DIRECT SOLUTION OF THE BOLTZMANN TRANSPORT EQUATION IN

NANOSCALE SI DEVICES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Kausar Banoo

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2000

To my mother for her amazing courage and love

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Banoo, Kausar, Ph.D., Purdue University, December, 2000. Direct solution of the Boltzmann Transport equation in nanoscale Si devices. Major Professor: Prof. Mark S. Lundstrom.

Predictive semiconductor device simulation faces a challenge these days. As devices are scaled to nanoscale lengths, the collision-dominated transport equations used in current device simulators can no longer be applied. On the other hand, the use of a better, more accurate Boltzmann Transport Equation (BTE) is hampered by the fact that it is a complicated integro-differential 6-dimensional kinetic equation and is extremely difficult to solve. Previous works on solving the BTE have used either a stochastic method or an approximate method, both of which do not have the suitable properties for practical device simulation. Therefore, this work describes the first direct numerical solution of the BTE for semiconductors that can be used for practical device simulation. This is done by using powerful mathematical techniques to discretise the BTE in energy and angle without making any approximations about the angular shape of the distribution function or the collision integral. Such a direct discretisation results in a very large matrix equation,with $N = 10^6$–10^7 unknowns. In order to address the need for efficient and fast solutions, this work also reports the first application of a preconditioned iterative method (GMRES) to the BTE. This method is not only fast (on the order of $N^{1.2}$) but also has low memory requirements because it does not require explicit storage of the matrix elements. The technique developed in this work is also highly suitable for self-consistent device simulations because it shows smooth and stable convergence when coupled to the Poisson equation. Finally, this method is applied to study transport in two representative nanoscale devices — a one-dimensional 50nm n^+-p-n^+ diode and a two-dimensional 50nm ultra-thin body dual-gate nMOSFET. The report ends with a summary and a discussion of possible future improvements in this field.

1. Introduction

This dissertation describes a direct numerical solution to the semi-classical Boltzmann Transport equation (BTE) in semiconductor devices. The objective of this work is to demonstrate, for the first time, that with powerful numerical techniques it is possible to solve the BTE directly without making any approximations. It will be shown that this method has important applications in predictive semiconductor device simulation (specifically for nano-scale Si MOS transistors) and greater accuracy over the common models in current use.

The organisation of this chapter is as follows. Section 1.1 starts with a brief background on semiconductor device simulation and attempts to put the various methods (models) of semiconductor simulation in perspective. In this regard, special attention is given to the BTE and the motivation behind this project. Section 1.2 describes the BTE and its important properties in detail. Section 1.3 presents a brief introduction to the finite volume method used to solve the BTE in this work. The concepts discussed in this section are simple and can be skipped by experts. Finally, Section 1.4 presents a chapter-wise overview of the dissertation and its key results.

1.1   Background

Semiconductor device simulation has seen considerable progress in the past 40 years. It has evolved from initial theoretical formulation into sophisticated tools that can simulate the electrical properties of a wide variety of semiconductor devices (such as [1, 2, 3]). There is no doubt that such device simulation has aided the growing semiconductor industry by predictive analysis and even trouble-shooting of device design [4]. At present, even as the semiconductor integrated circuits continue to increase in complexity and density, the underlying devices continue to get smaller, faster and more complex in their behaviour. Therefore, it is expected that device

simulation will continue to be the most effective way of understanding how to design devices with desirable properties, at least for the foreseeable future.

The desirable electrical properties of semiconductor devices fall into the categories of steady-state DC, large-signal, small-signal, high frequency (including RF), noise and reliability, to name a few. For a review, please see [5]. The models used to simulate these properties are basically a set of partial differential equations (PDEs) that describe the transport of the charged carriers through the device. The solution of this set of equations depends on the specific structure and geometry of the device and the applied voltage.

However, when it comes to transport models, there is not one but a hierarchy of models that can be used to simulate the electrical properties of devices. They can broadly be classified in the following order, according to decreasing complexity:
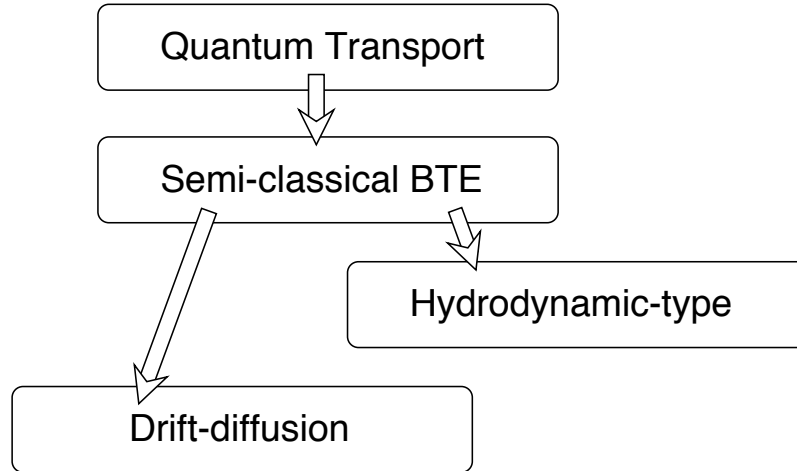
Fig. 1.1. Hierarchy of transport models.

The quantum transport equation and the semi-classical BTE are known as "kinetic models" because they describe the behaviour of the carriers in the entire state space (momentum, space and time). The drift-diffusion and hydro-dynamic equation are known as "balance" equations because they describe the behaviour of the ensemble of carriers only. Needless to say the more sophisticated set of equations (models) can capture complex transport effects more accurately than simple models, but at

the cost of greater computational burden and difficulty. The motivation for solving a higher-level transport model such as the BTE or quantum transport comes from the analysis of device designs currently being proposed or tested for IC production within the next 12 years [6]. Two examples can be given in this regard.

It has been known for a long time that carriers in the inversion layer of Si MOS-FETs must be confined in a region of < 50 Å thickness [7] and therefore must exhibit some quantisation effects. The interest in pursuing quantum models for Si transport was renewed when there were several reports of experimental observations of quantum confinement in MOSFETs inversion layers, particularly at low temperatures, low bias and under magnetic fields (review in [8]). These effects become stronger as the minimum device feature size shrinks and cannot be accounted for by classical models. Similarly, it is being increasingly reported that commonly-used drift-diffusion and hydrodynamic-type models are inadequate to describe nano-scale device operation [9, 10]. Recent experimental observations also demonstrate that when a reasonably conventional MOSFET design is scaled down to 30–60 nm, it operates in the quasi-ballistic regime [11]. So far, this effect can only be correctly described by kinetic transport models.

In general, the choice of a suitable model is determined by many factors including speed of solution, its numerical accuracy and its physical validity. Therefore, it is necessary to understand both the nature and range of transport models, in the context of the experimental observations, so that they can be applied correctly to device simulations. The following sections present a brief description of the quantum transport models, BTE, drift-diffusion and hydrodynamic models, with examples of their typical or most successful applications.

1.1.1 Quantum Transport

The most fundamental transport model of all is quantum transport. It is used to study the properties of variety of devices that exhibit "quantum" behaviour such as quantum interference, carrier confinement, tunnelling and single electron charging, to name a few. Such quantum devices can range from mesoscopic semiconductor

devices to superconductors, molecular conductors and carbon nanotubes. Quantum transport can describe the evolution of the trajectory of a carrier wave packet across the device under the influence of all possible interactions — coulombic, randomising and dissipative (scattering), and most importantly, phase interactions. In its full complexity, quantum transport is, indeed, a very difficult problem to solve.

Unlike the text book examples of the one-particle Schrödinger equation in the presence of a given independent Coulombic potential, quantum transport in semiconductors is a multi-particle interacting system with open boundaries and many types of scattering. It can only be done with great computational burden [12, 13] or by using approximations of limited applicability [14]. At present, many methods have been proposed to attack the problem of quantum transport in semiconductors, the most notable of which are Wigner transformations [15], Pauli Master Equation [16] and non-equilibrium Green's function method [17].

All the above methods are equivalent because they represent a different transformation of the same physical problem, but they differ in the ease with which they can include the effects of scattering, phase-breaking events and the transition from confined to unconfined and coherent to incoherent regions that must exist in any realistic semiconductor device. In this regard, the non-equilibrium Green's function (NEGF) formalism, or equivalently the Keldysh formalism, appears to be the most promising. It involves solving for the two-point correlation function or the Green's function in presence of all physical processes inside the device. The solution of the Green's function can then provide all the information about energy eigen-states, carrier concentration and current density from *ab initio* principles and statistical mechanics (for an excellent discussion, please see chapter 8 of [17]). The NEGF approach has been successfully applied to study quantum effects in III-V compound semiconductor devices [12, 18] but has only achieved limited success for traditional Si MOS devices, so far [19, 20].

1.1.2 Boltzmann Transport

The Boltzmann Transport Equation is a semi-classical approach to carrier transport. It describes the evolution of the trajectory of a particle by using a combination of Newtonian mechanics and quantum probabilistic scattering rates (Fermi-golden rule). The former account for the classical motion of the particle and the latter account for dissipative processes from one energy state to another. Unlike quantum transport, the energy eigen-states are not determined during the solution but are pre-computed by an independent method. The BTE can be derived from the quantum mechanical (Liouville-Von Neumann) transport equation under simplifying assumptions and ignoring all phase coherence [21, 22]. A good introduction to the BTE, its physical parameters and its application for device simulation can be found in [21].

Since the BTE does not include phase information, it is simpler to solve than quantum transport. It has been attacked by several methods of solution. Among the earlier approaches were approximate analytical methods based on the Legendre polynomial expansion [23]. These did not achieve much success because the drastic approximations used to simplify the problem and to obtain analytical solutions were valid only in the simplest cases and not for any practical devices. Other earlier methods were based on an iterative integration technique that appeared to work well only for low-field transport [24].

In the 1960s, a statistical integration method based on the Monte Carlo technique was suggested as a means to solve the BTE. It has achieved the most success among all other methods so far, due to its ease of programming, ease of including a variety of physical effects in the same framework, simple numerical algorithms and low memory requirements (review in [25]). The state-of-the-art Monte Carlo techniques can simulate transport in complicated device geometries with complicated band structures [26, 27]. However, the Monte Carlo technique suffers from several fundamental disadvantages *v.i.z.* statistical noise in low-bias near-equilibrium conditions and inability to handle rare events such as transport over a barrier and recombination-generation (although the latter are not thought to be critical in sub-micron devices). These con-

ditions involve events that occur at exponentially decreasing probabilities and cannot be detected by a stochastic method that has a well-known $1/\sqrt{N}$ convergence only for nearly uniform distributions. Some methods have been suggested to "enhance" the exponential tails of distribution functions so that they can be detected and that has provided some degree of respite [28]. However, a stochastic method inherently has lesser accuracy than a direct numerical method with controlled discretisation error.

Among other significant methods to solve the BTE, the Cellular Automata methods, the Scattering matrix method and the Spherical Harmonic method must be mentioned. For a more detailed review, please see [29].

1. Cellular Automata

   "A cellular automaton is a discrete dynamical system. Space, time, and the states of the system are discrete. Each point in a regular spatial lattice, called a cell, can have any one of a finite number of states. The states of the cells in the lattice are updated according to a local rule. That is, the state of a cell at a given time depends only on its own state one time step previously, and the states of its nearby neighbors at the previous time step. All cells on the lattice are updated synchronously. Thus the state of the entire lattice advances in discrete time steps" [30]. The BTE can also be viewed as a Cellular Automata problem [31, 32] by constructing cells momentum space and real space and ascribing a state (value of the distribution function) to each cell. The state of the cell is then updated in discrete time steps according to the above algorithm until the simulation is complete. The solution is, therefore, the value of the distribution function in all cells of the momentum and real space.

2. Scattering Matrix Approach

   This is also known as the *response* matrix approach. A scattering matrix of a given slab of semiconductor relates the out-going fluxes to the incident fluxes, due to the presence of all the scattering mechanisms and the electric field inside the slab [33, 34]. The flux in this case represents the flux carried by the distri-

bution function of the electrons. For better accuracy of the method, the flux (or equivalently the distribution function) is discretised in momentum space.

The advantage of this method is that the solution of the BTE easily be computed by breaking up the device into a series of slabs and then cascading the scattering matrices of the slabs together [35]. However, this requires that the scattering matrix for every slab must be available, where all slabs do not have the same size or scattering mechanisms and fields inside them. Hence a library of scattering matrices needs to be precomputed for slabs with all possible combinations of sizes, fields and scattering mechanisms. This has enormous memory requirements, especially if number of fluxes is large (or equivalently the discretisation in momentum space is fine). However, the size of the problem can be reduced using an exponential weighting in energy for the fluxes discretised in momentum space and a neat operator splitting technique [36].
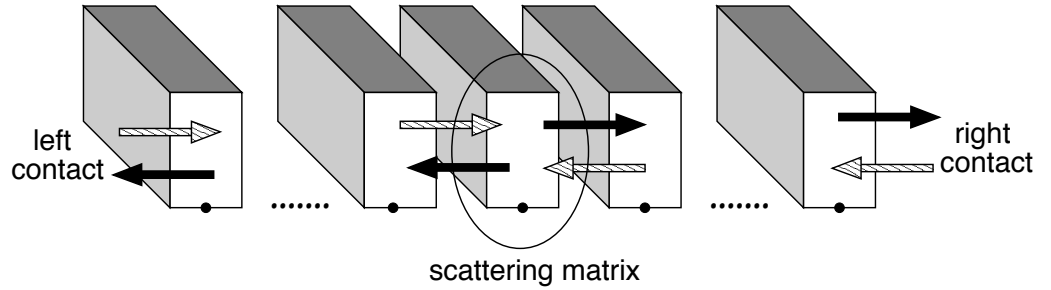
Fig. 1.2. Illustration of the scattering matrix approach in 1D. The solid arrows represent out-going fluxes and the hatched arrows represent incident fluxes.

The disadvantage of the scattering matrix method is that it gets more complicated in 2D and only square elements can be used in order to keep the size of the library small. Even then, the operator splitting technique does not generalise straightforwardly to 2D [29].

3. <u>Spherical Harmonic Method</u>

   The Spherical Harmonic Method is a method to solve the BTE by expanding the distribution function into a series of Spherical Harmonics. When the Spherical Harmonic expansion is substituted back into the BTE, the terms with the same Y_l^m can be separated on the basis of their orthogonality. This gives rise to a series of equations in terms of the spherical harmonics and can be solved for the coefficients, which are only functions of energy. This method is usually used with only the first two terms of the expansion [37, 38] which in general can be inadequate to describe highly asymmetric distribution functions (see Section 2.8)

4. <u>Relaxation time approximations</u>

   The relaxation time approximation (RTA) has been extensively studied since the 1950s [23] because it gives quick analytical solutions to the BTE that are approximately correct. The assumption in the RTA is that the collision integral can be represented by a single energy-dependent relaxation time, which is a severe approximation and is not valid for any realistic scattering mechanisms or devices (as discussed in [39]). However, it continues to be used to this day [40] because it is a simple and quick way to get approximate solutions to the BTE. This is a reasonable approach as long as the conclusions from the RTA analysis are not carried into the regime where they are not valid.

1.1.3 Drift-Diffusion and Hydrodynamic-type

Among all the above, drift-diffusion is the most basic of all device transport models. It was first proposed by van Roosbroeck[41] as a means of device simulation and a method to solve it for a practical device geometry was first demonstrated by Gummel [42]. Since then it has been extensively studied by mathematicians (for an excellent mathematical review, please see [43]) numerical analysts [44] and engineers [45], and is, by now, a fairly well-understood method. When the drift-diffusion equation is solved along with the current continuity equation for the particular conditions of the

device, it provides information about the carrier concentration and average velocity of the carriers. The parameters in the drift-diffusion model (such as mobility and diffusion coefficient) usually depend on the local electric field. In addition, transport effects due to changes in average energy of the carriers cannot be modelled within the framework of the drift-diffusion equations alone. This implies that drift-diffusion is valid only when the devices are fairly long and the electrical fields inside the device do not vary rapidly on the order of the mean free path [46]. As devices shrink, these conditions are no longer valid and the drift-diffusion equation can no longer provide an adequate description of the device operation.

The failure regime of the drift-diffusion model can be understood by examining its derivation from the BTE. The most common way to derive the drift-diffusion model is to take the first two moments of the BTE and truncate the set with an appropriate assumption [21]. There are other mathematical ways of deriving the drift-diffusion equation such as the Hilbert expansion of the BTE with certain assumptions to identify and retain leading order terms [47]. But the mathematical treatment essentially gives rise to the same set of the PDEs. The wealth of literature on the derivation of drift-diffusion and its solution is too great to go into detail in this report. However, an interested reader might find the references listed in the beginning of this section fairly comprehensive because they are written by experts in the field. As a side note, some researchers have also proposed an "augmented" drift-diffusion model, which is the same set of the PDEs as the drift-diffusion model but with an extra term related to the gradient of the field in the current equation [48]. Adding the extra term seemed to alleviate the restriction that the electric field must vary slowly on the order of the carrier mean free path and the augmented model could capture effects such as velocity overshoot [49]. However, it still lacked an appropriate energy flow model, which might account for the hot carrier effects that arise in submicron devices. This is the reason why augmented drift-diffusion is not commonly used for device simulation, at present.

The need to model hot carriers gave rise to the hydrodynamic-type models [50, 51].

They can be derived by taking the first four moments of the BTE and truncating the set with appropriate assumptions. When the hydrodynamic set of equations are solved along with the current continuity equation for the particular conditions of the device, they provides information average energy of carriers, average velocity and carrier density. Hydrodynamic models are gaining wider use these days but have not attained a status comparable to that of drift-diffusion because of several reasons:

1. The hydrodynamic equations are non-linear and belong the class of second-order PDEs known as "hyperbolic" [51]. This class of PDEs, particularly the non-linear class, is known to be numerically difficult to solve because it can give rise to non-smooth or discontinuous solutions (known as "shocks") even in the presence of smooth boundary conditions. Therefore, the numerical aspects of hydrodynamic problems are much more challenging than those of drift-diffusion [52].

2. The exact nature and values of the various parameters in the hydrodynamic model are still being debated (for example, see [53, 54]) and no consistent definition has been reached by the simulation community, so far. In general, the parameters of the hydrodynamic model depend on the distribution function which is the solution of the BTE. Since the solution of the BTE is not known, it is not possible to devise a set of consistent parameters that are valid for all conditions and some questionable assumptions need to be made. The most common practice is to "calibrate" the parameters by a comparison of simulations and measurements carried out on simple test devices.

3. Other versions of the hydrodynamic model have been proposed that neglect some non-linear terms and make the set of equations numerically more tractable [52, 55]. These are known as energy transport models and they belong to the elliptic class of the PDEs, similar to drift-diffusion. However, it is debatable whether neglecting the physics of the non-linear terms for the sake of numerical convenience is a valid proposition or not.

Therefore, the issue of the usefulness of hydrodynamic models is still a subject of some discussion [10, 56, 57].

1.2   The BTE for semiconductors

This dissertation is the first report of a direct numerical solution of the steady-state non-degenerate BTE in nanoscale semiconductor (Si) devices. The BTE here refers to the steady-state 6-dimensional integro-differential transport equation for semiconductors,

$$Hf - Q(f) = 0, \tag{1.1}$$

where

$$
\begin{aligned}
H &= H_r + H_k, \\
Q &= Q_{op} + Q_{ac} + Q_{imp} + Q_{ee} \dots .
\end{aligned}
\tag{1.2}
$$

H_r and H_k are first-order differential (also known as hyperbolic) operators, and Q is an integral collision operator. The BTE can describe the transport of either electrons or holes depending on the specific operators and parameters used. In this dissertation, only the BTE solution for electrons is discussed but transport for holes can be similarly treated.

The solution to the BTE is the distribution function of the carriers, $f(r, k)$, which describes the occupation of a state $\vec{k}$ at a position $\vec{r}$. Therefore, the solution belongs to the tensor product space $\mathbb{R}^3 \times \mathbb{R}^3$, where the former $\mathbb{R}^3$ denotes the real space of the device and the latter $\mathbb{R}^3$ denotes the momentum space (*i.e* the Brilliouin zone) of the semiconductor. The three terms of the BTE are the spatial term H_r, the field term H_k and the scattering (collision) term Q.

The first-order differential operators describe the electron motion according to Newton's laws:

$$
\begin{aligned}
H_r f &= \frac{1}{\hbar} \vec{\nabla}_k E(k) . \vec{\nabla}_r f, \\
H_k f &= \frac{e}{\hbar} \vec{\nabla}_r \phi(r) . \vec{\nabla}_k f,
\end{aligned}
\tag{1.3}
$$

where $E(k)$ is the energy-band structure of the semiconductor and $\phi(r)$ is the potential profile in the device.

The collision operator is an integral operator expressed as

$$Q_{scat}(f) = \int_{\mathbb{R}^3} S_{scat}(k, k')(1 - f(k))f(k') \ \mathrm{d}k' - \int_{\mathbb{R}^3} S_{scat}(k', k)(1 - f(k')) \ \mathrm{d}k' f(k),$$
(1.4)

where *scat* represents the type of scattering mechanism and could be any of the following: longitudinal optical phonons (*op*), longitudinal acoustic phonons (*ac*), impurities (*imp*), electron-electron (*ee*) etc. The collision operator sums over all the scattering mechanisms. In this work, only the non-degenerate version of the collision operator will be used. Non-degeneracy assumes that $f(k) \ll 1$, and allows Q to be written as

$$Q_{scat}(f) = \int_{\mathbb{R}^3} S_{scat}(k, k')f(k') \ \mathrm{d}k' - \int_{\mathbb{R}^3} S_{scat}(k', k) \ \mathrm{d}k' f(k).$$
(1.5)

This is typically not true in heavily-doped regions such as the source and drain of a MOSFET. In spite of that fact, the non-degenerate operator is used here because of the following reasons: including the degenerate operator makes the BTE non-linear and much more difficult to solve than the linear version and the error caused by neglecting degeneracy does not significantly alter the conclusions of the simulation studies done in this report.

1.2.1 Properties of the BTE

Much work has been done to understand the properties of the BTE [58, 59, 60] because that is the first step in finding a numerical solution of any PDE. In this work, several important properties of the BTE for semiconductors are invoked and are critical in the development of the numerical method. A summary of those properties is given below:

1. The BTE is a first-order PDE. For first-order PDEs, the concepts of "path integrals" or "characteristics" frequently arise in the solution. This makes the numerical solution of the BTE different from that of second-order PDEs found in most device simulation problems. However, even though this concept requires

special attention, the computation of the particle paths or trajectories, is, in practice, done straightforwardly by using "upwinding" techniques.

2. The BTE with a non-degenerate collision operator is linear.

3. The collision operator Q_{scat} is negative semi-definite for all types of scattering. This property arises directly from the very definition of scattering rates. In solving the BTE, the negative of the collision operator is used, $-Q_{scat}$, which is positive semi-definite. Positive definiteness is a favourable property when finding the numerical solution to any operator. The semi-definite aspect of this property implies that a null solution exists.

4. Momentum operators of the BTE conserve charge. This implies that:

$$
\begin{aligned}
\int_{\mathbb{R}^3} H_k f \ dk &= 0, \forall f \\
\int_{\mathbb{R}^3} Q_{scat}(f) \ dk &= 0, \text{ for all types of scattering and } \forall f.
\end{aligned} \tag{1.6}
$$

These conditions directly lead to current continuity in the BTE.

5. The null space, the span of independent null solutions, (also known as the *kernel*) of each operator of the BTE is a very important space. It has physical as well as mathematical significance. The physical significance is that given any distribution function, its eigenvector components will die out exponentially in time according to their eigenvalue and only the components of the null solutions will survive because their eigenvalue is 0. The mathematical significance is that a particular solution of the non-homogeneous PDE always contains some components that are solutions of the corresponding homogeneous PDE, *v.i.z.* null solutions.

   The definition of the null solution of an operator is a *non-zero* function that satisfies the homogeneous equation

$$
Rf = 0, \tag{1.7}
$$

   where R is any differential or integral operator.

6. All collision operators have collision invariants. An invariant is a quantity (such as charge, energy or momentum) that is preserved in the distribution function due to the scattering event and is defined as

$$\int_{\mathbb{R}^3} g(k) R f \; \mathrm{d}k = 0, \text{ for any operator } R \text{ and } \forall f, \tag{1.8}$$

where $g(k)$ is the invariant quantity such as 1, k or $E(k)$. Note that phonon collision operators have only charge invariance. Impurity or elastic collision operators have both charge and energy invariance. Electron-electron collision operators (not treated in this work) have all charge, energy and momentum invariance because they represent particle-particle (binary) collisions similar to those for gas molecules. The property of invariance is important because it gives rise to the space of null solutions of the collision operators.

7. The null solutions of non-degenerate phonon collision operators are given by the following conditions

$$\begin{aligned} \text{if } Q_{op}(f) &= 0 \text{ then } f = M(E), \\ \text{and if } Q_{ac}(f) &= 0 \text{ then } f = M(E), \end{aligned} \tag{1.9}$$

where $M(E)$ is a Maxwellian in energy. This is known as "detailed balance" and will be described in detail in Section 2.5. The null solution of elastic collision operators such as impurity scattering is given by the following condition

$$\text{if } Q_{imp}(f) = 0, \text{ then } f = f(E) \text{ only.} \tag{1.10}$$

8. The null solution of the differential operator H_k is 0.

9. The null solutions of the collision and field operators are the bulk solutions, *i.e.* distribution functions in a homogeneous semiconductor with uniform constant field

$$\text{if } H_k f - Q(f) = 0 \text{ then } f = f_{bulk}. \tag{1.11}$$

10. The null solution of only the *total* differential operator H is non-zero and is called the "ballistic" solution. This is expressed as:

$$\text{if } Hf = 0 \text{ then } f = f(E(k) - e\phi(r)). \tag{1.12}$$

This can be easily checked by substituting the above form of the solution into the left-hand side of the above equation. Physically this means that the ballistic solution of the BTE is a function of only the *total* energy $(E(k) - e\phi(r))$.

11. The zeroth moment of the distribution function is the carrier concentration:

$$\frac{1}{4\pi^3} \int_{\mathbb{R}^3} f(r,k) \, \mathrm{d}k = n(r). \tag{1.13}$$

12. The "first" moment of the distribution function is proportional to the current:

$$\frac{1}{4\pi^3} \int_{\mathbb{R}^3} \frac{1}{\hbar} \vec{\nabla}_k E(k) \, f(r,k) \, \mathrm{d}k = n(r)\langle \vec{v}(r) \rangle = -\frac{\vec{J}(r)}{e}, \tag{1.14}$$

where $\vec{\nabla}_k E(k)/\hbar$ is the group velocity in momentum space and $\langle \vec{v}(r) \rangle$ is the *ensemble* or *average* velocity of the carriers. Technically, the first moment of the distribution function is defined with respect to the momentum $\hbar\vec{k}$. In that case, the first moment need not be directly related to the current since the conversion between crystal momentum and velocity is not always linear. The relation between momentum and velocity is linear *only* for spherical parabolic bands. Hence, for the sake of consistency in this report, the first moment is defined as shown above.

13. The "second" moment of the distribution function is proportional to the average energy:

$$\frac{1}{4\pi^3} \int_{\mathbb{R}^3} E(k) f(r,k) \, \mathrm{d}k = n(r)\langle u(r) \rangle, \tag{1.15}$$

where $\langle u(r) \rangle$ is the *ensemble* or *average* energy of the carriers. Similar to the previous case, the second moment is technically defined with respect to the momentum k^2. In that case, the second moment need not be directly related to the average energy since the conversion between magnitude of k^2 and energy E

is not always linear. The relation between k^2 and E is linear *only* for spherical parabolic bands. Hence, for the sake of consistency in this report, the second moment is defined as shown above.

14. Just like any other PDE, the BTE has a unique solution for a given problem specification (device) only when the *correct* boundary conditions are defined. This requires that the boundary conditions for the device must themselves be solutions permitted by the BTE. This will be discussed in Sections 2.4 and 3.3.1.

1.3   Finite volume discretisation

This section describes some basic ideas behind the finite volume scheme used to discretise PDEs for numerical solution. There are two schemes that are commonly used to discretise PDEs for numerical solution. They are:

1. The finite element approach

   This approach is based on the variational or weak formulation of the PDE instead of the operational form. The solution is expressed as a linear combination of "basis" functions. Then the coefficients of the basis functions can be determined by using the Rayleigh-Ritz-Galerkin procedure on the PDE [61] . The result is a set of linear equations in terms of the coefficients that can be solved by any appropriate matrix solution technique. The advantages of the finite element approach are that there exist formal theories that can accurately estimate the discretisation error in the numerical solution with respect to the given PDE and the basis functions used. In fact, such error estimates are usually used in finite element grid refinement. Higher-order discretisations can be generated in finite element techniques simply by using higher-order basis functions.

   Although the finite element methods have been shown to work well for elliptic (second-order) PDEs such as the Laplace/Poisson equations, their disadvantage is that they have had less success with hyperbolic PDEs, strong convection-diffusion equations and upwinding in first-order PDEs. These problems can be

handled by some suitable modifications but at the expense of the power and generality of the method (see chapter 2 in [62]).

2. <u>The finite volume method</u>

This method is based on the operational form of the PDE and is closely related to the well-known finite difference method. Finite difference simply means replacing a differential with an appropriate difference operator. However, such a difference scheme can only be applied straightforwardly to rectangular grids. It cannot be applied so readily to triangular or arbitrary shaped grids. For such cases, the finite volume method was developed. Note that, this method is equivalently known as the "control volume" method and the "generalised box discretisation" method. In the finite volume method, the PDE is integrated over the arbitrary shaped volumes of the grid. The volume integral is converted to a surface integral by using Gauss' Law and the resulting expression is evaluated by finite difference (an example is shown below).

Similar to the finite element method, higher-order variants of the finite volume method are also available. But at the level of a first-order discretisation, it is well-known that the finite volume method and finite element method are *completely* equivalent [44] and there is a linear transformation between the two. Error estimates for finite volume methods cannot be derived as formally as those for finite element methods, but the advantage of finite volume is that it can be used where finite element might fail *v.i.z.* hyperbolic PDEs [51], drift-diffusion [62] and upwinding first-order PDEs (this work).

1.3.1 A finite volume example

For the reader not familiar with the finite volume discretisation, an example of the method applied to the Laplace equation is shown below. Consider the one-dimensional Laplace equation

$$\frac{\partial}{\partial x}\epsilon\frac{\partial}{\partial x}\phi = 0. \tag{1.16}$$

The application of the finite volume method proceeds according to the following three steps:

1. First a grid is generated in the problem domain (here $\mathbb{R}^1$). This step involves placing a finite number of nodes x_i, $i = 1 \ldots N$ appropriately inside the volume of the domain. This converts the unknown continuous function into a set of discrete unknowns given by

$$\phi(x = x_i) = \phi_i. \tag{1.17}$$

The common practice is to ensure that the nodes x_1, $x_2, \ldots, x_N$ are centroids of non-intersecting volumes Ω_1, $\Omega_2, \ldots, \Omega_N$ that cover the entire domain.

2. The given PDE (here eq. 1.16) is then integrated over each volume element,

$$\int_{\Omega_i} \frac{\partial}{\partial x} \epsilon \frac{\partial}{\partial x} \phi \; \mathrm{d}x = \int_{\Omega_i} 0 \; \mathrm{d}x = 0. \tag{1.18}$$

The idea here is to use mathematical or numerical manipulation to convert the above integral into an algebraic expression involving only the discrete values of the unknowns ϕ_i.

For the case of the above integral this can be easily done by invoking Gauss' Law. It converts the above volume integral to a surface integral as:

$$\int_{\Omega_i} \frac{\partial}{\partial x} \epsilon \frac{\partial}{\partial x} \phi \; \mathrm{d}x \;\; = \;\; \int_{\partial\Omega_i} \frac{\partial}{\partial x} \epsilon\phi \; \hat{x}.\hat{n} \; \mathrm{d}S$$

$$= \;\; \frac{\epsilon_{i+1}\phi_{i+1} - \epsilon_i\phi_i}{h_{i+1,i}} - \frac{\epsilon_i\phi_i - \epsilon_{i-1}\phi_{i-1}}{h_{i,i-1}} + \mathcal{O}(h^2), \quad (1.19)$$

where $h_{i+1,i} = x_{i+1} - x_i$. Note that finite *difference* is used to evaluate the differential in the last step. If the grid spacing is uniform, the local truncation error in this method is $\mathcal{O}(h^2)$.

3. The entire set of equation is then written down for each i as follows:

$$\frac{1}{h_{1,0}}\phi_0 - (\frac{1}{h_{1,0}} + \frac{1}{h_{2,1}})\phi_1 + \frac{1}{h_{2,1}}\phi_2 \qquad\qquad\qquad = \;\; 0,$$

$$\frac{1}{h_{2,1}}\phi_1 - (\frac{1}{h_{2,1}} + \frac{1}{h_{3,2}})\phi_2 + \frac{1}{h_{3,2}}\phi_3 \qquad\qquad = \quad 0,$$

$$\vdots$$

$$\frac{1}{h_{N-2,N-1}}\phi_{N-1} - (\frac{1}{h_{N,N-1}} + \frac{1}{h_{N+1,N}})\phi_N + \frac{1}{h_{N,N+1}}\phi_{N+1} \ = \ 0,$$

$$(1.20)$$

where ϕ_0 and ϕ_{N+1} are "ghost" points that don't really exist but may be used in case of Neumann boundary conditions. The final step is to include the boundary conditions. In case of Dirichlet boundary conditions, the value of ϕ_1 or ϕ_N is specified and the equation for the ghost point is dropped. However, in case of Neumann boundary conditions, the value of the derivative at the boundary is set to the specified value as follows:

$$\frac{\epsilon_{i+1}\phi_{i+1} - \epsilon_i\phi_i}{h_{i+1,i}} = \text{given} , \qquad\qquad (1.21)$$

where $i = 0$ or N (boundary node). Clearly the above set of equations is a matrix equation and can therefore be solved by any matrix solution technique. This solution gives the values of ϕ_i for the given problem and boundary conditions.

A critical aspect of the above method is the discretisation error. If the grid is coarse and its constituent volume elements are too large then the solution will not be very accurate. As the grid is refined and its elements are made smaller, the solution gets more accurate and the discretisation error decreases at the rate of h^2. This is the basic idea behind finite volume methods and will be used in this work.

1.4   Overview of the report

The main objective of this work is to demonstrate that with powerful numerical techniques, it is possible to solve the BTE directly without making any approximations about the angular shape of the distribution function or the collision integral. The techniques used here are not new but are well-established in the fields of computational science, particularly those related to numerical solutions of partial differential equations [63] and to numerical solutions of the Boltzmann Equation in neutron transport [64] and gas dynamics [65, 66]. However, this approach has never been used to

solve the BTE in semiconductors because of two reasons — the BTE in semiconductors is more complicated than that in neutron and gas transport and it is, at the very least, a 4–5 dimensional problem. The consequence of the above two factors is that a direct discretisation of the BTE usually results in a very large matrix equation with 10^6–10^7 unknowns. This work uses several key ideas to overcome the above difficulties. These include correct mathematical techniques for discretising the BTE and preconditioned iterative methods for solving large systems of equations, and will be described in detail in the course of this dissertation. A chapter-wise overview of this report is given below. Some background in device simulation is assumed on the part of the reader.

Chapter 2 describes the finite volume method used to discretise the BTE in momentum space $\mathbb{R}^3$. This involves computation of the discrete terms of the collision operator Q and the hyperbolic field operator H_k on any given grid in momentum space. Upon discretisation, the collision and field operators turn into corresponding collision and field matrices. In this process, it will be shown that particular attention needs to be paid to the conservation properties of the continuous operators so that these properties are preserved in their discrete versions. The momentum space discretisation is completed by imposing an appropriate asymptotic boundary condition on f in $\mathbb{R}^3$. The chapter ends with an analysis of the discretisation error in the solution with respect to the grid spacing in momentum space. This is estimated by comparing the discrete solutions with known analytical or independent solutions of the BTE.

Chapter 3 describes the finite volume discretisation of the hyperbolic spatial operator H_r in real space. At this stage, a one-dimensional device is used in order to demonstrate many important concepts of the numerical technique, such as imposing boundary conditions at the contacts of the device and setting up the system of equations as a matrix equation. It will be seen that the real space discretisation used in this method has a severe restriction on the grid spacing. But this restriction will be relaxed on the basis of an analysis of the discretisation error with respect to spatial

grid spacing.

Chapter 4 takes a small diversion from device physics in order to describe some of the iterative methods used to solve large linear systems of equations. This is done because the number of variables in solving the BTE can be as large as 10^6–10^7 and special fast iterative methods are required to solve this system. The choice of iterative method used in this work is described and justified. Using the concepts developed in this chapter, several important issues of numerical computation are addressed, such as, the condition number of the matrix and the best criterion for measuring the convergence of an iterative method, from both a physical and mathematical point of view. The first section in this chapter can be skipped by experts familiar with the field of numerical analysis and matrix computations.

Chapter 5 returns to the numerical solution of the BTE in devices. It first addresses the issue of solving a transport equation self-consistently with the Poisson equation for device simulations. In this regard, the properties of coupling a BTE solver to a (non-linear) Poisson solver are discussed *v.i.z.* stability, overall convergence of the BTE-Poisson loop and speed. In addition, this chapter also describes the application of the numerical technique developed in this work, to study transport in realistic nano-scale devices. Two devices are used in this regard — a one-dimensional 50nm n^+-p-n^+ diode and a two-dimensional 50nm ultra-thin body dual-gate nMOS-FET. A comparison of the solution of the BTE in these devices with the solutions from commonly-used macroscopic transport models shows that the latter can produce unphysical results because they are no longer applicable at nanometre length scales.

Finally, Chapter 6 summarises the important conclusions of this work and addresses the issues that remain for potential further development of this work.

2. Momentum space discretisation

2.1 Introduction

This chapter describes the finite volume method used to discretise the BTE in momentum space, $\mathbb{R}^3$. It involves computation of the discrete terms of the collision operator Q and the hyperbolic field operator H_k on any given grid in momentum space. Real space discretisation is treated separately and independently in the next chapter. This is done because of the following reasons. The discretisation in momentum space depends on the energy-band structure of the semiconductor and the type of collision mechanisms — properties which are, in general, completely independent of the type of real space grid in semi-classical devices. In addition, the strict conservation properties of momentum space operators must be preserved in order to reproduce experimentally observed material properties of the semiconductor (*i.e.* bulk mobilities and average energies [67]). Therefore, the discretisation in momentum space is a critical part of solving the BTE in semiconductors. The organisation of this chapter is as follows.

Section 2.3 describes the finite volume method used to discretise the collision integral on a grid in momentum space. In this regard, particular attention is paid to the energy, momentum and mass conservation properties of the collision operator.. Upon discretisation, the collision operator turn into corresponding collision matrix. This section ends with a list of collision mechanisms used in this report.

The above discretisation in momentum space is not complete because solutions cannot be found without a boundary condition. Hence Section 2.4 describes the imposition of an appropriate asymptotic boundary condition on the distribution function.

Section 2.5 describes the method used to estimate the discretisation error in the computed collision terms. This is done by comparing the discrete null solutions

of the collision matrices with the corresponding analytical solutions, known to be equilibrium Maxwellian distributions. This comparison shows that the discretisation error in the solution tends to zero as the grid spacing is reduced. The conclusion here is that it is not possible to reduce the error (grid spacing) all the way to zero and hence, it is only practical to use a grid spacing that gives a reasonably small error.

Section 2.6 describes the finite volume method used for discretising the field term H_k in momentum space. This is done by using upwinding that transforms the hyperbolic operator into a positive-definite matrix and preserves the stability of the discretisation.

Section 2.7 estimates the discretisation error in the field terms. This is done by computing the null solutions (also known as bulk solutions) of the collision and field matrices for different values of field. The average velocity and energy of these bulk distributions are compared with those from Monte Carlo simulations with the same parameters. This analysis shows that average quantities from these solutions approach their correct values as the grid spacing is reduced.

Section 2.8 compares the numerical method developed so far to a popular approximate analytical technique known as the Legendre polynomial expansion method. This is done by computing bulk solution from the two methods for the same conditions of scattering and band parameters. This analysis shows that the numerical method developed has potential for greater accuracy than the Legendre polynomial expansion.

Section 2.9 summarises the conclusions of this chapter.

2.2   Momentum space grid

The first step of the finite volume method is to generate a grid in the domain of the PDE, here $\mathbb{R}^3$ space of the Brilliouin zone. Grid generation and refinement are very important aspects of finding numerical solutions to a PDE but a detailed discussion of gridding algorithms is beyond the scope of this report. For the purpose of this work, simple physically-intuitive gridding algorithms will be used.

The energy band structure used in this work is a spherical non-parabolic type with

parameters calibrated to reproduce the average velocities and energies in Si shown in Table. 2.1. For details of the calibration please see Appendix A. The discretisation of the band-structure is done both in energy and angle, hence this method is not restricted to any specific band-structure and can also be applied to other ways of gridding the Si band-structure [68]. This method is also not restricted to a single energy band and can be generalised to multiple bands by extending the ideas shown here.

Parameter	Value
α	0.5 eV^{-1}
m_0	9.1×10^{-31} kg
m_{eff}^*	$0.26 \times m_0$
E_{max}	1.6 eV
ρ	2.329 kg/cm^3

Table 2.1
Energy band parameters used in this work.

The momentum space grid used in this work is shown in Fig. 2.1. It can be described as an "energy contour" mesh because it is built along the constant energy surfaces of the band-structure. The volume elements of this grid are constructed by partitioning the spherical shells between equally spaced constant energy surfaces into equally spaced angular segments. This is also known as the discrete ordinates method in neutron transport [64].

This is known as a spherical hexahedral grid whose volume elements are completely determined by three parameters: the energy spacing ΔE, the angular spacing $\Delta \theta$ and the angular spacing $\Delta \phi$. The grids used for the purposes of analysis in this work are tabulated in Table 2.2, in terms of their parameters ΔE, $N_\theta = \pi/\Delta\theta$ and $N_\phi = 2\pi/\Delta\phi$. The number of volume elements in each grid is $N_k = N_E \times N_\theta \times N_\phi$.
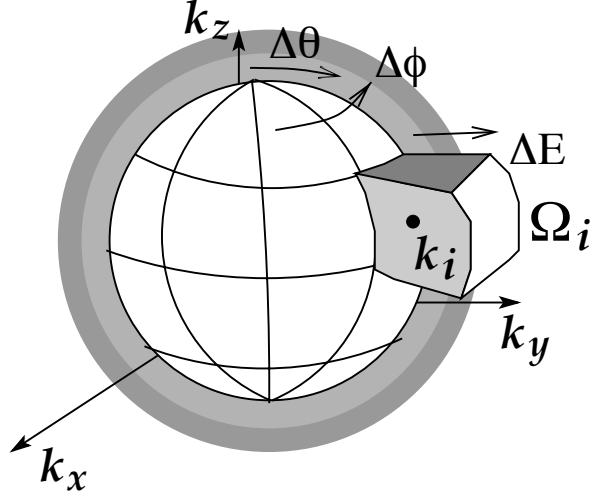
Fig. 2.1. Illustration of energy contour mesh used in this work.

| N_θ, N_ϕ | 4,4 | 8,8 |
ΔE		
0.1 eV		1088
0.05 eV	528	2112
0.025 eV		4160

Table 2.2

Tabulation of grid sizes (N_k) used in this work.

Now the unknown distribution function needs to be solved at the nodes $k_i \in \mathbb{R}^3$ that are centroids of the above volume elements Ω_i. This yields a discrete set of unknowns

$$f(k = k_i) = \tilde{f}_i, \ \forall \Omega_i. \tag{2.1}$$

The next step is to apply the control volume integration to the BTE.

2.3 Discretising the collision integral

This section describes the finite volume discretisation of the collisional form of the BTE

$$-Q(f) = 0, \tag{2.2}$$

where Q sums over all the types of scattering represented generically by Q_{scat}. Note that this is not the complete BTE. This is only the BTE describing transport in a uniform homogeneous semiconductor with no electric field. The remaining terms of the BTE will be added successively and in a physically meaningful manner, in the following sections and chapters.

Consider any one type of non-degenerate scattering operator Q_{scat} in eq. 2.2. The finite volume method begins by integrating the BTE over each element Ω_i of the momentum space grid as follows:

$$-\int_{\Omega_i} Q_{scat}(f) \, \mathrm{d}k = 0, \quad \forall \, \Omega_i. \tag{2.3}$$

Using the full form of $Q_{scat}(f)$, as shown in eq. 1.5, the above integral becomes

$$-\int_{\Omega_i} \left(\int_{\mathbb{R}^3} S_{scat}(k, k') f(k') \, \mathrm{d}k' - \int_{\mathbb{R}^3} S_{scat}(k', k) \, \mathrm{d}k' f(k) \right) \, \mathrm{d}k = 0, \quad \forall \, \Omega_i. \tag{2.4}$$

Note that the first integral on the left-hand side of eq. 2.4 refers to *in*-scattering rates and the second integral refers to *out*-scattering rates. The particular form of the scattering rates $S_{scat}(k, k')$ depends on the type of scattering mechanism. It is not the intention of this report to derive the scattering rates *ab initio* but only to use them in their commonly accepted forms. A discussion on the derivation of scattering rates from the Fermi-Golden rule and other related concepts can be found in specialised texts such as [21, 25, 69] and the references therein.

The integrals on the left-hand side of eq. 2.4 now need to be converted to an algebraic expression in terms of the discrete unknowns $\tilde{f}_i$. Note however that Gauss' Law cannot be used here because they are integrals and not differential operators. However, a technique for approximating integrals by sums can be used instead. In order to illustrate this, a more detailed but general form of the in-scattering rates is required and that is shown below:

$$\int_{\mathbb{R}^3} S_{scat}(k, k') f(k') \, \mathrm{d}k' = \int_{\mathbb{R}^3} |M_{scat}(q)|^2 \, \delta(E(k) - E(k') \mp \hbar w(q)) \, f(k') \, \mathrm{d}k',$$

$$\tag{2.5}$$

where *scat* represents a particular type of scattering mechanism, $q = \pm(k - k')$ and $\hbar w(q)$ is the dispersion relation of the scattering mechanism. This expression represents both phonon (optical and acoustic) and elastic (impurity) in-scattering rates, in their most general form. It does not represent electron-electron scattering quite as well, but *ee* scattering will not be treated in this report. The notation of the in-scattering rates $S(k, k')$ as scattering *from k' to k* is commonly used and has numerical significance because it refers to the *row* elements of collision matrix (shown later).

The out-scattering rates have a similar form. However in this report, the out-scattering terms will be computed from the in-scattering terms by invoking a relation between them which is derived from the the property of conservation of charge (property 4 in Section 1.2.1).

2.3.1 In-scattering terms

Applying the control-volume integration to the in-scattering rates gives

$$\int_{\Omega_i} \int_{\mathbb{R}^3} S_{scat}(k, k') f(k') \, dk' \, dk$$

$$= \int_{\Omega_i} \int_{\mathbb{R}^3} |M_{scat}(q)|^2 \, \delta(E(k) - E(k') \mp \hbar w(q)) \, f(k') \, dk' \, dk,$$

$$= \int_{\Omega_i} \int_{\mathbb{R}^3} |M_{scat}(q)|^2 \, \delta(E(k) - E(k \mp q) \mp \hbar w(q)) \, f(k \mp q) \, dq \, dk,$$

$$= \int_{\Omega_i} \int_{\Gamma(k \mp q)} |M_{scat}(q)|^2 \, \frac{1}{|\nabla_q(E(k \mp q) \pm \hbar w(q))|} \, f(k \mp q) \, dS \, dk, \qquad (2.6)$$

where the original form of the integral is transformed *twice*.

The first transformation converts the inner integral from that over the initial state k' to that over q, the change in momentum. $|M_{scat}(q)|$ is the matrix element obtained from Fermi Golden Rule that describes the overlap integral between the initial and final state and generally depends on the magnitude q for the mechanisms under consideration in this report. The delta function imposes the conservation of energy and

momentum. It is this delta function that makes collision integrals numerically difficult to evaluate [70]. The delta function cannot be either neglected or approximated because energy and momentum conservation is a critical aspect of scattering (see Section 2.5).

Therefore, the second transformation in eq. 2.6 transforms the integral over the delta function into a surface integral with a factor accounting for density of states on the surface (see Chapter 8 of [71]). The inner integration is done over $\Gamma(k \mp q)$ which is the surface of all possible states $(k \mp q)$ that can scatter into the state k by conserving both energy and momentum

$$\Gamma(k \mp q) = \{given\ k, \text{all } q \text{ s.t. } (E(k) - E(k \mp q) \mp \hbar w(q)) = 0\}. \qquad (2.7)$$

The outer integral is left unchanged. This procedure effectively converts the 6-dimensional integral into a 5-dimensional integral by taking the delta function *exactly* into account. The following figures illustrate the scattering surfaces for phonon and impurity collision mechanisms.

Fig. 2.2 illustrates the surfaces of all states k' that can scatter into state k due to optical phonon scattering, for spherical energy bands and a constant optical phonon dispersion relation. Fig. 2.2(a) shows all states that can emit a phonon of energy $\hbar\omega_{op}$ and scatter into k. Fig. 2.2(b) shows all states that can absorb a phonon of energy $\hbar\omega_{op}$ and scatter into k. These surfaces are computed from the energy and momentum conservation rule for optical phonon scattering. Note that optical phonon scattering surfaces are also spherical because they are constant energy surfaces $E(k) \mp \hbar\omega_{op}$. This makes it relatively easy to numerically compute the surfaces and perform the surface integral in eq. 2.6.

Similarly, Fig. 2.3 illustrates the surfaces of all states k' that can scatter into state k due to acoustic phonon scattering, for spherical energy bands and a *linear* acoustic phonon dispersion relation.
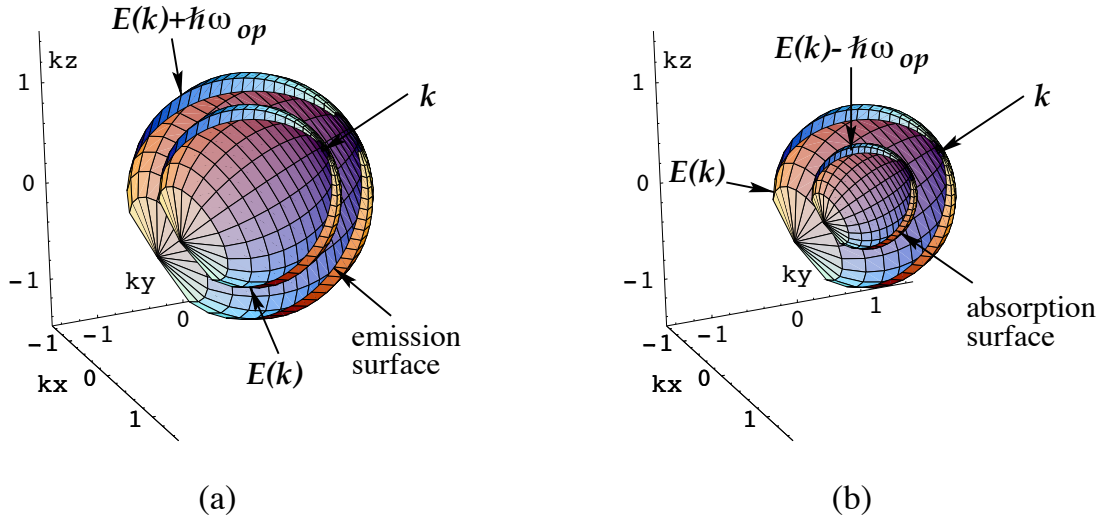
Fig. 2.2. Cut-away of spherical surfaces for optical phonon (a) emission and (b) absorption.
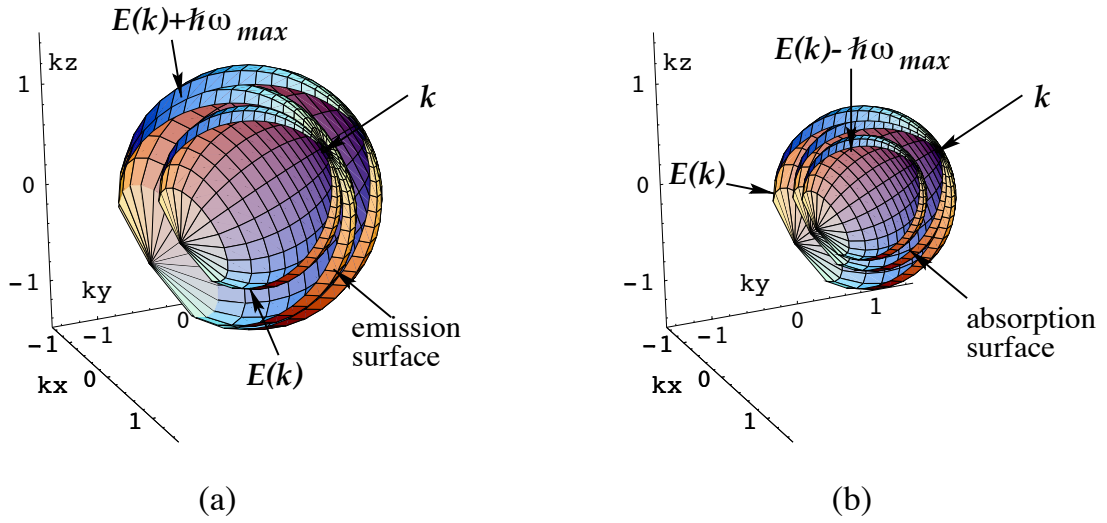
Fig. 2.3. Cut-away of distorted "spherical" surfaces for acoustic phonon (a) emission and (b) absorption.

The acoustic phonon surfaces are, in general, distorted because their dispersion relation is linear and is not approximated to be constant or elastic. This distortion makes the computation of the acoustic phonon surface integral in eq. 2.6 non-trivial. Finally for the case of elastic impurity scattering, the in-scattering and out-scattering surfaces are the constant energy surfaces themselves because energy does not change during an elastic event. Elastic surface integrals are not difficult to compute.

Going back to eq. 2.6, note that the integral still needs to be expressed in terms of the discrete unknowns $\tilde{f}_i$. The integral only refers to $f(k \mp q)$, which is the value of f on the scattering surface $\Gamma(k \mp q)$. But the value of f cannot be found or known everywhere, and so it must be approximated in some way. This is known as a "quadrature rule" in the language of numerical integration. In this work, the following simple quadrature rule is used:

$$\text{if } (k \mp q) \in \Omega_{i'}, \quad f(k \mp q) = \tilde{f}_{i'}. \tag{2.8}$$

This reduces to the familiar trapezoidal rule and is only of the many quadratures that can be used for numerical computation of integrals. Other higher-order quadratures may also be used (such as the Simpson's rule) but the trapezoidal rule turns out to be physically intuitive and reasonably accurate for the purposes of this work.

Using the above quadrature in eq. 2.6 gives

$$\int_{\Omega_i} \int_{\mathbb{R}^3} S_{scat}(k, k') f(k') \, dk' \, dk$$

$$= \int_{\Omega_i} \int_{\Gamma(k \mp q)} |M_{scat}(q)|^2 \frac{1}{|\nabla_q(E(k \mp q) \pm \hbar w(q))|} f(k \mp q) \, dS \, dk,$$

$$= \int_{\Omega_i} \left(\sum_{i'} \int_{\Gamma(k \mp q) \cap \Omega_{i'}} |M_{scat}(q)|^2 \frac{1}{|\nabla_q(E(k \mp q) \pm \hbar w(q))|} \tilde{f}_{i'} \, dS \right) dk + \mathcal{O}(h_k^n),$$

$$\approx \sum_{i'} \left(\int_{\Omega_i} \int_{\Gamma(k \mp q) \cap \Omega_{i'}} |M_{scat}(q)|^2 \frac{1}{|\nabla_q(E(k \mp q) \pm \hbar w(q))|} \, dS \, dk \right) \tilde{f}_{i'} \quad ,$$

$$\equiv \sum_{i' \neq i} \tilde{Q}_{scat}(i, i') \tilde{f}_{i'}, \tag{2.9}$$

where the $\tilde{f}_{i'}$ terms are moved outside the integrals because they are just a number and the $\mathcal{O}(h_k^n)$ terms are neglected. This is the error made when discretising the collision integral and is usually on the order of the grid spacing h_k^n, where $n \approx 2$, as will be shown in Section 2.5.

The discrete in-scattering rates $\tilde{Q}_{scat}(i, i')$, are defined as the complicated 5-dimensional integrals

$$\tilde{Q}_{scat}(i, i') \equiv \int_{\Omega_i} \int_{\Gamma(k \mp q) \cap \Omega_{i'}} |M_{scat}(q)|^2 \frac{1}{|\nabla_q(E(k \mp q) \pm \hbar w(q))|} \, dS \, dk. \tag{2.10}$$

They refer to the in-scattering rate from $\Omega_{i'}$ to Ω_i. Note that the terms for which $i' = i$ are *dropped* from the sum in eq. 2.9. This is not because terms for which $i' = i$ are zero. In fact, for very small values of q, it is possible that scattering into a state inside an element may occur from a state within the *same* element. In such a case, the integral for $i' = i$ in eq. 2.10 evaluates to a small but finite number. However, careful (but not obvious) book-keeping shows that, *within* the accuracy of the discretisation, these terms do *not* contribute to the collision integral because they are cancelled by the out-scattering rates between states inside the same element.

The index (i, i') comes about naturally and refers to the elements of the ith row of the collision matrix. Therefore, in-scattering rates are also referred to as the *off-diagonal* terms of the collision matrix. In addition, if $\Gamma(k \mp q) \cap \Omega_{i'}$ is empty, then $\tilde{Q}_{scat}(i, i') = 0$. This makes sense physically because if the scattering surface does not intersect the volume element $\Omega_{i'}$, then there is no event that can scatter an electron from i' to i. In fact, most of the numerical work in evaluating $\tilde{Q}_{scat}(i, i')$ is expended in computing the intersection of the scattering surface $\Gamma(k \mp q)$ with the volume elements $\Omega_{i'}$ of the grid. This can get quite complicated for full-band structures, *Umklapp* processes and higher-order evaluations of eq. 2.9. But although the mathematical aspects can get challenging, the terms $\tilde{Q}_{scat}(i, i')$ actually have a very simple physical interpretation:

$$\tilde{Q}_{scat}(i, i') = \Big(\text{average } |M_{scat}(q)|^2 \times \text{density of states on the}$$

$$\text{average surface } \Gamma(k \mp q) \cap \Omega_{i'}) \times (\text{average number of states in } \Omega_i),$$

$$(2.11)$$

which approaches the correct definition of $S_{scat}(k, k')$ as the volume elements $\Omega_i \rightarrow 0$. This physical interpretation is, in fact, used in detailed Monte Carlo methods as an algorithm for selecting final states with random numbers after a scattering event [27].

The accuracy of the $\tilde{Q}_{scat}(i, i')$ terms depends on the accuracy of the numerical computation of the 5-dimensional integral. This work again uses a trapezoidal rule for computing the collision terms with a very fine spacing that is much smaller than the spacing of the underlying momentum space grid. Hence the most dominant component of the discretisation error does not turn out to be the accuracy of the collision terms themselves but the spacing of the underlying momentum space grid. This is because of the approximation made in eq. 2.8, as it will be shown in Section 2.5.

2.3.2   Out-scattering terms

In this section, the out-scattering terms are computed. This is done by using the property of charge conservation, described in Section 1.2.1, which states that for any f,

$$\int_{\mathbb{R}^3} Q_{scat}(f) \, \mathrm{d}k = 0,$$

$$\therefore \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} S_{scat}(k, k') f(k') \, \mathrm{d}k' \, \mathrm{d}k = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} S_{scat}(k', k) \, \mathrm{d}k' f(k) \, \mathrm{d}k.$$

$$(2.12)$$

The above integral over the entire momentum space can be expressed as the sum of integrals over its volume elements Ω_i

$$\sum_i \int_{\Omega_i} \sum_{i'} \int_{\Omega_{i'}} S_{scat}(k, k') f(k') \, \mathrm{d}k' \, \mathrm{d}k - \sum_i \int_{\Omega_i} \int_{\mathbb{R}^3} S_{scat}(k', k) \, \mathrm{d}k' f(k) \, \mathrm{d}k = 0, \quad (2.13)$$

for any f.

Using the definition of in-scattering rates developed in the previous section, the first term can be expressed as

$$\sum_i \int_{\Omega_i} \sum_{i'} \int_{\Omega_{i'}} S_{scat}(k, k') f(k') \ \mathrm{d}k' \ \mathrm{d}k = \sum_i \sum_{i' \neq i} \tilde{Q}_{scat}(i, i') \tilde{f}_{i'}. \tag{2.14}$$

This leads to the *definition* of the second term (out-scattering rates) as

$$-\sum_i \int_{\Omega_i} \int_{\mathbb{R}^3} S_{scat}(k', k) \ \mathrm{d}k' f(k) \ \mathrm{d}k \equiv \sum_i \tilde{Q}_{scat}(i, i) \tilde{f}_i \tag{2.15}$$

where $f(k)$ is approximated in the same manner as it was for the in-scattering rates (eq. 2.8) and therefore, the discretisation error is the same. Note that $\tilde{Q}_{scat}(i, i)$ does not describe scattering from element Ω_i to itself, instead it describes out-scattering from element Ω_i into *all other* elements. The notation (i, i) only refers to its diagonal position in the collision matrix. Therefore the out-scattering terms are also referred to as the diagonal terms of the collision matrix.

Using eqs. 2.14 and 2.15 in eq. 2.13 gives

$$\sum_i \sum_{i' \neq i} \tilde{Q}_{scat}(i, i') \tilde{f}_{i'} + \sum_i \tilde{Q}_{scat}(i, i) \tilde{f}_i = 0, \tag{2.16}$$

for any $\tilde{f}_i$. Now interchanging the summation variables in the first summation and taking out the common factor of $\tilde{f}_i$, the above equation reduces to

$$\sum_i \left[\sum_{i' \neq i} \tilde{Q}_{scat}(i', i) \right] \tilde{f}_i + \sum_i \tilde{Q}_{scat}(i, i) \ \tilde{f}_i, = 0, \quad \forall \tilde{f}_i,$$

$$\therefore \tilde{Q}_{scat}(i, i) = -\sum_{i' \neq i} \tilde{Q}_{scat}(i', i). \tag{2.17}$$

This is the property of conservation of charge as applied to discrete collision matrices. It states that the diagonal out-scattering term must be negative of the sum of all other elements in its column or in other words, the sum of every *column* of the collision matrix must be 0. Eq. 2.17 is, in fact, used in this work to construct out-scattering terms from the in-scattering terms. Hence, all the collision matrices are guaranteed to conserve charge directly by their construction.

There are some important points to note here. Eq. 2.17 is also a physically expected result, because the sum of in-scattering rates into all possible states ($\Omega_{i'}$) from

an initial state (Ω_i) must be the same as the total out-scattering rate from the initial state. This property requires that the column sum of the collision matrices be 0, implying that at least one row of the collision matrix depends linearly on all the other rows, and hence every collision matrix must have at least one null solution. Also note that eq. 2.17 has nothing to do with "detailed balance", which actually describes the *form* of the null solution for the phonon collision matrices (Sec.2.5).

2.3.3 The complete collision integral

To summarise, the in-scattering terms are computed from 5-dimensional integrals by taking the energy and momentum conservation exactly into account,

$$\tilde{Q}_{scat}(i,i') = \int_{\Omega_i} \int_{\Gamma(k\mp q)\cap\Omega_{i'}} |M_{scat}(q)|^2 \; \frac{1}{|\nabla_q(E(k\mp q)\pm\hbar w(q))|} \; \mathrm{d}S \; \mathrm{d}k.$$

The out-scattering terms are computed in-directly from the in-scattering terms using the property of conservation of charge,

$$\tilde{Q}_{scat}(i,i) = -\sum_{i'\neq i} \tilde{Q}_{scat}(i',i).$$

In this procedure, the positive semi-definite property of $-Q_{scat}$ is retained in its discrete version $-\tilde{Q}_{scat}$.
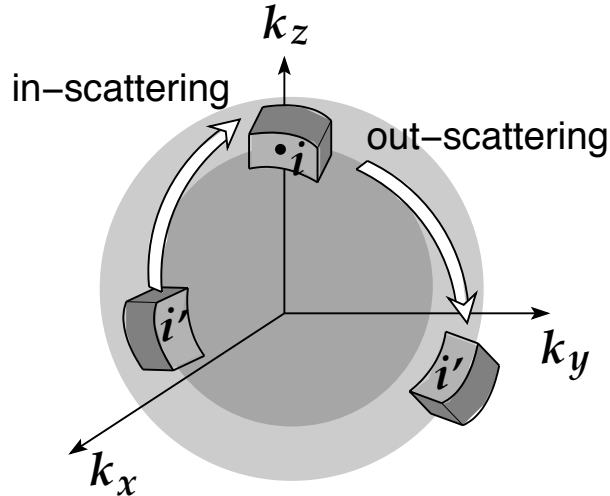
Fig. 2.4. Physical meaning of the terms of the collision matrix $\tilde{Q}$.

Putting them together, the control volume integration of the total collision operator becomes

$$-\int_{\Omega_i} Q_{scat}(f) \, \mathrm{d}k = - \left(\sum_{i' \neq i} \tilde{Q}_{scat}(i, i') \tilde{f}_{i'} + \tilde{Q}_{scat}(i, i) \tilde{f}_i \right), \qquad (2.18)$$

which is simply the ith element of a matrix-vector multiplication. Therefore, the matrix-vector product $\tilde{Q}_{scat}\tilde{f}$ will henceforth refer to the above expression (eq. 2.18). The size of the matrix $\tilde{Q}_{scat}$ is $N_k \times N_k$ but it is very sparse because scattering events can only connect momentum space elements that differ by one phonon energy. The vector $\tilde{f}$ is $N_k \times 1$ because it represents the distribution function being solved at the given nodes in momentum space.

An important point to note here is that although the *units* of Q are 1/s, the units of $\tilde{Q}$ are $1/(\text{cm}^3\text{s})$. The continuous distribution function $f(k)$ and its discrete version $\tilde{f}$ are both dimensionless.

2.3.4   List of collision mechanisms

The list of collision mechanisms used in this work and their properties [21] are given below:

1. Longitudinal optical phonons

   The dispersion relation for longitudinal optical phonons used here is

   $$\hbar\omega_{op}(q) = \hbar\omega_{op} = 0.062\text{eV}, \qquad (2.19)$$

   similar to previous works [25, 26]. Here q is the optical phonon wave-vector and $\hbar\omega_{op}$ is the optical phonon energy. The matrix element for the optical phonon scattering rates is

   $$|M_{op}(q)|^2 = \frac{1}{4\pi^2} \frac{D_{op}^2}{\rho\omega_{op}} \left(N(\hbar\omega_{op}) + \frac{1}{2} \pm \frac{1}{2} \right), \qquad (2.20)$$

   where D_{op} is the optical phonon deformation potential, ρ is the density of Si and $N(\hbar\omega_{op})$ is the phonon number.

2. Longitudinal acoustic phonons

   The dispersion relation for longitudinal acoustic phonon used here is the linear

approximation

$$\hbar\omega_{ac}(q) = \hbar v_\ell |q|, \tag{2.21}$$

where q is the acoustic phonon wave-vector, and $\hbar\omega_{ac}$ is the acoustic phonon energy and v_ℓ is the velocity of sound in Si, taken to be 9.04×10^5 cm/s. The matrix element for the optical phonon scattering rates is

$$|M_{ac}(q)|^2 = \frac{1}{4\pi^2} \frac{|q|^2 D_{ac}^2}{\rho\omega_{ac}(q)} \left(N(\hbar\omega_{ac}(q)) + \frac{1}{2} \pm \frac{1}{2} \right), \tag{2.22}$$

where D_{ac} is the acoustic phonon deformation potential and $N(\hbar\omega_{ac}(q))$ is the phonon number.

Parameter	Value
E_{op}	0.062 eV
D_{op}	15.56×10^8 eV/cm
$E_{ac,max}$	0.0435 eV
v_ℓ	9.04×10^5 cm/s
D_{ac}	8.4 eV

Table 2.3
Phonon scattering parameters used in this work.

Table 2.3 lists the phonon scattering parameters used in this work. These are used to compute the LO and LA collision matrices by following the procedure described in the previous sections. The out-scattering rates for each scattering mechanism can be computed from its corresponding collision matrix by dividing the diagonal terms with the volume of the element Ω_i. The result is plotted versus energy in Fig. 2.5. Note that the LO and LA scattering parameters in this work have been chosen such that their resulting out-scattering rates are the same as the phonon out-scattering rates used in Monte Carlo.
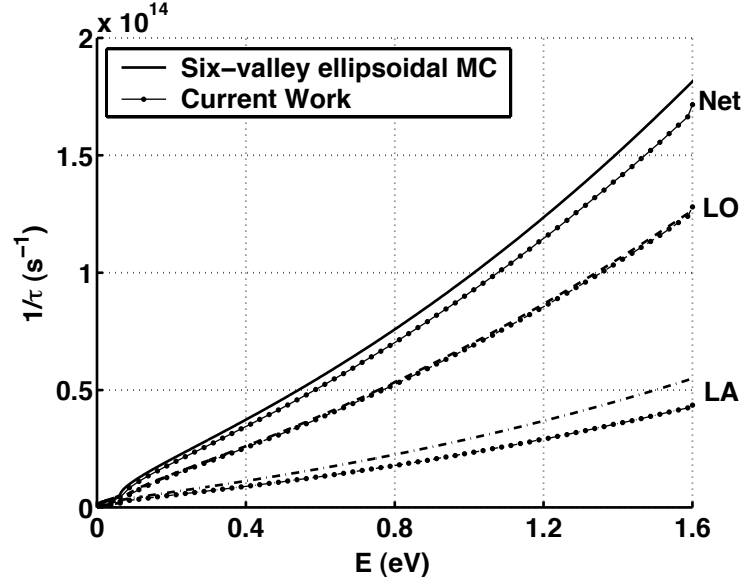
Fig. 2.5. Comparison of LO and LA out-scattering rates used in this work with Monte Carlo.

3. <u>Impurity scattering (Ridley approach)</u>

Impurity scattering is an elastic scattering mechanism and does not change the energy of the carriers, therefore,

$$\hbar\omega(q) = 0, \ \text{although} \ q \neq 0. \tag{2.23}$$

Several methods have been developed to describe the matrix element for impurity scattering rates [72, 73], but the Ridley approach is generally considered to be the most applicable because it can take into account both the screening by ionised impurities and free carriers.

The matrix element for impurity scattering in the Ridley third-model exclusion model [74] for spherical bands is given by

$$|M_{imp}(q)|^2 = \frac{1}{4\pi^2\hbar} \frac{Z^2 e^4}{\epsilon_{si}^2 \epsilon_0^2} \frac{N_I}{\left(|q|^2 + \frac{1}{L_D^2}\right)^2} \exp\left(-\frac{Z^2 e^4 N_I^{2/3} m_{eff}^{*\,2}\left(1 - \frac{|q|^2}{4|k|^2}\right)}{\pi\hbar^4\epsilon_{si}^2\epsilon_0^2\left(|q|^2 + \frac{1}{L_D^2}\right)\left(4|k|^2 + \frac{1}{L_D^2}\right)}\right)$$
$$\tag{2.24}$$
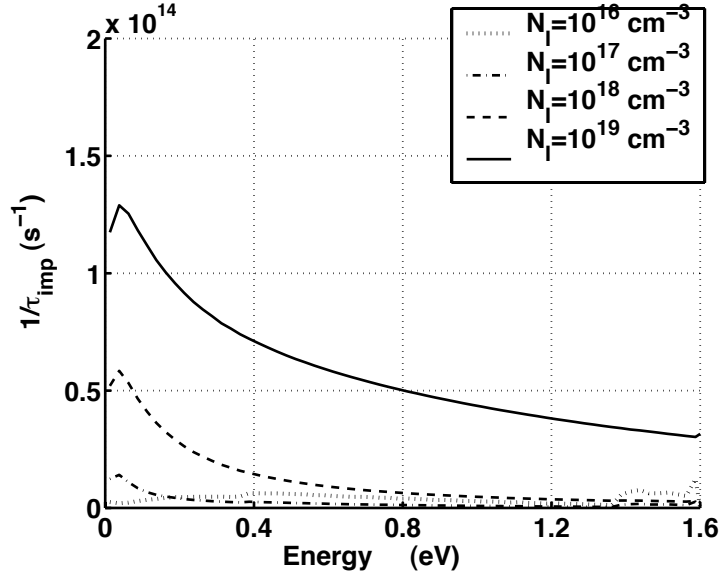
Fig. 2.6. Impurity out-scattering rates versus energy used in this work.

where L_D is the screening length

$$L_D = \left(\frac{\epsilon_{si}\epsilon_0 k_B T_L}{e^2 n_{free}} \right)^{1/2}.$$ (2.25)

For generalisation to other more detailed band-structures see [25, 26]. Unlike LO and LA phonon scattering which is inelastic and isotropic (the same in every direction), impurity scattering is elastic and anisotropic. In addition, the scattering rates for ionised impurities peak at low electron energies and decrease at high energies (see Fig. 2.6), which is also different from phonon scattering.

2.4 Boundary conditions in momentum space

This section describes the boundary condition for momentum space used in this work. The above section demonstrated that the discretisation of the collision integral turns it into a sum over i, representing the momentum space elements Ω_i. These elements are constructed in the Brilliouin zone between equally spaced constant energy surfaces of the band-structure. Naturally, this type of grid has a boundary at the maximum energy surface $E = E_{max}$ and hence requires a suitable boundary condition there.

The simplest boundary condition would be the truncation boundary condition, obtained by neglecting f at energies greater than E_{max},

$$f(k) = 0, \text{ if } E(k) > E_{max}. \tag{2.26}$$

However, this is valid only when E_{max} is large (usually > 3.0 eV). A large value of E_{max} is unacceptable for this work because of two reasons — the band-structure being used here is only valid for low energies $E \leq 1.6$ eV (Appendix A) and a large value of E_{max} increases the number of elements in the grid considerably. Therefore, in this work, the value of E_{max} is set at 1.6 eV for all r, which is high enough so that it can include all the physics of interest here and still be valid with respect to the band-structure.

In keeping with the above, a more appropriate asymptotic (or Robin) boundary condition [37] is used in this work,

$$\frac{\partial}{\partial E} f(E) = -\frac{1}{k_B T_L} f(E) \text{ when } E > E_{max}, \tag{2.27}$$

where E_{max} is the boundary of the grid. This condition implies that the distribution function must fall off exponentially at the rate of $k_B T_L$ (Maxwellian) at energies $E \geq E_{max}$. It is particularly well-suited for device simulations because Monte Carlo simulations of devices have shown that distribution function exhibit Maxwellian tails at energies $E \geq V_A$, the applied bias. Hence an asymptotic boundary condition can be used in devices for which $V_A \leq 1.6$ V. The trade-off of using an asymptotic boundary condition which is tied to the applied voltage across the device is that it produces some error in the bulk solutions because there is no concept of an applied bias in the bulk (Section 2.7). However, this error in the bulk solutions is not large.

In order to implement the above, boundary elements are added to the grid that lie in the spherical shell between the contours E_{max} and $E_{max} + 10k_B T_L$ (shown in Fig. 2.7). The value of 10 is chosen because $\exp(-10)$ is sufficiently small and anything beyond that can be neglected. The component of the distribution function inside the boundary elements is weighed with the exponential factor $\exp(-(E(k)-E_{max})/k_B T_L)$.
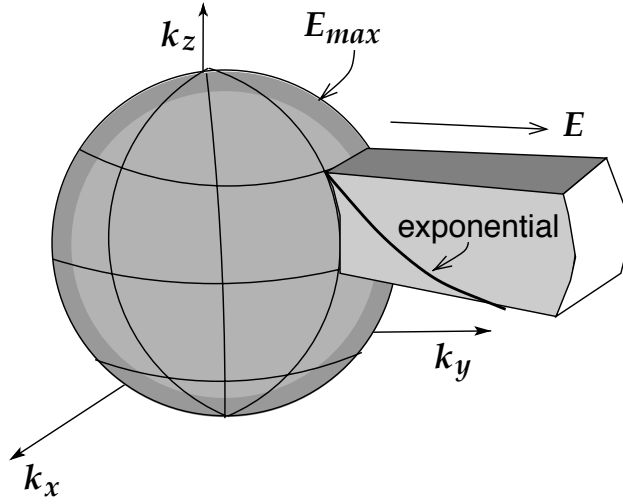
Fig. 2.7. Illustration of boundary elements used in this work.

This forces the distribution function to satisfy eq. 2.27. All scattering events that fall outside the range of $E_{max} + 10k_BT_L$ are neglected.

Note that the boundary elements also partitioned into angular segments similar to the interior grid. This allows the angular information to be retained in the boundary condition. Therefore, distribution functions that are distorted along a certain axis due to acceleration by the field will show the same anisotropy in the boundary elements and will not become isotropic. This is, of course, physically correct. The calculation of the collision terms remains the same as shown in eq.2.9. The only difference is that if Ω_i is a boundary element, a factor of $\exp(-(E(k) - E_{max})/k_BT_L)$ is included in the integral.

2.5 Discretisation error in collision integrals

This section estimates the error in the collision terms due to the discretisation described in the previous section. This is done by computing the null solutions of the collision matrices and comparing them with the corresponding analytical values.

1. Phonon collision matrix

   The null solution of phonon collision integrals are Maxwellians. This is known as

the property of "detailed balance" and arises because electrons exchange energy with phonons. Therefore, in a homogeneous system with no field, the electrons must attain equilibrium with the phonon bath such that rate of absorption of phonons equals their rate of emission. Mathematically, this property is stated in two equivalent ways:

(a) If $Q_{ph}(f) = 0$, then $f = M(E)$, where $M(E)$ is a Maxwellian in energy at the lattice temperature.

(b) $S(k, k') = S(k', k) \, M(E(k) - E(k'))$.

However, when the continuous scattering rates $S(k, k')$ are integrated over volumes in momentum space to get the discrete scattering rates, the Maxwellian relation between them also gets integrated. This causes an error in the null solution of phonon collision matrices. The following analysis shows that this is only a discretisation error and can be reduced by reducing the grid spacing.

Fig. 2.8 shows the null solutions for LO phonons obtained by solving

$$\tilde{Q}_{op}\tilde{f} = 0, \tag{2.28}$$

for different grid sizes. The expected analytical answer is a Maxwellian. From this figure, it is clearly seen that when the energy spacing is reduced such that $\Delta E < E_{op}$, the calculated solution approaches the expected solution. Note that the dot-dashed line lies on top of the dashed line indicating that angular refinement has little effect on reducing the error in equilibrium distributions.

Similarly, Fig. 2.9 shows the null solutions for LA phonons obtained by solving

$$\tilde{Q}_{ac}\tilde{f} = 0 \tag{2.29}$$

for different grid sizes. For acoustic phonons, the the convergence to Maxwellian distributions is slower than that for optical phonons. This is because acoustic phonons exchange energy from 0 to some maximum value $\sim 2\hbar k v_\ell$ (eq. 3.43 in [25]). At low energies (low k), the grid size $\Delta E \gg 2\hbar k v_\ell$ and hence, the error

is large. But at high energies (high k), the criterion $\Delta E < 2\hbar k v_\ell$ is satisfied, therefore the error decreases and the distributions approach Maxwellian.
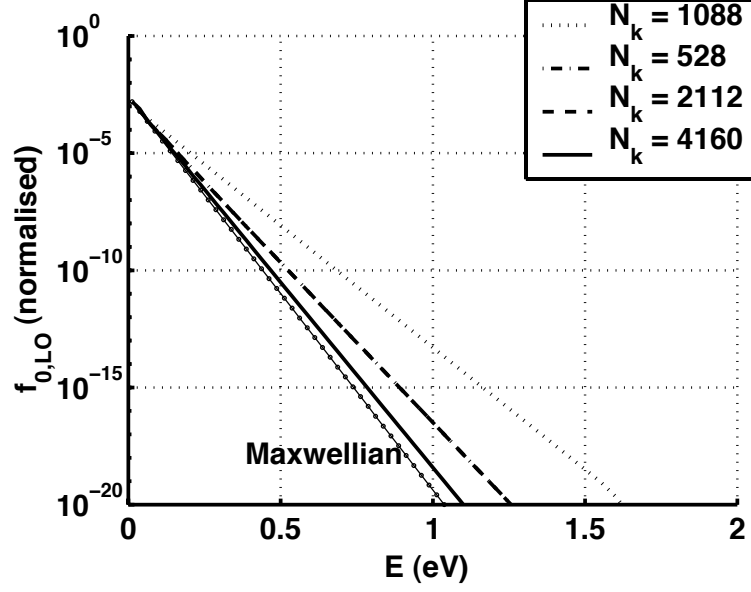
Fig. 2.8. Null solution for LO phonons with increasing grid size.
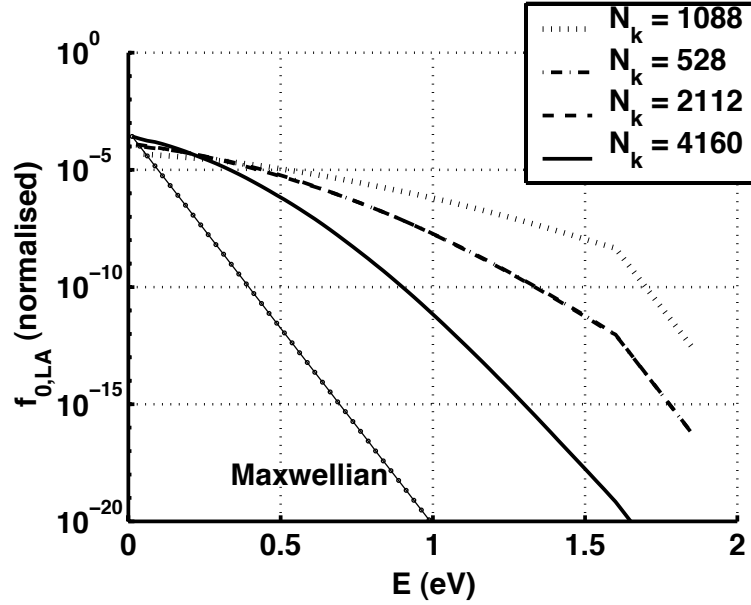
Fig. 2.9. Null solution for LA phonons with increasing grid size.

Fig. 2.10 shows the null solutions for both LO and LA phonons obtained by solving

$$(\tilde{Q}_{op} + \tilde{Q}_{ac})\tilde{f} = 0 \qquad (2.30)$$

for different grid sizes. The important point to note here is that the convergence behaviour is dominated by optical phonons, which is expected because optical phonons have higher scattering rates than acoustic phonons.
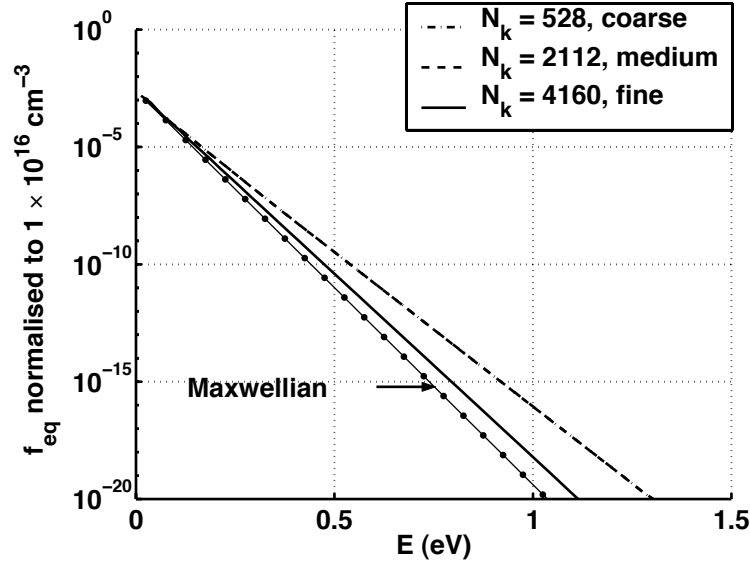
Fig. 2.10. Null solutions for LO and LA phonons with increasing grid size.

The thermal velocity calculated from the above null solutions as the velocity of *half* the equilibrium distribution is shown in Fig. 2.11(a). The calculated values converge quadratically to the expected values. The same behaviour is seen in Fig. 2.11(b) with the thermal energy of the above null solutions. Thermal energy, however, is calculated as the energy of the *total* distribution function. Again it is clearly seen,in both cases, that the effect of energy refinement is more significant than the effect of angular refinement. This conclusion holds only for equilibrium distributions because they are isotropic. It will be shown later that this conclusion does not hold for bulk distributions which occur in the presence of field and are not isotropic.

The analytical values used for comparison are computed for the appropriate spherical non-parabolic band as shown Appendix A.

(a)

(b)
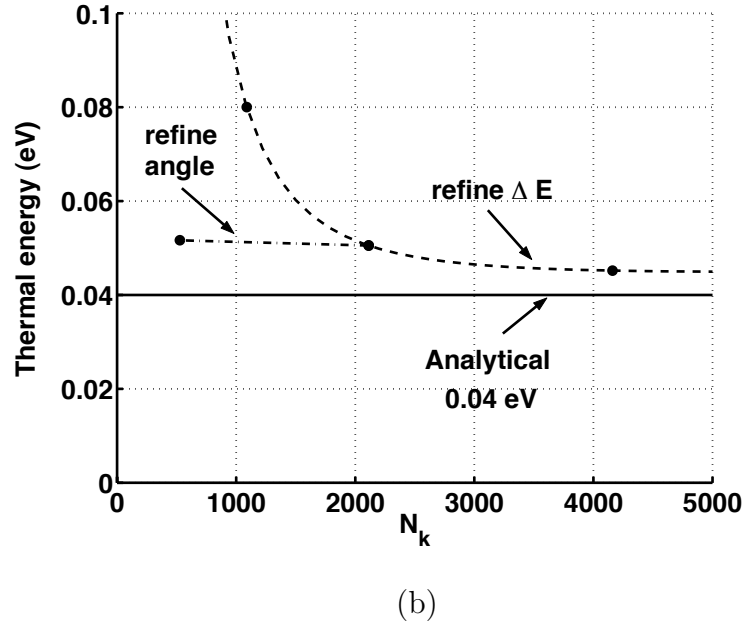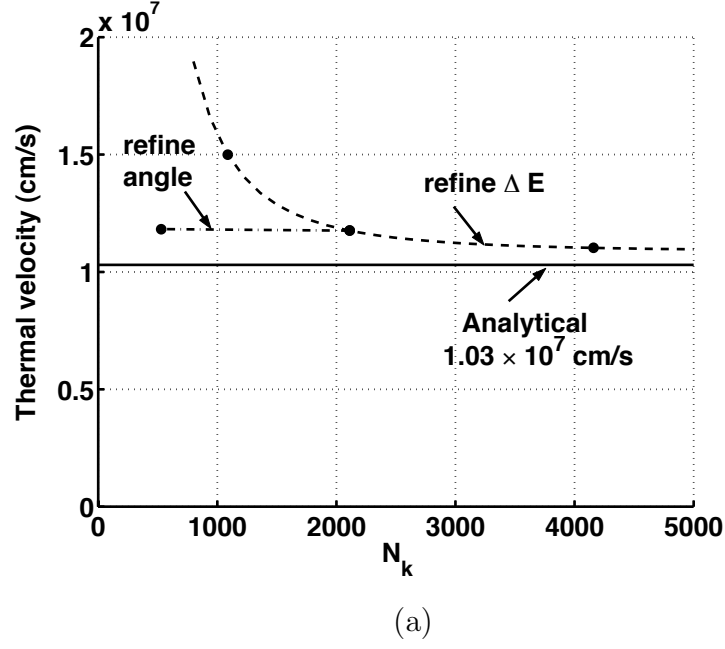
Fig. 2.11. Convergence of (a) thermal velocity and (b) thermal energy for the null solutions of Fig. 2.10 with increasing grid size.

2. <u>Impurity scattering</u>

The set of null solutions for impurity scattering are shown below for $N_k = 4160$. Note that each null solution is an independent delta function which is non-zero only at the energy nodes of the grid. Since there are 65 energy nodes in this grid, there are 65 null solutions. Therefore *any* function of energy can be constructed out of this set of null functions. This satisfies property 7 in Section 1.2.1 for impurity scattering.
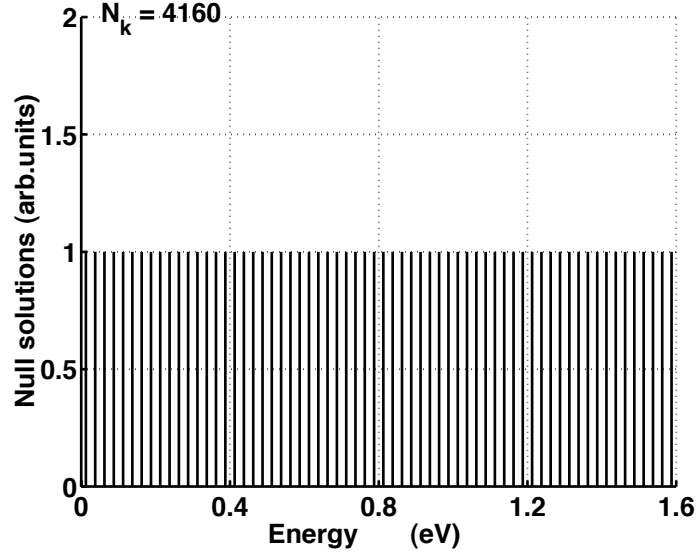
Fig. 2.12. Null solutions for impurity scattering for $N_k = 4160$.

2.6 Discretising the field term

This section describes the finite volume discretisation of the BTE by including the field term

$$H_k f - Q(f) = 0, \tag{2.31}$$

where Q, as usual, sums over all the types of scattering and H_k is given by

$$H_k f = \frac{e}{\hbar} \vec{\nabla}_r \phi(r) . \vec{\nabla}_k f = -\frac{e}{\hbar} \vec{\mathcal{E}}(r) . \vec{\nabla}_k f \tag{2.32}$$

The finite volume method begins by integrating the BTE over each element Ω_i of the momentum space grid as follows:

$$\int_{\Omega_i} H_k f \, \mathrm{d}k - \int_{\Omega_i} Q(f) \, \mathrm{d}k = 0, \quad \forall \, \Omega_i. \tag{2.33}$$

The collision integral has been treated in detail in the previous section, therefore, only the field term will be evaluated here.

Since the field term can be interpreted as a divergence, its control volume integral can be transformed by using Gauss' Law:

$$
\begin{aligned}
\int_{\Omega_i} H_k f \ \mathrm{d}k &= \int_{\Omega_i} \frac{e}{\hbar} \ \vec{\nabla}_k \cdot \left(-\vec{\mathcal{E}} f\right) \ \mathrm{d}k \\
&= \frac{e}{\hbar} \int_{\Gamma_i} f \ (-\vec{\mathcal{E}}) . \hat{n} \ \mathrm{d}S, \\
&= \frac{e}{\hbar} \sum_{i'} \int_{\Omega_i \cap \Omega_{i'}} f \ (-\vec{\mathcal{E}}) . \hat{n}_{i,i'} \ \mathrm{d}S,
\end{aligned}
\tag{2.34}
$$

where the integral is transformed *twice*. First the integral of the divergence of the flux $-\vec{\mathcal{E}} f$ over the volume element Ω_i is transformed into the integral of the flux leaving its bounding surface Γ_i. The generic outward normal to the bounding surface is represented by $\hat{n}$. Then the integral over the bounding surface is written as the sum of integrals over its sides, $\Omega_i \cap \Omega_{i'}$. Each side represents the part of the bounding surface of Ω_i that is shared with its adjacent element or neighbour, $\Omega_{i'}$. In case of the spherical hexahedral grid used in this work, there are 6 neighbours and therefore, 6 sides to each element. Therefore,

$$
\Gamma_i = \sum_{i'} \Omega_i \cap \Omega_{i'},
\tag{2.35}
$$

where i' only sums over the *neighbours* of Ω_i otherwise the intersection is 0. In addition, each side is considered to be a plane face such that it can be represented by a single outward normal $\hat{n}_{i,i'}$.

Now the integral over the sides of the volume element can be written as:

$$
\begin{aligned}
\int_{\Omega_i} H_k f \ \mathrm{d}k &= \frac{e}{\hbar} \sum_{i'} \int_{\Omega_i \cap \Omega_{i'}} f \ (-\vec{\mathcal{E}}) . \hat{n}_{i,i'} \ \mathrm{d}S, \\
&\approx \frac{e}{\hbar} \sum_{\nu} \sum_{i'} \int_{\Omega_i \cap \Omega_{i'}} f \ (-\mathcal{E}_\nu) \ \hat{\nu} . \hat{n}_{i,i'} \ \mathrm{d}S, \\
&= \frac{e}{\hbar} \sum_{\nu} \sum_{i'} (-\mathcal{E}_\nu) \ \hat{\nu} . \hat{n}_{i,i'} \int_{\Omega_i \cap \Omega_{i'}} f \ \mathrm{d}S,
\end{aligned}
\tag{2.36}
$$

where ν represents the components of the field in x,y or z directions. Note that this integral requires the value of the distribution function f on the surface Γ_i. All other terms are constant and therefore can be moved out of the integral.
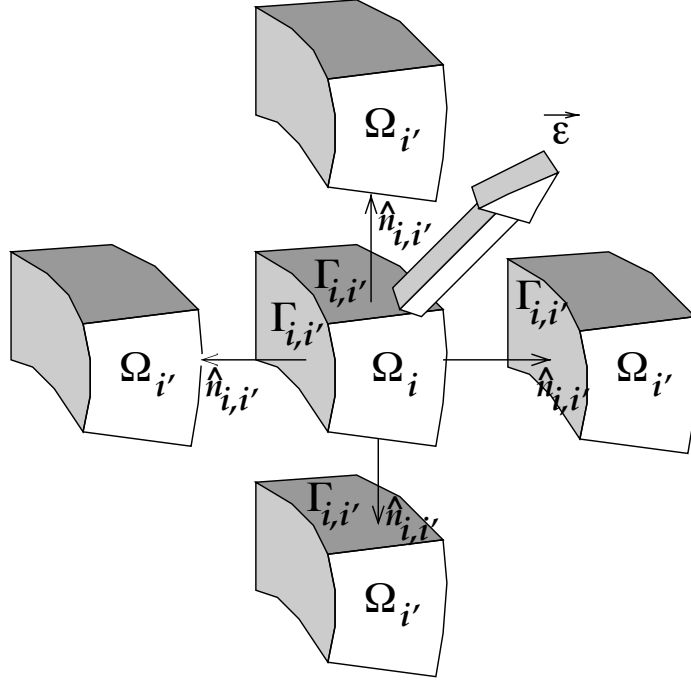
Fig. 2.13. Illustration of the notation used in momentum space upwinding. Elements in the third dimension are not shown but they are treated similarly.

In order to evaluate the above, the direction of "fluxes" needs to be taken into account (see Fig. 2.13). In the field term of the BTE, the "flux" is in the direction of the acceleration due to the field $(-e)\vec{\mathcal{E}}$. If $(-\mathcal{E}_\nu)\,\hat{\nu}.\hat{n}_{i,i'} > 0$, then the flux is *leaving* the element Ω_i and must come from *inside* the element. Therefore,

$$\text{if } (-\mathcal{E}_\nu)\,\hat{\nu}.\hat{n}_{i,i'} > 0, \ \int_{\Omega_i \cap \Omega_{i'}} f \ \mathrm{d}S \approx \tilde{f}_i \, \Gamma_{i,i'}, \tag{2.37}$$

where $\Gamma_{i,i'}$ is the numerical value of the area of the side $\Omega_i \cap \Omega_{i'}$. On the other hand, if $(-\mathcal{E}_\nu)\,\hat{\nu}.\hat{n}_{i,i'} < 0$, then the flux is *entering* the element Ω_i and must come from the *neighbouring* element. Therefore,

$$\text{if } (-\mathcal{E}_\nu)\,\hat{\nu}.\hat{n}_{i,i'} < 0, \ \int_{\Omega_i \cap \Omega_{i'}} f \ \mathrm{d}S \approx \tilde{f}_{i'} \, \Gamma_{i,i'}. \tag{2.38}$$

The error made in doing the above approximation is also on the order of the grid spacing $\mathcal{O}(h_k^n)$, where $n \approx 1$, as will be seen in Section 2.7. The boundary condition

can also be included in the field term by noting that f in the boundary elements is weighed by the exponential factor. Similar to the collision integral, any "flux" entering from or leaving into the states outside the boundary elements is neglected.

Note that this type of discretisation is known as "upwinding". Upwinding is critical for discretising first-order differential operators in order to to achieve stability (or positive-definiteness) in the discretised system. Without upwinding, it is well-known that the system can become unstable and negative numbers can arise in the solution that is expected to be positive. Using upwinding, however, produces a stable discretisation at the expense of accuracy. In the case of the BTE, note that upwinding is not only stable but also conserves mass because because the *incoming* flux for every element Ω_i is the *outgoing* flux when calculated for the element $\Omega_{i'}$.

Another advantage of upwinding in the field term of the BTE is that it is set by the direction of the force with respect to the volume elements of the grid, both of which are predetermined and fixed by the orientation of the crystal to the device geometry. This is *unlike* the case of upwinding in drift-diffusion where upwinding is required in the direction of the current, which is an unknown. Note also that the terms due to components of the field add and at any point, there can only be a field component along either the $+\nu$ or the $-\nu$ direction. Therefore, the application of upwinding is unambiguous.

2.6.1 The complete field term

To summarise, the final form of field term of the BTE can now be written as the multiplication of a matrix with a vector

$$\int_{\Omega_i} H_k f \ \mathrm{d}k = \sum_{i'} \tilde{H}_k(i, i') \tilde{f}_{i'}, \tag{2.39}$$

where

$$\begin{aligned}
\tilde{H}_k(i, i') &= \frac{e}{\hbar} \sum_{\nu \text{ s.t. } (-\mathcal{E}_\nu) \ \hat{\nu}.\hat{n}_{i,i'} < 0} (-\mathcal{E}_\nu) \ \hat{\nu}.\hat{n}_{i,i'} \ \Gamma_{i,i'}, \\
\tilde{H}_k(i, i) &= \frac{e}{\hbar} \sum_{\nu \text{ s.t. } (-\mathcal{E}_\nu) \ \hat{\nu}.\hat{n}_{i,i'} > 0} (-\mathcal{E}_\nu) \ \hat{\nu}.\hat{n}_{i,i'} \ \Gamma_{i,i'}.
\end{aligned} \tag{2.40}$$
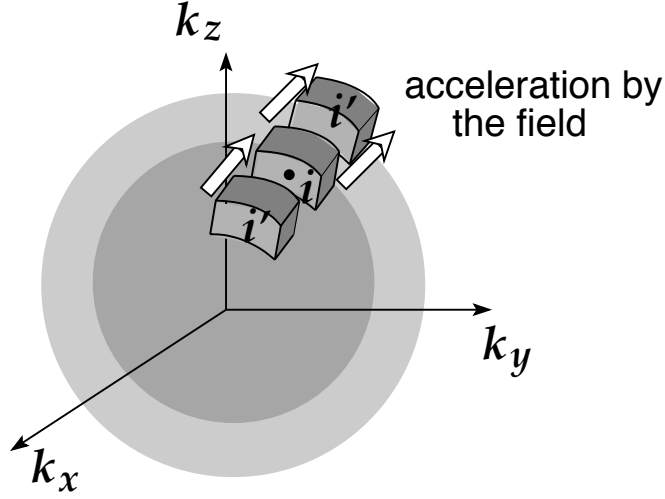
Fig. 2.14. Physical meaning of the terms of the field matrix $\tilde{H}_k$.

Henceforth, the matrix-vector product $\tilde{H}_k \tilde{f}$ will refer to the above expression (eq. 2.39). The size of the matrix $\tilde{H}_k$ is $N_k \times N_k$ but it is very sparse because acceleration by the field can only connect neighbouring momentum space elements. The vector $\tilde{f}$ is $N_k \times 1$ because it represents the distribution function being solved at the given nodes in momentum space. Two important observations about the discretised field matrix need to be made here. Firstly, $\tilde{H}_k$ conserves mass, its columns sum to zero and it is a positive semi-definite matrix. Secondly, the terms of $\tilde{H}_k$ are linear in the electric field (just like the BTE itself) and hence the effect of any field can be obtained simply by using the appropriate value of $\mathcal{E}_\nu$ in eq. 2.40. Also note here that although the *units* of H_k are 1/s, the units of $\tilde{H}_k$ are 1/(cm^{3}s). The continuous distribution function $f(k)$ and its discrete version $\tilde{f}$ are both dimensionless.

In conclusion, the advantage of a direct discretisation is that *both* the collision and field operators become matrices and can be treated simultaneously on the same footing. This is *unlike* the case of Monte Carlo where the collision events and the acceleration by the field are treated as occurring successively in time.

2.7 Discretisation error in the field term

This section estimates the error in the field terms due to the discretisation method described in the previous section. This is done by computing the null solutions of eq. 2.31 and comparing them with the corresponding analytical values.

The solutions of eq. 2.31 are known as bulk solutions because they describe transport in a uniform semiconductor with constant field. In the discretised system, these bulk solutions are the null solutions of the total matrix

$$(\tilde{H}_k - \tilde{Q}) \; \tilde{f} = 0, \tag{2.41}$$

where $\tilde{Q}$ sums over all the collision mechanisms. For the first part of this analysis, only phonon collision matrices are considered. Impurity scattering is added later. Unlike Monte Carlo, bulk solutions at different fields can be computed simply by varying the magnitude of the electric field in $\tilde{H}_k$. This computation is extremely fast and time for solution is between 2 seconds to 2 minutes on a 400 MHz UltraSparc2 workstation. This time depends on on the size of the grid and the magnitude of the field, where the former determines the size of the matrix and the latter determines the condition number of the matrix.
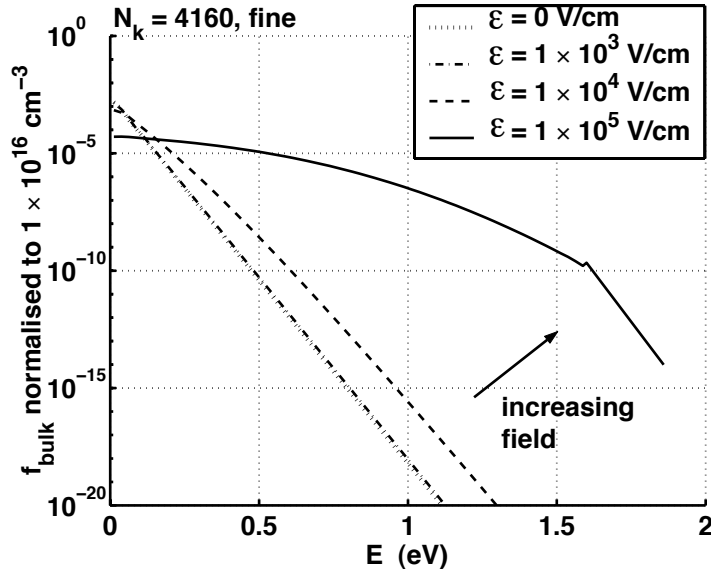
Fig. 2.15. Bulk distributions for increasing field for $N_k = 4160$.

An example of calculated bulk distributions for intrinsic Si (*i.e.* with only phonon collision matrices) using a fine grid $N_k = 4160$ is shown in Fig. 2.15. The bulk distributions for the other momentum space grids are not shown here instead their average velocities and average energies versus electric field are plotted in Fig. 2.16.
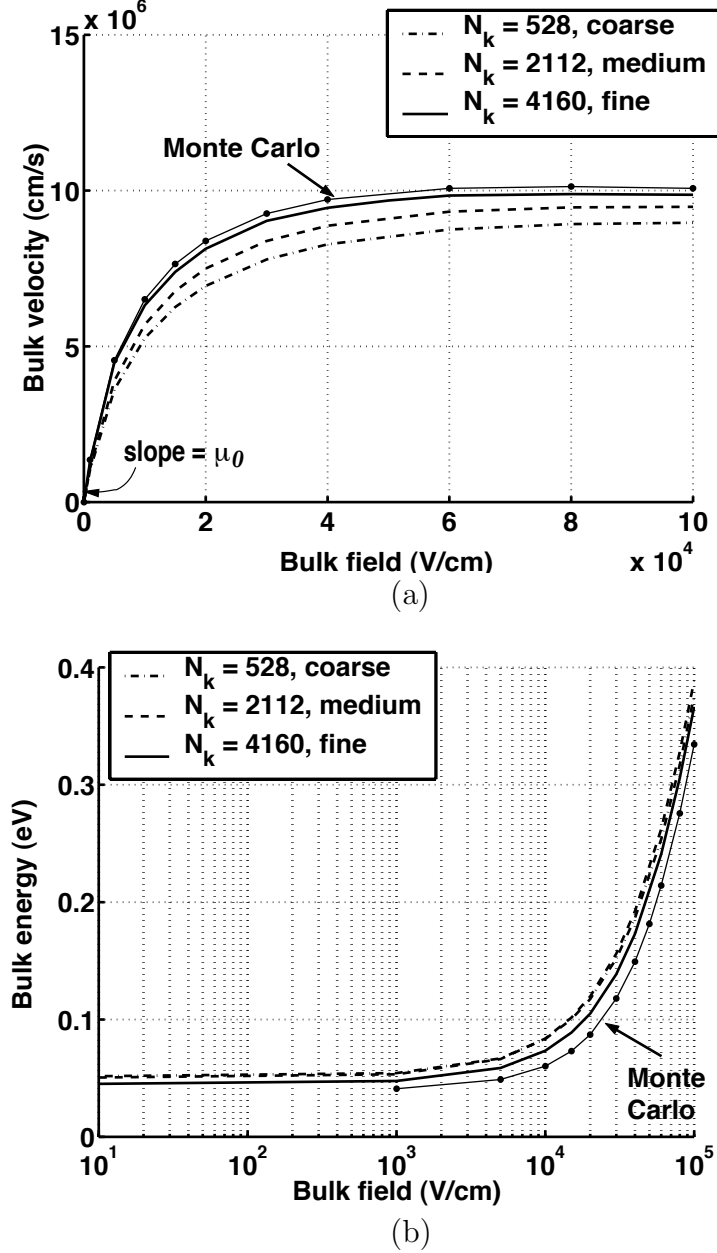
Fig. 2.16. Comparison of (a) average velocities and (b) average energies in the bulk for different N_k with bulk Monte Carlo.

Fig. 2.16 clearly demonstrates the effect of grid refinement on bulk solutions. It shows that as the grid is refined, the the bulk velocities and energies converge *linearly* to the values calculated by Monte Carlo simulation. It also shows that the angular refinement (dot-dashed line compared to the dashed line) is more critical for velocities than for energies. Energy refinement, however, still has the most important effect on the error (compare the dashed line and solid line).

The slope of the linear part of the velocity-field curve in Fig. 2.16 is the low-field mobility and comes out to be 1350 cm^2/V-s for the fine grid, which is the expected value in intrinsic Si. However, when impurities are added to bulk Si, the low-field mobility decreases. This is a well-established experimental result and hence can be used as a check for the accuracy of the impurity scattering rates. In order to do so, the impurity scattering matrix is added to $\tilde{Q}$ in eq. 2.41 and bulk solutions are computed at low fields for different impurity concentrations. The extracted low-field mobility is plotted versus impurity concentration in Fig. 2.17. It is seen that this shows the same characteristic decrease at high impurity concentrations as seen in experimental results [67]. Hence this demonstrates that impurity scattering rates have also been computed correctly.
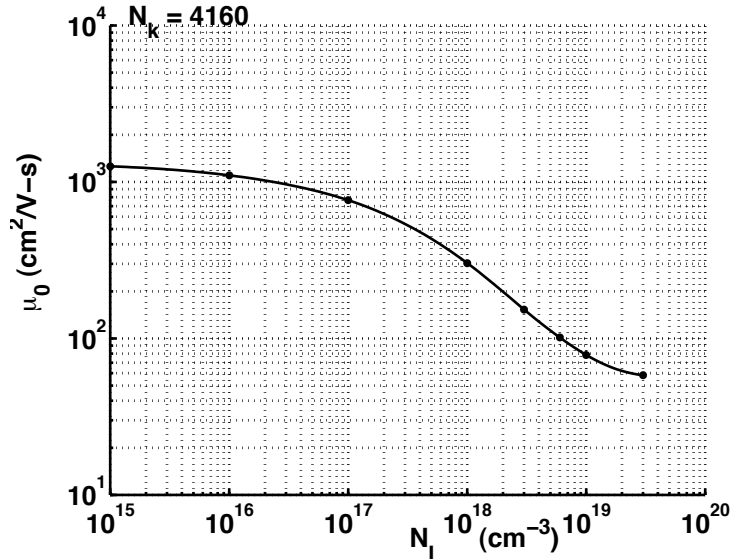
Fig. 2.17. Low-field mobility versus impurity concentration.

2.8   Comparison with Legendre expansion

In this section, the finite volume method is compared with a popular method of solving the BTE *v.i.z.* the Legendre polynomial expansion. This is done in order to show the advantages and disadvantages of the finite volume with respect to the series expansion methods used to solve the BTE.

The basis of comparison is the bulk averages (velocity and energy) computed from the finite volume method and from the Legendre polynomial expansion method. Care is taken to ensure that both the methods use the same a spherical non-parabolic energy band and the same optical phonon scattering mechanism. Acoustic phonons are not included because it is not clear if they can be treated consistently in both the Legendre polynomial method (which uses an approximation [38]) and the finite volume method.

Using the above mentioned parameters, the solution of the Legendre polynomial method proceeds as follows. First, the distribution function is expanded in terms of a Legendre polynomial series

$$f(k) = f_0(E) + k\cos\theta f_1(E) + \dots, \tag{2.42}$$

where only the first two terms of the series f_0, the symmetric 0th polynomial and f_1, the anti-symmetric 1st polynomial, are retained. The arguments for this and the first application of the Legendre expansion method to GaAs are in [23, 75].

Bulk solutions can then by found by substituting the above expansion for f into constant field homogeneous BTE, The resulting symmetric and anti-symmetric parts are projected out and the symmetric part is solved by eliminating the anti-symmetric part through a relaxation time approximation. The steps of this derivation can be found in [21]. The final set of equations are

$$f_1(E) = \left(\frac{2}{m^*_{eff}}\right)^{1/2} \frac{e\mathcal{E}\tau_0}{(\gamma'(E))^2} \frac{\partial}{\partial E} f_0(E),$$

$$-\frac{2}{3}\frac{(e\mathcal{E})^2\tau_0}{m^*_{eff}} \frac{(\gamma(E))^{1/2}}{(\gamma'(E))^3} \frac{\partial^2}{\partial E^2} f_0 + \frac{4}{3}\frac{(e\mathcal{E})^2\tau_0}{m^*_{eff}} \frac{(\gamma(E))^{1/2}\gamma''(E)}{(\gamma'(E))^4} \frac{\partial}{\partial E} f_0 = Q_{op}(f_0),$$

$$\tag{2.43}$$

where $\gamma(E) = E(1 + \alpha E)$ for the non-parabolic band and $\tau_0 = 4.049 \times 10^{-14}$ eV$^{1/2}$ s for optical phonon scattering, as defined in this report.
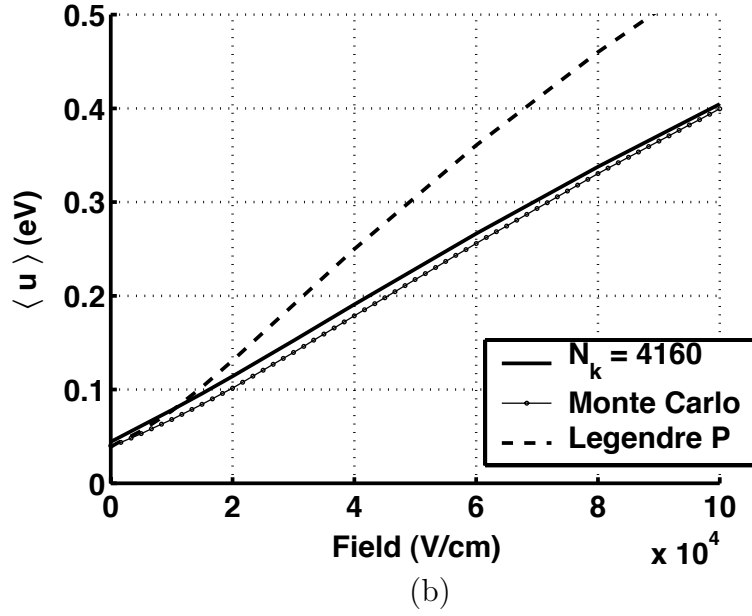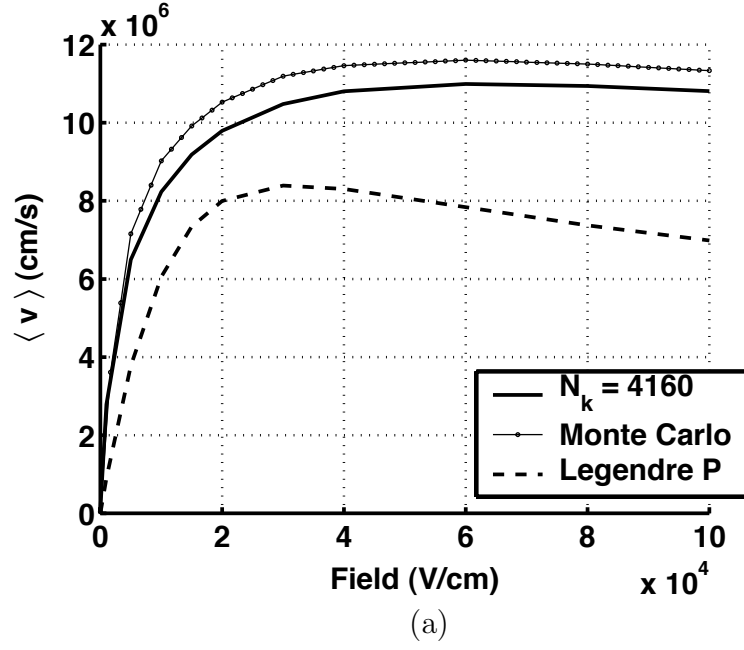
Fig. 2.18. Comparison of (a) average velocities and (b) average energies in the bulk with only optical phonons from the Legendre expansion method, finite volume method and Monte Carlo.

The two parts of eq. 2.43 are then solved for f_0 and f_1 for different electric fields and the corresponding bulk distributions are constructed. The resulting average velocities and energies are plotted versus field and compared to the values obtained from Monte Carlo and the finite volume method (for optical phonon scattering only) in Fig. 2.18.

It is seen that a Legendre polynomial method using only two terms of the expansion is inadequate to capture either the bulk velocity or the bulk energy profile. This is because a $\cos\theta$ variation in the angular direction is too smooth and cannot accurately represent distorted distribution functions at high fields. Distorted distributions, in general, have significant higher-order terms which are neglected by the Legendre polynomial method, as it is commonly used. Such an observation leads to the following "fix" for the Legendre polynomial method. If the optical phonon scattering rates are *reduced* to a smaller value $D_{op} = 14.55$ eV/cm, $E_{op} = 0.05$ eV and $\tau_0 = 3.73 \times 10^{-14}$ eV$^{1/2}$ s (as in [38]) then the magnitude of the asymmetric ($\cos\theta$) component of the distribution function reduces and the resulting bulk velocities increase to match the Monte Carlo values. The finite volume method does not suffer from the above limitation because it can represent distortions in the distribution function straightforwardly through its angular discretisation.

Finally, it must also be noted that the the Legendre polynomial expansion is a one-dimensional version of the spherical harmonics expansion [37] and in general, it is used with a Herring-Vogt transformation of the Si band-structure. However, the Herring-Vogt transformation is not used for the above analysis because the band-structure in this work is a single spherical non-parabolic band.

2.9 Summary

In conclusion, this chapter 2 described the finite volume method used to discretise the BTE in momentum space $\mathbb{R}^3$. It involved the computation of the discrete terms of the collision operator Q and the hyperbolic field operator H_k on any given grid in momentum space. Upon discretisation, the collision and field operators turned into corresponding collision and field matrices. In this process, care was taken to

preserve the properties of conservation of energy and momentum in the collision operator and conservation of mass in both collision and field operators. The momentum space discretisation was completed by imposing an appropriate asymptotic boundary condition on f in $\mathbb{R}^3$. In order to analyse the the discretisation error of this method with respect to the grid spacing h_k, the discrete null solutions of the collision and field matrices were compared with corresponding analytical solutions. This process demonstrated that a grid with 4160 unknowns has good accuracy and yet reasonable size. The chapter ended by showing that this method has potential for better accuracy than the Legendre polynomial method.

3. Real space discretisation

3.1 Introduction

In the previous chapter, the finite volume method used to discretise the collision and field term of the BTE was described. It addressed all the important issues of undertaking such a discretisation in momentum space. Therefore, this chapter describes the inclusion of the spatial term H_r in the BTE and its subsequent discretisation by using the finite volume method.

The organisation of this chapter is as follows. Section 3.2 describes the real space discretisation of the BTE using the finite volume method. The treatment is kept general, so that it can be applied to both 1D as well as 2D devices.

Section 3.3 describes the specific application of the above method to a one-dimensional device in order to demonstrate many important concepts of the numerical technique, such as imposing boundary conditions at the contacts of the device and setting up the system of equations as a matrix equation. This example also shows that the real space discretisation used in this method has a severe restriction on the grid spacing. However, an analysis of the discretisation error in the average velocity and energy of a one-dimensional device shows that this restriction can be relaxed because it comes from components in the distribution function that do not significantly affect the average velocity or energy of the carriers.

Finally, section 3.4 summarises the conclusions of this chapter.

3.2 Discretising the spatial term of the BTE

This section describes the finite volume discretisation of the BTE by including the spatial term

$$H_r f + H_k f - Q(f) = 0, \tag{3.1}$$

where Q and H_k are as defined earlier and H_r is given by

$$H_r f = \frac{1}{\hbar} \vec{\nabla}_k E(k).\vec{\nabla}_r f. \tag{3.2}$$

As shown earlier, the finite volume method begins by integrating the BTE over each element Ω_i of the momentum space grid as follows:

$$\int_{\Omega_i} (H_r f + H_k f - Q(f)) \ \mathrm{d}k = 0, \quad \forall \ \Omega_i. \tag{3.3}$$

Using a convenient quadrature, the collision integral and the field term evaluate to matrix-vector products. The spatial term, on the other hand, becomes a vector whose terms are

$$\begin{aligned} \int_{\Omega_i} H_r f \ \mathrm{d}k &= \int_{\Omega_i} \frac{1}{\hbar} \vec{\nabla}_k E(k).\vec{\nabla}_r f \ \mathrm{d}k \\ &\approx \left(\int_{\Omega_i} \frac{1}{\hbar} \vec{\nabla}_k E(k) \ \mathrm{d}k \right).\vec{\nabla}_r \tilde{f}_i \end{aligned} \tag{3.4}$$

where the integral refers to the group velocity of the electrons in the volume element Ω_i. This is denoted by a diagonal matrix

$$\vec{V}(i,i) = \int_{\Omega_i} \frac{1}{\hbar} \vec{\nabla}_k E(k) \ \mathrm{d}k, \tag{3.5}$$

whose units are $1/(\mathrm{cm}^2\mathrm{s})$.

Hence, the result of the momentum space discretisation on eq. 3.1 is

$$\vec{V}(i,i).\vec{\nabla}_r \tilde{f}_i + \sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i'} - \sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i'} - \tilde{Q}(i,i) \tilde{f}_i = 0. \tag{3.6}$$

However, the unknowns in the above equation are still a function of r.

Therefore, the above equation must be discretised in real space. This is done in a manner similar to momentum space discretisation. First, a real space grid is constructed by placing nodes r_j in real space such that they are centroids of non-overlapping volumes Ω_j that cover the entire problem domain. This yields a discrete set of unknowns

$$\tilde{f}_i(r = r_j) = \tilde{f}_{i,j}. \tag{3.7}$$

Note that the index i is used to denote volume elements in momentum space and the index j is used for real space. The final step is to integrate eq. 3.6 over the volumes Ω_j,

$$\int_{\Omega_j} \left(\vec{\tilde{V}}(i,i).\vec{\nabla}_r \tilde{f}_i + \sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i'} - \sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i'} - \tilde{Q}(i,i) \tilde{f}_i \right) \ \mathrm{d}r = 0. \qquad (3.8)$$

The integrals over the collision and field terms are approximated in a simple way as follows

$$\int_{\Omega_j} \left(\sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i'} \right) \ \mathrm{d}r \ \approx \ \left(\sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i',j} \right) \Omega_j,$$

$$\int_{\Omega_j} - \left(\sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i'} + \tilde{Q}(i,i) \tilde{f}_i \right) \ \mathrm{d}r \ \approx \ - \left(\sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i',j} + \tilde{Q}(i,i) \tilde{f}_{i,j} \right) \Omega_j,$$

$$(3.9)$$

where Ω_j is the numerical value of the volume of the element.

However, the space term is interpreted as a divergence and its control volume integral is transformed by using Gauss' Law. Therefore,

$$\begin{aligned}
\int_{\Omega_j} \vec{\tilde{V}}(i,i).\vec{\nabla}_r \tilde{f}_i \ \mathrm{d}r \ &= \ \int_{\Omega_j} \vec{\nabla}_r. \left(\vec{\tilde{V}}(i,i) \tilde{f}_i \right) \ \mathrm{d}r \\
&= \ \int_{\Gamma_j} \tilde{f}_i \ \vec{\tilde{V}}(i,i).\hat{n} \ \mathrm{d}S, \\
&= \ \sum_{j'} \int_{\Omega_j \cap \Omega_{j'}} \tilde{f}_i \ \vec{\tilde{V}}(i,i).\hat{n}_{j,j'} \ \mathrm{d}S, \qquad (3.10)
\end{aligned}$$

where the integral is transformed *twice*. First the integral of the divergence of the flux $\vec{\tilde{V}}(i,i) \tilde{f}_i$ over the volume element Ω_j is transformed into the integral of the flux leaving its bounding surface Γ_j. The generic outward normal to the bounding surface is represented by $\hat{n}$. Then the integral over the bounding surface is written as the sum of integrals over its sides, $\Omega_j \cap \Omega_{j'}$. Each side represents the part of the bounding surface of Ω_i that is shared with its adjacent element or neighbour, $\Omega_{i'}$. In case of a rectangular hexahedral grid used in this work, there are 6 neighbours and therefore, 6 sides to each element. Therefore,

$$\Gamma_j = \sum_{j'} \Omega_j \cap \Omega_{j'}, \qquad (3.11)$$

where j' only sums over the *neighbours* of Ω_j otherwise the intersection is 0. In addition, each side is considered to be a plane face such that it can be represented by a single outward normal $\hat{n}_{j,j'}$.

Now the integral over the sides of the volume element can be written as:

$$\int_{\Omega_j} \vec{\tilde{V}}(i,i).\vec{\nabla}_r \tilde{f}_i \ \mathrm{d}r \ = \ \sum_{j'} \int_{\Omega_j \cap \Omega_{j'}} \tilde{f}_i \ \vec{\tilde{V}}(i,i).\hat{n}_{j,j'} \ \mathrm{d}S,$$

$$\approx \ \sum_{\nu} \sum_{j'} \int_{\Omega_j \cap \Omega_{j'}} \tilde{f}_i \ \tilde{V}_\nu(i,i) \ \hat{\nu}.\hat{n}_{j,j'} \ \mathrm{d}S,$$

$$= \ \sum_{\nu} \sum_{j'} \tilde{V}_\nu(i,i) \ \hat{\nu}.\hat{n}_{j,j'} \int_{\Omega_j \cap \Omega_{j'}} \tilde{f}_i \ \mathrm{d}S, \qquad (3.12)$$

where ν represents the components of the velocity in x,y or z directions. Note that this integral requires the value of the distribution function f on the surface Γ_j. All other terms are constant and therefore can be moved out of the integral.
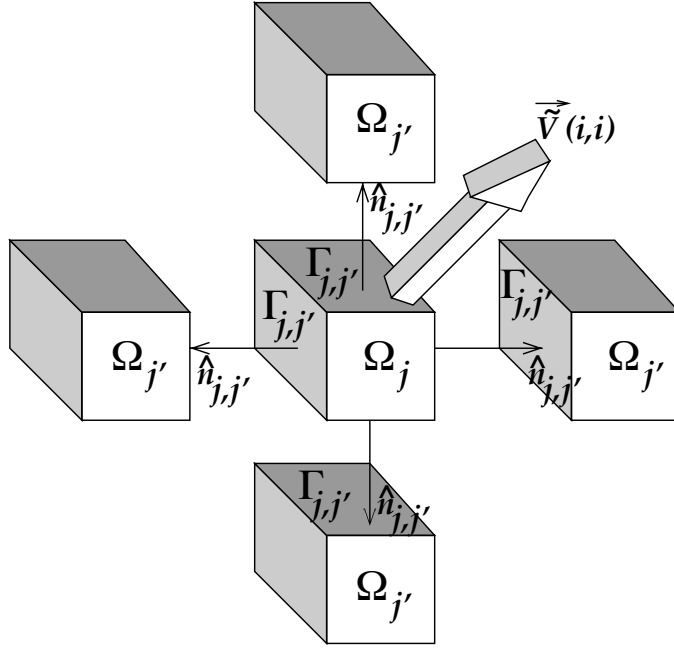
Fig. 3.1. Illustration of the notation used in real space upwinding. Elements in the third dimension are not shown but they are treated similarly.

In order to evaluate the above, the direction of "fluxes" needs to be taken into

account (see Fig. 3.1). In the space term of the BTE, the "flux" is in the direction of the group velocities of the electrons $\vec{\tilde{V}}(i,i)$ for every element Ω_i. If $\tilde{V}_\nu(i,i)\,\hat{\nu}.\hat{n}_{j,j'} > 0$, then the flux is *leaving* the element Ω_j and must come from *inside* the element. Therefore,

$$\text{if} \quad \tilde{V}_\nu(i,i)\hat{\nu}.\hat{n}_{i,i'} > 0, \quad \int_{\Omega_j \cap \Omega_{j'}} \tilde{f}_i \;\, \mathrm{d}S \approx \tilde{f}_{i,j} \; \Gamma_{j,j'}, \qquad (3.13)$$

where $\Gamma_{j,j'}$ is the numerical value of the area of the side $\Omega_j \cap \Omega_{j'}$. On the other hand, if $\tilde{V}_\nu(i,i)\,\hat{\nu}.\hat{n}_{j,j'} < 0$, then the flux is *entering* the element Ω_j and must come from the *neighbouring* element. Therefore,

$$\text{if} \quad \tilde{V}_\nu(i,i)\hat{\nu}.\hat{n}_{i,i'} < 0, \quad \int_{\Omega_j \cap \Omega_{j'}} \tilde{f}_i \;\, \mathrm{d}S \approx \tilde{f}_{i,j'} \; \Gamma_{j,j'}, \qquad (3.14)$$

The error made in doing the above approximation is also on the order of the grid spacing $\mathcal{O}(h_r^n)$, where $n \approx 1$, as will be seen in Section 3.3.4. Note that this type of discretisation is equivalent to the upwinding used for the field term. Therefore, all the properties of upwinding also apply here such as positive-definiteness, conservation of mass and stability (at the expense of lower accuracy).

In the spatial term, the direction of upwinding is set by the direction of the group velocity of the volume elements of the momentum space grid, which is predetermined and fixed by the band-structure. The terms due to components of the group velocity add and at any point, there can only be a velocity component along either the $+\nu$ or the $-\nu$ direction. Therefore, the application of upwinding is unambiguous.
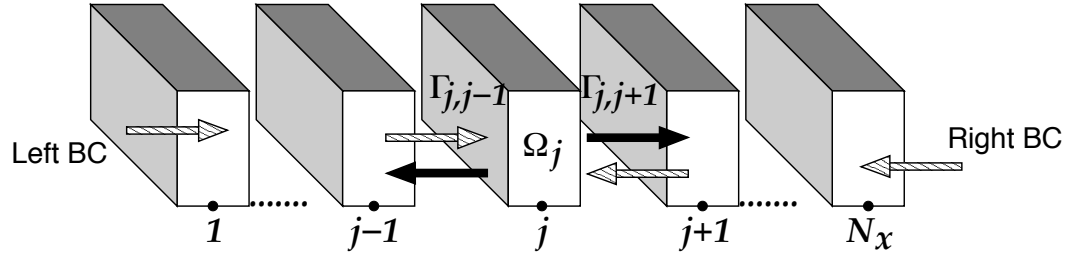
Fig. 3.2. Illustration of a one-dimensional device simulation. Notice the similarity to the scattering matrix approach.

3.2.1 The complete real space discretisation

The complete real space discretisation (eq. 3.8) can now be written in terms of matrix-vector products as:

$$
\sum_{j'} \left(\sum_{\nu \text{ s.t. } \tilde{V}_\nu(i,i)\ \hat{\nu}.\hat{n}_{j,j'}>0} \tilde{V}_\nu(i,i)\ \hat{\nu}.\hat{n}_{j,j'}\ \Gamma_{j,j'} \right) \tilde{f}_{i,j} +
$$

$$
\sum_{j'} \left(\sum_{\nu \text{ s.t. } \tilde{V}_\nu(i,i)\ \hat{\nu}.\hat{n}_{j,j'}<0} \tilde{V}_\nu(i,i)\ \hat{\nu}.\hat{n}_{j,j'}\ \Gamma_{j,j'} \right) \tilde{f}_{i,j'} +
$$

$$
\left(\sum_{i'} \tilde{H}_k(i,i')\tilde{f}_{i',j} \right) \Omega_j - \left(\sum_{i'\neq i} \tilde{Q}(i,i')\tilde{f}_{i',j} + \tilde{Q}(i,i)\tilde{f}_{i,j} \right) \Omega_j = 0.
$$

$$(3.15)$$

This is a mathematically complicated expression but physically it is simple. It states that the electrons that leave the element depend on the electrons that enter the element and the electrons that undergo scattering and acceleration inside the element. This is, in fact, the the *response* or *scattering* matrix formulation of the BTE.

When the above expression is written out for all i and j, a linear system of equations is generated. It can be written as a matrix equation whose size is $N \times N$ where $N = N_r \times N_k$. Although the size is very large, usually on the order of 10^6–10^7, the resulting matrix is very sparse because the space term can only connect neighbouring elements. In order to conceptually understand this formulation, an example of a one-dimensional device is given below.

3.3 One-dimensional device

Consider a one-dimensional device (as shown in Fig. 3.2) where the distribution function is only a function of x and does not depend on the transverse co-ordinates. This reduces eq. 3.15 to

$$
\sum_{j'} \left(\sum_{x \text{ s.t. } \tilde{V}_x(i,i)\ \hat{x}.\hat{n}_{j,j'}>0} \tilde{V}_x(i,i)\ \hat{x}.\hat{n}_{j,j'}\ \Gamma_{j,j'} \right) \tilde{f}_{i,j} +
$$

$$
\sum_{j'} \left(\sum_{x \text{ s.t. } \tilde{V}_x(i,i)\ \hat{x}.\hat{n}_{j,j'}<0} \tilde{V}_x(i,i)\ \hat{x}.\hat{n}_{j,j'}\ \Gamma_{j,j'} \right) \tilde{f}_{i,j'} +
$$

$$\left(\sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i',j} \right) \Omega_j - \left(\sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i',j} + \tilde{Q}(i,i) \tilde{f}_{i,j} \right) \Omega_j = 0,$$

$$(3.16)$$

where j can only be either $j+1$ or $j-1$. In addition, in 1D, the surface area and volumes of the elements are related as follows:

$$\Gamma_{j,j+1} = \Gamma_{j,j-1} = \frac{1}{\Delta x_j} \, \Omega_j,$$

$$(3.17)$$

and the outward normals to the surfaces are

$$\hat{n}_{j,j-1} = -\hat{x},$$
$$\hat{n}_{j,j+1} = \hat{x}.$$

$$(3.18)$$

Suppose the x-component of the group velocity of element Ω_i is positive, then $\tilde{V}_x(i,i) \, \hat{x}.\hat{n}_{j,j+1} > 0$ and $\tilde{V}_x(i,i) \, \hat{x}.\hat{n}_{j,j-1} < 0$. Therefore, eq. 3.16 becomes

$$|\tilde{V}_x(i,i)| \, \Gamma_{j,j+1} \, \tilde{f}_{i,j} - |\tilde{V}_x(i,i)| \, \Gamma_{j,j-1} \, \tilde{f}_{i,j-1} + \left(\sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i',j} \right) \Omega_j -$$

$$\left(\sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i',j} + \tilde{Q}(i,i) \tilde{f}_{i,j} \right) \Omega_j = 0.$$

$$(3.19)$$

Similarly, if the x-component of the group velocity of element Ω_i is negative, then $\tilde{V}_x(i,i) \, \hat{x}.\hat{n}_{j,j-1} > 0$ and $\tilde{V}_x(i,i) \, \hat{x}.\hat{n}_{j,j+1} < 0$. Therefore, eq. 3.16 becomes

$$|\tilde{V}_x(i,i)| \, \Gamma_{j,j-1} \, \tilde{f}_{i,j} - |\tilde{V}_x(i,i)| \, \Gamma_{j,j+1} \, \tilde{f}_{i,j+1} + \left(\sum_{i'} \tilde{H}_k(i,i') \tilde{f}_{i',j} \right) \Omega_j -$$

$$\left(\sum_{i' \neq i} \tilde{Q}(i,i') \tilde{f}_{i',j} + \tilde{Q}(i,i) \tilde{f}_{i,j} \right) \Omega_j = 0.$$

$$(3.20)$$

This implies that when the electrons are travelling in the positive direction, the node j and the *previous* node $j-1$ are related. But when the electrons are travelling in the negative direction, the node j and the *next* node $j+1$ ar related. This relation is not only mathematically correct but also physically correct because "information"

travels only in the direction of the velocity. This can also be generalised to 2D grids (rectangular or triangular) provided the proper book-keeping is done according to eq. 3.15.

When the above expression is written out for all i and all j, a linear system of equations is generated. It can be written as a matrix equation whose size is $N \times N$ where $N = N_x \times N_k$. Although the size is very large, usually on the order of 10^6–10^7, the resulting matrix is very sparse because the space term can only connect j to neighbouring elements, $j \pm 1$.

In conclusion, it seen that a direct discretisation of the BTE, by using the finite volume method, results in a very large matrix equation. When this is compared to solving most common PDEs with simple differential operators and 2 or 3 dimensions, it is not surprising that solving the BTE is much more difficult than solving any other PDE.

3.3.1 Boundary conditions at the contacts

The boundary conditions at the contacts of a device fall out naturally as a result of upwinding. Fig. 3.1 shows a 1D device that consists of N_x elements and therefore j runs from 1 to N_x. Eq. 3.19, shows that when $j = 1$, the element $\tilde{f}_{i,j-1}$ is at a node 0 which is not defined for this device. Therefore, $\tilde{f}_{i,0}$ must be *specified* and this constitutes the left boundary condition.

Note that the element $\tilde{f}_{i,j-1}$ occurs only in the term for which the velocity $\tilde{V}_x(i,i) > 0$, therefore this implies that we can only impose boundary conditions on the left boundary for the components of the distribution function that *enter* the device. The term for $\tilde{V}_x(i,i) < 0$ has the elements $\tilde{f}_{i,j}$ and $\tilde{f}_{i,j+1}$, both of which have nothing to do with the left boundary.

Similarly when $j = N_x$, the element $\tilde{f}_{i,j+1}$ is at a node $N_x + 1$ which is also not defined for this device. Therefore, $\tilde{f}_{i,N_x+1}$ must be specified and this gives rise to the right boundary condition. Eq. 3.20 shows that this element is required only for the term for which $\tilde{V}_x(i,i) < 0$. Therefore, again on the right boundary, only the components of the distribution function that enter the device need be specified.

In conclusion, physically correct boundary conditions on the device can be applied simply by using the framework of upwinding. The actual numerical values of $\tilde{f}_{i,0}$ and $\tilde{f}_{i,N_x+1}$ are calculated from the equilibrium distributions in the contacts and scaled according to the electron quasi-fermi level specified at the left and right boundaries respectively.

3.3.2   Setting up the matrix equation

In this section, the linear systems of equations for the 1D device (obtained as described in Section 3.3) is set up as a matrix equation $Ax = b$. In particular, the ordering of the unknowns and the structure of the matrix A are discussed.

The system of unknowns in eq. 3.16 is $\tilde{f}_{i,j}$. The indices i and j represent 4 dimensions in all, where the index i runs over all the volume elements in $\mathbb{R}^3$ momentum space and the index j runs over the real space volume elements of the 1D device. If the number of nodes in momentum space is N_k and the number of nodes in real space is N_x, then the number of unknowns is $N = N_x \times N_k$. The question is how to order the system of unknowns such that the bandwidth of A is small and the sparse nature of the matrix can be exploited. This is an advantage because iterative methods of solution work well on small bandwidth and structured sparse matrices.

Among the many schemes of ordering unknowns for optimal bandwidth and structure, the most common algorithms are red-black ordering, Cuthill-Mckee ordering and natural ordering. However, since it is beyond the scope of this work to investigate optimum ordering techniques, a convenient natural ordering system is used here. In this ordering system the inner ordering index is i and the outer ordering index is j. This ensures that A is a block tri-diagonal matrix (shown in Fig 3.3) where the positions of the blocks are determined by the outer index j. Within itself, each block is then ordered according to the inner index i such that it has a small band-width. Since these blocks consist of $\tilde{Q}$ and $\tilde{H}_k$ which have have strong correlations in energy, the inner ordering is done by arranging the elements Ω_i according to increasing energy. The resulting structure is shown in Fig. 3.4. The structure of $\tilde{V}_x$ is diagonal, of course.
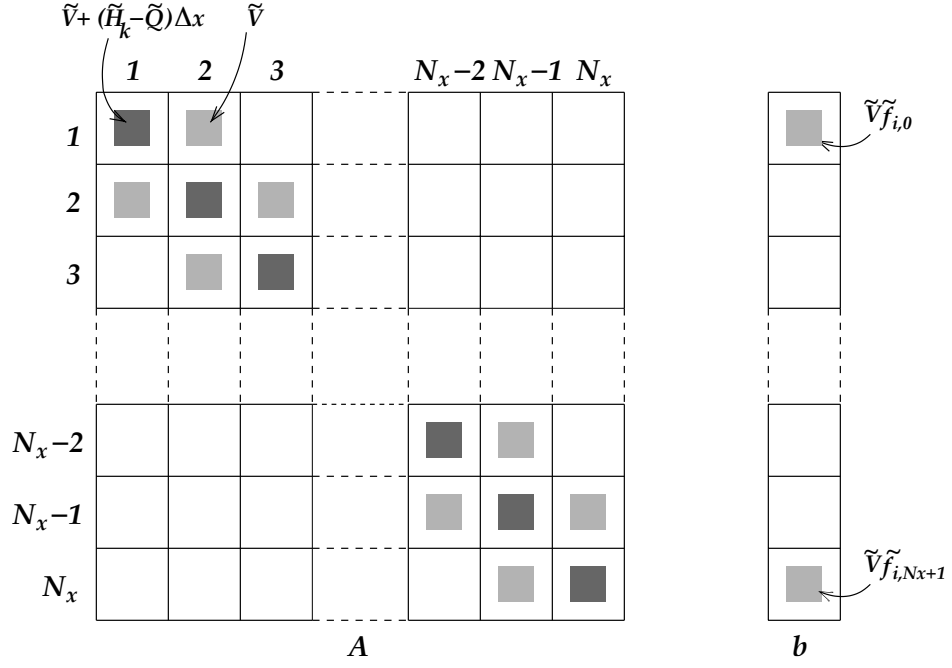
Fig. 3.3. Block tri-diagonal structure of matrix A (with right-hand side b) ordered such that the inner index is i and the outer index is j. All the blocks are sparse.
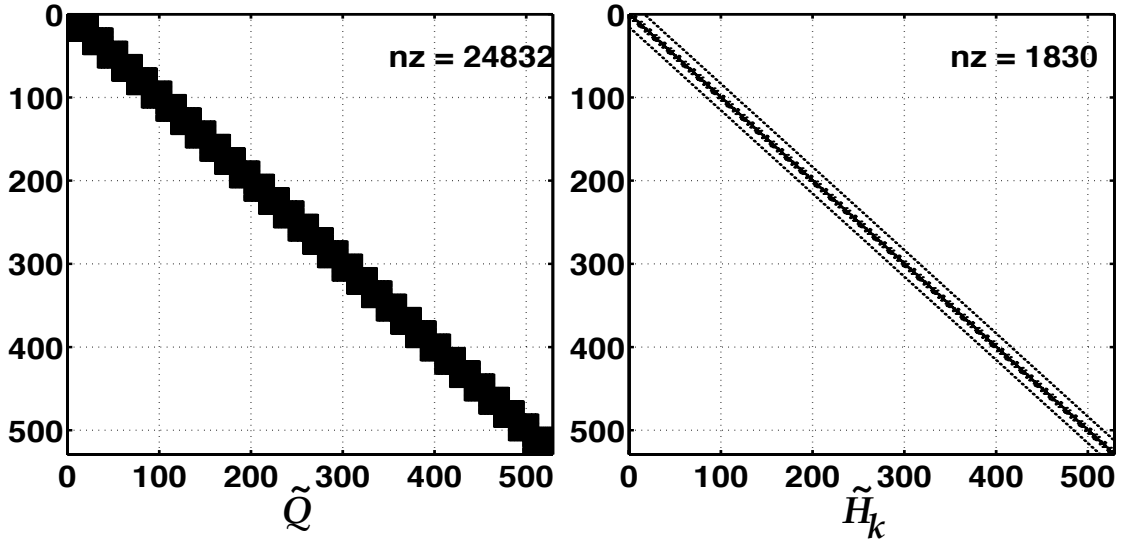
Fig. 3.4. Ordering in $\tilde{Q}$ and $\tilde{H}_k$ for an example grid size $N_k = 528$.

Finally, anything that is *specified* and not an unknown goes into the right-hand side vector b. These are the boundary conditions, which have been described earlier as the elements $\tilde{f}_{i,0}$ for $\tilde{V}_x(i,i) > 0$ and $\tilde{f}_{i,N_x+1}$ for $\tilde{V}_x(i,i) < 0$. The ordering of b must be the *same* as A. The program written to solve the BTE is organised such that it automatically orders A and b using the above algorithm.

In general, the number of terms in the matrix A is a very large number, on the order of $2 \times 10^8 - 2 \times 10^9$ for $N = 10^6 - 10^7$. However, the issue of storing all these terms does not arise in this work, because an iterative method is used to solve the above matrix equation. These iterative methods (see Chapter 4) typically require only a matrix-vector multiplication at each step. Therefore, the program written to solve the BTE is such that it computes the product of A with a given vector by constructing the elements of A on the fly.

3.3.3   Restrictions on grid spacing

Section 3.2 described a simple way to discretise the spatial term of the BTE. The validity of the discretisation and its limitations, if any, are analysed in this section.

In order to do so, the terms of eq. 3.16 must be examined carefully. It only has terms of the first degree in Δx_j. It is, in fact, a first-order *approximation* to the volume integral in eq. 3.8. This can be seen by considering a similar integral

$$\int_{x_{n-1/2}}^{x_{n+1/2}} a \,\frac{\mathrm{d}}{\mathrm{d}x}g(x) + b\, g(x) = 0, \tag{3.21}$$

whose exact solution is an exponential

$$a \,\exp\left(\frac{b}{a}(x_{n+1/2} - x_{n-1/2})\right) g(x_{n+1/2}) - a\, g(x_{n-1/2}) = 0, \tag{3.22}$$

where $(x_{n+1/2} - x_{n-1/2}) = \Delta x_n$. However, if the coefficient is small *i.e.* if $\frac{b}{a}\Delta x_n < 1$, then it can be approximated to first-order as

$$a \,\left(1 + \frac{b}{a}\Delta x_n\right) g(x_{n+1/2}) - a\, g(x_{n-1/2}) \approx 0. \tag{3.23}$$

Similarly, the first-order approximation in eq. 3.16 is valid only when its *diagonal* coefficient is small, that is when

$$\frac{(\tilde{H}_k(i,i) - \tilde{Q}(i,i))\,\Omega_j}{|\tilde{V}_x(i,i)|\,\Gamma_{j,j\pm1}} < 1.$$

or equivalently,

$$\frac{(\tilde{H}_k(i,i) - \tilde{Q}(i,i))}{|\tilde{V}_x(i,i)|} \Delta x_j < 1, \tag{3.24}$$

where the first term has units of inverse length. This is the origin of a characteristic length or a *mean free path* associated with every node i of the momentum space grid. For the case of 1D, it can be denoted as

$$\lambda_x(i) = \frac{|\tilde{V}_x(i,i)|}{(\tilde{H}_k(i,i) - \tilde{Q}(i,i))}, \tag{3.25}$$

where λ_x *does not* refer to eigen-value but to mean free path. In addition, it is not a point-wise quantity, but an average quantity obtained by integrating over the volume element Ω_i. Since $\tilde{H}_k(i,i)$ scales linearly with electric field, this mean free path depends inversely on the electric field when the scattering term is small.

Note that the above definition of "mean free path" is not the same as its conventional definition. The conventional definition refers to the mean free path of the electrons along their direction of travel due to *only* scattering

$$\lambda(i) = \left|\frac{\tilde{V}(i,i)}{\tilde{Q}(i,i)}\right|.$$

But this is *not* the critical length when solving the BTE with field in a one dimensional device. The critical length turns out to be the mean free path as defined by the *diagonal* term of the discretisation. It describes the rate at which *both* scattering and field influence the electrons travelling along the x-direction. At any given condition of scattering and field, the mean free path $\lambda_x(i)$ also depends on the energy $E(k_i)$ associated with the element i and the angle of its momentum with respect to the x-axis, $\theta_x(k_i)$. Note that, for the case of spherical bands, the angle θ_x is both the angle of the electron velocity and crystal momentum with respect to the x-axis.

Fig. 3.5 illustrates the typical values of mean free path versus energy and angle, for spherical non-parabolic bands and LO and LA scattering. Fig. 3.5(a) shows the value of λ_x in Å for zero electric field.
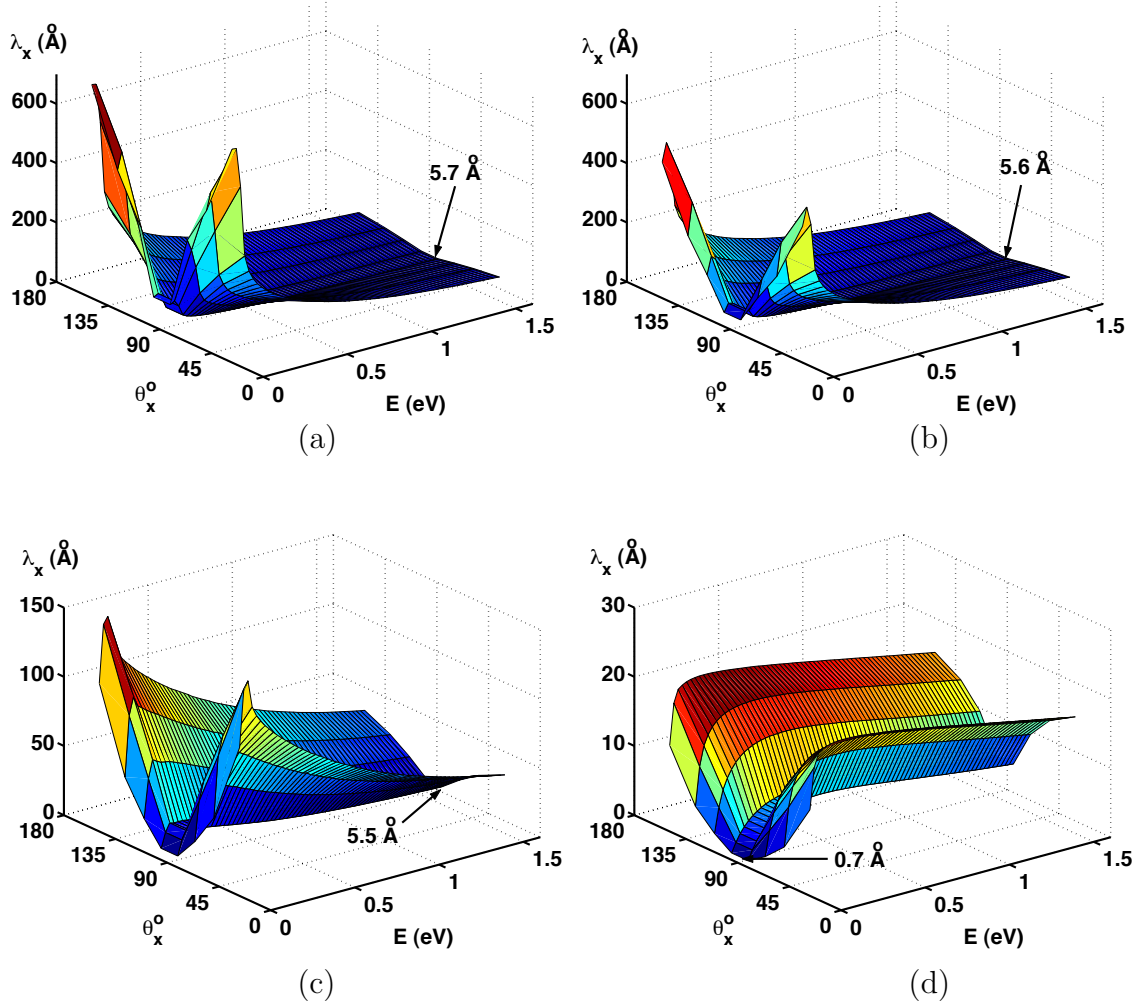
Fig. 3.5. Mean free paths λ_x versus energy E and direction θ_x for fields $\mathcal{E}_x$ of (a) 0 V/cm (b) 1×10^3 V/cm (c) 1×10^4 V/cm (d) 1×10^5 V/cm.

It is seen that at zero electric field, the highest energy electrons have the smallest mean free paths. The lower energy electrons tend to have longer mean free paths by comparison. However, low energy electrons that enter the 1D volume element at $\theta_x \approx 90°$ have smaller mean free path than those at enter at $\theta_x \neq 90°$. This occurs because low energy electrons have low velocities and if they travel along near-orthogonal trajectories, they tend to remain longer inside the 1D volume element and hence, scatter more often.

As the electric field is increased (Fig. 3.5 (b)), all the mean free paths get smaller, but their overall behaviour does not change significantly until fields of 1×10^4 V/cm. As the fields approach 1×10^4 V/cm, the behaviour of λ_x changes dramatically. The slowest moving electrons begin to show comparatively smaller mean free paths even though the high energy electrons still have the lowest mean free paths. Eventually, at very high electric fields (Fig.3.5 (d)), the slow moving electrons have the smallest mean free paths. As before, the reduction in the mean free path is more pronounced for normally-incident slow electrons and drops to 0.7 Å.
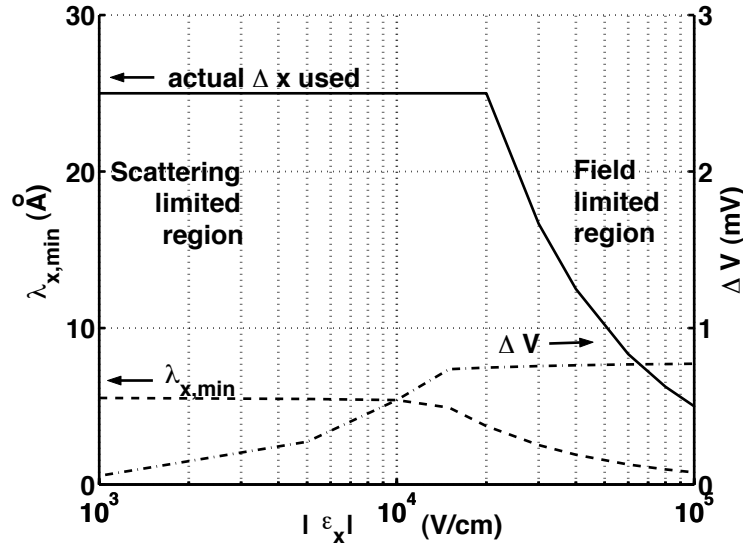
Fig. 3.6. Minimum mean free path and its associated potential drop versus electric field.

According to eq. 3.24, the grid spacing must be such that

$$\Delta x_j < \lambda_x(i), \quad \forall \ i. \tag{3.26}$$

This implies that the ultimate restriction on the grid spacing is set by the minimum mean free path $\lambda_{x,min}$. Fig. 3.6 shows the value of $\lambda_{x,min}$ versus increasing electric field. In addition, the potential drop across one minimum mean free path is plotted as $\Delta V = |\mathcal{E}_x|\lambda_{x,min}$ on the right axis. It is seen that at low electric fields, the restriction on grid spacing is nearly constant ≈ 5.7 Å and is independent of the field.

This is denoted as the scattering limited region. However, at high electric fields, the restriction on grid spacing varies inversely with field such that the potential drop across one grid spacing is limited to 0.6 mV. This is denoted as the field limited region.

If these values are used to grid a typical 1D 1μm device, then the resulting number of elements will be very large, approximately 500–1000. By contrast, a standard Scharfetter-Gummel discretisation of drift-diffusion equations produces about 100 elements. Indeed, these restrictions are very severe.

Therefore, the next section attempts to understand the error made when the above restriction is relaxed. That is known as the real space discretisation error. The goal is to find a criterion to increase the grid spacing and at the same time, maintain a reasonably small discretisation error. However, grid spacing in the BTE cannot be made as large as that in drift-diffusion. That is impossible because the grid spacing in drift-diffusion depends on the characteristic lengths of the macroscopic averages as defined by mobilities and diffusion constants, but the grid spacing in the BTE depends on the *microscopic* behaviour of the distribution function itself.

3.3.4   Discretisation error in the space term

Section 3.2 described a simple way to discretise the spatial term of the BTE. The previous section, however, indicated that there is a severe restriction on the grid spacing in doing so. In this section, the discretisation error in the average velocity in a device by using different grid spacing will be estimated. It will be seen that the error goes to zero as the grid spacing is reduced. But the grid spacing cannot be reduced all the way to 0, therefore, the objective here is to find a criterion for grid spacing that reasonable in size as well as accuracy.

For this study, consider a simple 1D low-high-low problem with a fixed potential profile as shown in Fig. 3.7. Since only the real space discretisation error is under study, the momentum space grid is fixed to $N_k = 4160$. However, this analysis holds for any type of grid in momentum space.

The real space grid is constructed in three ways, from coarse to fine, as shown in

Table 3.1. The grid spacing in the low field region is denoted as Δx_{lo} and that in the high field region is denoted as Δx_{hi}. The values are chosen in order to see the effect of varying the low and high field grid spacing independently. In each case, the total number of real space elements is denoted as N_x, and the total number of unknowns is $N = N_k \times N_x$.
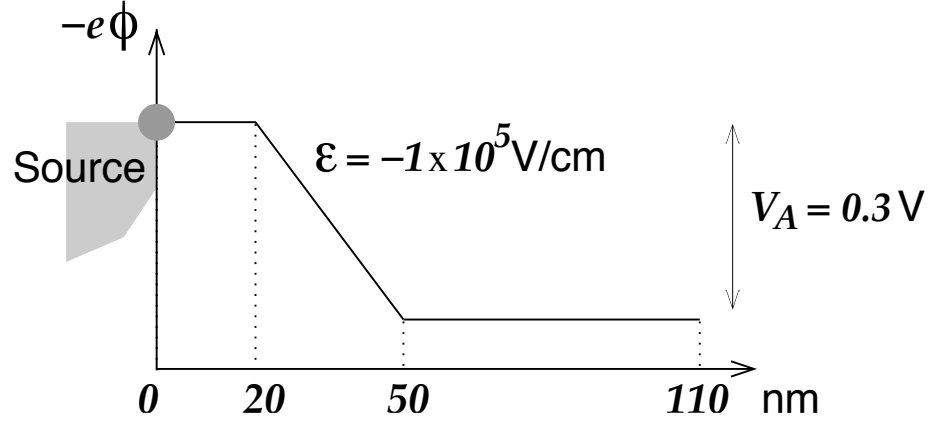
Fig. 3.7. One-dimensional device with fixed low-high-low potential profile.

$\Delta x_{lo}, \Delta x_{hi}$	N_x
50Å, 5Å	76
10Å, 10Å	110
10Å, 5Å	140

Table 3.1
Tabulation of grid sizes (N_x) used for error analysis.

The physical parameters of the device are LO and LA scattering, which correspond to a low-field mobility of $\mu_0 = 1350$ V/cm^2. Impurity scattering is not included here. The matrix equation for this device is set up as described in Section 3.3.2. The boundary conditions on the device are defined as described in Section 3.3.1

with one important difference. The left and right boundary distributions are scaled such that the *total* carrier concentration is 1×10^{16} cm^{-3} at both contacts. This requires two solutions of the matrix equation. The first solution is calculated for an arbitrarily scaled left boundary distribution and right absorbing contact. Then the second solution is calculated for an arbitrarily scaled right boundary distribution and left absorbing contact. Finally, the two solutions are superimposed appropriately so as to give the correct carrier concentration at the two contacts.
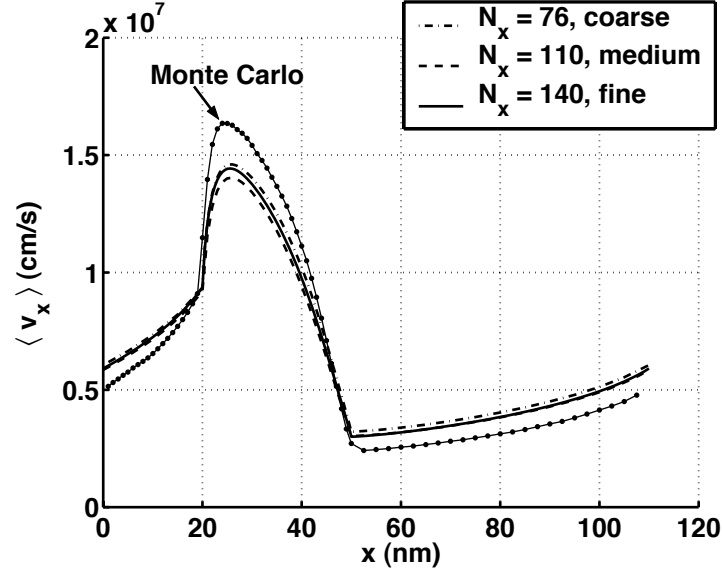
Fig. 3.8. Average velocity from the BTE solution of the potential profile in Fig. 3.7.

Fig. 3.8 shows the average velocity in the device calculated from the solutions for increasing N_x. It can be seen that as the grid is refined, the calculated velocity profile approaches the Monte Carlo velocity profile. However, the more interesting point to note is that the refinement from $\Delta x_{hi} = 10$ Å to $\Delta x_{hi} = 5$ Å does not produce an appreciable increase in accuracy. In fact, it appears that $\Delta x_{hi} = 5$ Å is a reasonably good value to use at $|\mathcal{E}| = 1 \times 10^5$ V/cm even though Fig. 3.6 prescribes a limit of 0.7 Å.

The reason for this is actually simple. The components of the distribution function that require the smallest grid spacing are also those that *do not* significantly affect the value of the average velocity. These components are either high energy electrons

where the distribution function is small or low energy normally-incident electrons whose x-directed velocity is small. Therefore a larger grid size that produces an error in these components of the distribution function will not necessarily produce a large error in the average velocities and energies. Hence the actual gridding criterion used in this work is relaxed to the solid line shown in Fig. 3.6. This choice of grid spacing is made on the basis of keeping the number of unknowns low and at the same time, achieving reasonable accuracy in the average quantities across the device.

3.4   Summary

This chapter described the finite volume discretisation of the spatial operator H_r in real space. This was done by using upwinding that transformed the hyperbolic operator into a positive-definite linear system of equations and preserved the stability of the discretisation. At this stage, a one dimensional device was used in order to demonstrate many important concepts of the numerical technique, such as imposing boundary conditions at the contact of the device and setting up the system of equations as a matrix equation. It was seen that there is a severe restriction on the grid spacing in this method but this restriction came from components of the distribution function that do not affect the average velocity or energy of the carriers. This conclusion could be drawn only by doing an analysis of the discretisation error in the average velocity in an example one-dimensional device with different grid spacings. Finally, note that in this chapter, a matrix equation was set up for the device and a solution was "found". No mention was made of *how* this solution was actually computed. This is, in fact, a non-trivial issue and will be discussed in detail in the following chapter.

4. Iterative methods for linear systems

4.1 Introduction

There are a variety of methods available for solving large linear systems of equations. What is the basis for choosing between them? In this chapter, several concepts are discussed that can guide a numerical analyst in making a knowledgeable choice of iterative method to apply to a specific problem. However, only a broad overview of the issues is given here without going into the details of the theorems and proofs. For excellent and complete discussions on all the issues discussed in this chapter, please refer to [76, 77].

An overview of this chapter is as follows. Section 4.2 presents a brief survey of iterative methods available for solving large linear systems of equations. This includes the characteristic behaviour of each method, the speed of solution and restrictions (if any). Finally, the section ends with a short description of the iterative method used in this work. This section can be skipped by those familiar with the field of matrix computations and numerical analysis.

Section 4.3 describes the preconditioned GMRES method used for finding solutions of the matrix equations that were described in Section 3.3.2. This includes a discussion on the performance of the preconditioner as well.

Section 4.4 describes the criterion for determining when an iterative method has reached its solution to within a prescribed degree of accuracy. This will be used to solve a model problem in order to demonstrate that the criterion is also justified from a physical point of view.

Section 4.5 then describes the meaning of the term "convergence" as used in this report. This is necessary because this report refers to not one but three different types of convergences. A brief discussion at this stage can help avoid future confusion.

Section 4.6 summarises the conclusions of this chapter.

4.2   A brief survey

Over the many years of advances in numerical analysis, many methods have been developed to solve the system of linear equations $Ax = b$. All of them *do not* compute A^{-1} directly because computing the inverse is an extremely slow N^3 process. This implies that even though these methods are fast for $N = 100$ problems, they slow down on the order of N^3 and become unacceptable even for reasonable size problems $N \approx 4000$. Needless to say, computing the inverse is also extremely wasteful because there is no need to calculate N^2 terms of A^{-1} when only the N terms of x are needed.

The types of methods developed for attacking linear systems of equations are of two types — direct methods, and iterative methods. These are described below.

4.2.1   Direct methods

Under direct methods, the two most important algorithms are Gaussian elimination and LU decomposition. Both these methods are of the order N^2 for dense matrices. Hence they are too slow to tackle large problems $N > 500$. However, they do not suffer from any failures or breakdown mechanisms that can plague iterative methods and can be applied to non-symmetric, ill-conditioned and indefinite matrices. They only suffer from zero pivot (occurrence of a zero on the diagonal) breakdown which can usually be corrected by partial or full pivoting. The sparse versions of these algorithms have lower complexity, of the order of order Nlu, where l is the lower bandwidth and u is the upper bandwidth. These can be used to solve problems of size up to $N \approx 4000$ but are severely limited by the high number of fill-ins for larger problems. For more details as well as sparse versions of these algorithms, please see [78].

4.2.2   Iterative methods

Under iterative methods, there are many algorithms that are significantly faster than direct methods. However, care must be taken to understand a particular method and all its restrictions before applying it to any problem. The following section presents a brief discussion on concepts of iterative methods required for the purposes of this project.

In any iterative scheme, the objective is to compute an estimate x^k that is closer to the solution x than all the previous estimates. The reasoning is that if the estimate is continuously "improved", then it will eventually approach the true solution. In order to find how close the estimate x^k is to the true solution, the concept of a residual $r^k = b - Ax^k$, is used. But does a small residual really imply a small error? This question will be addressed in Section 4.4. The most important and/or popular iterative methods are:

1. Jacobi iteration

   This is the simplest of all algorithms. It proceeds by decomposing $A = D+L+U$ where D = diagonal part and $L + U$ = the lower and upper triangular parts respectively. Hence when the starting guess is x^0, the updates can be computed using

   $$D \, x^{k+1} = -(L + U)x^k + b. \tag{4.1}$$

   The algorithm finishes when $x^{k+1} = x^k = x$. Of course, practically speaking, the above condition will only happen at times tending to infinity. As described earlier, the size of the residual can be used as a test for stopping the iterations. In the Jacobi method, instead of computing the residual directly as $b - Ax^k$, one of the following options can be used. One option is

   $$r^k = D(x^{k+1} - x^k), \tag{4.2}$$

   which is an inexpensive computation but is one step behind the update. Another option is

   $$r^{k+1} = -(L + U)(x^{k+1} - x^k), \tag{4.3}$$

   which is almost as expensive as a direct computation but estimates the current residual.

   The restriction on this algorithm is that the spectral radius of the Jacobi iteration matrix $\rho(D^{-1}(L+U))$ must be less than 1 or the residual will not decrease with iteration. Such a condition is satisfied automatically by a diagonally dominant matrix. There are no requirements on symmetry or positive-definiteness

but the algorithm is the *slowest* iterative method among all the iterative methods (see section on Gauss-Siedel and SOR).

2. Gauss-Siedel and Successive over-relaxation

   Gauss-Siedel and Successive over-relaxation (SOR) are included in the same section here because they are very similar in implementation even though their speed of solution can be significantly different. The iteration scheme for both is defined as

$$(\frac{1}{\omega}D + L)x^{k+1} = (-U + \frac{(1-\omega)}{\omega}D)x^k + b. \tag{4.4}$$

where ω is the over-relaxation parameter. For Gauss-Siedel, $\omega = 1$ and for SOR, $\omega \in (0,2)$. Even though $0 < \omega < 1$ is technically under-relaxation, these days the term "over-relaxation" is used for both $\omega \in (0,1) \cup (1,2)$. The relation for the residual in terms of the updates is

$$\begin{aligned} r^{k+1} &= (-U + \frac{(1-\omega)}{\omega}D)(x^{k+1} - x^k), \\ r^k &= (\frac{1}{\omega}D + L)(x^{k+1} - x^k). \end{aligned} \tag{4.5}$$

In general, Gauss-Siedel with $\omega = 1$ is slower than SOR with an optimal ω. The optimal value of ω can be determined from the spectral radius of the corresponding Jacobi iteration matrix $\rho(D^{-1}(L+U))$,

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho^2}}. \tag{4.6}$$

A natural assumption here is that all the the eigenvalues of a matrix need to be computed so that the spectral radius can be determined as $\rho = \max\{|\lambda_i|\}$. This is a virtually impossible task for a large matrix. However, the amount of work can be significantly reduced by the use of a very clever algorithm known as the *Lanczos/Arnoldi* method which can compute the extremal eigenvalues of a matrix without having to compute those in between. But finding the spectral radius of a matrix is always a computationally expensive task. Therefore, in cases of dire need, the value of ω_{opt} can be determined experimentally by running

iterations with different values of ω and extracting the value that reduces the number of iterations significantly. There is also a bound on the spectral radius of any matrix

$$\rho(D^{-1}(L+U)) \leq ||D^{-1}(L+U)||_\infty, \tag{4.7}$$

but it is not always useful because it is not a tight bound.

The restriction on this algorithm is the same condition on the spectral radius as that for the Jacobi iteration matrix. Again diagonally dominant matrices are well-suited for these methods. Gauss-Siedel and SOR are faster than Jacobi but usually too slow for large problems. This is because in all the above three methods, the residual decreases at a slower rate with each iteration such that

$$||r^{k+1}|| = c_k||r^k||, \tag{4.8}$$

where c_k is a monotonically decreasing function of k. Therefore, although the initial decrease in the residual is impressive in all the above methods, the algorithms stagnate very quickly and take a very long time to knock out the remaining error. This behaviour is not mysterious but is a natural consequence of the spectral decomposition of these methods. This is explained below and used to introduce the concept of multi-grid methods.

First the iteration scheme of the Jacobi and SOR-type methods, is represented in a general form, as

$$Mx^{k+1} = Nx^k + b, \tag{4.9}$$

where M and N are the left and right matrices of the iterative step. With a little linear algebra, the correction at each step $x^{k+1} - x^k$ is obtained as

$$x^{k+1} - x^k = (M^{-1}N)(x^k - x^{k-1}) = (M^{-1}N)^k(x^1 - x^0). \tag{4.10}$$

Since A is taken to be a diagonally-dominant matrix here, the spectral radius of the iteration matrix $(M^{-1}N)$ will be less than 1.

Suppose the eigenvalues of $(M^{-1}N)$ are λ_i corresponding to the eigenvectors v_i. In this case, they can be ordered as $1 > \lambda_1 \geq \lambda_2 \ldots \geq \lambda_N > 0$. Note that

this assumes that the smallest eigenvalue is strictly > 0 (*i.e.* matrix is positive-definite). Now, if the first correction is expressed in terms of the eigenvectors as $(x^1 - x^0) = \sum_i c_i v_i$, then the $(k+1)$th correction is

$$x^{k+1} - x^k = (M^{-1}N)^k \sum_i c_i v_i = \sum_i c_i \lambda_i^k v_i. \qquad (4.11)$$

The question now is, how does the correction go towards zero as the iterations proceed? The answer lies is in the above equation — the terms corresponding to the smallest eigenvalues disappear quickly as k increases but the terms corresponding to the largest eigenvalues decrease much more slowly. This is a characteristic behaviour of Jacobi and SOR-type methods. Hence, in order to accelerate the speed of these methods, *Multi-grid* procedures need to be used.

Briefly, a multi-grid method can be described as solving the *same* problem for different discretisations in order to quickly reduce the error components along the large eigenvalues. If the eigenvalues of the iteration matrix on a coarse grid $M_c^{-1}N_c$ are calculated and compared against those of a fine grid iteration matrix $M_f^{-1}N_f$, then it will be seen that as the grid spacing h increases, the largest eigenvalues *decrease* and the smallest eigenvalues (at the lower end of the spectrum) simply get left out. This leads to two conclusions. Firstly, the error that has a large eigenvalue on a fine grid will have a small eigenvalue on the coarse grid and therefore, can be reduced quickly by performing iterations on the coarse grid. Secondly, the errors along the small eigenvalues on a fine grid can be reduced quickly by performing iterations on the fine grid itself, because they are not captured by the coarse grid at all. Thus, a multi-grid method alternately performs fine and coarse grid iterations until the error reduces globally along all eigenvalues. In fact, this method has had much success with Jacobi and SOR-type methods because it is well suited to their particular spectral behaviour. An example of a multi-grid method is the use of 1-flux grid to accelerate the Jacobi iterative solution of an M-flux grid in the 1D-SMA program [79].

Although, multi-grid methods have been proven to work well for systems arising

out of discretisation of elliptic partial differential equations, they may not be the optimal method for solving the BTE. The reason is because there is only a limit to which Jacobi and SOR-type methods can be accelerated. By nature they are the slowest iterative methods compared to methods that we will discuss below. However, these methods must not be entirely discounted. In cases of small ($N < 1000$), reasonably sparse matrices or large, extremely sparse ($< 0.01\%$) matrices, simple methods such as SOR with optimal ω (with or without multigrid acceleration) turn out to be quick to program and reasonably efficient.

3. Conjugate-Gradient methods

From the Conjugate-Gradient (CG) method onwards, the iterative methods described below differ significantly in philosophy from the previous methods. The CG method is constructed in the same way out of two different approaches — both of which have historical significance and will be described below. The first approach is to solve the system of equations $Ax = b$, of size N, by forming the functional

$$\phi(x) = \frac{1}{2}x^T Ax - x^T b, \qquad (4.12)$$

and try to minimise it. If A is symmetric and positive definite, then the minimum value of $\phi(x)$ is $-\frac{1}{2}b^T A^{-1}b$ which occurs when $x = A^{-1}b$.

In such a problem, the objective is to search for directions that minimise the error. Suppose the CG method is at some some estimate for the solution x^k then the minimisation direction is computed as the direction of the gradient

$$-\nabla\phi(x^k) = -Ax^k + b = r^k, \qquad (4.13)$$

which is, in fact, the residual. However, if this new residual only has components along the direction of residuals that have been found previously then there will be no reduction in the error. Therefore, the solution is to build a set of directions that are *orthogonal* to all the previous directions. This is done by calculating a

new A-orthogonal set of vectors p^k such that

$$p^{jT} Ap^k = 0 \text{ if } j < k. \tag{4.14}$$

These vectors are calculated from the residuals so that they satisfy the minimisation condition and the orthogonality condition using the following rules

$$\begin{aligned}
\beta^{k+1} &= r^{kT} r^k / (r^{k-1T} r^{k-1}) \text{ if } k > 1 \text{ else } \beta^1 = 0, \\
p^{k+1} &= r^k + \beta^{k+1} p^k.
\end{aligned} \tag{4.15}$$

Therefore, the new updates are calculated by finding the magnitude α^{k+1} of the direction p^{k+1} that will improve the solution and decrease the residual as follows

$$\begin{aligned}
x^{k+1} &= x^k + \alpha^{k+1} p^{k+1}, \\
r^{k+1} &= r^k - \alpha^{k+1} Ap^{k+1},
\end{aligned} \tag{4.16}$$

where The required value is found to be $\alpha^{k+1} = r^{kT} r^k / (p^{k+1T} Ap^{k+1})$. Note that every step requires only *one* matrix vector multiplication and *two* previous residuals to compute the updated solution and the current residual. The value of the current residual actually falls out automatically as a part of the algorithm without explicitly calculating it.

As a result of the above set of rules, the residuals themselves are also orthogonal to all the previously calculated residuals

$$r^{jT} r^k = 0 \text{ if } j < k. \tag{4.17}$$

Hence, the residuals form an *orthogonal* basis for the set $[r^0 \ Ar^0 \ \dots \ A^{k-1} r^0]$ which is the space of vectors due to the repeated action of A on a vector r^0. This set is called as the *Krylov* sub-space and will play an important role in all the following discussions. The proofs of the above relations are not obvious at all but they cannot be discussed in detail because they go beyond the scope of this report. Interested readers can refer to [76, 80].

The above definition is the traditional CG method (first proposed by Hestenes and Stiefel). An equivalent but more powerful definition is the *Lanczos* method. The Lanczos method was developed to calculate the eigenvalues of symmetric positive definite matrices and works as follows.

It can be seen from eq. 4.15 that the relation between the set of k vectors $P_k = [p^1 \ldots p^k]$ and the set of k residuals $R_k = [r^0 \ldots r^{k-1}]$ is

$$R_k = B_k P_k, \tag{4.18}$$

where B_k is a bi-diagonal matrix of elements $B_k(i,i) = 1$ and $B_k(i, i+1) = -\beta_{i+1}$. Note that the vectors in P_k are A-orthogonal such that $P_k^T A P_k = \text{diag}(p^{i^T} A p^i) = D_P$, a diagonal matrix. Also the vectors in R_k are an orthogonal basis of the Krylov sub-space such that $R_k^T R_k = \text{diag}(r^{i^T} r^i) = D_R$. However, it is more convenient to work with an orthonormal basis, therefore the normalised vectors $Q_k = R_k D_R^{-1}$ will be used instead. Now the product $Q_k^T A Q_k$ has a very significant form

$$
\begin{aligned}
Q_k^T A Q_k &= D_R^{-1} R_k^T \ A \ R_k D_R^{-1} = D_R^{-1} \ B_k^T (P_k^T A P_k) B_k \ D_R^{-1} \\
&= D_R^{-1} \ (B_k^T D_P B_k) \ D_R^{-1} = T_k, \tag{4.19}
\end{aligned}
$$

where T_k turns out to be a $k \times k$ tri-diagonal matrix. Therefore, the action of the Krylov basis vectors on the matrix A converts it to a tri-diagonal matrix. This is the Lanczos method.

Several important observations can be made about this transformation. First of all, all the above derivations (and their proofs) require a symmetric matrix. Secondly, important eigenvalue information about A is obtained from T_k. The eigenvalues of T_k are found to be a *subset* of the eigenvalues of A (unless $k = N$, in which case the eigenvalues of T_k are exactly the eigenvalues of A). However, the usefulness of this algorithm lies in the fact that as k increases, the extremal (smallest and largest) eigenvalues of T_k approach the extremal eigenvalues of

A with greater and greater accuracy. This allows the estimation of many important properties of the matrix such as condition number and spectral radius without the prohibitive expense of computing all its eigenvalues.

But what is the speed of the CG for finding the solution of the linear system $Ax = b$? It is clear that the CG method can reduce the error components in the extremal eigenvalues very quickly but how long does it take to reduce the error in the intermediate eigenvalues? Detailed analyses have shown that the number of iterations required to reduce the error globally is proportional to $\sqrt{\kappa_2(A)}$, where κ_2 is the spectral condition number of A. This in fact turns out to be rather slow even for reasonably well-conditioned problems. In fact, there were problems reported for which convergence took place only at the last step $k = N$.

But suppose instead of $Ax = b$, an equivalent system $K^{-1}Ax = K^{-1}b$ is solved. The number of iterations required is now proportional to $\sqrt{\kappa_2(K^{-1}A)}$. If the matrix K is chosen such that it is symmetric and somewhat crudely approximates A, then $K^{-1}A \approx I$ and its condition number will be much smaller (< 100). This increases the speed of the conjugate-gradient method considerably and is known as *Preconditioning*. The Jacobi and SOR-type methods do not have such an advantage and can only be accelerated up to a limit. The trade-off is the simplicity because firstly, finding a suitable preconditioner is not trivial and secondly, a preconditioned-CG type algorithm performs a lot more computations per iteration than Jacobi and SOR-type methods.

Another point to note is that the mention of an inverse (K^{-1}) is always of some concern. Can there be any gain from an algorithm that requires an inverse of a matrix as big as A (even if somewhat simpler)? The answer is that the actual implementation of a preconditioned method requires only two matrix operations per iteration — the product $v = Ar$ and the solution of $Kw = v$. The explicit inverse of K is never computed.

Finally, the restrictions on conjugate gradient method are that A must be symmetric and well-conditioned (or suitably preconditioned). The matrix also must be positive-definite otherwise the method breaks down (diagonal entry of T_k becomes 0).

4. Generalised Minimal Residual method

Preconditioned CG had much success with solving problems with symmetric matrices but it cannot be applied to non-symmetric matrices. For this purpose, many variants of the CG method were developed — namely conjugate gradient square (CGS) which applies the Lanczos method to $A^T A$ and bi-conjugate gradient method (Bi-CG) which applies the Lanczos method to generate a bi-orthogonal set of vectors with respect to A and A^T. The problem with these methods is that the residual is not guaranteed to decrease with each iteration and could oscillate wildly from one iteration to the next, although it would eventually decrease. Indeed it has been said that "Bi-CG is not for the faint-hearted" [81]. In addition, there is the danger of generating a zero or near-zero pivot in the algorithm which breaks down the method. There can also be instances when the underlying Lanczos process might break down. Hence stabilisation methods were developed to handle this problem — namely, Bi-CGSTAB (Bi-CG stabilised) and Bi-CG with look-ahead (composite step Bi-CG) [82].

For the case of solving the BTE in this work, a different approach was chosen. It is known as the Generalised Minimal Residual (GMRES) method and was originally developed by Saad and Schulz [83]. This method is described briefly here by using the concepts introduced in the previous section. The objective in GMRES is also to find a set of vectors that form an orthonormal basis for the Krylov subspace $[r^0 \ Ar^0 \ \dots A^{k-1}r^0]$ which are represented as $V_k = [v^1 \dots v^k]$. Every step, therefore, requires one matrix multiplication, similar to CG. Now, when the transformation using V_k is applied to A, the result is

$$V_{k+1}^T AV_k = \tilde{H}_{k+1}, \qquad (4.20)$$

where $\tilde{H}_k$ is a $(k + 1) \times k$ upper Hessenberg matrix (and not a tridiagonal matrix). For the purposes of this section and to retain the notation of Saad and Schulz, $\tilde{H}_k$ does not refer to the hyperbolic field operator as it does in the rest of the report. Now, the new update is computed as

$$x^k = x^0 + V_k y_k, \tag{4.21}$$

where y_k is the *least squares* solution of $\tilde{H}_{k+1} \ y_k = ||r^0||e_1$. This is known as the *Arnoldi* method. Again the residual is obtained automatically from

$$r^k = r^0 - V_{k+1}\tilde{H}_{k+1}y_k. \tag{4.22}$$

without explicitly computing it. In this case, it is necessary to store *all* the vectors in V_{k+1} in order to be able to compute the updated solution and residual at each step. Compare this with CG where only two previous vectors are needed to do the same. Of course, this significantly increases the memory requirements of GMRES. Therefore, the most common implementations of GMRES stop after a number of iterations $k = m$, construct the most recent update x^m and use it as a first guess to restart the next set of iterations. This is known as the restarted GMRES(m) method.

The advantage of the GMRES method is that the residual is *guaranteed* to decrease with each iteration. Its convergence behaviour is similar to that of CG. The extremal eigenvalues of $\tilde{H}_k$ approximate the extremal eigenvalues of A with greater accuracy as the iterations proceed. The speed at which the residual decreases also depends on the condition number of A (same as CG). Hence, the speed of GMRES can be increased considerably by preconditioning. This is done by transforming the original problem into $K^{-1}Ax = K^{-1}b$ where it is expected that the eigenvalues of $K^{-1}A$ form a more well-conditioned spectrum than those of A. In this case, the preconditioner K need not be symmetric but it must approximate A in some sense. This is not always easy to do but it must be noted that under the right conditions and the optimal K, GMRES

produces near-constant rate of reduction of the residual with every iteration and converges in < 50 iterations.

Another advantage is that GMRES is a very robust algorithm; in fact, Saad and Schulz also prove that GMRES does not break down even when applied to indefinite matrices. The speed of GMRES is comparable to the many Bi-CG type methods but its indisputable disadvantage is the memory required to store m previous residuals, where m is the restart value. In cases where memory is an issue, it would be more advantageous to use a Bi-CG type method (such as [82]) which has much lower memory requirements.

This concludes the discussion of iterative methods. In this report, three of the above mentioned methods are used to solve all the matrix equations, namely, sparse LU [76], SOR with optimal ω and preconditioned GMRES(m). Sparse LU is used for small systems of equations arising from the discretisation of drift-diffusion and Poisson equations (both in 1D and 2D devices). SOR with optimal ω is used for preconditioner solves. Finally, preconditioned GMRES(m) is used to solve the matrix arising from the direct discretisation of the BTE. Therefore, it is necessary that the discretisation of the BTE must be done correctly so that the underlying matrices are diagonally dominant and positive-definite, and the *best* performance can be obtained from the iterative methods.

4.3   Preconditioned GMRES

In this work, preconditioned GMRES(50) was chosen as a matrix solver for the BTE, where 50 denotes the number of iterations before restart. This section discusses the memory, speed and performance of this method.

In this work, the GMRES method is implemented in such a way that no memory is required to store the elements of the matrix itself. This is because they are computed on the fly whenever they are required in the matrix-vector multiplication step (as discussed in Section 3.3.2). Hence, the only memory requirement comes from the maximum number of vectors that need to be stored in order to compute the solution

update. This is given by the restart value of the GMRES method and is 50, in this case. Therefore, for a problem size of $N \approx 10^6$, the memory requirement turns out to be on the order of 0.5 GB, which can be handled easily by modern computers. The largest problem size considered in this work $N = 10^7$ has a memory requirement of 2.2GB, all of which is used only to store the intermediate vectors that are used to compute the solution update.

The speed of the preconditioned GMRES method, on the other hand, depends mainly on the type of preconditioner. Significant improvements can be made in solution speeds simply by choosing the right preconditioner. However, in many areas of numerical analysis (including this work), the choice of preconditioner is still a matter of experience and is not an exact science. Hence in this section, four potential preconditioners are investigated. They are

$$
\begin{aligned}
K &= I, \quad \text{unpreconditioned,} \\
K &= D_A, \quad \text{the diagonal of A,} \\
K &= H, \quad \text{only the hyperbolic terms in A and,} \\
K &= H - D_Q, \quad \text{the hyperbolic terms and the diagonal} \\
&\qquad\quad \text{of the scattering terms in A.}
\end{aligned}
\tag{4.23}
$$

The sample problem chosen for analysis is the same at that described in Section 3.3. The sample grid used here is $N_x = 140$ and $N_k = 528$ giving a total of $N = 73920$ unknowns. The matrix equation is therefore solved with the above four preconditioners and the speed of solution is plotted as the norms of the residuals versus iteration in Fig. 4.1. It is seen that the unpreconditioned method is unacceptably slow whereas the fastest preconditioner $(H - D_Q)$ takes only 15 iterations.

This result also makes sense physically because $(H - D_Q)$ is, in fact, the relaxation time approximation to the BTE, where the collision integral is approximated by a diagonal matrix with energy dependent relaxation times. Hence even though the relaxation time approximation is not a good method to solve the BTE itself, it makes an extremely effective preconditioner for solving the BTE.
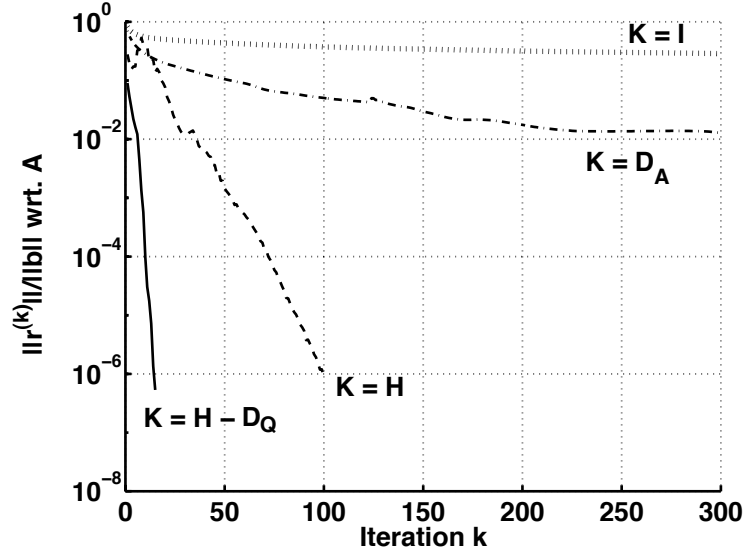
Fig. 4.1. Residual versus iteration for $N = 73920$ unknowns.
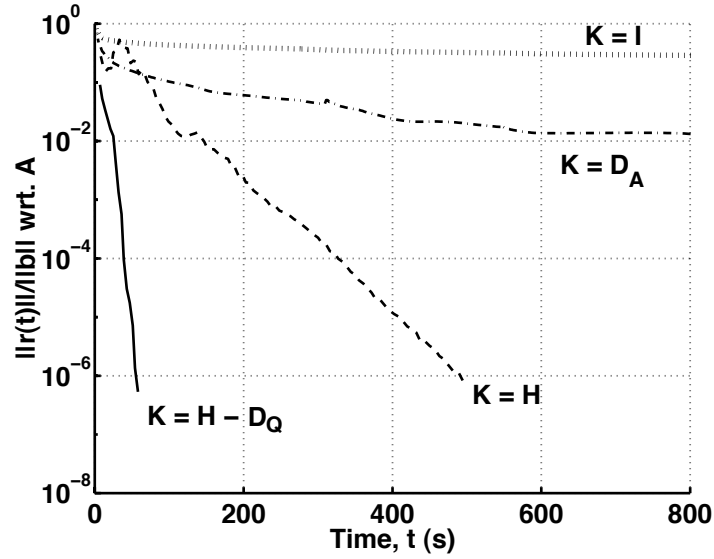
Fig. 4.2. Residual versus time for $N = 73920$ unknowns.

The same residual is plotted versus computation time in Fig. 4.2. The fastest preconditioner $(H - D_Q)$ takes about 50 seconds on a 400 MHz, UltraSparc2. However, this only shows that the preconditioner $(H - D_Q)$ is fast for one particular grid size. As the grid size is increased (for the same problem), the number of unknowns N increases and the time for solution also increases. Therefore, the performance of the

above preconditioner is plotted as the time taken for solution versus N in Fig. 4.3.

Fig. 4.3. Time for solution versus number of unknowns N for the low-high-low problem. The numbers on the plot denote the number of iterations

The above figure shows that the performance of the preconditioner $(H - D_Q)$ scales as $N^{1.2}$. It takes about 35 minutes to solve a problem with $N = 10^6$ on a 400 MHz, UltraSparc2. The dashed lines on the plot are used to show that the time taken for solution for other types of problems could either be higher or lower depending on their condition number.

The condition number (see Section 4.4) is a very critical issue when solving large systems of equations. Although the condition number of the discretisation used in this work cannot be analysed in detailed here, an important observation can still be made. This is done mainly by looking at the conditioning analysis of the Scharfetter-Gummel discretisation of drift-diffusion equations (for an excellent treatment see [84]). The results of that analysis indicate that the condition number of the discretised drift-diffusion system depends on the type of problem in the following way:

1. At a $(p\text{-}n)$ junction with contacts at both sides, the condition number of the discretised drift-diffusion system is small and nearly constant for moderate doping and moderate bias.

2. At a (p-n) junction with contacts at both sides, the condition number of the discretised drift-diffusion system increases when the doping is high and/or a large reverse bias is applied.

3. For two back-to-back (p-n) junctions with contact at two sides, the condition number depends exponentially on the difference between the maximum and the minimum of the potential profile, $\exp(e(\phi_{max} - \phi_{min})/(k_B T_L))$. This only happens when the region between the junctions is left floating. Large condition numbers imply that the discretised drift-diffusion system cannot be solved (even with Scharfetter Gummel). And even if a solution could be found, the large round-off errors in its computation would make it meaningless. Physically, this means that the device is highly unstable and needs to be stabilised by placing a contact at its floating regions.

Therefore, it is expected that condition number of the discretised BTE will also depend on the type of problem in the same way. However, due to the lack of a suitable Scharfetter-Gummel discretisation (in real space) and the presence of detailed scattering rates that vary over large orders of magnitude, the condition number of the discretised BTE ends up being larger than that of drift-diffusion for the same type of problem. Hence, this implies that the the discretisation developed in this work cannot be used to solve the BTE in devices such as n^+-p^+ junctions, strongly reverse-biased junctions and of course, floating regions.

To summarise, in this section, preconditioned GMRES with $(H - D_Q)$ as pre-conditioner was devised to solve the BTE. It was chosen because it was the optimal combination of speed and memory requirement, and will henceforth be used to solve the BTE for all the simulations in this report. However, it was also noted that there are cases where the process of discretisation does not automatically guarantee a solution, particularly those in which the underlying problem is physically ill-posed or ill-conditioned.

4.4 Error versus residual

The error and residual are two critical concepts in solving any linear system of equations. So far, the residual $r^k = b - Ax^k$ has been referred to as a "good" measure of the distance from the true solution. When the residual is deemed "small" enough, the iterations can stop. Note that the term "small" is not absolute but refers to the size of ratio $||r^k||/||b||$, where the norms can be computed in either the ℓ_1, ℓ_2 or ℓ_∞ sense.

Here the residuals need to be contrasted with errors. For Jacobi and SOR-type methods, in which the iteration step can be written as $Mx^{k+1} = Nx^k + b$, the error and the residual are closely related

$$e^{k+1} = x^{k+1} - x = M^{-1}N(x^k - x) = M^{-1}Ne^k,$$

where

$$r^{k+1} = b - Ax^{k+1} = NM^{-1}(b - Ax^k) = NM^{-1}r^k. \tag{4.24}$$

For any general method, the error is related to the residual as follows:

$$e^k = A^{-1}(Ax - Ax^k) = A^{-1}(b - Ax^k) = A^{-1}r^k, \tag{4.25}$$

which is difficult to compute because of the presense of the inverse. But this relation can be used to find bounds on the error by invoking the property of a vector norm and its sub-ordinate matrix norm ([80] p.166), which gives

$$\frac{1}{\kappa(A)}\frac{||r^k||}{||b||} \leq \frac{||e^k||}{||x||} \leq \kappa(A)\frac{||r^k||}{||b||}, \tag{4.26}$$

where $\kappa(A)$ is the condition number of a matrix calculated as

$$\kappa(A) = ||A|| \times ||A^{-1}|| \tag{4.27}$$

for any choice of matrix norm. It does not matter what norm is used to estimate the condition number, the condition number behaves consistently in all matrix norms. In the 2-norm sense, the condition number has a special name, it is called the *spectral*

condition number and can also be determined from the eigenvalues of the matrix AA^T,

$$\kappa_2(A) = \frac{|\lambda_{max}(AA^T)|^{1/2}}{|\lambda_{min}(AA^T)|^{1/2}}. \tag{4.28}$$

Note that eq. 4.26 is a very important relation which states that if a matrix is well-conditioned *i.e.* $\kappa(A) <= 10^3$, then a small residual automatically implies a small error. But if the condition number is extremely large $\kappa(A) \gg 10^4$, then a small residual is meaningless and it may not be possible to get a solution of the matrix $Ax = b$ by an iterative method in the first place.

Therefore, the size of the residual is a mathematically reasonable criterion for determining if an iterative method has reached its solution. But what does the size of the residual mean in terms of *physical* quantities such as carrier concentration and average velocity? This is analysed below.
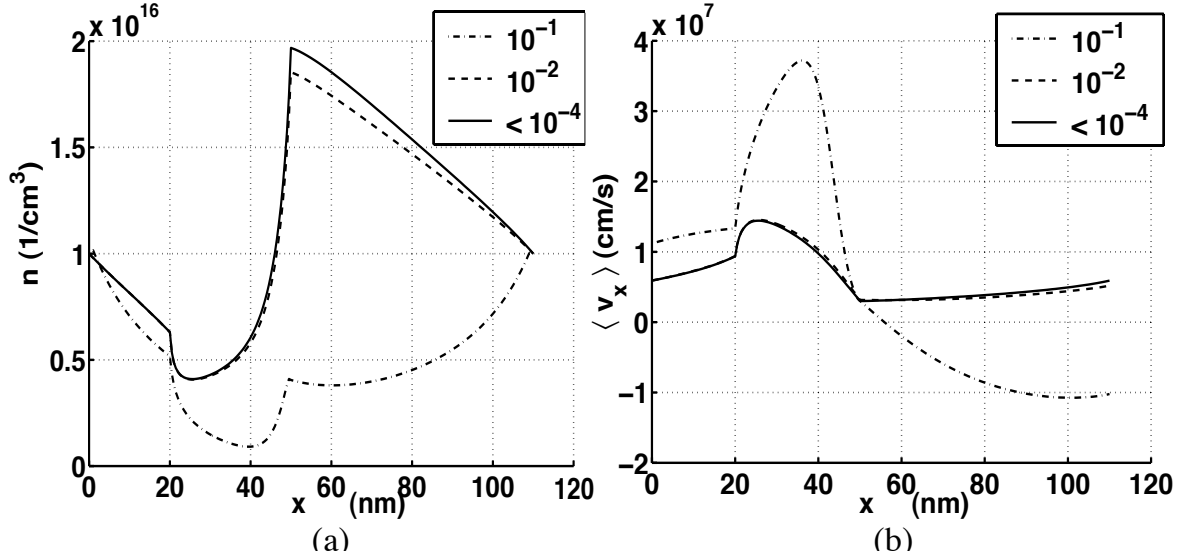
Fig. 4.4. Plot of (a) carrier concentration and (b) average velocity at decreasing values of residual norm for $N_x = 140$ and $N_k = 4160$.

Consider the example problem described in Section 3.3 with $N_x = 140$ and $N_k = 4160$. Fig. 4.4 shows the solution for this problem, in terms of carrier concentration and average velocity, plotted at three *decreasing* values of the residual norm ($||r^k||/||b||$). Here it is seen that the residual is also a good indication of the

accuracy of the average quantities and hence can be used as a check for convergence.

The conclusion here is that it is acceptable to use the residual as a check for error, as long as the matrix is well-conditioned. All the above iterative methods work well only on well-conditioned (or well pre-conditioned) matrices with sparse structure. In cases of an ill-conditioned matrix, there is no other option but to use a direct method because direct methods can tolerate larger condition numbers.

4.5   Three types of convergence

In any numerical scheme, there are many different types of convergence that must be analysed and that can become quite confusing. Therefore, this section lists all the "convergences" that will be discussed in this report and the context in which they arise.

1. Iterative solution

   When the BTE is discretised (for a given grid spacing), a large matrix equation is obtained. This system is solved by using an appropriate iterative method and monitoring the reduction in the residual at each step of the iteration. When the residual at step k, $||r^k||_2$, falls below a prescribed accuracy $\epsilon||b||_2$ (usually $\epsilon \approx 10^{-4}$ to 10^{-6}), then the iterative method is said to have converged to the solution.

2. Discretisation error

   When the grid spacing in either real space or momentum space or both is reduced, then the discretised BTE produces an even larger matrix equation. This matrix equation can be solved by using the same iterative method described above. Now, if the solution on the new grid does not change appreciably from that on the old grid, then the it is said to have converged to the true solution of the BTE.

3. Coupling to Poisson equation

   The procedure to couple the transport equation (BTE) self-consistently with the Poisson equation will be discussed in detail in a later section. Here it

can be briefly described as follows. Using a given potential profile, the BTE is solved for carrier concentration in the device. This carrier concentration is converted into a mathematically equivalent quasi-Fermi level and fed into to the Poisson solver. The Poisson solver then finds a *new* potential profile. When the difference between the old potential profile and the new potential profile is small, say $||\psi^{new}(x) - \psi^{old}(x)||_2 \leq 1$ mV, then the BTE-Poisson equations are said to have converged.

4.6   Summary

To summarise, this chapter presented a brief survey of iterative methods available to solve large matrix equations. The iterative method used in this work was described and a suitable preconditioner was chosen in order to obtain fast solution speeds. Finally the convergence criterion was chosen in terms of the residual, which is meaningful from both a mathematical and a physical point of view. Finally, the different contexts of the term "convergence" as used in this work were discussed.

In the next chapter, this method will be applied to solve the BTE self-consistently with Poisson equation, such that it can be used for actual device simulations.

5. Device simulations

5.1 Introduction

This chapter describes the application of the numerical technique developed in this work to device simulation. In order to so, Section 5.2 describes the method used to solve the BTE self-consistently with the (non-linear) Poisson equation. This must be done because the carriers move under the influence of the potential (field) and in turn, they carry charge which changes the potential inside the device. Therefore, a solution to the transport equation is not complete with being consistent with the solution of the Poisson equation in the device. The method used here to couple the BTE-Poisson equations is shown to have smooth and stable convergence, comparable to drift-diffusion.

Section 5.3 describes the application of this method to study transport in a one-dimensional 50nm n^+-p-n^+ diode. It will be compared to the solution from three macroscopic transport models, drift-diffusion (DD), hydrodynamic (HD), and energy transport (ET). This study will show that nanoscale devices operate in a quasi-ballistic regime, where the carriers travel nearly ballistically across a small region of the channel near the source. Standard macroscopic models fail in these regimes because they are based on collision-dominated assumptions.

Similarly, Section 5.3 describes the application of this method to study transport in a two-dimensional 50nm ultra-thin body dual-gate nMOSFET. In this case, the presence of quasi-ballistic transport in nanoscale transistors can clearly be seen.

Section 5.5 summarises the conclusions of this work.

5.2 Coupling with the Poisson equation

This section describes the the method used to solve the BTE self-consistently with the Poisson equation. The method used is known as Gummel iteration [42] and it is
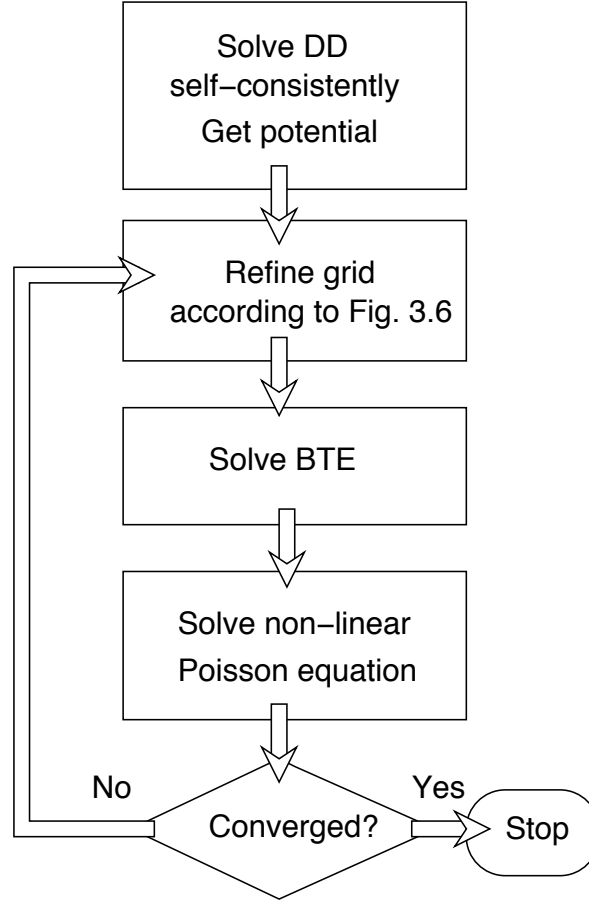
illustrated in Fig. 5.1.

Fig. 5.1. Coupling BTE to Poisson equation.

At any given bias, *first* the drift-diffusion equation is solved for its self-consistent potential profile. Using this profile as a starting guess, the BTE is solved for carrier concentration in the device. This carrier concentration is converted into a mathematically equivalent quasi-Fermi level and fed into to the non-linear Poisson solver. The Poisson solver then finds a *new* potential profile. This entire process is referred to here as one BTE-Poisson loop. The coupling between the BTE and the Poisson equation is said to stable if the difference between the old potential profile and the new potential profile decreases monotonically with each loop for all conditions of bias. This is

shown in Fig. 5.2 for an example one-dimensional device (described in Section 5.3).
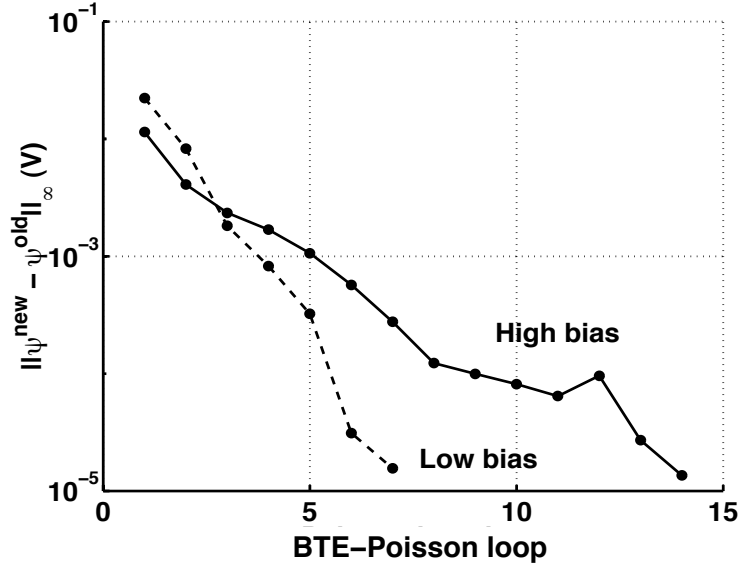

Fig. 5.2. Correction in the potential profile versus the number of the BTE-Poisson loop.

Note that the convergence is smooth and stable down to very small errors of 10^{-5} V, which is comparable to the convergence of self-consistent drift-diffusion simulations. This is unlike Monte Carlo which typically shows noise in its convergence and finds it difficult to reduce the error below $k_B T_L/e$. Also it must be noted here that every time the loop goes back to solve the BTE, the previous solution of the BTE is used as starting guess. This reduces the number of iterations required to solve the BTE in each loop shown in Fig. 5.3 (for the same device). Hence there is a substantial saving in the time taken for self-consistent solutions because the transport (BTE) step is the most time-consuming step in the loop.
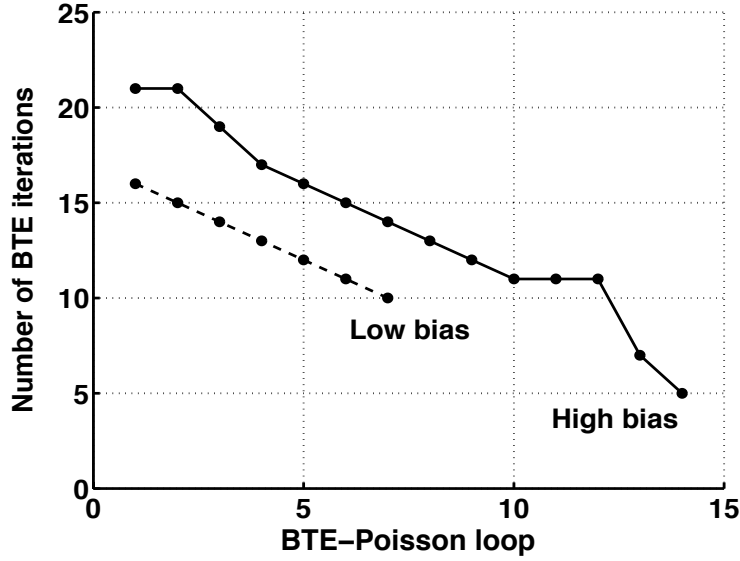
Fig. 5.3. Number of iterations taken to solve the BTE versus the number of the BTE-Poisson loop.

This analysis shows that the discretisation approach developed in this work is stable when it is coupled with the Poisson equation and can be used for realistic device simulations. The following sections describe the application of this method to solve the BTE in nanoscale 1D and 2D devices.

5.3   Application to one-dimensional devices

This section describes an application of the numerical technique developed in this work to find the solution of the BTE in a nano-scale one-dimensional device. The number of unknowns are $N_x \approx 300$, $N_k = 4160$ and $N \approx 1.2 \times 10^6$, the memory requirement is about 0.5 GB and the time taken for one self-consistent solution at one bias point is about 6–8 hours on a 400 MHz, UltraSparc2.

5.3.1   Objective of the study

It has been demonstrated recently that current-day MOS transistors operate at roughly 50% of the ballistic limit [85] and that when they are scaled to nanometer sizes, they can operate at over 80% of the ballistic limit [11]. This is a surprising

result because carriers in these devices have finite mobilities and low mean free paths. The reason for this is that there is a critical region of the channel near the source [86] where the carriers can travel nearly ballistically (*i.e.* with little scattering). Hence, this is known as "quasi-ballistic" transport. Although this effect has been studied before, with approximate solutions to the BTE [39, 57], this work is the first report of a systematic analysis of quasi-ballistic transport in a realistic one-dimensional device by solving the BTE without any approximations. The solutions of the BTE are then compared to those from commonly-used macroscopic transport models and the solution from the ballistic BTE (with a zero collision term). The aim of this analysis is two-fold — one, to study the impact of quasi-ballistic transport on the performance of nano-scale 1D devices and second, to test the validity of commonly-used macroscopic models in the nanometer regime.

5.3.2 Analysis of a 50nm n^+-p-n^+ diode

The physical parameters used in the BTE are spherical non-parabolic energy band and LO and LA scattering mechanisms, as described before. This gives an effective intrinsic Si low-field mobility of 1350 cm^2/V-s. The three macroscopic models used in this study are drift-diffusion (DD), hydrodynamic (HD), and energy transport (ET) [87]. All these models are also calibrated to the same low-field mobility, velocity versus field curve and, in the case of HD and ET, energy versus field curve of bulk Si. Hence this ensures that all the transport models represent the same *same* physical problem but only based on different assumptions. The DD model is solved using the Scharfetter-Gummel discretisation and the HD and ET models are solved using the essentially non-oscillatory method [88]. For a given bias, all transport models are also solved self-consistently with the Poisson equation.

The 1D device used in this study is an n^+-p-n^+ diode with a 50 nm channel. This is shown in Fig. 5.4. The doping in this device is $N_D = 5 \times 10^{18}$ cm^{-3} in the source and drain and $N_A = 5 \times 10^{16}$ cm^{-3} in the channel with a smooth transition at the junctions.
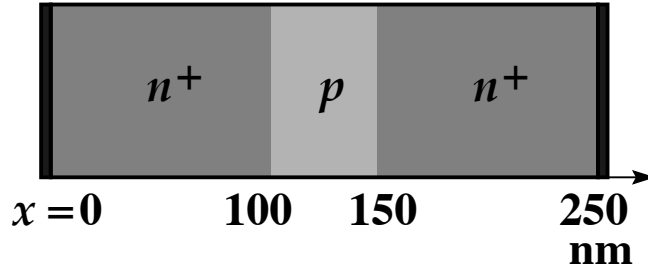
Fig. 5.4. Illustration of the n^+-p-n^+ diode used in this work.

The *I-V* results of the five transport models are plotted in Fig. 5.5.
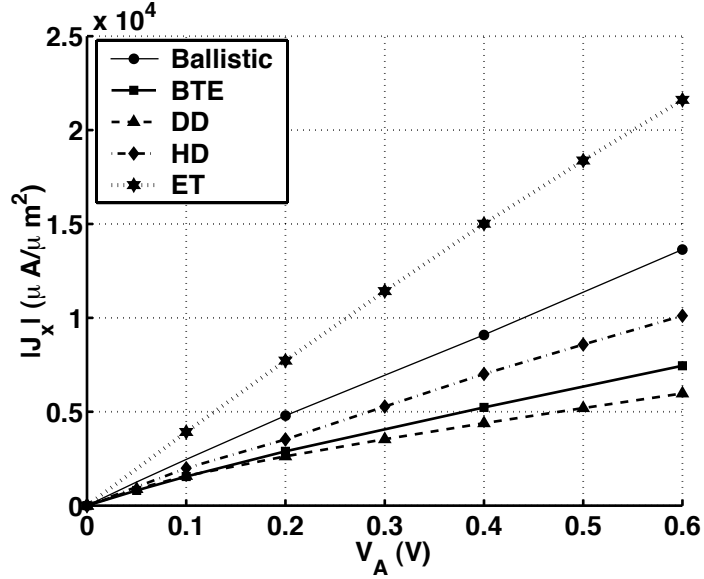
Fig. 5.5. Comparison of *I-V* from ballistic, BTE, DD, HD and ET simulations of n^+-p-n^+ diode.

It is seen that the current from DD is remarkably close to that from the BTE, even though the DD model is a completely wrong description of transport at these length scales. The HD model and ET models give higher currents and the ET model is even higher than the current from the ballistic solution. The reason for this can be seen from the average carrier velocities inside the device. The average velocity from the

BTE solution for increasing bias is shown in Fig. 5.6. It shows that although there is significant velocity overshoot in the channel, the velocity at the source-channel junction $x \approx 108$ nm, approaches the thermal velocity, 1×10^7 cm/s. This is a characteristic signature of quasi-ballistic (and ballistic transport), as will be seen below.

Fig. 5.6. Average velocity from the BTE solution of n^+-p-n^+ diode at increasing bias.

The comparison of average velocity from all the five transport models is shown in Fig. 5.7 for a high value of applied bias 0.6 V. It is seen that the velocity from DD is highly unphysical because it gets clamped to the saturation velocity inside the channel. At the source channel junction, the saturation velocity from DD turns out to be numerically the same as the thermal velocity from the BTE (both $\approx 1 \times 10^7$ cm/s in Si). Hence the values of current from DD and BTE turn out be nearly equal.

The HD and ET velocities are higher than the BTE solution. The HD model shows a spurious velocity overshoot near the drain which is not physical but a mathematical artifact of the HD equations [89]. ET does not show any spurious effects but it allows the velocity near the source-channel junction to increase much higher than the thermal velocity, which is also unphysical. Note that although the ballistic solution
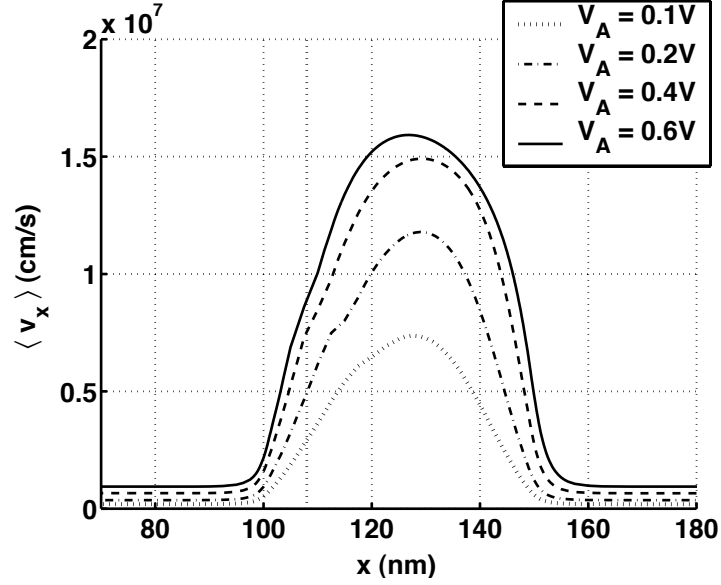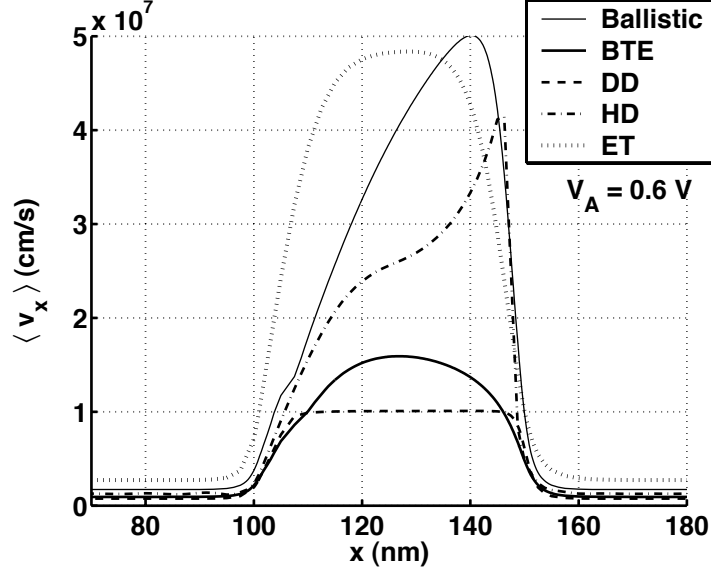
Fig. 5.7. Comparison of average velocity from ballistic, BTE, DD, HD and ET simulations of n^+-p-n^+ diode at high bias.

shows extremely high velocities inside the channel, it still is limited by the thermal velocity at the source-channel junction.

At this point, it seems to be clear that DD, HD and ET fail to describe the behaviour of nano-scale devices. The case against DD is obvious but the case against HD and ET models is not so clear-cut because the parameters in these models are typically not physical constants and can be "tuned" to match the characteristics of a wide variety of devices [55, 53]. However, the point of this analysis is not to suggest a better tuning of parameters but to prove that macroscopic models are fundamentally incapable of describing ballistic or quasi-ballistic transport.

Hence a better way of examining examining the quasi-ballistic nature of transport in a device is used here. This is done by plotting the reflection coefficient r_c [90] inside the device. The reflection coefficient is the ratio of the amount of current carried by the distribution function in the *opposite* direction of the total current to the amount of current carried by the distribution function in the *same* direction as the total current,

$$r_c = \frac{j^-}{j^+}, \tag{5.1}$$

where the total current is $J = j^+ - j^-$. A value of r_c close to 1 indicates that the

negatively and positively directed parts of the distribution functions are nearly symmetric. This happens when there is a lot of scattering that drives the the distribution function towards symmetry in all directions. A value of r_c close to 0 indicates that the negative part of the distribution function is very small compared to the positive part. This happens under conditions of ballistic or quasi-ballistic transport when there is not enough scattering and the two parts of the distribution are highly asymmetric.

The reflection coefficient for the above device is plotted in Fig. 5.8 for high bias. Note that this the same information as that in Fig. 5.6 but just plotted in a different way.
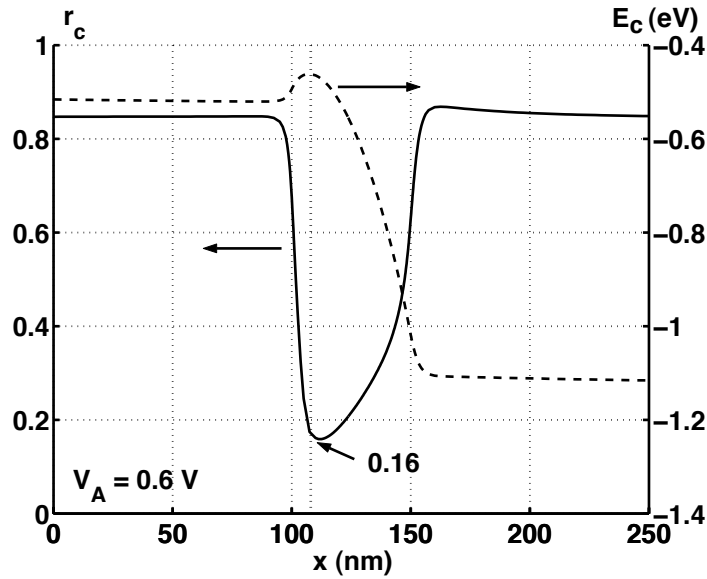
Fig. 5.8. Reflection coefficient from the BTE solution of n^+-p-n^+ diode at high bias.

This figure clearly shows where the least amount of scattering occurs in the device. It occurs at the source-channel junction where the reflection coefficient is minimum. This information can only be obtained from the solution of a kinetic model such as the BTE and cannot be obtained from any macroscopic model such DD, HD and ET. All macroscopic models combine the two oppositely directed populations of carriers *a priori* to compute ensemble averages and often use collision-dominated parameters such as temperature to simplify their equations. The result is that macroscopic models lose the fundamental nature of ballistic transport by their very formulation. It must

be mentioned that although this conclusion is not new and is well-known, this work is the first conclusive proof that nanoscale devices exhibit quasi-ballistic transport, whose characteristic signature is a low reflection coefficient. This behaviour simply cannot be captured by macroscopic models due the limitation of their formulation.

5.4   Application to two-dimensional devices

This section describes an application of the numerical technique developed in this work to find the solution of the BTE in a nano-scale two-dimensional device. The BTE can be generalised to 2D in the same way as described in Chapter 3. In this case, the matrix equation becomes block penta-diagonal instead of block tri-diagonal. The number of unknowns becomes $N = N_x \times N_y \times N_k$, which is typically a very large number. Therefore, in order to keep the size of the problem small, a ultra-thin body dual-gate nMOSFET is chosen for the purposes of this study. The number of unknowns for this device are $N_x \approx 300$, $N_y \approx 20$, $N_k = 2112$ and $N \approx 1.2 \times 10^7$, the memory requirement is about 2.2 GB and the time taken for one self-consistent solution at one bias point is about 12–18 hours on a 400 MHz UltraSparc2.

5.4.1   Objective of the study

The aim of this analysis is two-fold — one, to demonstrate that the technique generalises straightforwardly to two-dimensions and second, to show that the essential observations about quasi-ballistic transport in one-dimensional devices also hold for MOSFETs. A detailed analysis with respect to different transport models is not done here.

5.4.2   Analysis of a 50nm dual-gate ultra-thin body nMOSFET

The dual gate ultra-thin body nMOSFET used in this study is shown in Fig. 5.9. The channel length of this device is chosen to be 50 nm with oxide thickness of $t_{ox} = 2$nm and silicon film thickness of $t_{si} = 10$nm. The source and drain doping is high but the channel is left undoped in order to avoid the effects of dopant fluctuations. In such a structure, the threshold voltage, is adjusted to 0.3V by using a mid-gap metal with a suitable work function for the gate. The normal operation of this device

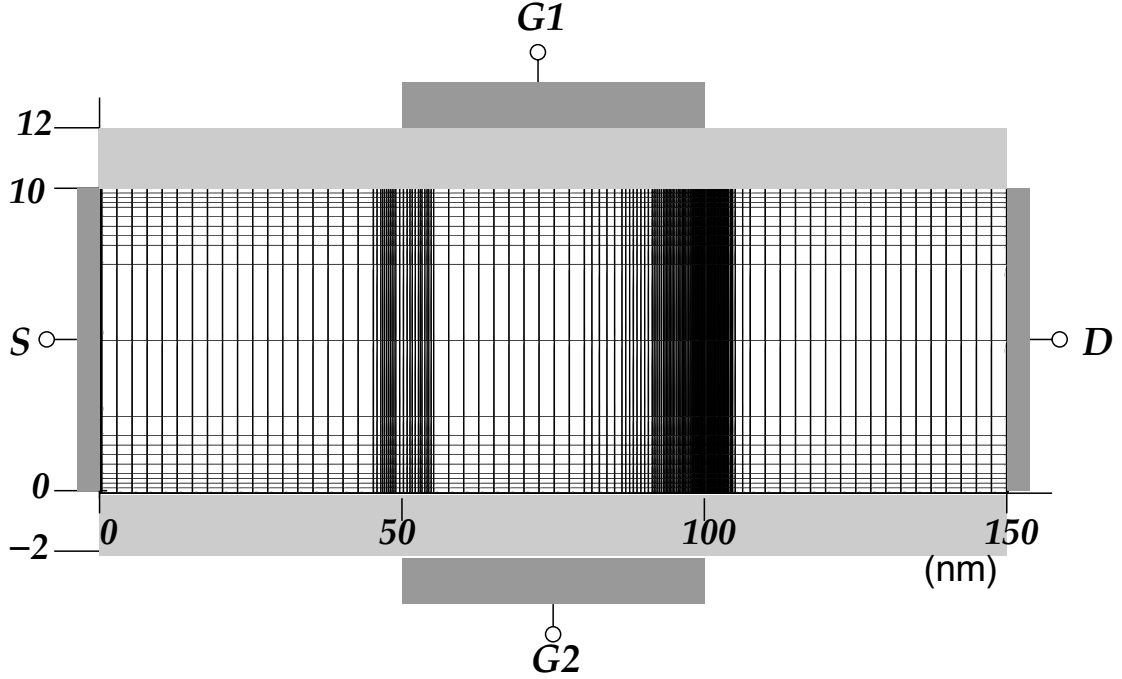is with both the gates tied, hence $V_{G1} = V_{G2} = V_G$.

Fig. 5.9. Structure of the dual-gate ultra thin body nMOSFET used in this work. A typical grid at high bias ($V_{GS} = V_{DS} = 0.6$V) is shown.

The physical parameters used in the BTE are spherical non-parabolic energy band and LO and LA scattering mechanisms, as described before. This gives an effective intrinsic Si low-field mobility of 1350 cm^2/V-s. In addition, a complete definition of a 2D device requires a boundary condition on the insulating surface in terms of surface scattering [91].

Surface mobility can be computed using the set-up shown in Fig. 5.10. Non-self consistent 2D simulations are done on a 50 nm by 10 nm block of semiconductor with a left injection boundary, a right absorbing boundary, a reflective lower boundary and an oxide upper boundary. The oxide boundary condition is specified by a combination of specular and diffuse surface scattering. The scattering mechanisms inside the slab are the same as described above. The parallel field inside the block is zero but the transverse field is in the confining direction and varies from 0 to 1×10^5 V/cm. Under these conditions carriers diffuse across the slab from left to right and are absorbed

in the right contact. The low-field mobility can then be computed from the diffusion coefficient by using Fick's Law and the Einstein relation (chapter 9 in [21]).

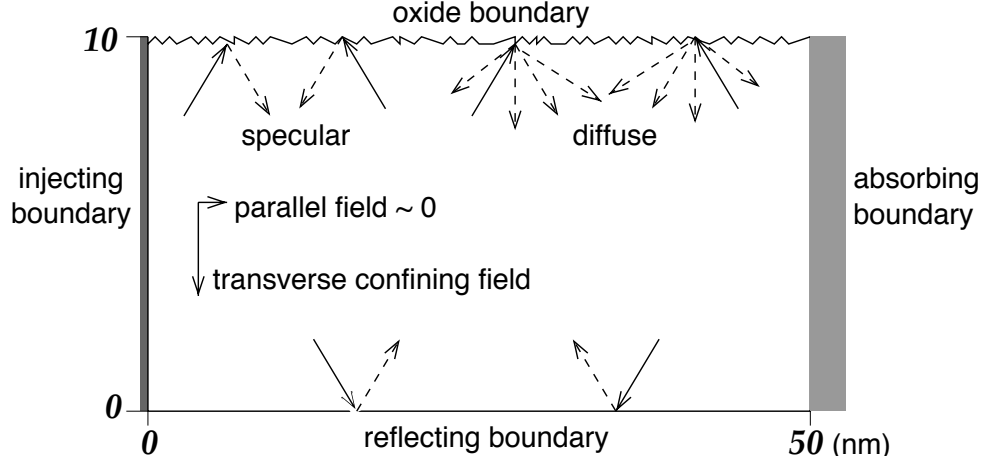Fig. 5.10. Set-up used to compute effect of surface scattering on low-field mobility.

The extracted low-field mobility is plotted versus transverse field for different percentage of diffuse scattering in Fig. 5.11. For the case of the dual-gate MOSFET in this study, 6% diffusive scattering is chosen similar to previous works [92].
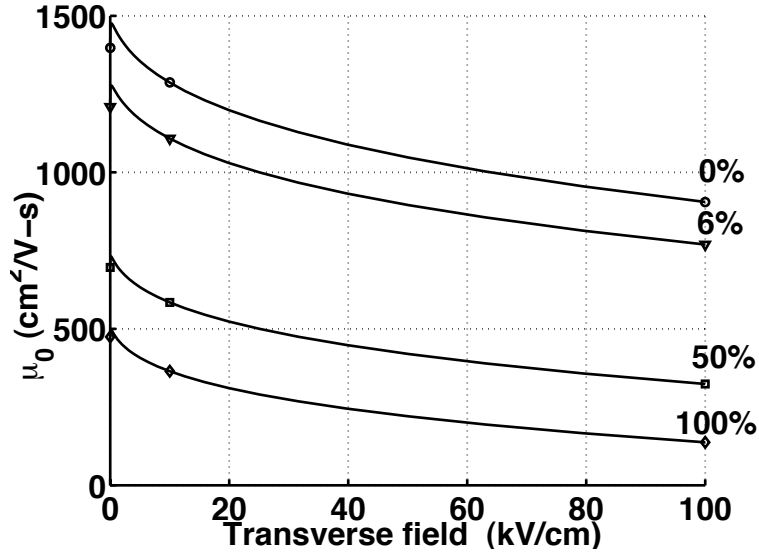
Fig. 5.11. Surface mobility versus transverse field for different percentage of diffuse scattering.

Fig. 5.12 shows the BTE solution of the distribution function inside the device under high bias $V_{GS} = V_{DS} = 0.6$ V. The distribution functions shown are at $y = 5$ nm and at x located in the source, barrier channel and drain respectively.
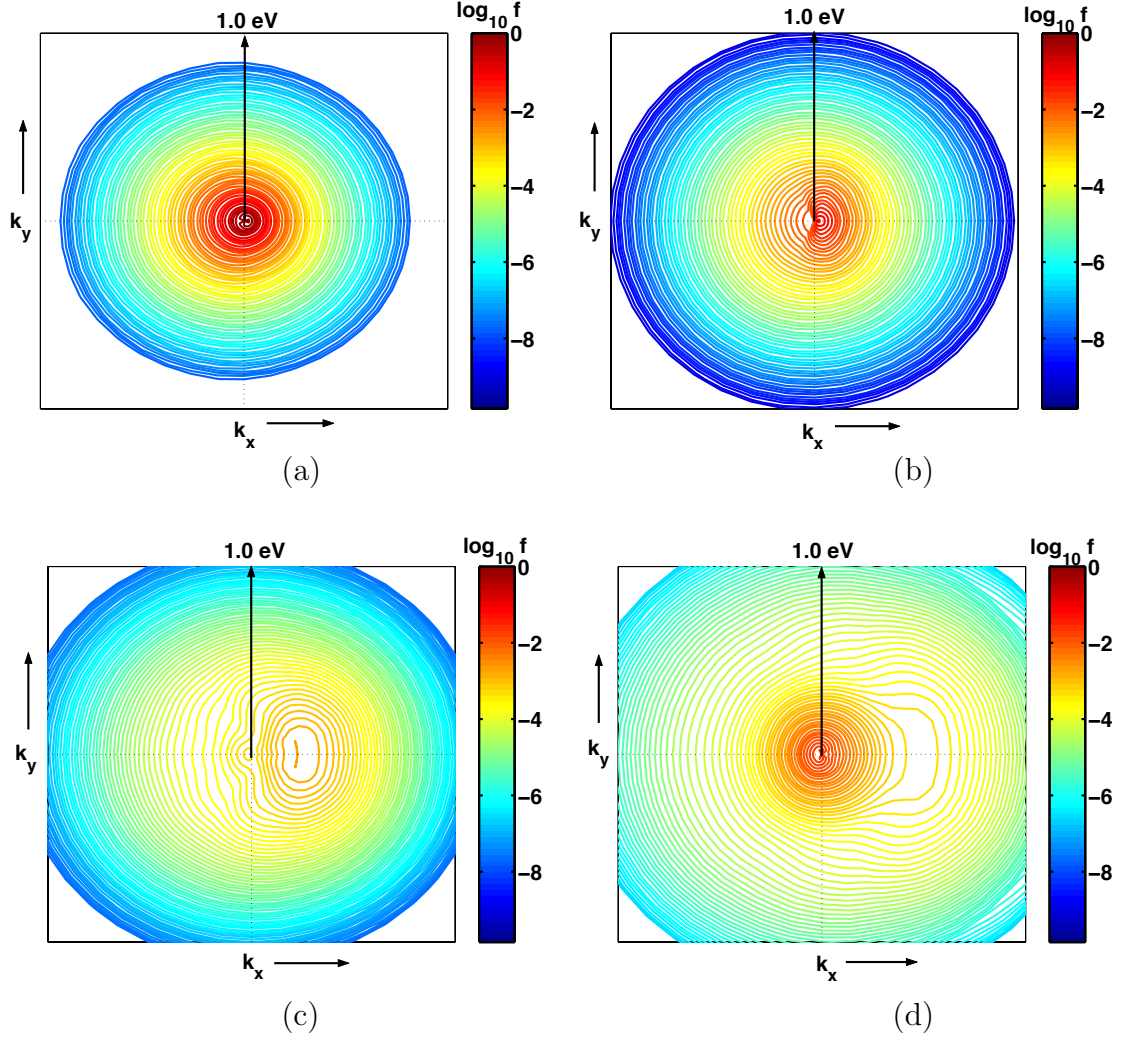
Fig. 5.12. Distribution functions for the device in Fig. 5.9 at $y = 5$ nm and at x located in the (a) source (b) barrier (c) channel and (d) drain.

Using the above solution, the average velocity inside the device is calculated and plotted in Fig. 5.13. In this case, only a moderate velocity overshoot is seen at the

drain end of the channel.
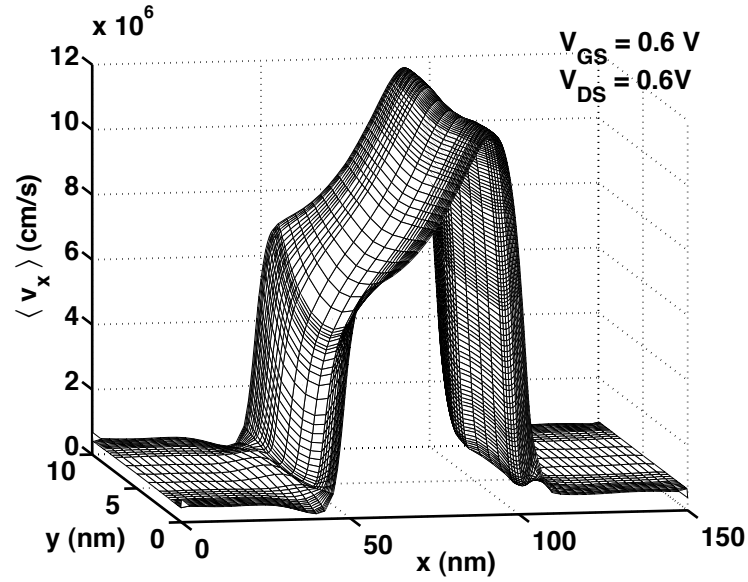
Fig. 5.13. Average velocity from the BTE solution of the nMOSFET at $V_{GS} = V_{DS} = 0.6\text{V}$.

Fig. 5.14 shows the same information as that in Figs 5.12 and 5.13, but plotted in terms of the reflection coefficient inside the device.
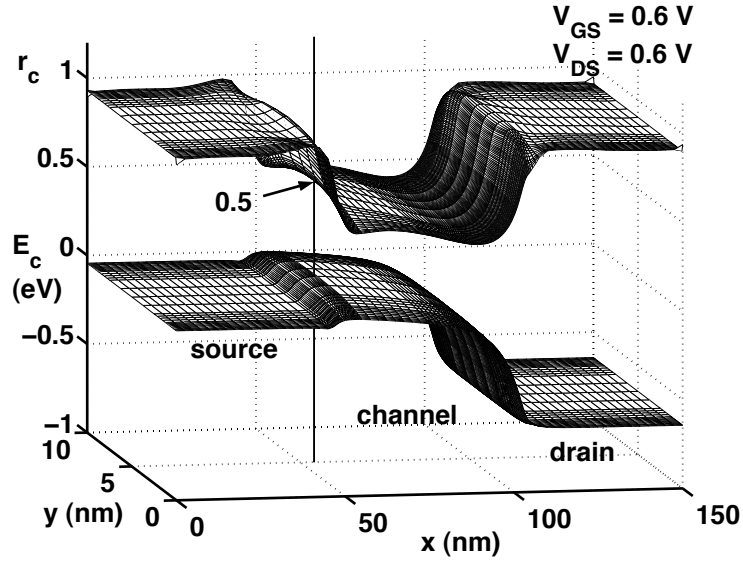
Fig. 5.14. Reflection coefficient and conduction band profile from the BTE solution of the nMOSFET at $V_{GS} = V_{DS} = 0.6\text{V}$.

It is observed, as before, that the reflection coefficient is high in the source region but drops sharply at the source-channel junction. Therefore, this is consistent with earlier works [93, 94] that present day devices operate at about 50% of the ballistic limit.

5.5 Summary

This chapter described the application of the numerical technique developed in this work to device simulation. It first addressed the properties of solving the BTE self-consistently with the (non-linear) Poisson equation, including stability, overall convergence of the BTE-Poisson loop and speed. In addition, this chapter also described the application of the numerical technique developed in this work, to study transport in realistic nano-scale devices. Two devices were used in this regard — a one-dimensional 50nm n^+-p-n^+ diode and a two-dimensional 50nm ultra-thin body dual-gate nMOSFET. This study demonstrated that that nanoscale devices operate in a quasi-ballistic regime, where the carriers travel nearly ballistically across a small region of the channel near the source. Standard macroscopic models fail in these regimes because they are based on collision-dominated assumptions.

6. Conclusion

6.1   Summary

This work described the first direct numerical solution of the BTE for semiconductors that can be used for practical device simulation. This was done by using powerful mathematical techniques to discretise the BTE in energy and angle without making any approximations about the angular shape of the distribution function or the collision integral. Such a direct discretisation resulted in a very large matrix equation,with $N = 10^6$–10^7 unknowns. Therefore, in order to address the need for efficient and fast solutions, this work also reported the first application of a preconditioned iterative method (GMRES) to the BTE. This method is not only fast (on the order of $N^{1.2}$) but also has low memory requirements because it does not require explicit storage of the matrix elements. The technique developed in this work was also found to be highly suitable for self-consistent device simulations because it shows smooth and stable convergence when coupled to the Poisson equation. Finally, this method was applied to study transport in two representative nanoscale devices — a one-dimensional 50nm n^+-p-n^+ diode and a two-dimensional 50nm ultra-thin body dual-gate nMOSFET. This study demonstrated that that nanoscale devices operate in a quasi-ballistic regime, where the carriers travel nearly ballistically across a small region of the channel near the source. Standard macroscopic models fail in these regimes because they are based on collision-dominated assumptions.

6.2   Future work

Future work that can be done to improve the performance and applicability of this method is as follows:

1. It was seen that the real space discretisation used in this work had a severe restriction on grid spacing (although this restriction could be somewhat relaxed).

This is because of the lack of a suitable Scharfetter-Gummel type discretisation for the BTE. This type of discretisation has been very successful in drift-diffusion equations, allowing large grid spacings and smaller problem sizes and yet maintaining high accuracy. An equivalent method for the BTE, possibly along the lines of [95] could prove to be very useful.

2. There is enormous potential to speed up the BTE (matrix) solution step by application of a suitable parallel algorithm. Note that the iterative step itself cannot be parallelised because it proceeds in a sequence of steps. But its internal operations are mainly based on matrix and vector algebra, which can definitely be parallelised. This can help speed up the transport solution step immensely.

3. This technique can also be extended to other more complex band structures. However, care must be taken when doing so because, in general, more complex band structures require a larger number of momentum space elements and hence imply a larger problem size.

A. Calibration of the electron density of states, thermal velocity and thermal energy for spherical non-parabolic bands

This appendix describes the consequences of using a single non-parabolic energy band with a conductivity effective mass for describing the electron band-structure of Si at low energies. There is a trade-off when using a single non-parabolic energy band structure for electrons in Si because it can accurately model either the ensemble velocities and energies or the density of states but *not* both.

The electron band structure of Si, at low energies, consists of six ellipsoids centered at the X points which respond with different effective masses depending on their orientation. Their cumulative effect is such that when they are reduced to an equivalent single non-parabolic band, the ensemble velocities (both thermal and bulk) depend on a conductivity effective mass m_c^* whereas the density of states depends on a density of states effective mass m_d^* [71]. These two effective mass are not equal [21] because

$$
\begin{aligned}
m_c^* &= \frac{3 m_\ell^* m_t^*}{m_t^* + 2 m_\ell^*} = 0.258 \ m_0, \\
m_d^* &= 6^{2/3} \left(m_\ell^* m_t^{* \, 2} \right)^{1/3} = 1.0598 \ m_0,
\end{aligned}
\tag{A.1}
$$

where m_ℓ^* is the longitudinal effective mass of the ellipsoids, m_t^* is their transverse effective mass and m_0 is the free electron rest mass. Note that the above values of effective mass apply only to low energies, $E < 1.6$ eV. At high energies and with the inclusion of multiple energy bands, the electron band structure of Si itself can no longer be reproduced by a single non-parabolic band [26]. Also note that the above values apply to a 3-dimensional band structure only.

In this work, the conductivity effective mass m_c^* is used for defining a non-parabolic energy band structure for electrons in Si. This gives the correct values of ensemble electron velocities and energies but it does not give the correct electron concentration because the density of states is too low (see Fig. A.1). In order to reconcile this difference, a correction factor is used for the electron concentrations, as shown below.
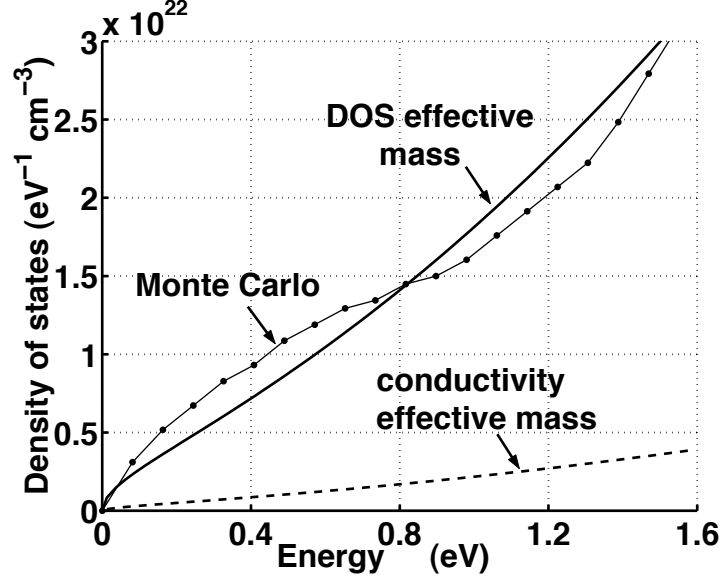
Fig. A.1. Density of states of a single spherical non-parabolic band with conductivity effective mass and density of states effective mass. Approximate density of states in full band Monte Carlo calculations is also known.

Suppose the distribution function is at equilibrium, then it is given by

$$f(k) = \exp\left(\frac{\psi_n - E_c}{k_B T_L}\right) \exp\left(-\frac{E(k)}{k_B T_L}\right), \tag{A.2}$$

where the first exponential is a scaling factor that depends on the electron quasi-fermi level and the second exponential is a Maxwellian in energy. Note that the Maxwellian involves information about the band structure because it depends on energy. In the case of this work, the band structure is

$$E(k)(1 + \alpha E(k)) = \frac{\hbar^2 k^2}{2m_{eff}^*}, \tag{A.3}$$

where $m_{eff}^* = m_c^*$. Now the electron quasi-Fermi level is set at the intrinsic level $E_i = E_c - 0.5615$ eV (for discussion of this value please see Chapter 2 of [96] and

references therein).

It is well-known that the electron concentration from a distribution function at equilibrium in intrinsic Si is $n_i = 1.1 \times 10^{10}$ cm^{-3} at 300 K. But when the electron concentration is computed from eq. A.2, the value obtained is

$$
\begin{aligned}
n &= \frac{1}{4\pi^3} \int_{\mathbb{R}^3} f(k) \; \mathrm{d}k \\[2mm]
&= \left[\frac{2}{(4\pi\alpha k_B T_L)^{1/2}} \exp\left(\frac{1}{2\alpha k_B T_L}\right) K_2\left(\frac{1}{2\alpha k_B T_L}\right) \right] \frac{1}{4\pi^3} \left(\frac{2\pi m_c^* k_B T_L}{\hbar^2}\right)^{3/2} \\
&\quad \times \; \exp\left(\frac{E_i - E_c}{k_B T_L}\right) \\[2mm]
&= 1.04907 \; \frac{1}{4\pi^3} \left(\frac{2\pi m_c^* k_B T_L}{\hbar^2}\right)^{3/2} \exp\left(\frac{E_i - E_c}{k_B T_L}\right) \\
&= 1.3 \times 10^9 \; \text{cm}^{-3},
\end{aligned} \tag{A.4}
$$

where $K_n(x)$ is a modified Bessel function of the second kind [97]. This is *lower* than the expected value 1.1×10^{10} cm^{-3} by a factor of $(m_d^*/m_c^*)^{3/2}$. Hence in order to get the correct electron concentration, all computed concentrations in this work are *scaled* by a

$$
\text{correction factor} = \left(\frac{m_d^*}{m_c^*}\right)^{3/2}. \tag{A.5}
$$

This makes sense because density of states obtained by using a conductivity effective mass is *lower* than the actual density of states by the above factor.

However, the thermal velocity obtained from the same distribution function is

$$
\begin{aligned}
v_{th} &= \left(\frac{2}{n}\right) \frac{1}{4\pi^3} \int_{-\pi/2}^{\pi/2} \int_{-1}^{1} \int_{0}^{\infty} \frac{1}{\hbar} \frac{\partial}{\partial k_x} E(k) \; f(k) \; k^2 \mathrm{d}k \; \mathrm{d}\sin\theta \; \mathrm{d}\cos\phi \\[2mm]
&= \frac{(\pi\alpha k_B T_L)^{1/2}}{\exp(\frac{1}{2\alpha k_B T_L}) K_2(\frac{1}{2\alpha k_B T_L})} \left(\frac{2k_B T_L}{\pi m_c^*}\right)^{1/2} \\[2mm]
&= 0.953 \left(\frac{2k_B T_L}{\pi m_c^*}\right)^{1/2} \\
&= 1.036 \times 10^7 \; \text{cm/s},
\end{aligned} \tag{A.6}
$$

which is very close to the correct value of 1.059×10^7 cm/s at 300 K.

Similarly, the thermal energy obtained from the above distribution function is

$$
\begin{aligned}
u_{th} &= \left(\frac{1}{n}\right) \frac{1}{4\pi^3} \int_{\mathbb{R}^3} E(k)\ f(k)\ \mathrm{d}k \\
&= \frac{1}{3\alpha k_B T_L} \left(\frac{1}{4} \frac{K_{-1}(\frac{1}{2\alpha k_B T_L}) + 2K_1(\frac{1}{2\alpha k_B T_L}) + K_3(\frac{1}{2\alpha k_B T_L})}{K_2(\frac{1}{2\alpha k_B T_L})} + 4\alpha k_B T_L - 1 \right) \\
&\quad \times \frac{3}{2} k_B T_L, \\
\\
&= 1.031\ \frac{3}{2} k_B T_L, \\
&= 0.04\ \text{eV}. \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (\text{A.7})
\end{aligned}
$$

This is again close to the correct value of 0.0388 eV at 300 K.

Hence, the use of the conductivity effective mass for defining a non-parabolic energy band structure for electrons in Si gives the correct values of ensemble electron velocities and energies. However, it produces lower values of electron concentration which must be corrected by an appropriate density of states factor.

B. Computation of discrete ensemble averages

This appendix describes the computation of ensemble averages or "moments" when the distribution function is discretised under the finite volume scheme.

The moments of a continuous distribution function have already been defined in properties 11, 12 and 13 of Section 1.2.1. They are the carrier concentration, average velocity and average energy respectively. Similarly, when the distribution function is discretised, the corresponding moments can be can be calculated as follows:

1. Carrier concentration

$$n = \frac{1}{4\pi^3} \sum_i \left(\int_{\Omega_i} \mathrm{d}k \right) \tilde{f}_i = \frac{1}{4\pi^3} \sum_i \Omega_i \, \tilde{f}_i, \tag{B.1}$$

   where the factor Ω_i refers to the numerical value of the volume of the momentum space element i.

2. Average velocity

$$n\langle \vec{v} \rangle = \frac{1}{4\pi^3} \sum_i \left(\int_{\Omega_i} \frac{1}{\hbar} \vec{\nabla}_k E(k) \, \mathrm{d}k \right) \tilde{f}_i, \tag{B.2}$$

   where there is no factor of Ω_i because it is included in the integral over the element.

3. Average energy

$$n\langle u \rangle = \frac{1}{4\pi^3} \sum_i \left(\int_{\Omega_i} E(k) \, \mathrm{d}k \right) \tilde{f}_i. \tag{B.3}$$

LIST OF REFERENCES

[1] Integrated Systems Engineering AG, "Dessis release 4.1", 1994, Zürich, Switzerland.

[2] Technology Modeling Associates Inc., "Medici version 2.2", 1996, Sunnyvale, California.

[3] TU Insititute for Microelectronics, "Minimos version 6.0", 1987, Vienna, Austria.

[4] D. Cole, E. Buturla, S. Furkay, K. Varahramyan, J. Slinkman, J. Mandelman, D. Forty, O. Bula, A. Strong, J. Park, T. Linton, J. Johnson, M. Fischetti, S. Laux, P. Cottrell, H. Lustig, F. Pileggi, and D. Katcoff, "The use of simulation in semiconductor technology development", *Solid-St. Electron.*, vol. 33, pp. 591–623, 1990.

[5] C. H. Snowden, editor, *Semiconductor Device Modelling*, Springer-Verlag, Berlin, 1st edition, 1989.

[6] SIA (Semiconductor Industry Association), "The National Technology Roadmap for Semiconductors", San Jose, CA, 1994.

[7] F. Stern and W. E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit", *Phys. Rev.*, vol. 163, pp. 816–835, 1967.

[8] T. Ando, A.B. Fowler, and F. Stern, "Electronic Properties of two-dimensional systems", *Rev. Mod. Phys.*, vol. 54, pp. 437–672, 1982.

[9] K. Banoo and M.S. Lundstrom, "Electron Transport in a model Si transistor", to be published *Solid-St. Electron.*, 2000.

[10] K. Banoo, J.-H. Rhew, M.S. Lundstrom, C.-W. Shu, and J.W. Jerome, "Simulating Quasi-Ballistic Transport in Si nanotransistors", to appear in proceedings of *Seventh International Workshop on Computational Electronics*, 2000.

[11] G. Timp, J. Bude, K.K. Bourdelle, J. Garno, A. Ghetti, H. Gossmann, M. Green, G. Forsyth, Y. Kim, R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampfor, W. Mansfield, S. Moccio, T. Sorsch, D. M. Tennant, W. Timp, and R. Tung, "The ballistic nano-transistor", *in IEDM Technical Digest*, vol. Cat. 99CH36318, pp. 55–58, Electron Devices Society of IEEE, Piscataway NJ, 1999.

[12] D. K. Blanks, G. Klimeck, R. Lake, D. Jovanoic, R.C. Bowen, C. Fernando, W. R. Frensley, and M. Leng, "NEMO: general release of a new comprehensive quantum device simulator", *in Proceedings of the IEEE 24th International Symposium on Compound Semiconductors*, pp. 639–642, New York, NY, 1998.

[13] G. Klimeck, R. K. Lake, M.J. Mclennan, and S. Datta, "QUEST user's manual", Technical Report TR-EE 93-17, Purdue University, 1993.

[14] D. K. Ferry and H. L. Grubin, *Modeling of quantum transport in semiconductor devices*, Arizona State University, Tempe, AZ, 1994.

[15] P. Bordone, M. Pascoli, R. Brunetti, A. Bertoni, C. Jacoboni, and A. Abramo, "Quantum transport of electrons in open nanostructures with the Wigner-function formalism", *Phys. Rev. B*, vol. 59, pp. 3060–3069, 1999.

[16] M. V. Fischetti, "Master-equation approach to the study of electronic transport in small semiconductor devices", *Phys. Rev. B*, vol. 59, pp. 4901–4917, 1999.

[17] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, New York, NY, 1st edition, 1995.

[18] R. Lake and S. Datta, "Nonequilibrium Green's-function method applied to double-barrier resonant tunneling diodes", *Phys. Rev. B*, vol. 45, pp. 6670–6685, 1992.

[19] A. Svizhenko, M. P. Anantram, and T. R. Govindan, "2D quantum simulation of MOSFET using the non-equilibrium Green's function method", presented at *IWCE 2000*.

[20] D. Jovanovic, "Computational techniques for the nonequilibrium quantum field theory simulation of MOSFETs", proceedings of *IWCE 2000*.

[21] M. S. Lundstrom, *Fundamentals of Carrier Transport*, Addison-Wesley, Reading, Massachusetts, 1st edition, 1990.

[22] G. Baccarani, F. Odeh, A. Gnudi, and D. Ventura, *Semiconductors, Part II*, chapter "A critical review of the fundamental semiconductor equations", Springer-Verlag, New York, NY, 1994.

[23] E. M. Conwell, *High field transport in semiconductors*, Academic Press, New York, 1967.

[24] D. L. Rode, *Semiconductors and Semimetals*, vol. 10, chapter "Low-field transport in semiconductors", Academic Press, New York, 1972.

[25] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials", *Rev. Mod. Phys.*, vol. 55, pp. 645–705, 1983.

[26] M.V. Fischetti and S.E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", *Phys. Rev. B.*, vol. 38, pp. 9721–9745, 1988.

[27] T. Kunikiyo *et. al.*, "A Monte Carlo simulation of anisotropic electron transport in Silicon including full band structure and anisotropic impact-ionization model", *J. Appl. Phys.*, vol. 75, pp. 297–312, 1994.

[28] A. Pacelli and U. Ravaioli, "Analysis of variance reduction schemes for ensemble Monte Carlo simulation of semiconductor devices", *Solid-St. Electron.*, vol. 41, pp. 599–605, 1997.

[29] C. Huster, *Two-dimensional scattering matrix simulations of Si MOSFETs*, PhD thesis, Purdue University, May 2000.

[30] H. Gutowitz, "A life online", 1996, http://alife.santafe.edu/alife.

[31] G. Zandler, A. DiCarlo, K. Kometer, P. Lugli, and E. Gornik, " A comparison of Monte Carlo and Cellular Automata approaches for semiconductor device simulation", *IEEE Elec. Dev. Lett.*, vol. 14, pp. 77, 1993.

[32] K. Kometer, G. Zandler, and P. Vogl, "Lattice-Gas Cellular-Automaton method for semi-classical transport in semicondcutors", *Phys. Rev. B.*, vol. 46, pp. 1382–1394, 1992.

[33] A. Das and M. S. Lundstrom, "" A scattering matrix approach to device simulation"", *Solid-St. Electron.*, vol. 33, pp. 1299–1307, 1990.

[34] M. A. Alam, M. A. Stettler, and M. S. Lundstrom, ""Formulation of the Boltzmann Equation in terms of scattering matrices"", *Solid-St. Electron.*, vol. 36, pp. 263–271, 1993.

[35] M. A. Stettler and M. S. Lundstrom, "Self-consistent scattering matrix calculation of the distribution function in semiconductor devices", *Appl. Phys. Lett.*, vol. 60, pp. 2908–2910, 1992.

[36] M. A. Alam, *Computational studies of the physics of opto-electronic devices*, PhD thesis, Purdue University, August 1995.

[37] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh, "Multi-dimensional spherical harmonics expansion of Boltzmann Equation for transport in semiconductors", *Appl. Math. Lett.*, vol. 5, pp. 85–90, 1992.

[38] N. Goldsman, L. Henrickson, and J. Frey, "A physics-based analytical/numerical solution to the Boltzmann Transport Equation for use in device simulation", *Solid-St. Electron.*, vol. 34, pp. 389–396, 1991.

[39] H. U. Baranger and J. W. Wilkins, "Ballisitic structure in the electron distribution function of sma ll semiconducting structures: General features and specific trends", *Phys. Rev. B.*, vol. 36, pp. 1487–1502, 1987.

[40] C. Cercignani, I. M. Gamba, J. W. Jerome, and C.-W. Shu, "Device benchmark comparisons via kinetic, hydrodynamic and high-f ield models", *Comput. Methods Appl. Mech. Engrg.*, vol. 181, pp. 381–392, 2000.

[41] W. V. van Roosbroeck, "Theory of flow of electrons and holes in germanium and other semiconductors", *Bell Syst. Tech. J.*, vol. 29, pp. 560–607, 1950.

[42] H. K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculations", *IEEE Trans. Electron Devices*, vol. 11, pp. 455–465, 1964.

[43] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser, *Semiconductor Equations*, Springer-Verlag, Wien, Austria, 1990.

[44] R. E. Bank, D. J. Rose, and W. Fichtner, "Numerical methods for semiconductor device simulation", *SIAM J. Sci. Stat. Comput.*, vol. 4, pp. 416–435, 1983.

[45] S. Selberherr, *Analysis and simulation of semiconductor devices*, Springer-Verlag, Wien, Austria, 1984.

[46] K. Banoo F. Assad and M. S. Lundstrom, "The Drift-Diffusion equation revisited", *Solid-St. Electron.*, vol. 42, pp. 283–295, 1998.

[47] P. A. Markowich and C. Schmeister, "The Drift-Diffusion limit for electron-phonon interactions in semiconductors ", *Math. models & methods in Appl. Sci.*, vol. 7, pp. 707–729, 1997.

[48] K. K. Thornber, "Current equations for velocity overshoot", *IEEE Elec. Dev. Lett.*, vol. 3, pp. 69–71, 1982.

[49] E. C. Kan, U. Ravioli, and D. Chen, "Multi-dimensional augmented current equation including velocity overshoot", *IEEE Elec. Dev. Lett.*, vol. 12, pp. 419–421, 1991.

[50] B. Meinerzhagen and W.L. Engl, "The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon sub-micrometer MOS transistors", *IEEE Trans. Electron Devices*, vol. 35, pp. 689–697, 1988.

[51] M. Rudan and F. Odeh, "Multi-dimensional discretisation scheme for the hydrodynamic model of semiconductor devices", *COMPEL*, vol. 5, pp. 149–183, 1986.

[52] M. C. Vecchi and L. G. Reyna, "Generalized energy transport models for semiconductor device simulations", *Solid-St. Electron.*, vol. 37, pp. 1705–1716, 1994.

[53] T.-W. Tang, S. Ramaswamy, and J. Nam, "An improved hydrodynamic transport model for Silicon", *IEEE Trans. Electron Devices*, vol. 40, pp. 1469–1477, 1993.

[54] J. O. Bark and G. Gildenblat, "A simple transport model for submicron semiconductor device analysis", *J. Appl. Phys.*, vol. 80, pp. 2904–2907, 1996.

[55] D. Chen, E. C. Kan, U. Ravioli, C.-W. Shu, and R .W. Dutton, "An improved energy transport model including non-parabolicity and non-Maxwellian distribution effects", *Electron Device Lett.*, vol. 13, pp. 26–28, 1992.

[56] M. A. Stettler, M. A. Alam, and M. S. Lundstrom, ""A critical examination of the assumptions underlying macroscopic transport equations for silicon devices"", *IEEE Trans. Electron Devices*, vol. 40, pp. 733–740, 1993.

[57] M. Nekovee, B. J. Geurts, H. M. J. Boots, and M. F. H. Schuurmans, "Failure of extended-moment-equation approaches to describe ballistic transport in submicrometer structures", *Phys. Rev. B.*, vol. 45, pp. 6643–6651, 1992.

[58] N. Bellomo, P. LeTallec, and B. Perthame, "The Solution of the Nonlinear Boltzman: A Survey of Analytic and Computational Methods ", *Computers Math. Applic.*, vol. 7, pp. 21–30, 1995.

[59] N. B. Abdallah and P. Degond, "On a hierarchy of macroscopic models for semiconductors", *J. Math. Phys.*, vol. 37, pp. 3306–3333, 1996.

[60] P. A. Markowich and C. Schmeiser, "Relaxation time approximation for electron-phonon interaction in semiconductors", *Math. models and methods in Appl. Sci.*, vol. 5, pp. 519–527, 1995.

[61] K. H. Huebner, *The finite element method for engineers*, John Wiley & Sons, USA, 1975.

[62] M. Pinto, *Comprehensive semiconductor device simulation for Silicon ULSI*, PhD thesis, Stanford University, August 1990.

[63] M. Guidry and M. Strayer, "The Computational Science Education Project at University of Tennessse", 1995, http://csep10.phys.utk.edu/guidry/phys594/.

[64] E. E. Lewis and Jr. W. F. Miller, *Computational methods of Neutron Transport*, American Nuclear Society Inc., La Grange Park, Illinois, 1993.

[65] C. Buet, "Conservative and entropy schemes for Boltzmann Collision Operator of polyatomic gases", *Math. models and methods in Appl. Sci.*, vol. 7, pp. 165–192, 1997.

[66] C. Buet, "A discrete velocity scheme for the Boltzmann operator of rarefied gas dynamics", *Transport theory and stat. phys.*, vol. 25, pp. 33–60, 1996.

[67] C. Jacoboni, C. Canali, G. Ottaviani, and A. Alberigi Quaranta, "A review of some charge transport properties of Silicon", *Solid-St. Electron.*, vol. 20, pp. 77–89, 1977.

[68] J. Bude and R. K. Smith, "Phase-space simplex Monte Carlo for semiconductor transport", *Semiconductor Science & Technology*, vol. 9, pp. 840–843, 1994.

[69] J. Bude, *Monte Carlo device simulation: full band and beyond*, chapter "Scattering mechanisms for semiconductor transport calculations", Kluwer Academic Publishers, Norwell, Massachussetts, 1991.

[70] P. Degond *et. al.*, "Semiconductor Modelling via the Boltzmann Equation", *in Lectures in Applied Mathematics – Vol. 25*. American Mathematical Society, 1990.

[71] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College Publishing, 1976.

[72] E. Conwell and V. F. Weisskopf, "Theory of impurity scattering in semiconductors", *Phys. Rev.*, vol. 77, pp. 388–390, 1950.

[73] T.G. Van de Roer and F. P. Widdershoven, "Ionized impurity scattering in Monte Carlo calculations", *J. Appl. Phys.*, vol. 59, pp. 813–815, 1986.

[74] B. K. Ridley, "Reconciliation of the Conwell-Weisskopf and Brooks-Herring formulae for charged-impurity scattering in semiconductors: Third-body interference", *J. Phys. C:Solid State Phys.*, vol. 10, pp. 1589–1593, 1977.

[75] G. A. Baraff, "Maximum anisotropy approximation for calculating electron distributions; application to high field transport in semiconductors", *Phys. Rev.*, vol. 133, pp. A26–A33, 1964.

[76] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 3rd edition, 1996.

[77] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, "Templates for the solution of linear systems : building blocks for iterative methods", http://www.netlib.org/linalg/html_templates/Templates.html.

[78] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C : the art of scientific computing*, Cambridge Univerisity Press, 2nd edition, 1993.

[79] F. Assad, *Scattering theory of Si MOSFETs*, PhD thesis, Purdue University, August 2000.

[80] D. Kincaid and W. Cheney, *Numerical analysis*, Brooks Cole, Belmont, California, 1st edition, 1991.

[81] R. Kent Smith, "private communication".

[82] R. E. Bank and T. F. Chan, "A composite step bi-conjugate gradient algorithm for nonsymmetric linear systems", *Numer. Algorithms*, vol. 7, pp. 1–16, 1994.

[83] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems", *SIAM J. Sci. Stat. Comput.*, vol. 7, pp. 856–869, 1986.

[84] U. Ascher, P. A. Markowich, C. Schmeiser, H. Steinrück, and R. Weiss, "Conditioning of the steady state semiconductor device problem", *SIAM J. Appl. Math*, vol. 49, pp. 165–185, 1989.

[85] F. Assad, Z. Ren, S. Datta, M. S. Lundstrom, and P. M. Bendix, "Performance limits of silicon MOSFETs", *in IEDM Technical Digest*, vol. Cat.99CH36318, pp. 547–550, Electron Devices Society of IEEE, Piscataway NJ, 1999.

[86] M.S. Lundstrom, "Elementary scattering theory of the Si MOSFET", *IEEE Elec. Dev. Lett.*, vol. 18, pp. 361–363, 1997.

[87] J. W. Jerome and C.-W. Shu, *Semiconductors, Part II*, chapter "Energy models for one-carrier transport in semiconductor devices", Springer-Verlag, New York, NY, 1994.

[88] C.-W. Shu and S. J. Osher, "Efficient implementation of essentially non-oscillatory shock capturing schemes", *J. Comp. Phys.*, vol. 77, pp. 439–471, 1988.

[89] J. W. Jerome and C.-W. Shu, "Transport effects and characteristic modes in the modeling and simulation of submicron devices", *IEEE Trans. CAD*, vol. 14, pp. 917–923, 1995.

[90] M. S. Lundstrom, "Scattering theory of the short-channel MOSFET", *in IEDM Technical Digest*, vol. Cat. 96CH35961, pp. 387–390, Electron Devices Society of IEEE, Piscataway NJ, 1996.

[91] J.R. Schrieffer, "Effective carrier mobility in surface-space charge layers", *Phys. Rev.*, vol. 97, pp. 641–646, 1955.

[92] E. Sangiorgi and M. R. Pinto, "A semi-empirical model of surface scattering for Monte Carlo simulation of Silicon nMOSFETs", *IEEE Trans. Electron Devices*, vol. 39, pp. 356–361, 1992.

[93] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. S. Lundstrom, "On the performance limits for Si MOSFETs: a theoretical study", *IEEE Trans. Electron Devices*, vol. 47, pp. 232–240, 2000.

[94] M. S. Lundstrom and F. Assad, "Scattering parameter assessment of device performance and transport models", *in 2nd NASA Ames Device Modeling Workshop*, Moffet Field, CA, 1997.

[95] C. Ringhofer, "Space-time discretization of series expansion methods for the Boltzmann Transport Equation", to be published, 1998.

[96] H. C. Casey, *Devices for Integrated Circuits: Silicon and III-V Compound Semiconductors*, John Wiley & Sons, 1999.

[97] G. Arfken, *Mathematical methods for physicists*, Academic Press Inc., San Diego, California, 3rd edition, 1985.

VITA

Kausar Banoo was born in Dhoraji, India, on January 2, 1972. She received her B.Tech (Bachelor of Technology) degree in Electrical Engineering from I.I.T. (Indian Institute of Technology), Mumbai in 1993, her M.S. degree in Electrical Engineering from Duke University, NC in 1995 and her Ph.D. in the same from Purdue University, IN in 2000. Since July 2000, she has been working for Lucent Technologies, Murray Hill, NJ as a Member of Technical Staff. Currently, she is working on compact modelling of SiGe bipolar transistors for wireless and data networking applications.