried out by a root transputer that sits separately from the allocated network of nodes.

The newer X'plorer systems have fixed links and so have a much more restricted set of topologies available. There is no special root node in these systems. Each node comprises of an 80 MHz PowerPC 601 with 8 megabytes of RAM and a T805 transputer to manage the communications interface through the four serial links. Each node is much faster than the old T805 based nodes.

5.2 The parallel implementation of our models

The Parsytec's MPP architecture requires some modification to the serial code to enable its execution in parallel. The nodes are physically linked in a rectangular two-dimensional network and so it would be most efficient to map the same to the problem of interest.

5.2.1 Partitioning the problem on a network

The first problem is to distribute the problem on to the computing network. This task involves the slicing of large data arrays onto the network so as to give each node a similar computational load. It would also be beneficial if the

distribution scheme minimised the amount of data that needed to be transferred between the nodes during the computation. For a three dimensional rectangular array distributed on a three dimensional rectangular network of nodes the simplest solution is to divide up the data so that each node has a rectangular array of similar size. When a problem of size $n_i \times n_j \times n_k$ is distributed on to a $N_i \times N_j \times N_k$ array of nodes, the number of elements held on each node, with coordinate (I, J, K), is given by

$$n_{\mathrm{elements}} = \zeta_i \zeta_j \zeta_k \tag{5.1}$$

where

$$\zeta_i = n_i \text{ div } N_j + H\left((n_i \bmod N_i) - I\right) \tag{5.2}$$
$$\zeta_j = n_j \text{ div } N_j + H\left((n_j \bmod N_j) - J\right) \tag{5.3}$$
$$\zeta_k = n_k \text{ div } N_k + H\left((n_k \bmod N_k) - K\right) \tag{5.4}$$

and the step function

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} .$$

(Note that the indices start from zero and so

$$\begin{aligned} i &\in \{0, 1, \ldots, n_i - 1\} \\ j &\in \{0, 1, \ldots, n_j - 1\} \\ k &\in \{0, 1, \ldots, n_k - 1\} \end{aligned}$$

and

$$\begin{aligned} I &\in \{0, 1, \ldots, N_i - 1\} \\ J &\in \{0, 1, \ldots, N_j - 1\} \\ K &\in \{0, 1, \ldots, N_k - 1\}. \end{aligned}$$

The Parsytec systems at the Nanoelectronics Research Centre have two-dimensional rectangular topologies the distribution is carried out in two dimensions. The total volume of data tends to be very large and so it would be most efficient to create algorithms that enable each node to generate its

own initial data based on its location. Each node would then deal with the sub-array which maps the indices

$$\begin{aligned} i &\in \{0,\ldots,n_i-1\} \\ j &\in \begin{cases} \{J\zeta_j,\ldots,(J+1)\zeta_j\} & J < M_j \\ \{M_j+J\zeta_j,\ldots,M_j+(J+1)\zeta_j\} & J \geq M_j \end{cases} \\ k &\in \begin{cases} \{K\zeta_k,\ldots,(K+1)\zeta_k\} & \text{if } K < M_k \\ \{M_k+K\zeta_k,\ldots,M_k+(K+1)\zeta_k\} & K \geq M_k \end{cases} \end{aligned}$$

on the main array, assuming that the problem is distributed over the j- and k-indices on an $N_j \times N_k$ network.

$$M_x = n_x \bmod N_x$$

where $x = i, j, k$

In some instances, where the simulation has to be started with a specified initial condition, the data has to be read from an input file and one of the nodes is given the task of reading the file and distributing the data appropriately. This is a simple process where the data is read (a row at a time) and is dealt to each node in the manner of a croupier at the card table. This process proves to be invaluable when long simulations have stopped prematurely due to unforeseen problems.

This simple distribution scheme has some drawbacks — as seen from figure 5.3 where node $(0,0)$ has twice the load of node $(3,4)$. ($\Leftrightarrow$ represent the communications links between the computational nodes represented by the rectangles containing the mesh points denoted by the solid pips.) There have been many solutions to this problem of load balancing and the one favoured in this group is the use of simulated annealing [128, 129] to find the optimal distribution of data that minimises both the variations in the load and the volume of data traffic between the nodes. Figure 5.4 shows a better distribution of this problem. An additional prerequisite is to ensure that the data on each node is as rectangular as possible to keep the program as simple as possible and that data is only redistributed to adjacent nodes. These considerations minimise the amount of information that has to be stored in order to implement this problem and speeds up execution.

Fig. 5.3: The distribution of a 3D array on a 2D network is equivalent to mapping a 2D array of vectors on the network. The partition of a 10×10 array of vectors on a 3×4 mesh.

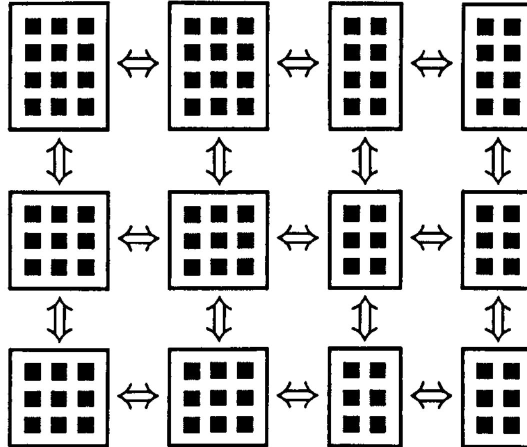
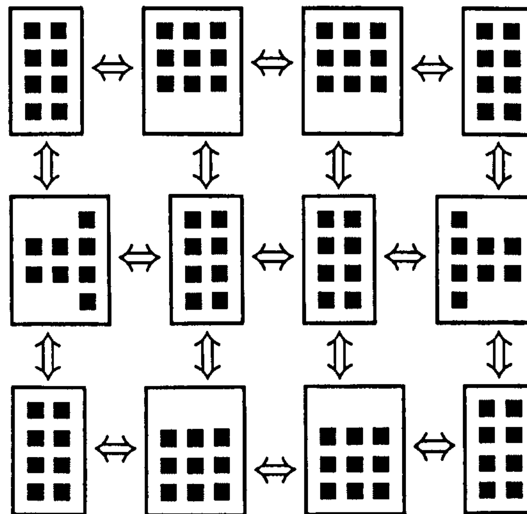


Fig. 5.4: The partition of the 10×10 array of vectors on a 3×4 mesh that achieves a better load distribution.



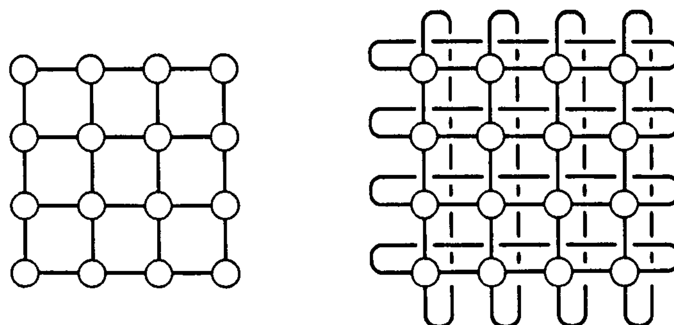
In real devices there is also the added complication due the fact that there is often a pronounced variation in the distribution of carriers. There are a high in the contact regions where the mesh tends to be relatively coarse since there is a small spatial variation in the electric fields and so a more complicated strategy may be required that also weights for this additional computational load. The simulation of small devices often requires the use of fine meshes to

study the trajectories of a small number of particles and so much of the CPU time is spent solving the Poisson equation. This fact makes the distribution of the Poisson mesh the main concern in these simulations.

5.2.2 Topologies and Message passing

Now that the problem has been partitioned onto the rectangular array of computing nodes the present task is to consider the structure of the interconnections between these nodes and the process of data transfer between them. The most useful topologies for implementing the solvers are the rectangular mesh and the torus (as shown in figure 5.5) because they make the best use of the physical interconnects between the Parsytec nodes. The SOR iterations and Gauss-Seidel smoothing steps of the multigrid method can only be implemented in parallel because of the black-white ordering of the updates. (Updating only the alternate squares ensures that only the old value of the data is used during each update.)

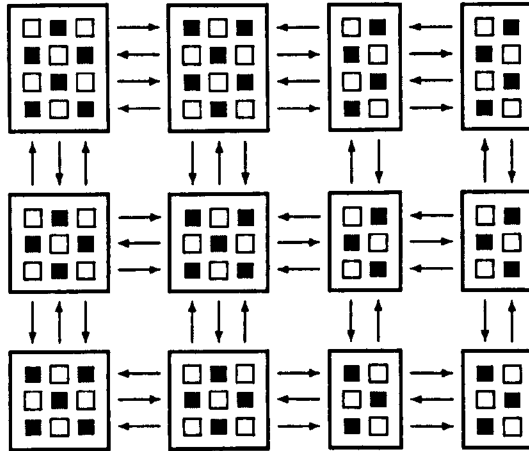
Fig. 5.5: The 2D mesh and torus



In order to ensure the reliability of code, the structure of the serial version was maintained and only the necessary portions were modified to enable concurrent execution. Cross products of column vectors with the square matrices involve the exchange of boundaries from adjacent nodes. The data from diagonally opposing nodes can be exchanged after one set of horizontal and vertical exchange. Only 'white' squares are exchanged between the nodes when the new 'black' values are computed in order to keep down the volume of data traffic (figure 5.6). Since all the data is exchanged between local nodes this op-

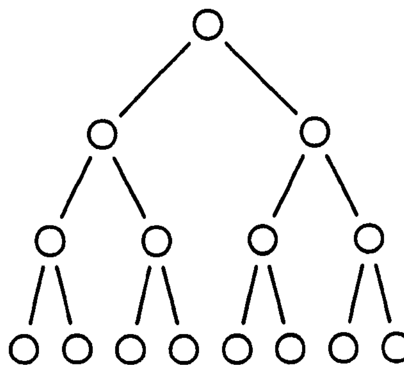
eration is not sensitive to the size of the network but only on the transmission rate of the links.

Fig. 5.6: The data exchanged every partial update for the simple problem.



The bi-conjugate gradient method also uses the dot products between row and column vectors. This inner product can be computed for each node and because these values have to be summed across the entire network, this operation suffers some penalty in speed on very large networks. The most optimum topology to carry out this summation is the tree (figure 5.7). If there are L

Fig. 5.7: The simple binary tree.



links per node then the number of levels (z) in a network of N nodes would be

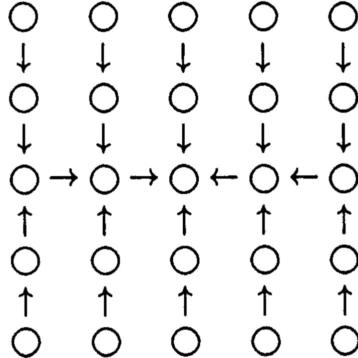
$$z = \text{ceil}\left(\frac{\log(N(L-1)+1)}{\log(L)}\right). \tag{5.5}$$

The function 'ceil' just gives the lowest integer that is not smaller than its argument. The levels can be labelled

$$l \in \{0, 1, \ldots, z - 1\}$$

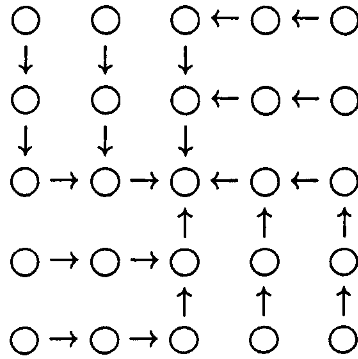
and so at each step the appropriate nodes would read the values from their more distal nodes and, after adding their own contribution would send the result to the proximal node. These steps are repeated until the root node has received the partial sum and this then sends the down through the tree. The total number of data transfers that are required to evaluate this inner product is $2(Z - 1)$. However, although the Parsytec systems can support arbitrary tree topologies, the actual data flow occurs through a rectangular network of links and so this may lead to bottlenecks during the routing of the messages. An efficient method (that has also been used by some authors [130, 131]) is to use data flow in the shape of an 'II' as shown in figure 5.8.

Fig. 5.8: The 'H'-shaped data path that is suited for rectangular topologies..



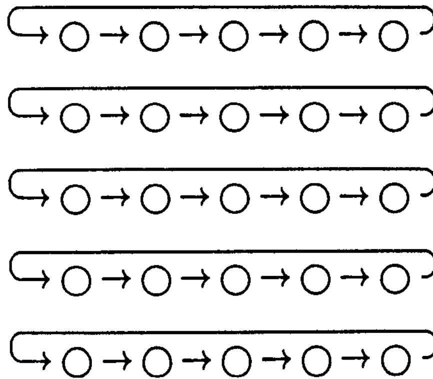
This method allows for the parallel accumulation of the sum while avoiding the bottleneck imposed by the physical topology. The number of steps taken by a $N_i \times N_j$ mesh is $2\,[(N_i \text{ div } 2) + (N_j \text{ div } 2)]$. (This calculation gives the number of steps required to get the intermediate result from the most distant node and then to return the total sum to that node.) The number of data paths to the central node can be increased to 4 using a swastika pattern (figure 5.9). The swastika requires the same number of steps as the 'II'-shaped topology

Fig. 5.9: The fylfot offers more data paths to the centre than an 'II'-shaped topology.



but by providing more data paths it is more useful where large quantities have to be moved.

Fig. 5.10: The wholesale shifting of data across the network.



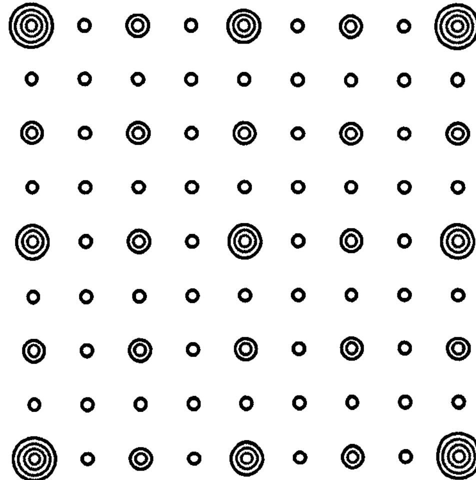
A toriodal topology (see figure 5.10) provides an alternate way of carrying out this sum. Data can be shifted both horizontally and vertically and all the nodes accumulate their own value for the total sum. This method is the most efficient if the nodes can send and receive data simultaneously through all their links. The number of transfer steps is now only $\max(N_i, N_j)$. This pattern is also very useful when large amounts of data need to be updated across the

whole the network.

It must be noted that although the discussion up till now has used a two-dimensional network for clarity, it is easy to extend these ideas to an n-dimensional hypermesh (or hyper-torus).

The multigrid method faces some further problems on MPP architectures. As the defect is restricted onto coarser meshes, careful consideration has to be given to the size of the network so as to achieve the optimal balance between the advantages of parallelism and the costs of communication. The minimum size of mesh on each node will be particular to that particular machine and this number may be different for different mesh sizes of network dimensions. Fast processors and slow communication speeds would make it inefficient to have few mesh points per node. The obvious solution is to restrict the meshes as shown in figure 5.11. Communications between the nodes is now no longer

Fig. 5.11: The multigrid method implemented on a rectangular network. The small circles represent the entire network that is used the finest meshes.



local, but the problem sizes are very much smaller and the volume of data traffic is greatly reduced for coarser meshes on the reduced topologies.

A major concern with the transputer networks is the communications bottleneck. An experiment was carried out to demonstrate the effects of latency on performance. Table 5.1 gives the timings for $30 \times 30 \times 20$ block with von Neumann boundary conditions implemented on various network topologies. It is clear that distributing the problem on many nodes results in a loss of

Tab. 5.1: A 30 × 30 × 20 mesh with von Neumann boundary conditions [with the (0,0,0) corner fixed to 0 V] distributed on a 2D rectangular network over the first two indices.

Network	t (s)	Speed up	Efficiency
1 × 1	131.8573	1.0	1.0
2 × 2	39.43108	3.34399413	0.83599853
3 × 3	18.43339	7.15317693	0.79479744
4 × 4	12.55658	10.501052	0.65631575
5 × 5	8.228488	16.0244871	0.64097948
6 × 6	6.444421	20.46069	0.5683525
6 × 6	6.5306048	20.1906721	0.560852
4 × 9	7.12109983	18.5164235	0.5143451
3 × 12	7.0440892	18.7188572	0.51996826
2 × 18	8.491329	15.5284644	0.43134623
12 × 3	7.9004606	16.6898244	0.46360623
18 × 2	8.78483375	15.0096523	0.41693479

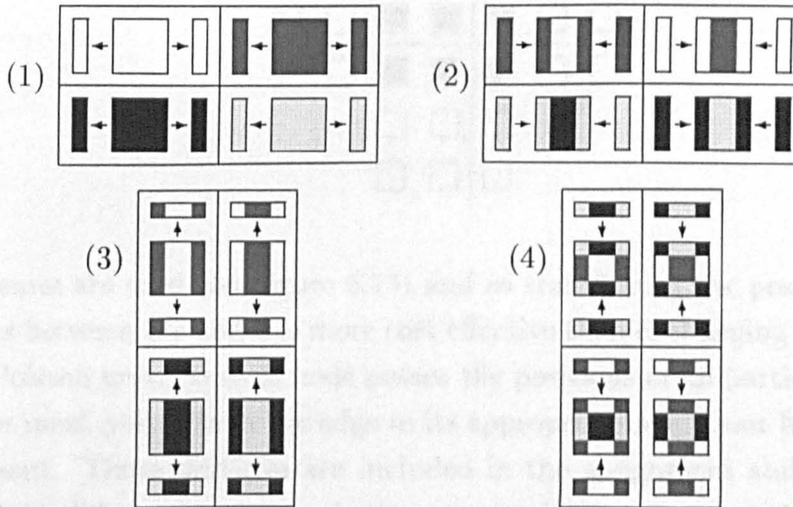
efficiency. The long latency of the interconnects in these transputer systems appears to account for much of this loss of efficiency. The study of the random fluctuations in device characteristics require the characterisation of many devices in order to derive the statistical distribution of these quantities. Therefore the optimal solution for networks with slow communications is to employ each node to simulate an entire device with a given distribution of impurities. The limitations in memory and speed necessitate a simple description of the device and the multigrid method was chosen because of its speed and low memory requirements.

5.2.3 *Partitioning and particles*

A significant problem in load balancing is due to the fact that the main computational burden for most Poisson solvers occur in places where there are large variations in the potential distribution, such as under the gate toward the drain, but often there are few carriers in this region. The bulk of carriers are in the contact regions. So the load imposed by the integration of the equations of motion must be considered when devising a partitioning scheme. This load obviously depends on the number of particles that are held in the

node. Other factors are also important since the scattering rates are higher where there are high electric fields and where the particles have high energies.

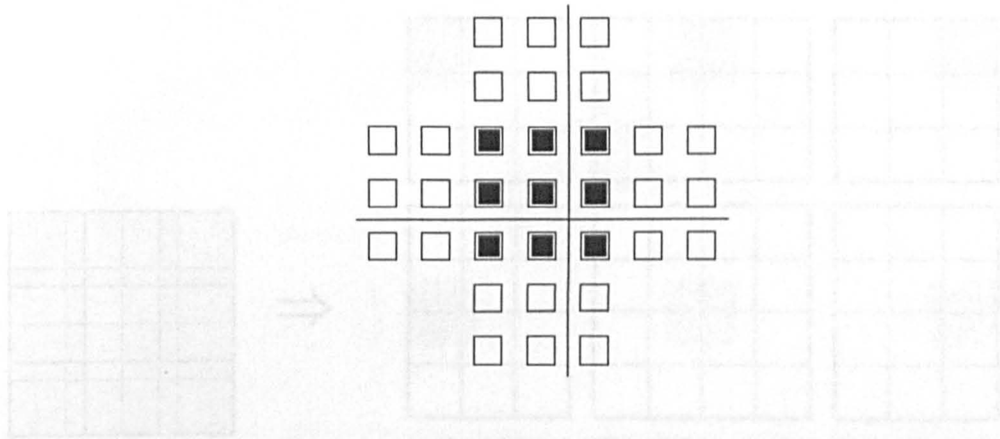
Fig. 5.12: The use of buffers to exchange particles, in four steps, between nodes. The data is represented by four coloured squares. The arrows indicate the direction of copying between the buffers (shown as thin rectangles) and the local data area. Note that the the corners and the sides have been exchanged at the end of the process.



Transferring the particles between the nodes is a simple matter if it is carried out through buffers. This is shown in figure 5.12 that shows four adjacent nodes whose data have been given different colours. For a 2D rectangular array 2 buffers (denoted by the the thin rectangles) can be used and at the end of a time step any carrier having an x-coordinate outside the range for that node is placed in the appropriate bin. The bins are then exchanged along the horizontal direction. This procedure is then repeated using the y-component before exchanging vertically. This method will correctly transfer particles between adjacent nodes, including those that are diagonally adjacent. Care must be taken to ensure that no particle has travelled more than the dimension of any node at each time step.

The spatial device decomposition as described above also requires information about the particles, to be exchanged to enable proper charge assignment and differencing. This problem becomes more acute when high order differenc-

Fig. 5.13: The problem of high order differencing and assignment schemes is that each node may need data that is found deep in the mesh of an adjacent node.



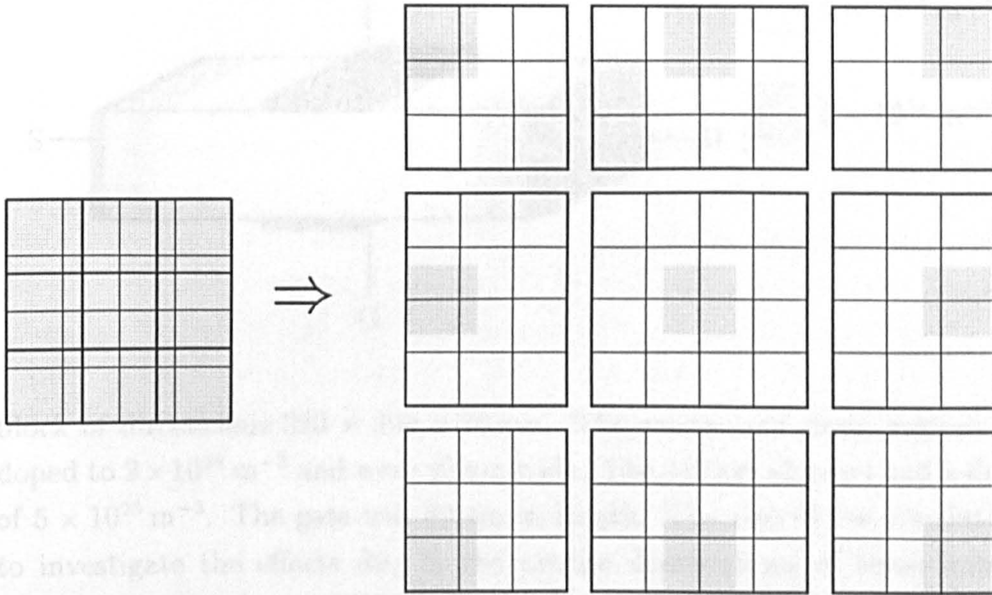
ing schemes are used (see figure 5.13) and so transferring the positions of the particles between the nodes is more cost effective than exchanging the portions of the Poisson grid. So each node passes the positions of all particles that are within n mesh points from the edge to its appropriate neighbour before charge assignment. These particles are included in the assignment and the partial sums of the differencing scheme is then returned when the mesh derived forces are calculated. The 4-point differencing scheme, that is used to obtain the forces, can necessitate the need to get data that can be three layers in, from the surface, on the facing and diagonally opposite nodes.

The simulation is divided into cells which are used to calculate the short range force. The particles in cells on adjacent nodes need to be accessed and a cell may straddle two or more nodes. Here it may be advisable store the contents of all adjacent cells on each node in order to minimise the data traffic between the nodes (as shown in figure 5.14).

5.3 The atomistic simulation of a dual-gate MESFET

The final part of this chapter is concerned with bringing all these various approaches together into a simple FET simulation. A full study of the of the influence of the random placements of atomistic dopants on the IV-characteristics will not be performed due to insufficient time available to carry out an ade-

Fig. 5.14: It is advisable to keep a copy of the adjacent cells on adjacent nodes in order to alleviate communications bottlenecks. A 4×4 array of cells mapped onto a 3×3 network on computing nodes results in the following partition.

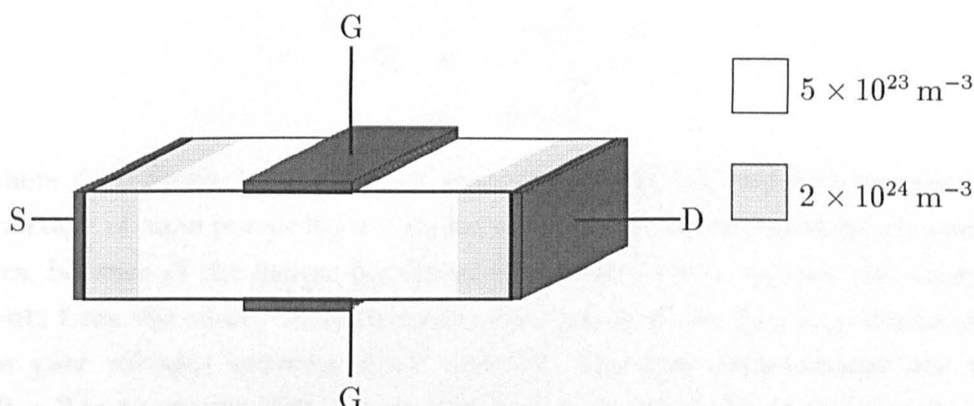


quate number of simulations to obtain statistically acceptable data. A simpler problem is to compare the results for simulations using the standard continuous doping, that is used in conventional device simulations, with the new atomistic treatment that has been developed over the course of this work. The main objective is to verify the simulation methods and so the results will be compared with those produced by a commercial drift-diffusion simulator (MEDICI).

Studies on double-gate Si MOSFETs [12, 132, 133, 134, 135, 136, 137] have found an ideal subthreshold (S) factor, high transconductance, and a lower susceptibility to short channel effects. A double gated Si-MESFET (see figure 5.15) was chosen because it provides a much better control over the channel when using very short gates. There is the added advantage that the active region of this structure occupies a large fraction of the simulation domain.

The software was capable of simulating any rectangular device geometry which had contacts, for the source and drain, on the opposing faces on the xz-planes. The width, length, and depth refers to the spatial measurements along the x, y and z axes respectively. The simulation domain was a rectangular

Fig. 5.15: The double gate MESFET.



block of dimensions $320 \times 320 \times 80\,nm$. The source and drain regions were doped to $2 \times 10^{24}\,\mathrm{m}^{-3}$ and were $40\,nm$ wide. The $240\,nm$ channel had a doping of $5 \times 10^{23}\,\mathrm{m}^{-3}$. The gate was $80\,nm$ in length. The aim of the simulation is to investigate the effects due to the precise distributions of these impurity ions. The results from MEDICI was used as a baseline where all charges were represented as continuous distributions. Brownian dynamics has been shown to be a good approximation to the drift-diffusion model and so it was used to model the dynamics of the particles. Two simulations were carried out: the first where the dopants were represented by a uniform charge distribution, and the second that treated them at discrete entities. The same same mesh was used in both Brownian simulations.

The source contact was grounded and the drain bias was 0.1 V. The electrostatics were studied by solving Poisson's equation using a multigrid solver on a $129 \times 129 \times 33$ mesh with a mesh spacing of $2.5\,nm$. This mesh spacing is much smaller than $17.1\,nm$ — the mean separation between the impurity ions in the channel — and so it is reasonable to assume that any effects due to the atomistic nature of the impurities would be evident. The carriers and ions (in the second simulation) were placed at random in the sample according to the doping profile but their numbers were in accordance with the level of doping in each of the source, channel and drain regions. There were 3072 donors in the channel impurities — with around 1024 under the gate — and 16384 ions

in the source and drain.

The calculated threshold potential for this device was,

$$V_T = -\frac{\left(\frac{L_z}{2}\right)^2 qN_I}{2\epsilon} = -0.61\,\mathrm{V} \tag{5.6}$$

where $L_z = 80\,n\mathrm{m}$ is the distance separating of the top and bottom gates. The shortage of time prevented a detailed analysis of the subthreshold characteristics, because of the longer simulation times required to extract the small currents from the noise. Measurements were made of the I_{DS}/V_{GS}-characteristic for gate voltages between -0.6 V and 0 V. The first measurement was taken after $2\,ps$ to ensure that the system had attained thermal equilibrium. The subsequent current measurements were made after 1 ps after the altering the gate bias to allow for the transients to decay.

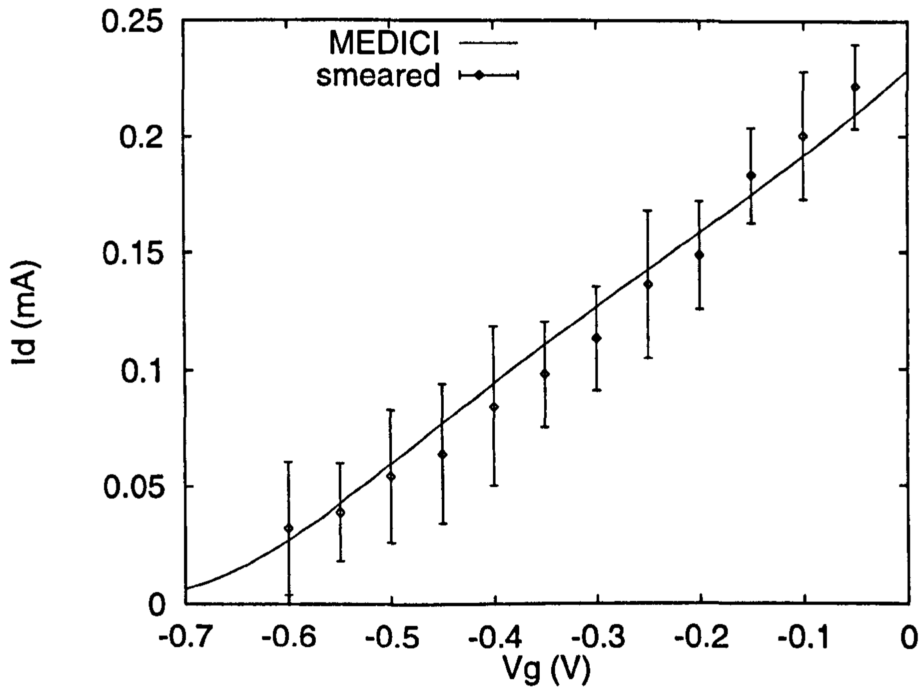


Fig. 5.16: The I-V characteristics of the $80\,n\mathrm{m}$ double gated MESFET with a smeared doping.

The simulation was verified by comparing its results with that of MEDICI (as shown in figure 5.16). Figure 5.16 shows the results for the doping repre-

sented as a continuous distribution of charge. The error-bars show the standard deviation in the current due to thermal and shot noise. There is broad agreement between both sets of results as expected from the close relationship between the drift diffusion approach and Brownian dynamics as as shown in chapter 2.

The statistical variation in the number of dopants under the gate is 32 (*i.e.* a standard deviation of 3 %) and so this is not expected to have a significant influence on the threshold characteristics. The fluctuation in threshold voltage due to variation in dopant number is expected to be only 0.02 V for this device — neglecting all other parameters. However the gate length is less than five times the donor separation and so there should be conducting channels forming due to their chance alignment in the channel. Chapter 4 has shown that the percolation threshold is reduced by the order of $k_B T$ for the samples containing 125 impurities at a mean density of $10^{24}\,m^{-3}$.

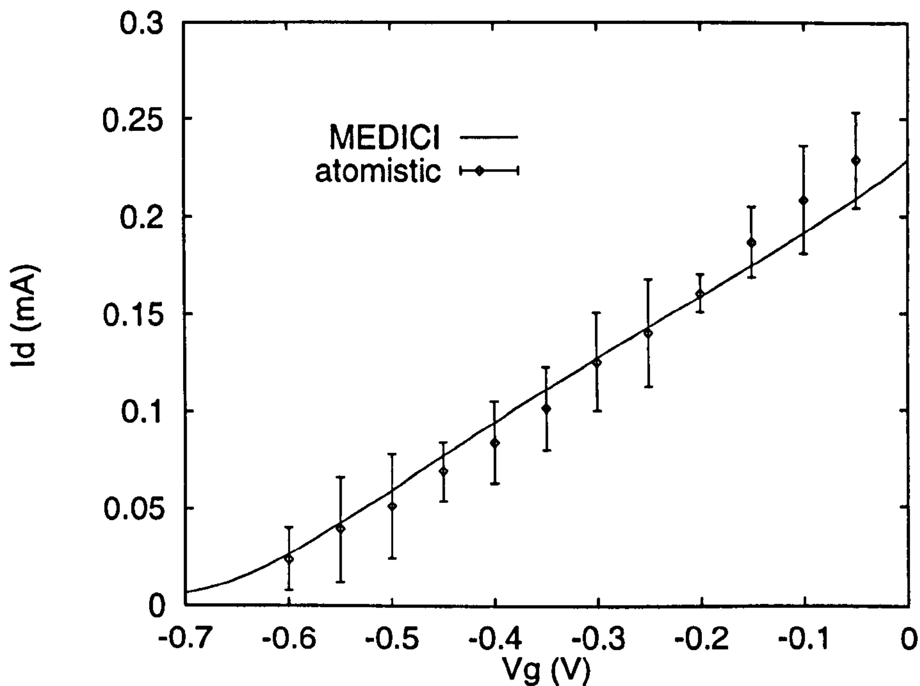


Fig. 5.17: The I-V characteristics of the 80 *nm* double gated MESFET with atomistic donors.

The results from figure 5.17, however, show very similar results to the continuous doping distribution. These results indicate that the use of partic-

ulate ionised impurities has little effect on the overall IV characteristics of the device. The lack of time has prevented a more detailed examination of the subthreshold characteristics. Studies of the channel region over time, when the threshold gate bias is applied (see figures 5.18 and 5.19), show the development of channels. The figures show the potential averaged over 20 *ps* and so blurring out the electrons. The results were accumulated after 2 *ps* so that only the steady state was included in the averaging. The conduction channels form at around -0.64 V which is 30 mV below the calculated threshold. This lowering of threshold approximates the value for the percolation threshold in section 4.4 but much longer simulation times may be needed to show up the small alterations in the subthreshold current.

It is evident that the initial channels are narrow and tortuous and so they are not easily crossed by particulate carriers. Carriers are expected to interact and impede each other's progress they try to pass through the channels. Their motion would then be akin to the kinematics of granular materials, such as sand travelling in a narrow funnel. There has not been sufficient time to carry out a proper study of this device but the good agreement between the simulation results and those obtained from a commercial simulation is sufficient to validate the methods that have been developed.

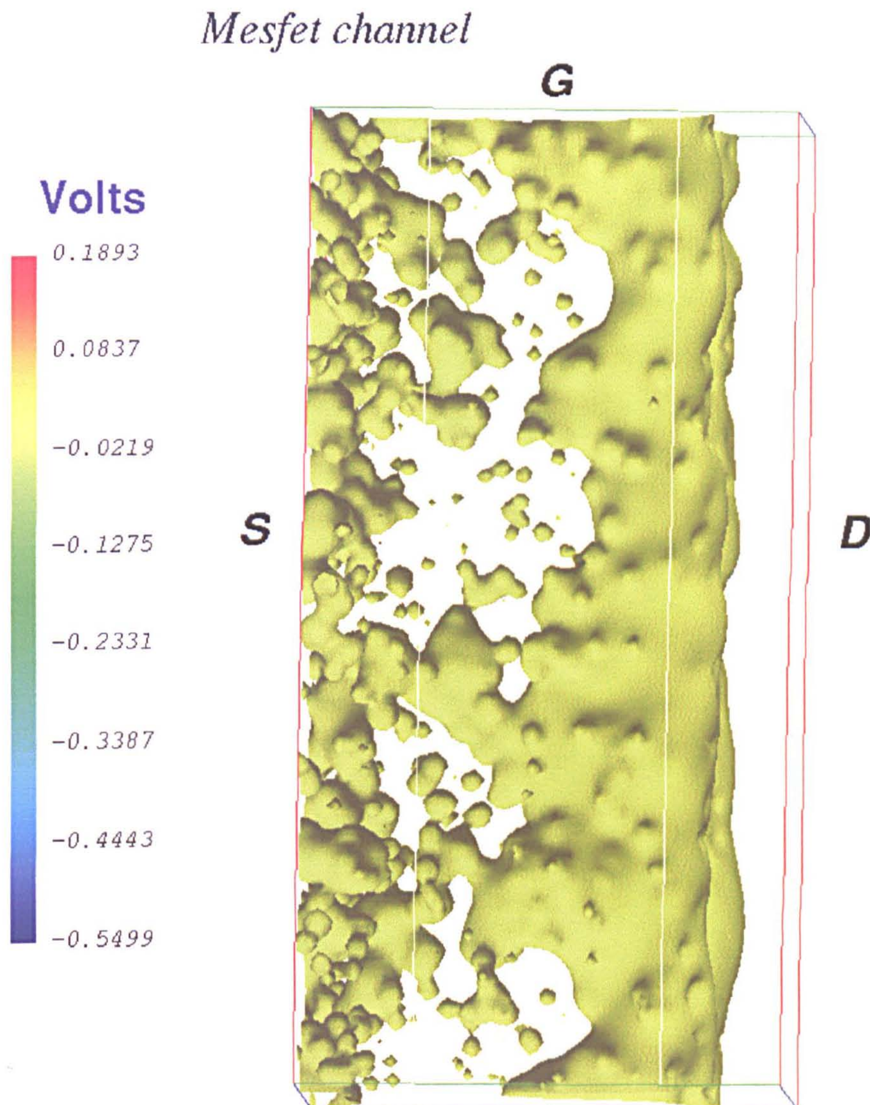


Fig. 5.18: A cross section of the centre of the channel through the xy-plane at threshold.

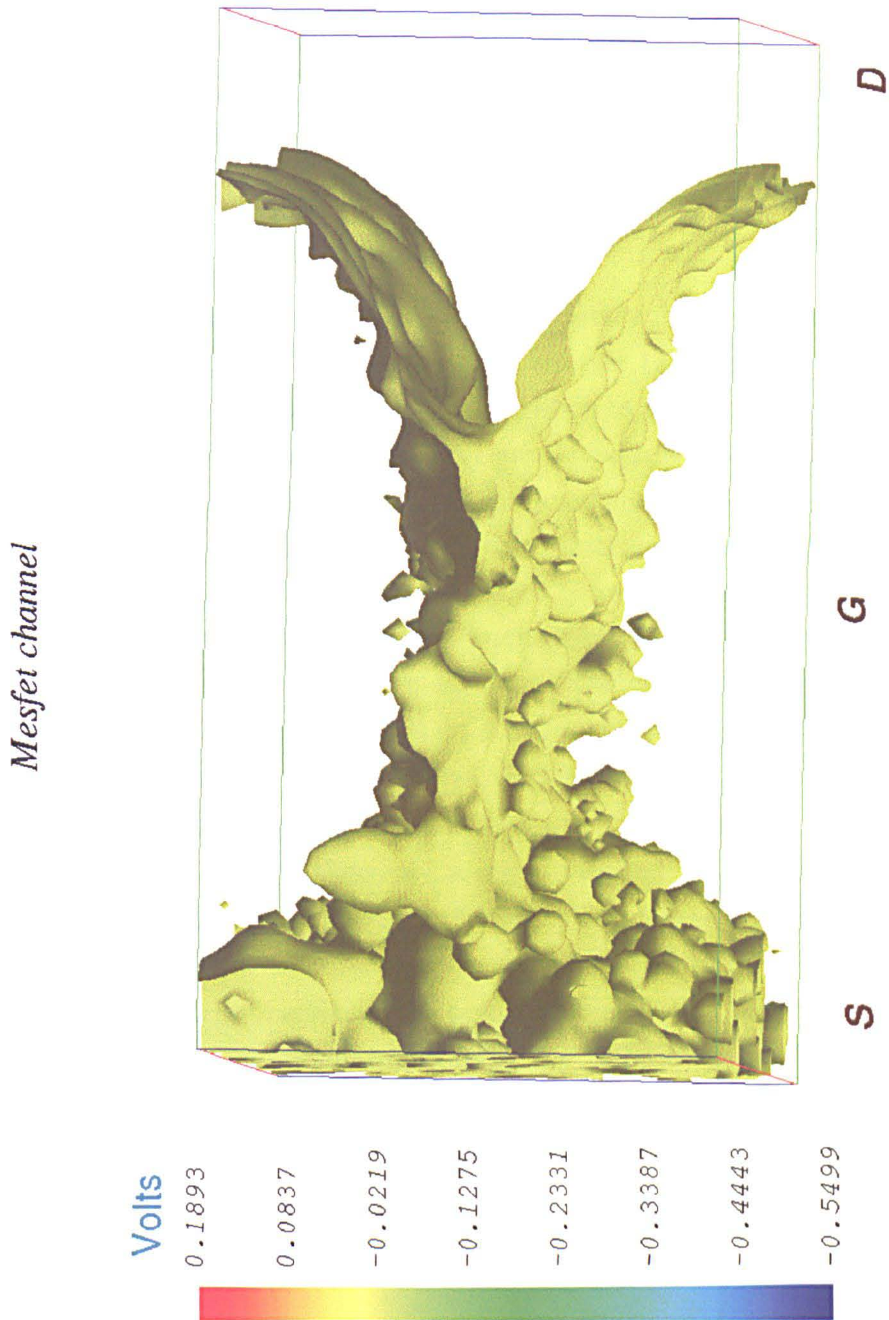


Fig. 5.19: A cross section of the centre of the channel through the yz-plane at threshold.

6. CONCLUSIONS

The current trends in the semiconductor industry indicate that memories and CPUs will use FETs with gates that are shorter than 0.1 μm within the next five years. The main aim of this work has been to demonstrate viable methods to study the behaviour of these sub-0.1 *um* devices, where the quantised nature of electronic charge begins to exert a significant influence on the threshold, and subthreshold, IV-characteristics. The introductory chapter outlined the new behavior that emerge as devices are scaled below 0.1 μm and their impact on the process of miniaturisation. The fluctuation in the numbers and distributions of dopants would results in an unacceptable variation in the threshold IV-characteristics between devices on a wafer and so impede the design of ICs. The past work carried out by others, in studying the effects of the random dopant fluctuations, have been discussed but they failed to treat carriers as discrete particles. Since there are a small number of carriers present in the channel region, the discrete nature of the carriers has to be considered in order to study effects such as screening or noise. Therefore, it is essential to use models that describe carriers as discrete particles over models which rely on the solution of the moment equations of the Boltzmann equation. The validity of the Boltzmann equation may be questioned in the time scales and fine spatial resolution of the device model. The standard 'Monte Carlo' models, whilst retaining the particulate view of the carriers, are computationally expensive and so ways of speeding up and simplifying the problem have been investigated.

Conventional Monte Carlo simulations are very time consuming and so a new, and simpler, model — based on the theory of Brownian motion — has been introduced in the second chapter that is applicable in systems that are not driven far from equilibrium. The method has been validated through studies of bulk materials under constant electric fields and through the successful simu-

lation of a short diode. Brownian dynamics has some similarities to the Monte Carlo method using the relaxation time approximation. The main difference here is that the large discontinuous changes in momentum, at the end of the free flights on the Monte Carlo method, are replaced by a continuous process. This computationally efficient method ensures that the system relaxes to wards the appropriate Maxwell-Boltzmann distribution (for the lattice temperature) and does not introduce any artefacts into the steady state distribution functions — unlike typical Monte Carlo models. The application of the Langevin equation, with its δ-correlated stochastic force in conjunction with a frictional force, describes the motion of the ensemble and not any individual particle. It therefore requires fewer flights to get a statistically acceptable set of data — it has been found to be around seventy times faster than a conventional two dimensional ensemble Monte Carlo model. The only input parameters are the mobility, effective mass and temperature. The first two can be derived from Monte Carlo simulations or experimental measurements and so this method has many similarities with the drift-diffusion model. The heavily doped contacts can use up a large fraction of the computational resources in a Monte Carlo simulation, and so the Brownian model can be used to save much of this effort.

The general aim of this work is to include all electrostatic interactions, with the exception of phonon scattering, in the equations of motion. Much of the CPU time, in these molecular dynamics simulations, is spent calculating and integrating the electrostatic forces and so any improvements here would pay back great dividends. A very efficient multi-grid method is presented in chapter 3 for the solution of Poisson's equation on very large meshes, while coping with the sharp variations in potential arising from the point like charge distributions. The problems arising from the use of a mesh are also discussed and a new efficient technique to overcome these problems is introduced. The force, as derived from the mesh, deviates significantly from the Coulomb force between two point charges when they approach to within about three mesh spacings and so a new and quick method has been developed to tackle this problem. The cubic symmetry of the Poisson grid has been used to devise a method that provides a reasonable solution to this problem without an excessive additional computational overhead. The discrepancy can be mapped along the three

directions of symmetry and a simple interpolation can be made between them. The procedure has been verified by comparing the estimated mesh force with the actual results on large meshes.

The verification of the molecular dynamics technique was carried out in chapter 4. This is a complex many body problem that cannot be described very accurately by the the standard models that have been reviewed at the start of the chapter. The study of an orbit around of an electron around a solitary ionised donor shows the artifacts (showing up as oscillations in the angle of scatter) due to the simple differencing schemes that are used in many conventional simulations. However these artifacts do not have any great effect on the mobilities of the carriers at room temperature as demonstrated by the the comparison of the results for meshes and Ewald summation. The results from Brownian dynamics adequately describe ionised impurity scattering across a wide range of doping levels.

High order integration schemes have to be used when using 'free flights' integrate the equations of motion in samples containing discrete charges. Brownian dynamics has been shown to be very useful for integrating these sharply varying electric fields because it allows the carriers to reach thermal equilibrium and erodes errors that creep into integration of motion. This method allows a for a simple initialisation procedure and provides results that broadly agree with experimental results.

The final section of chapter 4 dealt with the percolation of carriers through small samples. The influence of the distribution of the discrete ions on the percolation threshold was studied in section 4.4. The percolation threshold is the value of the lowest energy surface where the carriers could cross the sample. It explains the reduction in threshold and partly explains the fluctuations in the gate threshold. Conducting paths appeared when around 19 % of the mesh points were in the conducting state, in a 50^3 sample containing 125 random donors, and this value was considerably lower than the percolation threshold value (of 31 %) for a cubic lattice. This was because the presence of an impurity has some influence over the long range potential and so makes neighbouring cells more likely to conduct. The random distribution of impurities results in a lowering of the percolation threshold by $24 \pm 5\,mV$. This figure is of the same order of magnitude with the result (of $30\,mV$) from the MESFET simulations.

It must be noted that the region between the gate contacts had the dimensions $320 \times 80 \times 80\,nm^3$ and contained approximately 1024 ions.

The final task was to put together the methods described previously and formulate an efficient simulation tool. This is a very large computational problem and so it would be beneficial to use parallel or vector computing and so the practical details of the implementation on these systems was discussed in the beginning of the last chapter. The study of the $80\,nm$ dual-gate MESFET was used to validate the simulation approach. The results from the standard approach, that represented the doping profile using continuous charge distributions, were compared with those using the new atomistic treatment of ionised dopants. Studies of the channel showed that different parts regions turned on at different potentials and the conduction paths formed as described earlier. The simulations using discrete impurities showed a reduction in percolation threshold on 30 mV that agreed with the previous measurements on $50\,\mathrm{nm}^3$ samples. This reduction in threshold is only a twentieth of the turn-on voltage of -0.6 V and so discernable effects were detected at threshold since any other effects would have been hidden in the uncertainty in the measurements.

6.1 Scope for further study

This work has taken developed and tested the techniques needed for the study of mesoscopic devices. The first step would be to carry out a comprehensive study of the threshold, and subthreshold, characteristics of this example device. This would require the simulation of about a thousand devices with random dopant distributions to get statistically acceptable results. The number of dopants should follow the appropriate Poisson distribution for the mean doping level.

The percolation threshold is affected by the precise distribution of impurities in the channel and so it would be beneficial to investigate the effects of this on the threshold characteristics of the FET. Here the distributions would vary — in the region between the gates — between each simulation but the number of dopants would be kept constant. At around the threshold, one would expect the size, shape and tracks of the conducting channels to have some effects on the crossing of carriers and hence the observed current.

Other physical processes — such as interface charges and traps whose effects become more pronounced as the number of carriers is reduced — can be incorporated into the model. The goal will be to model more realistic devices — a task that becomes increasingly affordable due to the increasing computational resources. The staple of the semiconductor industry is the MOSFET and this would the the obvious candidate. The advantages of silicon–on–insulator (SOI) MOSFETs would be have to investigated. Accurate models of surface scattering may also be needed to account for the mobility in these small devices.

Process simulations can be carried out, such as the implantation of ions to obtain realistic doping profiles, and variations in patterning or growth can be included. Dislocations can produce variations in strain which in turn produce undulations in the conduction band can be included in the electrostatics.

The multigrid solver would have to be modified in order to include the different dielectric constants and more complex geometries. This can be accomplished, in part, by the creation of arrays to label each grid point as fixed or free and to store the dielectric constants at or between each grid point. Restriction and prolongations have to be applied to these arrays so as to define the problem on the meshes of different scales.

The solvers can be upgraded to use arbitrary meshes with nonuniform spacings and adaptive grid generation would be useful for complex geometries. This would also be useful in the presence of discrete particles where the grid can be locally refined near the vicinity of these charges. This would result in large savings in memory usage. The new method, developed for the treatment of short range forces, requires a constant mesh spacing but the mesh can be refined locally to obviate the need to treat these close interactions separately.

The parallel implementation of the simulation methods may be studied in greater detail. It may be more efficient to implement a "task farming" model where each computing node simulates an entire device under different operating conditions or with different device parameters in the systems that are currently available in the department. The main problem lies in the distribution of the problem on the computational nodes and then the maintenance of synchronisation and data exchange between the nodes. Modern compilers, such as the High Performance Fortran (HPF), can perform these

functions — albeit with some help from the programmer. Smart compilers, found in workstations with multiple processors can distribute the load between the CPUs and, coupled to the fact that prices of these systems will eventually fall, this probably represents the future platform for carrying out these molecular dynamics simulations. Massively parallel architectures will then become the preserve of specialist centres.

The problem with Brownian dynamics is that it it is only applicable to systems in equilibrium, and this is not guaranteed in small devices with high electric fields fields. It does not describe ballistic transport due the presence of the fluctuating and frictional forces on any time scale. It must be noted that this method cannot be described as an '*ab initio*' method since the dynamics are governed not by any physical description of the scattering processes but the fluctuation dissipation and the central limit theorems using input parameters. It would then be beneficial to implement a full Ensemble Monte Carlo model and validate the Brownian simulations. The combination of these two models would enable the investigation of these very small devices through the full range of their operating conditions. The Brownian method is unlikely to completely replace Monte Carlo simulations in the study of ultrasubmicron devices but it can provide a fast simulation tool where the cost of a full Monte Carlo method is unnecessary. In the longer run since the time duration of the collision events becomes no longer negligible, the high fields causing an appreciable change in energy — in comparison to the energy difference of the scattering event — over this time and distances (such as the gate length) approach the de Broglie wavelength of the electron a full quantum transport model may become necessary.

In short this is an up and coming field of study with bright prospects whose results are needed within the next few decades as the sizes of devices shrink to the order of tens of nanometres. This work has covered the initial groundwork by addressing the key practical considerations that must be considered when tacking the huge computational task of modelling small devices:

- simple and effective modelling of dynamics,
- fast solution of Poisson's equation on very large meshes with discrete charges,

- overcoming the band-limiting of meshes.

APPENDIX

A. THE LANGEVIN EQUATION

The scattering of carriers by phonons in the Brownian model is represented by a drag force in conjunction with a fluctuating force. In practical simulations the basic task is to determine the Weiner process from the stochastic acceleration. The first part of this section shows that the distribution of velocities is not affected by the external forces resulting from interactions with other charges and boundary conditions. This simplifies the mathematics that is used later to determine the magnitude of the autocorrelation function (*i.e.* the standard deviation) of the Weiner process.

The discrete-time Langevin equation, can be written as

$$\mathbf{v}_{n+1} = \gamma \mathbf{v}_n + \alpha \left(\mathbf{a}_n \Delta t + \mathbf{V}_n\right) \tag{A.1}$$

for a time $t = n\Delta t$, where $\mathbf{a}_n$ is the acceleration due to the local electric field and $\mathbf{V}_n$ is the Weiner process. If a simple Euler integration method is used then the frictional term is just $-\mathbf{v}_n/\tau$, where τ is the momentum relaxation time. A trapezoidal integration scheme would provide a better approximation by using the average velocity over the timestep for the frictional term $((\mathbf{v}_{n+1} - \mathbf{v}_n)/\tau)$.

Variable	Euler integration	Trapezoidal integration
γ	$1 - \dfrac{\Delta t}{\tau}$	$\dfrac{1 - \frac{\Delta t}{2\tau}}{1 + \frac{\Delta t}{2\tau}}$
α	1	$\dfrac{1}{1 + \frac{\Delta t}{2\tau}}$

Both methods are only accurate to second order in time.

The general solution to this equation with the initial condition $\mathbf{v}_0$ is

$$\mathbf{v}_n = \gamma^n \mathbf{v}_0 + \alpha \sum_{i=0}^{n-1} \gamma^{n-i-1} \left(\mathbf{a}_i \Delta t + \mathbf{V}_i\right) \tag{A.2}$$

Assuming a constant electric field, so $\mathbf{a}_i = a\ \forall i$ this can be written as

$$\begin{aligned}\mathbf{v}_n &= \gamma^n \mathbf{v}_0 + \alpha\left(a\Delta t\frac{1-\gamma^n}{1-\gamma} + \sum_{i=0}^{n-1}\left(\gamma^{n-i-1}\mathbf{V}_i\right)\right) \\ &= a\tau + (\mathbf{v}_0 - a\tau)\,\gamma^n + \alpha\sum_{i=0}^{n-1}\left(\gamma^{n-i-1}\mathbf{V}_i\right) \qquad (A.3)\end{aligned}$$

The deviation of velocities from the mean at time $n\,\Delta t$ is

$$\mathbf{v}_n - \langle \mathbf{v}_n\rangle = \alpha\sum_{i=0}^{n-1}\left(\gamma^{n-i-1}\mathbf{V}_i\right) \qquad (A.4)$$

since $\langle \mathbf{V}_i\rangle = 0$. The electric field has no bearing on this distribution: it merely shifts it by the drift velocity. This is illustrated in figure 2.2 where the same set of random numbers were used and the distribution profiles for the zero field and $1\,\mathrm{MVm}^{-1}$ cases were identical.

Since the motion along the three orthogonal directions are independent, a one dimensional approach will be taken for the sake of brevity. The magnitude of the acceleration due to the fluctuating force can be obtained from the variance of the velocity distribution $(\Delta v_n)^2$.

$$(\Delta v_n)^2 = \langle v_n^2\rangle - \langle v_n\rangle^2 \qquad (A.5)$$

and

$$\begin{aligned}\langle v_n^2\rangle &= \Bigg\langle \gamma^{2n}v_0^2 + 2\alpha v_0\sum_{i=0}^{n-1}\gamma^{2n-i-1}\left(a_i\Delta t + V_i\right) + \\ &\quad \alpha^2\left\{(\Delta t)^2\sum_{i=0}^{n-1}\left(\gamma^{n-i-1}\left(a_i\Delta t + V_i\right)\sum_{j=0}^{n-1}\gamma^{n-j-1}\left(a_j\Delta t + V_j\right)\right]\right\}\Bigg\rangle \\ &= \gamma^{2n}\langle v_0^2\rangle + 2\alpha\langle v_0\rangle\,\Delta t\sum_{i=0}^{n-1}\gamma^{2n-i-1}\langle a_i\rangle + \\ &\quad \alpha^2\left[(\Delta t)^2\left(\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\gamma^{2n-i-j-2}\langle a_i a_j\rangle\right) + \sum_{i=0}^{n-1}\gamma^{2n-i-1}\langle V_i^2\rangle\right] \qquad (A.6)\end{aligned}$$

The correlation function of the stochastic term has the property

$$\langle V_i V_j\rangle = \delta_{ij}\sigma^2 \qquad (A.7)$$

and so only the diagonal term has remained in equation A.6.

$$\langle v_n\rangle^2 = \gamma^{2n}\langle v_0\rangle^2 + 2\alpha\langle v_0\rangle\,\Delta t\sum_{i=0}^{n-1}\gamma^{2n-i-1}\langle a_i\rangle + \alpha^2(\Delta t)^2\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\gamma^{2n-i-j-2}\langle a_i a_j\rangle \tag{A.8}$$

Hence

$$(\Delta v_n)^2 = \gamma^{2n}\left(\langle v_0^2\rangle - \langle v_0\rangle^2\right) + \sigma^2\alpha^2\frac{1-\gamma^{2n}}{1-\gamma^2} \tag{A.9}$$

Equation A.9 gives the energy relaxation time for the Brownian particle as $\tau/2$. The particles relax towards thermal equilibrium and so

$$\lim_{n\to\infty}(\Delta v_n)^2 = \frac{k_B T}{m} = \sigma^2\frac{\alpha^2}{1-\gamma^2} \tag{A.10}$$

with

$$\frac{\alpha^2}{1-\gamma^2} = \begin{cases} \dfrac{\tau}{\Delta t\,(2-\Delta t/\tau)} & \text{Euler integration} \\ \dfrac{\tau}{2\Delta t} & \text{Trapezoidal integration} \end{cases}. \tag{A.11}$$

This gives the strength of the fluctuating acceleration to be

	Euler integration	Trapezoidal integration
σ	$\sqrt{\frac{k_B T}{m}\frac{\Delta t}{\tau}\left(2-\frac{\Delta t}{\tau}\right)}$	$\sqrt{2\frac{k_B T}{m}\frac{\Delta t}{\tau}}$

This result is in keeping with the fluctuation-dissipation theorem, that relates the autocorrelation of the stochastic term to the deterministic frictional term. It is also clear that $D_v = D_x\tau^{-2}$.

A final check on this method is to confirm that this fluctuating acceleration gives the correct diffusion coefficient in real space (D_x). Using the velocity autocorrelation function

$$D_x = \int_0^\infty \langle v(t)v(t+s)\rangle ds \equiv \int_0^\infty \langle v(0)v(s)\rangle ds \tag{A.12}$$

The Euler integration scheme gives:

$$\begin{aligned} D_x &= \sum_{n=0}^{\infty}\langle v_0 v_n\rangle\,\Delta t \\ &= \Delta t\sum_{n=0}^{\infty}\langle v_0\gamma^n v_n\rangle \end{aligned}$$

$$
\begin{aligned}
&= \frac{\langle v_0^2 \rangle}{1-\gamma}\,\Delta t \\
&= \langle v_0^2 \rangle \tau = \frac{k_B T \mu}{e}
\end{aligned}
$$

and the trapezoid integration results in

$$
\begin{aligned}
D_x &= \lim_{n\to\infty} \frac{1}{2}\left[\langle v_0 v_0 \rangle + 2\left(\sum_{i=1}^{n-1} \langle v_0 \gamma^i v_i \rangle\right) + \langle v_0 \gamma^n v_n \rangle\right] \Delta t \\
&= \frac{1}{2}\langle v_0^2 \rangle\,\Delta t \left(1 + 2\sum_{i=1}^{n-1} \gamma^i\right) = \frac{1}{2}\langle v_0^2 \rangle\,\Delta t \left\{1 + 2\left[\left(\sum_{i=0}^{n-1} \gamma^i\right) - 1\right]\right\} \\
&= \frac{1}{2}\langle v_0^2 \rangle\,\Delta t \left(1 + 2\frac{\gamma}{1-\gamma}\right) \\
&= \frac{1}{2}\langle v_0^2 \rangle \frac{1+\gamma}{1-\gamma}\,\Delta t = \langle v_0^2 \rangle \tau
\end{aligned}
$$

where the Einstein relation has been used to express the final result in terms of mobility. It has been assumed that the particles were in thermal equilibrium at time t and that there was no external field.

A.1 The relation to the Drift-diffusion equation

One of the fastest and most widespread modelling techniques is the drift-diffusion method. The simulation of the Si diode resulted in good agreement between Brownian dynamics and the drift-diffusion approach. Therefore it will now be shown that the Langevin equation is equivalent to the drift-diffusion model in the steady-state limit. When studying systems using time-scales greater than the momentum relaxation time, it can be assumed that $\dot{\mathbf{v}} \approx 0$ and $\mathbf{v} \approx \mu\mathbf{E}$.

The Langevin equation in real space is just

$$\dot{\mathbf{x}} = \mu\mathbf{E} + \Delta\mathbf{V} \tag{A.13}$$

The occupation of space can be represented using a probability distribution function $f(\mathbf{x}, t)$ and a transition probability density (or propagation operator) $f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_0, t_0)$ can be defined in order to study the evolution of the system from an initial state $\mathbf{x}_0, t_0)$.

$$f(\mathbf{x}, t + \Delta t) = \int f(\mathbf{x}, t_1 \,|\, \mathbf{x}_0, t_0)\, f(\mathbf{x}_0, t_0)\, d\mathbf{x}_0 \tag{A.14}$$

The Fokker-Plank equation for equation A.13 is

$$\frac{\partial}{\partial t} f(\mathbf{x}, t|\mathbf{x}_0, t_0) + A\nabla f(\mathbf{x}, t|\mathbf{x}_0, t_0) = \frac{1}{2} B \nabla^2 f(\mathbf{x}, t|\mathbf{x}_0, t_0) \tag{A.15}$$

$$A = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (\mathbf{x}' - \mathbf{x}) f(\mathbf{x}, t|\mathbf{x}_0, t_0) d\mathbf{x}' = \lim_{\Delta t \to 0} \frac{\langle \Delta \mathbf{x} \rangle}{\Delta t} = \mathbf{a}\tau \tag{A.16}$$

$$B = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (\mathbf{x}' - \mathbf{x})^2 f(\mathbf{x}, t|\mathbf{x}_0, t_0) d\mathbf{x}' = \lim_{\Delta t \to 0} \frac{\langle \Delta \mathbf{x}^2 \rangle}{\Delta t} = 2D_x \tag{A.17}$$

The probability density function at time t can be obtained after integrating this transition probability density over all the initial state. And so

$$\frac{\partial}{\partial t} f(\mathbf{x}, t) + \mathbf{a}\tau \nabla f(\mathbf{x}, t) = D_x \nabla^2 f(\mathbf{x}, t) \tag{A.18}$$

The time derivative can be eliminated by considering the conservation of particles. In the absence of generation or recombination

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} + \nabla \mathcal{F}(\mathbf{x}, t) = 0 \tag{A.19}$$

where $\mathcal{F}(\mathbf{x}, t)$ is the flux per unit area and $\mathcal{F}(\mathbf{x}, t) = f(\mathbf{x}, t)\mathbf{v}$. Therefore, after integrating over space we have

$$-f(\mathbf{x}, t)\mathbf{v} + \mathbf{a}\tau f(\mathbf{x}, t) = D_x \nabla f(\mathbf{x}, t) \tag{A.20}$$

and using $n = Nf(\mathbf{x}, t)d\mathbf{x}$ where N is the total number of particles

$$-n\mathbf{v} + n\tau\mathbf{a} = D_x \nabla n \tag{A.21}$$

After multiplying by the carrier charge q and using $\mathbf{a} = q\mathbf{E}/m$, the current continuity equation is obtained, after a little rearranging.

$$\mathbf{J} = \rho\mu\mathbf{E} - D_x \nabla \rho \tag{A.22}$$

This clearly demonstrates that the real-space Langevin equation is equivalent to the drift-diffusion model.

B. THE BICONJUGATE GRADIENT METHOD.

These methods, in general, try to solve the system of N linear equations in N variables

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{B.1}$$

in the case of $\mathbf{A}$ being symmetric and positive definite, by minimising the function

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \cdot \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \cdot \mathbf{x} \tag{B.2}$$

whose minimum occurs when

$$\nabla f = \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \tag{B.3}$$

The minimum is found by choosing a striking out in some direction $\mathbf{p}_k$ with a weighting α_k from each kth iterate $\mathbf{x}_k$.

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{p}_k \tag{B.4}$$

At each iteration a residual, given by,

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A} \cdot \mathbf{x}_k \tag{B.5}$$

is obtained. These vectors must satisfy the bi-orthogonality condition

$$\mathbf{r}_i^T \cdot \mathbf{r}_j = 0, \text{ if } i \neq j \tag{B.6}$$

and the bi-conjugacy condition

$$\mathbf{p}_i^T \cdot \mathbf{A} \cdot \mathbf{p}_j = 0, \text{ if } i \neq j. \tag{B.7}$$

This recurrence relation results in the termination of the series after N iterations, at most, since there cannot be more than N orthogonal directions in an N-dimensional vector space. However, this termination is only guaranteed when exact arithmetic is used.

Unfortunately, $\mathbf{A}$ is rarely symmetric and so the generalised minimum residual (GMRES) algorithm is often employed to circumvent this problem, by pre-multiplying by $\mathbf{A}^T$.

The particular bi-conjugate gradient method that was applied was the Bi-CGSTAB(2). The Bi-CGSTAB method relies on the fact that the residual can be constructed recursively as

$$\mathbf{r}_k = P_k(\mathbf{A})\, Q_k(\mathbf{A})\, \mathbf{r}_0 \tag{B.8}$$

Were P_k is a polynomial, of degree k, giving the residual of bi-conjugate gradient from the initial $\mathbf{r}_0$. Q_k is constructed from the product of k linear factors $(1 - \omega_j \mathbf{A})$ under the condition that $Q_k(0) = 1$. The ω_j are taken to minimise a residual locally, and so if it equals zero (or even becomes very small) at any point, this method breaks down. In fact, Bi-CGSTAB can be viewed as the combined effect of bi-conjugate gradient and GMRES(1) steps, and so when the GMRES(1) step [nearly] stagnates, then the bi-conjugate gradient part in the next iteration cannot (or only poorly) be constructed.

The optimal reduction polynomials for matrices with complex eigenvalues may also have complex roots, and so this construction for Q_k can made to allow for this by obtaining the Q-polynomial, for each successive even-numbered iteration form the previous one, by using the quadratic expansion,

$$1 - \omega_j^1 \mathbf{A} - \omega_j^{(2)} \mathbf{A}^2.$$

This Bi-CGSTAB2 algorithm does possess a better convergence, but it is still hampered by its reliance on the odd-numbered steps. In Bi-CGSTAB(2), the Q-polynomial is constructed out of quadratic factors, and so obviates the need for the linear factor of the odd step.

The Bi-CGSTAB(2) algorithm can be written as follows:
Let $\mathbf{x}_0$ be an initial guess and $\mathbf{r}_0 = \mathbf{b} - \mathbf{A} \cdot \mathbf{x}_0$.
Choose an arbitrary vector $\mathbf{q}$ such that $\mathbf{q} \cdot r \neq 0$, say $\mathbf{q} = \mathbf{r}$.

$$\begin{aligned}
&\rho_0 = 1; \alpha = 0; \omega_2 = 1;\\
&\mathbf{u}_0 = 0;\\
&\text{for } i = 0, 2, 4, \ldots\\
&\qquad \rho_0 = -\omega_2 \rho_0;
\end{aligned}$$

even Bi-CG step:

$$\rho_1 = \mathbf{q}\cdot\mathbf{r}; \beta = \alpha\tfrac{\rho_1}{\rho_0}; \rho_0 = \rho_1;$$
$$\mathbf{u} = \mathbf{r}_i - \beta\mathbf{u};$$
$$\mathbf{v} = \mathbf{A}\mathbf{u}$$
$$\gamma = \mathbf{v}\cdot\mathbf{q}; \alpha = \frac{\rho_0}{\gamma};$$
$$\mathbf{r} = \mathbf{r}_i - \alpha\mathbf{v};$$
$$\mathbf{s} = \mathbf{A}\mathbf{r};$$
$$\mathbf{x} = \mathbf{x}_i + \alpha\mathbf{u};$$

odd Bi-CG step:

$$\rho_1 = \mathbf{q}\cdot\mathbf{s}; \beta = \alpha\tfrac{\rho_1}{\rho_0}; \rho_0 = \rho_1;$$
$$\mathbf{v} = \mathbf{s} - \beta\mathbf{v};$$
$$\mathbf{v} = \mathbf{A}\mathbf{u}$$
$$\mathbf{w} = \mathbf{A}\mathbf{v}$$
$$\gamma = \mathbf{w}\cdot\mathbf{q}; \alpha = \tfrac{\rho_0}{\gamma};$$
$$\mathbf{v} = \mathbf{s} - \beta\mathbf{v};$$
$$\mathbf{r} = \mathbf{r} - \alpha\mathbf{v};$$
$$\mathbf{s} = \mathbf{s} - \alpha\mathbf{w};$$
$$\mathbf{t} = \mathbf{A}\mathbf{s};$$

GCR(2) - part:

$$\omega_1 = \mathbf{r}\cdot\mathbf{s}; \qquad \sigma = \mathbf{s}\cdot\mathbf{s};$$
$$\zeta = \mathbf{s}\cdot\mathbf{t}; \qquad \lambda = \mathbf{t}\cdot\mathbf{t};$$
$$\omega_2 = \mathbf{r}\cdot\mathbf{t}; \qquad \lambda = \lambda - \zeta^2/\sigma;$$
$$\omega_2 = (\omega_2 - \zeta\omega_1/\sigma)/\tau;$$
$$\omega_1 = (\omega_1 - \zeta\omega_2)/\sigma;$$
$$\mathbf{x}_{i+2} = \mathbf{x} + \omega_1\mathbf{r} + \omega_2\mathbf{s} + \alpha\mathbf{u};$$
$$\mathbf{r}_{i+2} = \mathbf{r} - \omega_1\mathbf{s} - \omega_2\mathbf{t};$$

if $\mathbf{x}_{i+2}$ is accurate enough then exit!

$$\mathbf{u} = \mathbf{u} - \omega_1\mathbf{v} - \omega_2\mathbf{w};$$

end loop.

C. MATERIAL PARAMETERS

This appendix gives all the relavant material parameters that are used in the simulations using GaAs and Si.

	GaAs	Si
$\frac{m_e}{m_0}$	0.067^a	m_l^b 0.98 m_t^b 0.19
ϵ_r	13.18^a	11.9^c
μ_e^c ($\mathrm{m^2\,V^{-1}\,s^{-1}}$)	0.85	0.135
n_i^c ($\mathrm{m^{-3}}$)	1.5×10^{16}	10^{13}

where n_i is the intrinsic electron number density and the references are
(a) Ridley [117],
(b) Reggiani [138],
(c) Cooke [139].

BIBLIOGRAPHY

[1] T. R. Halfbill, BYTE **21**, 101 (1996).

[2] T. R. Halfbill, BYTE **21**, 89 (1996).

[3] Y. Taur *et al.*, Proc. IEEE **85**, 486 (1997).

[4] H. Iwai, in *Proceedings of the* 26th *European Solid State Device Research Conference* (Editions Frontières, Gif-sur-Yvette, France, 1996), pp. 45–52.

[5] J. Chung *et al.*, IEEE Electron Device Lett. **EDL-9**, 186 (1988).

[6] B. Hoeneisen and C. A. Mead, Solid-State Elect. **15**, 819 (1972).

[7] R. H. Dennard *et al.*, IEEE J. Solid State Circuits **SC-9**, 256 (1974).

[8] G. Baccarini, M. R. Worderman, and R. H. Dennard, IEEE Trans. Electron. Dev. **ED-31**, 452 (1984).

[9] S. M. Serra, *Semiconductor devices, physics and technology* (John Whiley & Sons, New York, 1986).

[10] J. R. Brews, W. Fichtner, E. H. Nicollcan, and S. M. Sze, IEEE Electron. Dev. Lett. **EDL-1**, 2 (1980).

[11] T. Morimoto *et al.*, in *Ext. Abs. of Int. Conf. on Solid State Devices and Materials* (IEEE, New York, 1991), pp. 23–25.

[12] K. Suzuki *et al.*, IEEE Trans. Electron. Devices **40**, 2327 (1993).

[13] H. S. Momose *et al.*, IEDM Tech. Dig. 593 (1994).

[14] D. R. Wolters and J. J. van der Schoot, Philips J. Res. **40**, 115,137,164 (1985).

[15] I.-C. Chen and C. Hu, IEE Electron. Device Lett. **EDL-8**, 140 (1987).

[16] R. Moazzami, J. Lee, I.-C. Chen, and C. Hu, Tech dig. Int. Electron Dev. Meet. 710 (1987).

[17] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, Tech dig. Int. Electron Dev. Meet, 718 (1987).

[18] C. Duvvury, D. J. Redwine, and H. J. Stiegler, IEE Electron. Device Lett. **EDL-9**, 579 (1988).

[19] I.-C. Chen, C. W. Teng, D. J. Coleman, and A. Nishimura, IEE Electron. Device Lett. **EDL-10**, 216 (1989).

[20] R. Harata *et al.*, IEEE Electron Dev. Lett. **EDL-10**, 27 (1989).

[21] P. J. Wright, N. Kasai, S. Inoue, and K. C. Saraswat, IEEE Electron Dev. Lett. **EDL-10**, 347 (1989).

[22] T. Hori and H. Iwasaki, IEEE Electron Dev. Lett. **EDL-10**, 64 (1989).

[23] E. Takeda *et al.*, IEEE Trans. Electron Devices **ED-30**, 652 (1983).

[24] Y. Toyoshima, N. Nihira, and K. Kanzaki, in *Symp. VLSI Tech. Dig.* (IEEE, New York, 1985), p. 118.

[25] S. Jain, W. T. Cochran, and M. L. Chen, IEEE Electron Dev. Lett. **EDL-9**, 539 (1988).

[26] T. Hori, in *Proc. Int. Electron. Devices Meeting* (IEEE, New York, 1989), p. 777.

[27] R. Izawa, T. Kube, and E. Takeda, IEEE Trans. Electron Devices **ED-35**, 2088 (1988).

[28] I. Chen, C. C. Wei, and C. W. Teng, IEEE Electron Dev. Lett. **EDL-11**, 78 (1990).

[29] A. Tasch, T. Shin, T. J. Bordelon, and C. M. Maziar, IEEE Electron Dev. Lett. **EDL-11**, 517 (1990).

[30] F. L. Duan, S. P. Sinhà, D. E. Ioannou, and F. T. Brady, IEEE Trans. Electron Devices **ED-44**, 972 (1997).

[31] C. Zener, Proc. R. Soc. A **145**, 523 (1934).

[32] D. K. Ferry and R. O. Grondin, *Physics of Submicron Devices* (Plenum Press, New York, 1986).

[33] D. Hisamoto, K. Umeda, Y. Nakamura, and S. Kimura, IEEE Trans. Electron. Dev. **ED-44**, 951 (1997).

[34] Y. Ushiku, H. Kushibe, H. Ono, and A. Nishiyama, in *Proc. VMIC Conf.* (IEEE, New York, 1990), pp. 413–414.

[35] S. P. Murarka and S. W. Hymes, Critical Reviews in Solid State and Materials Sciences **20**, 87 (1995).

[36] J. Li, Y. Shachamdiamand, and J. W. Mayer, Materials Science Reports **9**, 1 (1992).

[37] T. Ohmi and K. Tsubouchi, Solid State Technology **35**, 47 (1992).

[38] W. Maly, A. Strojwas, and S. Director, IEEE Trans. Computer-Aided Design **CAD-5**, 114 (1986).

[39] R. S. Guo and E. Sachs, IEEE Trans. Semicond. Manuf. **6**, 41 (1993).

[40] E. Sachs, R. S. Guo, S. Ha, and A. Hu, IEEE Trans. Semicond. Manuf. **4**, 134 (1991).

[41] P. K. Mozumder, C. R. Shyamsundar, and A. Strojwas, IEEE Trans. Semicond. Manuf. **1**, 62 (1988).

[42] C. R. Shyamsundar, P. K. Mozumder, and A. Strojwas, IEEE Trans. Semicond. Manuf. **1**, 72 (1988).

[43] R. Sitte, S. Dimitrijev, and H. B. Harrison, Electronics Letters **29**, 1345 (1993).

[44] R. Sitte, S. Dimitrijev, and H. B. Harrison, IEEE Trans. Semicond. Manuf. **7**, 92 (1994).

[45] R. Sitte, S. Dimitrijev, and H. B. Harrison, IEEE Trans. Electron. Dev. **41**, 2210 (1994).

[46] J. T. Horstmann, U. Hilleringmann, and K. Goser, in *Proceedings of the* 26th *European Solid State Device Research Conference* (Editions Frontières, Gif-sur-Yvette, France, 1996), pp. 253–256.

[47] R. W. Keyes, IEEE J. Solid-State Circuits **41**, 245 (1975).

[48] W. Shockley, Solid-State Electron. **2**, 35 (1961).

[49] R. W. Keyes, Appl. Phys. **19**, 251 (1975).

[50] B. J. Last and D. J. Thouless, Phys. Rev. Letters **27**, 1719 (1971).

[51] T. Hagiwara, K. Yamaguci, and S. Asai, Tech. Dig. VLSI Symp. 46 (1982).

[52] K. Nishinohara, N. Shigyo, and T. Wada, IEEE Trans. Electron. Dev. **39**, 634 (1992).

[53] H. Wong and Y. Taur, IEEE IEDM 705 (1993).

[54] J. T. Watt and D. Plumber, IEEE Trans. Electron. Dev. **41**, 2222 (1994).

[55] E. H. Nicollian and A. Goetzberger, Bell Syst. Tchn. J. **46**, 1055 (1967).

[56] T. Mizuno, J. Okamura, and A. Toriumi, IEEE Trans. Electron. Dev. **41**, 2216 (1994).

[57] T. Mizuno, Jap. J. Appl. Phys., Part 1 **35**, 842 (1996).

[58] J.-R. Zhou and D. K. Ferry, in *Proc. 3rd International Workshop on Computational Electronics* (Plenum Press, New York, 1994), pp. 74–77.

[59] J. H. Davies and G. Timp, *Heterostructures and Quantum Devices* (Academic Press Inc., Oxford, 1994), p. 385.

[60] K. K. Hung, P. Ho, C. Hu, and Y. C. Cheng, IEEE Electron. Dev. Lett. **11**, 90 (1990).

[61] G. Ghibaudo, IEEE Electron. Dev. Lett. **11**, 352 (1990).

[62] K. S. Ralls *et al.*, Phys. Rev. Lett. **52**, 228 (1984).

[63] M. J. Kirton, M. J. Uren, and S. Collins, Appl. Surf. Sci. **30**, 148 (1987).

[64] K. Yano, D. K. Ferry, and K. Seki, IEDM Tech. Dig. 557 (1992).

[65] D. K. Ferry, A. M. Kriman, M. J. Kann, and R. P. Joshi, Computer Physics Commun. **67**, 119 (1991).

[66] J. H. F. Scott-Thomas *et al.*, Phys. Rev. Lett. **62**, 583 (1989).

[67] C. Fiegna *et al.*, IEEE Trans. Electron. Dev. **41**, 941 (1994).

[68] F. Venturi, R. K. Smith, E. C. Sangiorgi, and B. Pinto, M. nd Riccó, IEEE Transactions on Computer-Aided Design **8**, 360 (1989).

[69] K. Tarnay *et al.*, in *Software Applications in Electrical Engineering* (Computational Mechanics Publications Ltd., Ashurst Lodge, Southampton, UK, 1993), Chap. 46, pp. 345–358.

[70] C. Moglestue and S. J. Beard, in *Numerical Analysis of Semiconductor Devices*, edited by B. T. Browne and J. J. H. Miller (Boole Press, Dublin, Ireland, 1979), pp. 232–236.

[71] R. A. Warriner, IEE Proc 1: Solid State and Electron Devices **1**, 92 (1977).

[72] C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).

[73] R. M. Yorston, J. Comp. Phys. **64**, 177 (1986).

[74] M. G. Ancona, Solid-State Electronics **33**, 1633 (1990).

[75] K. Kometer, Phys. Rev. B **46**, 1382 (1992).

[76] K. Kometer, G. Zandler, and P. Vogl, Semiconductor Science and Technology **7**, B559 (1992).

[77] K. Kometer, Ph.D. thesis, Technical University, Munich, 1992.

[78] A. Rein, Ph.D. thesis, Technical University, Munich, 1995.

[79] G. Zandler *et al.*, IEEE Electron Device Letters **14**, 77 (1993).

[80] H. Budd, J. Phys. Soc. Japan Suppl. **21**, 420 (1966).

[81] P. J. Price, in *Proc.* 9th *Int. Conf. Phys. Semiconductors* (University of British Columbia, Vancouver, 1968), p. 753.

[82] H. D. Rees, J. Phys. Chem. Solids **30**, 643 (1969).

[83] S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer-Verlag, Wein - New York, 1984).

[84] R. K. Frolich and P. A. Blakey, in *Proc. NASECODE II Conf.* (Boole Press, Dublin, 1981), p. 208.

[85] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, Wein - New York, 1989).

[86] C. J. G Baccarini and A. M. Mezzone, Solid-State Electron. **20**, 5 (1977).

[87] R. Brown, Ann. d. Phys. u. Chem. **14**, 294 (1828).

[88] A. Einstein, Ann. d. Phys. **17**, 549 (1905).

[89] A. Einstein, *Investigations on the theory of Brownian movement* (Dover Publications, Inc., New York, 1956).

[90] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North, Holland, 1983).

[91] R. Serra, M. Andretta, M. Campiani, and G. Zanarini, *Introduction to the Physics of Complex Systems* (Pergamon Press, Oxford, 1986).

[92] J. W. Eastwood, J. Comp. Phys. **19**, 215 (1980).

[93] J. W. Eastwood, R. W. Hockney, and D. N. Lawrence, Computer Physics Communications **19**, 215 (1980).

[94] C. Temperton, J. Comput. Phys. **34**, 314 (1980).

[95] R. W. Hockney and J. W. Eastwood, *Computer Simulation using Particles* (McGraw-Hill, Maidenhead, 1981).

[96] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge University Press, Cambridge, UK, 1992).

[97] D. M. Young, *Iterative solution of large linear systems* (Academic Press, New York, 1971).

[98] J. Stoer and R. Bulirsch, *Introduction to numerical analysis* (Springer-Verlag, New York, 1980).

[99] R. Varga, *Matrix Iterative Analysis* (Academic Press, New York, 1971).

[100] M. R. M. R. Hestenes, *Conjugate direction methods in optimization* (Springer-Verlag, New York, 1980).

[101] H. A. van der Vorst, SIAM J. Sci. Stat. Comput. **13**, 631 (1992).

[102] A. Yoshi *et al.*, IEEE Trans. Electron. Dev. **ED-29**, 184 (1982).

[103] F. Y. Chang and L. F. Wagner, Electronics Letters **18**, 658 (1982).

[104] T. Wada and R. L. M. Dang, Electronics Letters **18**, 265 (1982).

[105] G. L. G. Sleijpen and D. R. Fokkema, Technical Report No. 772, University Utrecht, Dep. of Math. (unpublished).

[106] Y. Saad and H. M. Schultz, SIAM J. Sci. Stat. Comput. **7**, 856 (1986).

[107] H. A. van der Vorst and C. Vuik, Technical Report No. 91-80, Delft University of Technology, Faculty of Tech. Math. (unpublished).

[108] H. A. van der Vorst, D. R. Fokkema, and G. L. Sleijpen, Sim. Semicond. Devices and Process. **5**, 89 (1993).

[109] A. Brandt, Phys Rev. B **31**, 333 (1977).

[110] W. Hackbusch, *Multi-grid Methods and Applications* (Springer-Verlag, Wein - New York, 1985).

[111] K. Stuben and U. Trottenberg, in *Multigrid Methods*, edited by W. Hackbusch and U. Trottenberg (Springer-Verlag, New York, 1982).

[112] A. Brandt, in *Multigrid Methods*, edited by W. Hackbusch and U. Trottenberg (Springer-Verlag, New York, 1982).

[113] R. W. Hockney, in *The potential calculation and some applications*, edited by B. Alder, S. Fernbach, and M. Rotenberg (Academic Press, New York, 1980).

[114] C. Temperton, J. Comput. Phys. **31**, 1 (1979).

[115] A. Brandt, Applied Mathematics and Computation **80**, 73 (1996).

[116] P. M. Mathews and K. Venkatesan, *A Textbook of Quantum Mechanics* (Tata McGraw-Hill, New Delhi, 1976).

[117] B. K. Ridley, *Quantum Processes in Semiconductors* (Clarendon Press, Oxford, 1993).

[118] D. Chattopadhyay and H. J. Queisser, Rev. Mod. Phys. **53**, 745 (1981).

[119] E. Conwell and V. F. Weisskopf, Phys. Rev. **77**, 388 (1950).

[120] H. Brooks, Phys. Rev. **83**, 879 (1951).

[121] H. Brooks, in *Advances in Electronics and Electron Physics*, edited by L. Marton (Academic Press, New York, 1955), Vol. 7, p. 85.

[122] B. K. Ridley, J. Phys. C **10**, 1589 (1977).

[123] N. Takimoto, J. Phys. Soc. Jpn. **14**, 1142 (1959).

[124] S. Chandresekhar, Rev. Mod. Phys. **15**, 1 (1943).

[125] P. P. Ewald, Ann. Phys. **64**, 253 (1921).

[126] D. J. Adams and G. Dubey, J. Comp. Phys. **72**, 156 (1987).

[127] T. Thompson, BYTE **21**, 45 (1996).

[128] S. Roy, A. Asenov, A. Brown, and J. R. Barker, in *Proc. 4th ACME Conference* (Faculty of Engineering, University of Glasgow, Glasgow, 1996), pp. 41–44.

[129] S. Roy, A. Asenov, and J. R. Barker, in *Proceedings of the EUROSIM 1996 International Conference* (Elsevier, Amsterdam, 1996).

[130] R. D. da Cunha and T. Hopkins, *Transputing for Numerical and Neural Network Applications* (IOS Press, Amsterdam, 1993).

[131] R. D. da Cunha and T. Hopkins, *Transputing for Numerical and Neural Network Applications* (IOS Press, Amsterdam, 1993).

[132] D. J. Frank, S. E. Laux, and M. V. Fischetti, IEDM Tech. Digest 553 (1992).

[133] T. Sekigawa and Y. Hayashi, Solid-State Electron. **27**, 827 (1984).

[134] H. Horie *et al.*, SSDM Tech. Digest 165 (1991).

[135] T. Tanaka, H. Horie, S. Ando, and S. Hijiya, IEDM Tech. Digest 165 (1991).

[136] F. Balestra, S. Cristoloveanu, M. Benachir, and T. Elewa, IEEE Electron Device Lett. **EDL-8**, 410 (1987).

[137] S. Venkatesan, G. W. Neudeck, and R. F. Pierret, IEEE Electron Device Lett. **13**, 44 (1992).

[138] L. Reggiani, J. Phys. Soc. Jpn. Suppl. A **49**, 317 (1980).

[139] M. J. Cooke, *Semiconductor Devices* (Prentice Hall, London, 1990).

GLASGOW UNIVERSITY LIBRARY