Simulation of Hot Carriers in Semiconductor Devices

Khalid Rahmat

RLE Technical Report No. 591

February 1995

The Research Laboratory of Electronics
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139-4307

This work was supported in part by the U.S. Navy under Contract N00174-93-C-0035.

Simulation of Hot Carriers in Semiconductor Devices

by

Khalid Rahmat

Submitted to the Department of Electrical Engineering and Computer Science
on January 20, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Two approaches to calculate the population of hot carriers in semiconductor devices are studied in this thesis. Hot carriers are of interest because they have an energy significantly higher than the mean carrier energy and can thus cause device degradation by injection into the oxide and by interface state generation.

In the first approach the hydrodynamic model was used to calculate the first three moments of the carrier distribution using the Boltzmann transport equation. The calculated mean electron energy was then used to estimate the hot carrier population and the substrate current for MOSFET's. The drain current obtained from the hydrodynamic model was accurate for all biases and channel lengths down to 0.16 μm, but the calculated substrate current had a large error when compared with experimental data, near the threshold voltage of the MOSFET. The error is attributable to the use of an ensemble average, the temperature, to describe the details of the distribution, which is quite anisotropic at this bias condition.

To overcome this error which is inherent in a moments method, the second part of the thesis studies a method that solves the Boltzmann equation directly by expanding the distribution function in surface spherical harmonics in momentum space. Using the spherical harmonic expansion in momentum space and a standard finite difference discretization in real space, the Boltzmann equation is solved after incorporating the relevant scattering mechanisms and an appropriate band structure. The novelty of this thesis lies in the use of a Galerkin method which allows the spherical harmonic formulation for arbitrary order expansions. Results are presented for expansions up to third order in both one and two real space dimensions. It is shown that the higher order harmonics are significant at high fields. In two dimensions a rotated coordinate system approach was proposed and demonstrated. This method minimizes the number of harmonics that are needed for a given level of accuracy by always aligning the polar direction of expansion with the electric field direction. Important details regarding the the discretization of the spherical harmonic coefficients as well as computational efficiency issues are also addressed.

Thesis Supervisor: Jacob White
Title: Associate Professor

Acknowledgement

I would like to thank my thesis advisor, Prof. Jacob White and my thesis committee member, Prof. Dimitri Antoniadis for their guidance, help and support throughout my doctoral program. Prof. Jim Chung is also appreciated for agreeing to serve on my thesis committee.

Thanks to all the students of the 8th floor VLSI group, especially those who kept the computer systems running. Part of the thesis used code written by Mark Reichelt; these programs and many helpful discussions with him are greatfully acknowledged. Keith Nabors and Ricardo Telichevesky deserve a special mention for letting me wander into their office for impromptu discussions on all topics under the sun. Two office mates, Filip Van Aelten and José Carlos Monteiro made our office a bit more lively and kept me informed about exotic far away lands.

The deepest thanks must of course go to the United States tax payers who have spent over $200,000 on this work. I hope this document goes some way in convincing them that it hasn't been a complete waste.

... Much as he had accomplished, she could not but observe that his most splendid successes were almost invariably failures, if compared with the ideal at which he aimed. His brightest diamonds were the merest pebbles, and felt to be so by himself, in comparison with the inestimable gems which lay hidden beyond his reach.

The Birthmark

Nathaniel Hawthorne

He that has improved the Virtue or advanced the Happiness of one Fellow-Creature, he that has ascertained a single moral Proposition, or added one useful Experiment to natural Knowledge, may be contented with his own Performance, and with respect to mortals like himself, may demand, like *Augustus*, to be dismissed at his departure with Applause.

Idler #88

Dr. Samuel Johnson

I hate set dissertations,— and above all things in the world, 'tis one of the silliest things in one of them to darken your hypothesis by placing a number of tall, opake words, one before another, in a right line, betwixt your own and your reader's conception.

Tristram Shandy

Laurence Sterne

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Hot Carrier Problem

The goal of this thesis is to calculate the population of high energy carriers in semiconductor devices. High energy or hot carriers are of interest because they are implicated in a type of device degradation known as hot-carrier degradation. Device degradation is a reliability concern as it may, over a long period of device operation, significantly alter the electrical characteristics of the device, which in turn could lead to unintended changes in both the static and dynamic circuit performance.

Typical examples of hot carrier effects are an increase in the threshold voltage (for NMOS transistors); a decrease in drain current drive, both in the linear and saturation region; and asymmetry in the source and drain of an MOS transistor [1] [2]. These effects have been traced to the creation of some combination of interface states and trapped charge in the gate oxide of a MOSFET by hot carriers from the silicon that can overcome the potential barrier at the silicon-silicon dioxide interface [3] [4] [5] [6] [7]. The magnitude of the degradation depends on the applied biases which control the electric field in the device and therefore the carrier energy. For a given bias condition, the degradation increases with the duration of the stress and also depends on the device structure and the morphology of the oxide. Under normal operating conditions the device lifetime, which is usually defined somewhat arbitrarily as a 3% (or 10%) reduction in the linear current or a 0.1V shift in the threshold voltage, is expected to be years. For this reason device degradation is usually measured empirically at elevated drain voltages to accelerate degradation. Then the measured lifetime is extrapolated to obtain the lifetime at nominal bias. This kind of extrapolation works reasonably well because the degradation essentially depends on the integrated flux of hot carriers: thus a higher stress for a short time can be used to estimate a lower stress for a long time.

An even simpler way of estimating the robustness of a MOS transistor is to calculate or measure the substrate and gate currents produced at different bias conditions and use them as a measure of the susceptibility of the device to hot carrier degradation. The argument for the use of the substrate current is that it consists primarily of holes (in NMOS transistors) which

are generated as electron-hole pairs under impact ionization by hot electrons which are also the primary source for hot carrier injection into the oxide.[1] The gate current is of course an even more direct estimate of the number of highly energetic carriers. Although, the details of the correlations between the substrate and gate currents and the degradation are beyond the scope of this thesis as they are discussed extensively in the literature cited above, we will consider the ability to accurately calculate the substrate current as a necessary test for any method that attempts to calculate the hot carrier population.

Although we have only cited examples of degradation in MOS transistors the effects have also been reported in the literature for bipolar devices [8] [9].

The details of the degradation mechanism and the exact role played by interface states vs. trapped charge are still under debate and an active area of research. But all the theories begin by assuming an electron or hole flux from the silicon into the oxide i.e. a flux of highly energetic carriers with sufficient energy to overcome the oxide barrier. Thus knowledge of the number and momentum of hot carriers is an essential ingredient in quantitatively predicting device degradation from first principles. This task is made difficult by the fact that the energy required for impact ionization in silicon is about 1.5 eV and over 3.0 eV for injection into the oxide.[2]. These energies are of course much higher than the energies usually encountered by most electrons which are closer to 0.1–0.2 eV. Put another way, the mean energy for electrons, say, in a device operating under normal circumstances is very much less than the energy of the electrons that cause hot carrier degradation. Thus the hot carriers are a small fraction of the total number of carriers and can be thought to form the tail of the distribution of electrons as a function of energy. Hence calculating the hot carrier population correctly is usually a much harder problem than calculating the drain current which is an ensemble average over the whole distribution and is therefore dominated by carriers with energies near thermal equilibrium. In the future this statement may not hold, specially if devices are aggressively scaled down to below 0.15 μm. The reason for this is that at those ultra-short channels, a much larger fraction of carriers have energies significantly above the thermal energy and therefore even macroscopic averages may depend on band structure and scattering effects in a more complex way than heretofore seen. Therefore the hot carrier problem may become more central to the broader device simulation problem which is primarily focused on calculating terminal currents.

1.2 Hot Carrier Modeling

A hierarchy of techniques exist for simulating the hot carrier population in a semiconductor device. These range from elaborate full band structure Monte-Carlo simulations [11] to

[1]There is a small contribution to the substrate current from the diode leakage current of the drain/substrate junction but this is usually many orders of magnitude smaller than the hole current due to impact ionization.

[2]The threshold energy for impact ionization is given by $\frac{1+2\mu}{1+\mu} E_{\text{gap}}$ where μ is the ratio of the conduction band to valence band effective masses and E_{gap} is the bandgap [10].

macroscopic methods based on the drift-diffusion equations which are commonly used to calculate terminal currents [12]. As one would expect the most accurate and predictive methods require sophisticated physical models and substantial computational resources while the simplest models, though economical with computational time, tend to have limited accuracy and use a large number of fitting parameters.

The state of the art in modeling hot carriers is the Monte-Carlo method which solves the Boltzmann equation using a stochastic method. The essence of the Monte-Carlo method consists in tracking a large number of particles as they undergo lattice collisions while accelerating in an electric field (calculated self-consistently) in a semiconductor. Thus at the microscopic level one just imposes the semiclassical dynamics on each carrier with due regard to the frequency of the different scattering events. In principle, complex band structures can be readily included if the $E(k)$ function is tabulated, and after each scattering event the new momentum vector chosen based on conservation of energy and crystal momentum. In the Monte-Carlo method all types of phonons and impurity scattering (transverse/longitudinal acoustic and polar/non-polar optical etc.) are considered, although electron-electron scattering is not easily handled [13]. But as electron-electron scattering is not a dominant scattering mechanism except at very high concentrations neglecting it does not lead to signifianct inaccuracy. What is a major problem is that to get statistically meaningful data a large number of carrier histories must be simulated. This problem is exacerbated when carriers travel into regions of high doping concentrations where an even larger number of carriers must be simulated or sophisticated techniques used to correctly weight the carriers.

Another consequence of the stochastic approach used in the Monte-Carlo method is the difficulty of simulating rare events such as the ones that produce very high energy carriers, which may be a small fraction of the total number of carriers. Clearly, to get this small probability right an even larger sample space is needed.

Thus at present multidimensional Monte Carlo simulators are too slow to be used as engineering tools. Running times for the more sophisticated Monte Carlo simulators can be many hours per bias point on the most powerful computers, though their unmatched flexibility and sophistication make them the yardstick by which other simulation techniques are measured.

At the other end of sophistication are macroscopic approaches that do not attempt to calculate directly the distribution function but try to estimate it from average quantities such as mean energy and velocity. The simplest of these are post-processing methods that use the results obtained from the drift-diffusion model as a basis for calculating the electron energy using a simplified energy balance equation [12] [14] . These approaches though efficient, are usually limited in their predictive capability as they use a number of fitting parameters which are used to match experimental data but lack theoretical justification.

In the next section we present a summary of the contribution of the work presented in this thesis to the field of hot carrier modeling.

1.3 Thesis Contribution

The work described in this thesis comprises of two parts. In the first part we explore the ability of macroscopic methods to calculate the distribution of hot carriers. The primary emphasis here is on the energy balance or the hydrodynamic model which is a natural extension of the drift-diffusion model. Although the energy balance method, which solves an energy conservation equation akin to the charge and momentum balance equations already included in the drift-diffusion model, has been studied actively for the last few years there exist a number of different formulations and numerical implementations. The work in this thesis focuses on one physically justified formulation and its discretization. We also compare the terminal currents obtained from this implementation with experimental data for a range of devices down to $0.16\,\mu$m. Most of the work using the hydrodynamic model used data for comparison from devices with longer channels lengths and therefore did not fully test the accuracy of the model. More importantly we use a simple method, which uses a minimum of fitting parameters, to estimate the hot carrier population and compare the substrate currents predicted by the simulation with the experimental data for highly scaled devices. The calculated substrate current obtained by our technique while accurate for a range of bias voltages, has large errors near the threshold voltage of the MOSFET. We attribute this behavior of the simulation results obtained from the hydrodynamic model to the use of the electron energy or temperature which is an ensemble average and therefore cannot in principle provide detailed information about the distribution of the hot carriers. If the distribution is a Maxwellian or a shifted Maxwellian then the electron temperature can be used effectively to estimate the hot carrier population but for more complex distributions the temperature does not yield sufficient information to estimate the hot carrier population.

From this observation we were led to consider approaches that allow direct calculation of the distribution function of carriers, rather than some average quantity, while at the same time they are more efficient than the Monte-Carlo method. The second (and the larger) part of the thesis explores a method that may indeed meet both the criteria of accuracy and efficiency. In this method the Boltzmann equation, which is the fundamental description for transport in the semiclassical model, is solved by using a basis function expansion for the distribution function. The distribution function which describes the carrier density in phase space (real and momentum space) is expanded in surface spherical harmonics in momentum space. The choice of spherical harmonics is guided by the physical argument that the distribution has some spherical symmetry because in equilibrium the distribution is a Maxwellian which is perfectly spherical. Also many types of scattering have a randomizing nature and would induce some degree of spherical symmetry. Using the spherical harmonic expansion in momentum space and a standard finite difference discretization in real space we solve the Boltzmann equation after incorporating scattering mechanisms and a simplified band structure. Although the above approach has been used in the past, the novelty of this thesis lies in the use of a Galerkin method which allows the spherical harmonic expansion

formulation for arbitrary order expansions. Results are presented, for the first time, up to third order in both one and two dimensions and it is shown that the higher order harmonics can significantly alter the distribution at high fields. In two dimensions we also propose and demonstrate a rotated coordinate system approach that minimizes the number of harmonics that are needed for a given level of accuracy. The rotated coordinate system approach can reduce the computational cost of the problem by factors of two or three with only a minor decrease in accuracy. Important details regarding the choice and stability of the discretization of the spherical harmonic coefficients are presented. Also numerical issues, which are specially significant in two real space dimensions, are addressed fully.

1.4 Thesis Organization

We begin, in the next chapter, by describing the macroscopic approach to solving the hot carrier problem. The hydrodynamic model and the associated parameters that were implemented as part of this thesis are explained and the simulation results obtained from them are compared with experimental data. The ability of the hydrodynamic model to successfully predict the hot carrier distribution based on these results and the underlying theory is then discussed.

The next two chapters focus on the solution of the Boltzmann equation using a spherical harmonic expansion. In chapter three, the basic theory, physical mechanisms and the band structure are described. Results are presented in this chapter for the one dimensional real space problem. Chapter four extends the approach to multiple dimensions and also provides greater details of the numerical methods that were used as they become more significant in higher dimensions.

The last chapter gives a summary of the thesis results and discusses the future directions that may be pursued to enhance the work presented in this thesis.

Chapter 2

Macroscopic Methods

2.1 Introduction

The earliest work in device simulation was focused on calculating macroscopic parameters such as electron and hole concentrations and currents. This approach is of course not surprising because it allowed the calculation of the data most important to the device technologist with the minimum possible computing resources. This work was thus an extension of the analytic work that had been done in device modeling since the 50's. The basic problem was the solution of a set of partial differential equations, the drift-diffusion equations, which represented the conservation laws for electron and hole charge and momentum. The analytic solution to these equations was restricted to the one-dimensional case (with constant coefficients) though by clever approximations it could be extended to obtain solutions to the two-dimensional MOSFET problem as well. The numerical solution of these same problems allowed the use of much more realistic parameters (such as mobility and non-ideal contacts) and true two dimensional solutions. This kind of drift-diffusion model is embodied in many simulators such as MINIMOS [15], PISCES [16], FIELDAY [17] etc. which came into significant use in the early to mid 80's.

Beginning in the late 80's, shrinking device dimensions and the concomitant increase in electric fields (even with reduced supply voltages more recently), uncovered two major problems with the drift-diffusion based simulators. The first was that the macroscopic mobility models that had been used became less accurate, thus producing erroneous terminal currents. The second was that a new class of physical phenomena, the so-called hot-carrier effects, manifested themselves in devices with channel lengths shorter than one micron. Accurately simulating these phenomena, which are caused by the high energy non-equilibrium carrier populations, was a challenge for the drift-diffusion simulators because the models incorporated in them assumed local thermal equilibrium with the lattice and therefore they were in principle unequipped to model the hot carriers problem.

Attempts were made to solve the mobility modeling problem by refining and elaborating the mobility models used in the drift-diffusion simulators. Indeed, significant improvement

was made in the accuracy of the currents calculated by these simulators but at the cost of logical coherence and the lack of physical justification for the new models.

Two distincts approaches were taken to attack the hot carrier modeling issue. The first relied on what may be called a post-processing approach in which the results obtained from the drift-diffusion simulator were used as initial data to some fairly simple but physically intuitive models to estimate the hot carrier populations [18] [19] [20] [14]. In most of these cases, an *ansatz* regarding the energy dependence of the distribution was required to actually calculate the hot-carrier populations given the electron concentrations from the drift-diffusion simulator. Typically a Maxwellian or an exponential in energy (which is estimated from the local electric field) was used as a guess for the distribution function. Further work [21] though, showed that even the electron concentration obtained using the drift-diffusion model can be surprisingly inaccurate near the drain of a MOSFET under high field conditions. For example, in the high field region below the surface near the drain in a MOSFET, the electron concentration calculated by the drift-diffusion simulators could be orders of magnitude smaller than that calculated by more accurate methods such as Monte-Carlo. Therefore any attempt at naive post-processing was highly questionable. This led to the second approach to rectify the shortcoming of the drift-diffusion simulators: the hydrodynamic or energy balance model.

The essential idea of the hydrodynamic model is to explicitly include carrier energy (or equivalently the temperature) as one more unknown to be solved for, besides electron concentration and current. This requires the formulation of an energy conservation equation akin to the charge and momentum balance equations already included in the drift-diffusion model. This approach, therefore, has the advantage of being the next logical step in refining the drift-diffusion equation; it does not resort to an *ad hoc* technique. Another, perhaps more illuminating, way of thinking about the hydrodynamic model is to consider it as the next element in a a sequence of successively better estimates of the distribution function in momentum space. The electron concentration gives just the integral of the distribution function over momentum space, the current provides information of the relative shift of the distribution from the origin while the electron temperature provides information about the relative spread of the distribution. A high carrier temperature implies a broad distribution function whereas a low carrier temperature is associated with a more sharply peaked distribution. These qualitative notions are shown schematically in Fig. 2-1.

Just as in the case of the drift-diffusion model there are many implementations of the hydrodynamic model. In fact because of its greater complexity and the lesser accuracy of the parameters used in the model, a bestiary of theoretical models and implementations can be compiled [22] [23] [24] [25] [26]: each purporting to be the minimal sufficient model necessary to accurately predict the hot carrier population. Therefore, we implemented a version of the hydrodynamic model to actually test the accuracy of this technique by comparison with experimental data and to obtain a measure of its effectiveness.

In the following sections of the chapter we will first describe the hydrodynamic model, then give details of our implementation of this method and the results obtained from it along

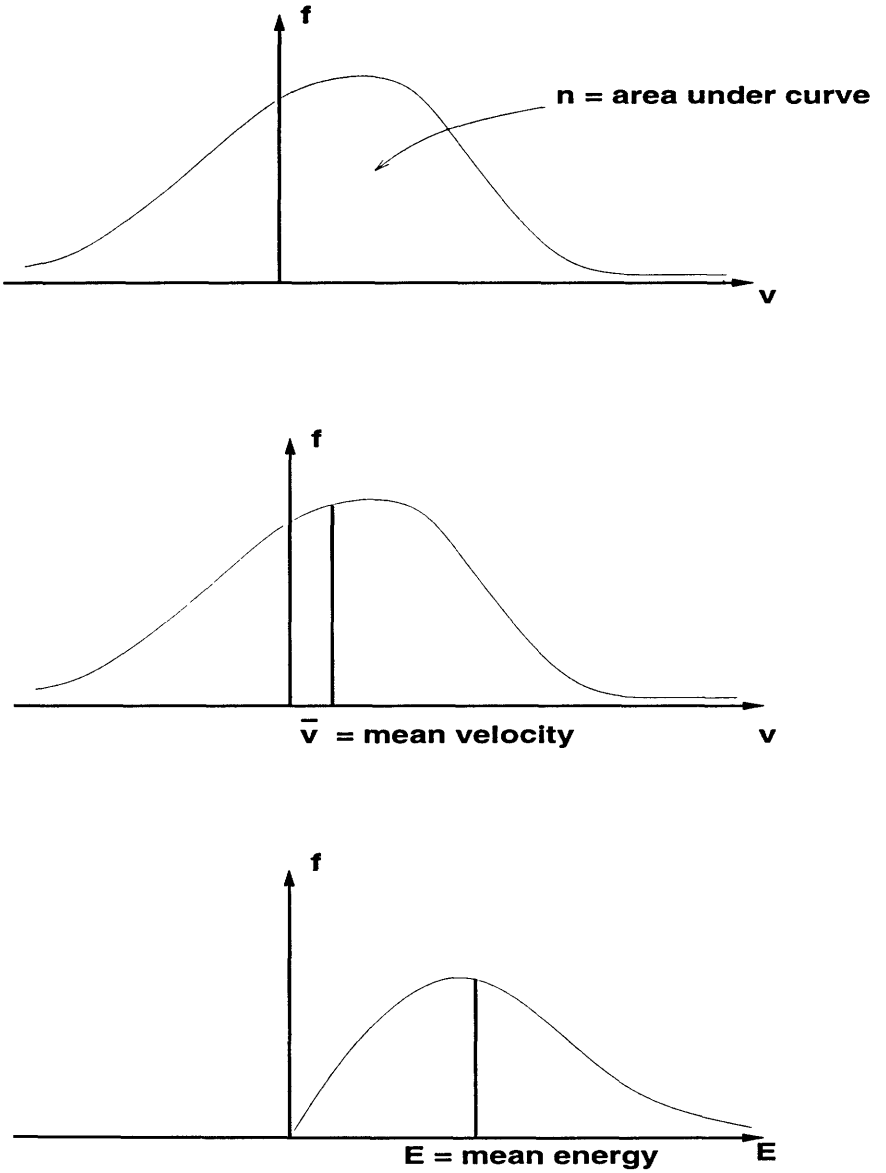

Figure 2-1: A schematic representation of the first three moments of the distribution function.

with comparison with experimental data. Finally a critique of the hydrodynamic model and some assessment of its usefulness will be presented.

2.2 Mathematical Formulation

A considerable body of work exists on the energy balance model [22] [23] [25] [24] [26]. Although the original formulation in [22] was presented more than twenty years ago, only in the last few years have there been sufficient computing resources to implement this method. The focus of most of the earlier work had been on calculating drain currents and only a few studies actually used the information about the electron energy to obtain substrate currents and compare the simulation results with experimental data for short-channel MOSFET's. The work that is discussed in this chapter of the thesis is aimed toward developing a stable numerical implementation of the energy balance formulation and determining the limits of its validity by comparing with experimental data for ultra-short-channel devices. We will focus on modeling transport in n-channel MOSFET's and therefore restrict the formulation and solution to a single carrier, electrons. The theory for holes is essentially the same after the appropriate sign changes.

To develop the energy balance model we begin by taking moments of the Boltzmann Transport Equation:

$$\mathbf{v}(\mathbf{k}) \cdot \nabla_r f(\mathbf{r},\mathbf{k}) - \frac{q\mathcal{E}(\mathbf{r})}{\hbar} \cdot \nabla_{\mathbf{k}} f(\mathbf{r},\mathbf{k}) = \int S(\mathbf{k}',\mathbf{k}) f(\mathbf{r},\mathbf{k}')\, d^3\mathbf{k}' - f(\mathbf{r},\mathbf{k}) \int S(\mathbf{k},\mathbf{k}')\, d^3\mathbf{k}' \quad (2.1)$$

where $f(\mathbf{r},\mathbf{k})$ is the distribution function for electrons which is a function of real space, $\mathbf{r}$ and momentum space $\mathbf{k}$, $\mathbf{v}(\mathbf{k})$ is the electron velocity which is determined from the band structure for the semiconductor in consideration and $\mathcal{E}(\mathbf{r})$ is the electric field. The right hand side of the BTE contains the in-scattering and out-scattering terms where the scattering rate is denoted by $S(\mathbf{k},\mathbf{k}')$ which denotes scattering from $\mathbf{k}$ to $\mathbf{k}'$.

Formally, the moments of the distribution function are defined as

$$n = \int f(\mathbf{r},\mathbf{k})\ d^3\mathbf{k} \quad (2.2)$$

$$n\bar{\mathbf{v}} = \int f(\mathbf{r},\mathbf{k})\mathbf{v}(\mathbf{k})\ d^3\mathbf{k} \quad (2.3)$$

$$n\bar{w} = \int f(\mathbf{r},\mathbf{k})\ m_n^* \mathbf{v}(\mathbf{k}) \cdot \mathbf{v}(\mathbf{k})\ d^3\mathbf{k} \quad (2.4)$$

In the above equations, n is the electron concentration, $\bar{\mathbf{v}}$ is the mean electron velocity, $\bar{w}$ is the mean electron energy and m_n^* is the electron effective mass. For each of the above moments we can write conservation laws starting from the Boltzmann Equation. For the scattering terms we will assume the relaxation time approximation which allows us to parametrize the scattering terms using two scattering times: the momentum relaxation time, τ_n and the energy relaxation time, τ_w. In this so-called relaxation time approximation the collision terms are

given as:

$$\left(\frac{\partial \bar{\mathbf{v}}}{\partial t}\right)_c = -\frac{\bar{\mathbf{v}}}{\tau_n} \tag{2.5}$$

and

$$\left(\frac{\partial w}{\partial t}\right)_c = -\frac{\bar{w} - w_0}{\tau_w} \tag{2.6}$$

where $w_0 = \frac{3}{2}k_B T_0$ is the energy of an electron in equilibrium. Using the above definitions after a lengthy derivation which is shown in detail in Appendix A, one obtains the three moment equations:

$$\mathbf{J}_n - \frac{\tau_n}{q}(\mathbf{J}_n \cdot \nabla)\left(\frac{\mathbf{J}_n}{n}\right) = qD_n \nabla n - q\mu_n n \nabla\left(\psi - \frac{k_B T}{q}\right) \tag{2.7}$$

$$\nabla \cdot \left[-\kappa \nabla T - \frac{\mathbf{J}_n}{q}(k_B T + w)\right] = \mathcal{E} \cdot \mathbf{J}_n - n\left(\frac{\bar{w} - w_0}{\tau_w}\right) - \bar{w}U \tag{2.8}$$

$$\nabla \cdot \mathbf{J}_n = qU \tag{2.9}$$

These along with Poisson's equation:

$$\nabla \cdot \mathcal{E} = \frac{q}{\epsilon}\left(n_i e^{\frac{q(\phi_p - \psi)}{k_B T_0}} - n + N_D - N_A\right) \tag{2.10}$$

form our hydrodynamic model. In the above equations, n is the electron concentration, $\mathbf{J}_n$ is the electron current and w is the average electron energy which is given by

$$w = \frac{1}{2}m_n^* v_n^2 + \frac{3}{2}k_B T \tag{2.11}$$

where v_n is the electron velocity, T is the electron temperature and k_B is Boltzmann's constant. In (2.7), τ_n, D_n and μ_n are the electron momentum relaxation time, diffusion constant and mobility. In (2.8), τ_w, κ, and U are the energy relaxation time, thermal conductivity, and the net recombination rate per unit volume for the electrons. w_0 is the thermal energy of the electrons in equilibrium and equals $\frac{3}{2}k_B T_0$ where T_0 is the lattice temperature. In Poisson's equation, (2.10), $\mathcal{E}$ is the electric field, q is the absolute value of the electronic charge and ϵ the permittivity of silicon. ψ and ϕ_p are the electrostatic potential and hole quasi-Fermi level respectively. The hole quasi-Fermi level is assumed to be constant and is determined by the biases on the contacts. Finally, the electrostatic potential and electric field $\mathcal{E}$ are related by $\mathcal{E} = -\nabla\psi$.

As mentioned earlier, the above system of equations can be considered an extension of the familiar drift-diffusion system which had been the mainstay for modeling transport. This is most clearly seen in the case of the momentum conservation equation, (2.7), where the right hand side is essentially the same as that in the drift-diffusion system except for the addition of a term which incorporates the effects of gradients in electron temperature. Also the left hand side has a correction to the current which is similar to that arising in fluid mechanics $(v \cdot \nabla v)$ and is called a convective term there, though as the current in MOSFET's tends to be dominated by electron drift we expect this term to be small and neglect it.

The energy balance equation (2.8) is, of course, a completely new equation that is not present in the drift-diffusion model. The right hand side of (2.8) represents the sources and sinks of carrier heating: the first term is Joule heating, the second term is energy relaxation through collisions and the last term is energy loss from net recombination. The left hand side of the energy balance equation represents the flux due to thermal gradients and an energy flux through carrier transport or flow.

Given mobility, thermal conductivity, and energy relaxation time models, the above set of equations can be solved to determine current and temperature distributions in MOS devices. We used the standard Weidemann–Franz law thermal conductivity model [27], and a mobility model which is a function of doping and vertical field near the device surface [28], and inversely proportional to the electron temperature [29]. The details of these parameters are given in Appendix B. For the energy relaxation time we used a constant value of 0.1 ps.

2.3 Numerics

2.3.1 Discretization Scheme

To solve the system of PDE's shown in the previous section we need to first discretize the equations over a mesh in space. We will solve the problem in two dimensions and consider only a simple rectangular mesh, though the methods carry over to a more general triangular mesh fairly easily.

The most convenient way of discretizing the PDE's is to put them in a conservation law form such as Poisson's equation or the current conservation equation. For example for Poisson's equation we enforce the condition that the net flux leaving a 'control volume' must equal the integral of the charge in that control volume as shown in Fig. 2-2. Thus the discretized version of Poisson's equation is given as:

$$\sum_j \left(\frac{\psi_i - \psi_j}{d_{ij}} \right) l_{ij} = \mathrm{dA}\, \rho_i. \tag{2.12}$$

This discretization scheme can obviously be applied to the current conservation equation as well, the only difficulty being that the value of the flux at the edges, the current in this case, is not as simply stated as in the case of the electric field. This difficulty arises from the nature of the current equation itself, which in the case of the simpler drift-diffusion model in the x dimension is:

$$J = qD_n \frac{\partial n}{\partial x} - q\mu_n \frac{\partial \psi}{\partial x} n. \tag{2.13}$$

The obvious discretization method for the above equation: a two point discretization of the electron concentration and the electric potential can, unfortunately, become unstable. The source of this problem is that to ensure stability, the direction of the two point discretization of the electron concentration (i.e. backward or forward from the point around which the fluxes are being computed) must depend on the sign of the electric field. The electric field can obviously

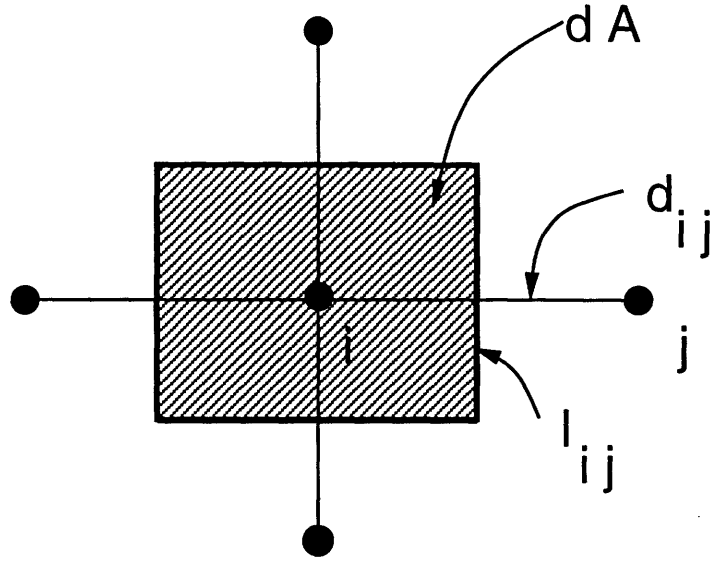

Figure 2-2: Discretization using a control volume method for Poisson's equation

change sign and consequently, for a fixed two point scheme the method can go unstable. One solution would be to change the direction of discretization depending on the electric field, but a more elegant solution was suggested in [30], which incorporated this 'winding' almost automatically by exploiting the exponential variation of the electron concentration and the relative smoothness of the current.

In the Scharfetter-Gummel discretization method, instead of just using a two point discretization to calculate the current along an edge we actually solve the one-dimensional ordinary differential equation Eq. 2.13, along that edge. This can be done if we assume a constant current and average mobility along the edge and a constant electric field between the points (i.e. a linear potential variation). For boundary conditions we use the electron concentration at the two nodes connecting the edge we are considering. The solution to this ordinary differential equation can then be written as:

$$\tilde{J}_{ij} = qD_n \frac{1}{d_{ij}} \left[B(u_{ji}) n_j - B(-u_{ji}) n_i \right] \tag{2.14}$$

where $B(x) = \frac{x}{e^x - 1}$ is the Bernoulli function; $\tilde{J}_{ij}$ is the assumed constant current along this edge; u_{ji} is the normalized electron potential difference defined by $\frac{\psi_j - \psi_i}{k_B T}$. The Bernoulli function has a peculiar feature that it is almost linear for negative arguments and an inverse exponential for large positive arguments as shown in Fig. 2-3. Thus when the potential between two nodes is positive ($u_{ji} > 0$), then the current is $J^+ \approx -qD_n \frac{u_{ji}}{d_{ij}} n_i$; whereas when the potential difference is negative ($u_{ji} < 0$), the current is $J^- \approx qD_n \frac{u_{ji}}{d_{ij}} n_j$. When the potential difference is small, $u_{ji} \approx 0$, then the current is approximately is $J^0 \approx qD_n \frac{n_j - n_i}{d_{ij}}$ which implies that current flow is by diffusion as expected from physical reasoning. Therefore it is evident that the Scharfetter-Gummel discretization method captures the physics of the one-

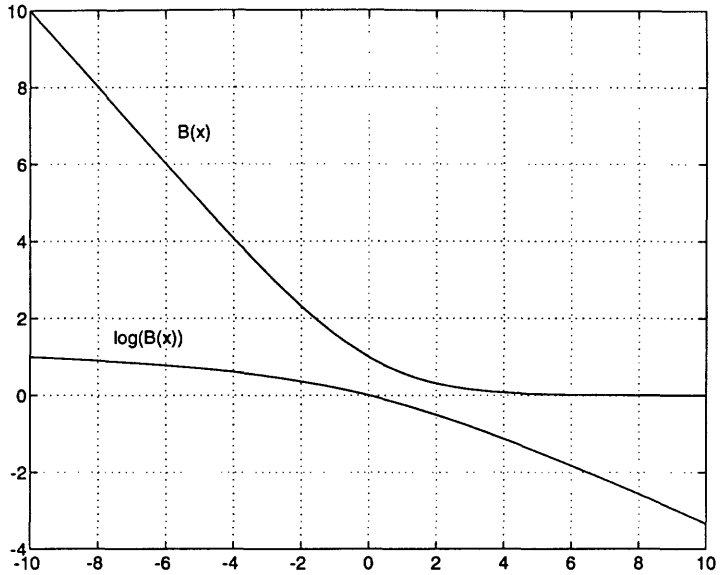

Figure 2-3: Bernoulli function and its logarithm

dimensional flow problem in a remarkably succinct manner. This discretization method can also be used for the current equation of the hydrodynamic model (2.7), with the modification that the argument of the Bernoulli function now is $(u_{ji} - r_{ji})$ where r_{ji} is the difference in the normalized electron temperatures $r_i = T_i/T_0$ and $r_j = T_j/T_0$.

In the discretization scheme for the energy equation used in [31], which is an extension of the work in [26], the problem is cast in terms of an energy flow density, $\mathbf{S}$, defined by

$$\mathbf{S} = -\kappa \nabla T - \left(\frac{5}{2} k_B T\right) \frac{\mathbf{J}_n}{q}. \tag{2.15}$$

Hence the energy equation can be written as

$$\nabla \cdot \mathbf{S} = \nabla \cdot \left(\frac{m_n^* |\mathbf{J}_n|^2}{2q^2 n^2} - q\psi\right) \frac{\mathbf{J}_n}{q} + (q\psi - w)U - n\left(\frac{w - w_0}{\tau_w}\right). \tag{2.16}$$

One advantage of this formulation is that a Scharfetter-Gummel or exponentially-fit discretization scheme can be applied. This can be seen by projecting $\mathbf{S}$ onto an edge between nodes i and j,

$$S_{ij} = -\kappa \frac{dT}{dx_{ij}} - \frac{5}{2} \frac{k_B T}{q} J_{ij}. \tag{2.17}$$

Treating S_{ij}, J_{ij} and κ as constant along the edge, (2.17) can be integrated analytically to obtain:

$$S_{ij} = -\frac{\kappa_{ij}}{d_{ij}} \left[B(\omega_{ij}) T_j - B(-\omega_{ij}) T_i\right], \tag{2.18}$$

where

$$\omega_{ij} = -\frac{5}{2} k_B \frac{J_{ij}}{q} \frac{d_{ij}}{\kappa_{ij}}, \tag{2.19}$$

B is the Bernoulli function as defined earlier, κ_{ij} is an average thermal conductivity between the two nodes, and d_{ij} is the distance between nodes i and j.

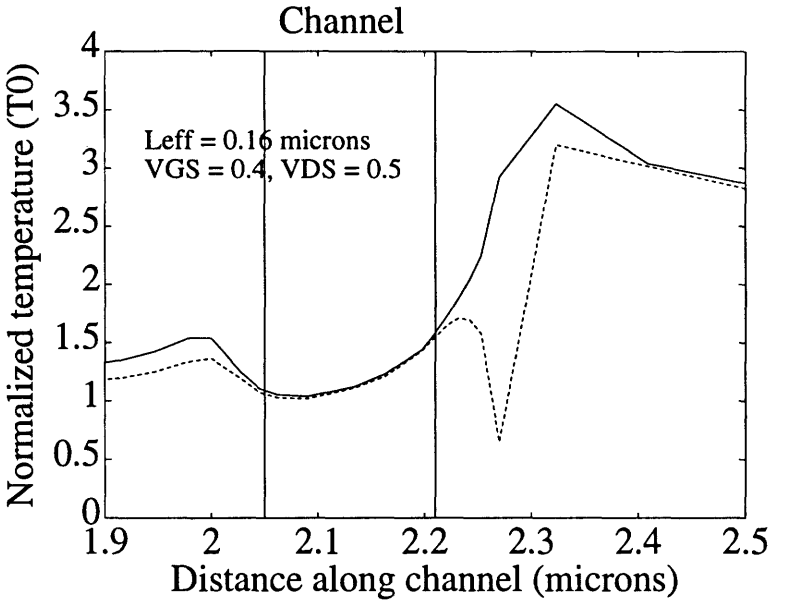

Figure 2-4: Stable and Unstable discretization using the discretization of (2-15) (dashed) and (2-27) (solid) respectively.

2.3.2 Temperature Instabilities

The above discretization technique was implemented in a two-dimensional finite-box based device simulator and used to simulate a short channel MOSFET. To solve the nonlinear algebraic problem generated by the discretization, a Newton's method was used combined with sparse Gaussian elimination to solve for the Newton updates. We observed that the temperatures computed using a coarse rectangular mesh with the discretization method described above exhibited numerical instabilities in certain regions of the device. In particular the computed temperatures oscillated in space, occasionally dipping *below* the lattice temperature. An example of this anomalous behavior is shown in Fig. 2-4. Although such instabilities can be eliminated by refining the mesh, this may not always be practical as the instability may artificially make the discretized problem more nonlinear, which in turn worsens the convergence of any iterative nonlinear solver, such as Newton's method. Without a converged solution, it may not be obvious where to add additional mesh points.

The source of this numerical instability is that the discretization of the energy equation in (2.18) and (2.19) inappropriately assumes that the thermal conductivity is a constant. To see why such an approximation leads to coarse-grid instability, consider computing the divergence of (2.15) assuming $\mathbf{J}_n$, but *not* κ, is constant. The result is

$$\nabla \cdot \mathbf{S} = -\kappa \nabla^2 T - \left(\nabla \kappa + \left(\frac{5}{2} k_B \right) \frac{\mathbf{J}_n}{q} \right) \nabla T. \tag{2.20}$$

To be stable for coarse grids, a method for discretizing (2.20) must *upwind* the ∇T term, that is discretize ∇T in the upwind direction given by the sign of

$$\left(\nabla \kappa + \left(\frac{5}{2} k_B \right) \frac{\mathbf{J}_n}{q} \right). \tag{2.21}$$

Equation (2.19) does not include the $\nabla\kappa$ term, and therefore the resulting Scharfetter-Gummel scheme will not upwind correctly unless the $\nabla\kappa$ term can be ignored. This is not the case, as can be seen if we write the thermal conductivity as

$$\kappa = \left(\frac{5}{2} + c\right) k_B D_{n0} n \tag{2.22}$$

where c is a constant which depends on the dominant scattering mechanism [29] and substitute this relation in (2.21) to yield

$$\left(\frac{5}{2}k_B\right)\left(\left(1 + \frac{2}{5}c\right) D_{n0}\nabla n + \frac{\mathbf{J}_n}{q}\right)\nabla T. \tag{2.23}$$

Clearly, $\left(1 + \frac{2}{5}c\right) D_{n0}\nabla n$ and $\frac{\mathbf{J}_n}{q}$ will be comparable when diffusion contributes significantly to current flow. In particular, this implies that the ∇T term in (2.20) may not be discretized in the upwind direction when ∇n is large. Our numerical experiments verify this, as the temperatures computed with the above approach oscillate in device regions where the electron concentration gradients are large.

2.3.3 Modified Energy Discretization

In this section, we develop a better stabilized discretization scheme for the energy equation. The convective term in (2.7)

$$\frac{\tau_n}{q}\left(\mathbf{J}_n \cdot \nabla\right)\left(\frac{\mathbf{J}_n}{n}\right) \tag{2.24}$$

is explicitly neglected, which makes it possible to substitute the expressions for the thermal conductivity, κ, and electron current density, $\mathbf{J}_n$, into the energy equation as suggested by [32]. The justification for neglecting the convective term is that in MOSFET's where current flow is by majority carriers, this term is generally small compared to $\mathbf{J}_n$ in regions where the electron concentration is large.

Neglecting the convective term, (2.7) becomes

$$\mathbf{J}_n \simeq q\frac{k_B T_0}{q}\mu_n\left[r\nabla n + n\nabla\left(r - u\right)\right] \tag{2.25}$$

where $r = \frac{T}{T_0}$ is the electron temperature normalized by the lattice temperature and u is the normalized electrostatic potential given by $u = \frac{q}{k_B T_0}\psi$. Note (2.25) is identical to (6.4) in [26], once the convective term is neglected. The temperature dependence of the mobility model can be included explicitly into the current equation, as in

$$\mathbf{J}_n = q\frac{D_{n0}}{r}\left[r\nabla n + n\nabla\left(r - u\right)\right]. \tag{2.26}$$

Substituting both the expression for the current (2.26) and the thermal conductivity (2.22) into the energy flux equation yields

$$\mathbf{S} = -\frac{5}{2}k_B T_0 D_{n0}\left[r\nabla n + n\left((2 + \frac{2}{5}c)\nabla r - \nabla u\right)\right]. \tag{2.27}$$

The above expression for S has the same form as that for $\mathbf{J}_n$, (2.26), but with a different coefficient in front of the ∇r term. Hence, the Scharfetter–Gummel method can be just as easily applied to this equation as to the current equation, with presumably equal success. Just as in the equation for $\mathbf{J}_n$, we have assumed that the electron temperature and electrostatic potential vary linearly between the two nodes. This assumption more naturally captures the physical variation of these variables, as it is the electron concentration which needs to be "exponentially fitted" rather than the electron temperature. It should be noted that the method is consistent with respect to temperature, in both the current and energy equation temperature is assumed to vary linearly, but is inconsistent with respect to electron concentration, the assumed form of the spatial variation is different in the two equations.

The discretization of the right hand side of (2.16) poses no special difficulties, and is handled in a conventional manner.

Note that the above modified discretization has a disadvantage if the equation system is solved with an iterative scheme which decouples the current-continuity from the energy equation. That is, given the electron concentration and the electric field, the original discretization scheme yields an energy equation which is linear in temperature, and therefore is easy to solve. However, the modified discretization scheme yields an energy equation which is exponentially nonlinear in temperature, and may produce solutions which will not converge in a decoupled scheme.

Using the modified discretization scheme with the same mesh spacing and biases as used for the unstable case, the solution shown in Fig. 2-4 does not display any instability. Of course, this solution should not only be stable, but also accurate in the sense that it should be close to the "correct" solution. To check that this is the case, we computed solutions with a much finer mesh (in both dimensions) using both the stable and unstable discretization techniques. The results obtained on this mesh are shown in Fig. 2-5. Note that the temperature profiles obtained from both methods on the finer mesh are similar, and that the stable discretization solution on the coarse mesh is at least qualitatively similar to the finer mesh solutions. The mesh size in the coarse case was 31x24 while it was 37x38 in the finer case, where most of the added mesh lines were placed in the region where the instability appeared.

2.4 Simulation Results

In this section we compare the results obtained from our two-dimensional simulator for devices with effective channel lengths from 0.16 μm to 0.90 μm with experimental data reported in [33]. Gate oxide thickness for these devices is 52 Å, the junction depth is about 0.09 μm and the device width is 10 μm for all the simulated MOSFET's. One set of parameters was used for all devices and a constant series resistance of 30 Ω was added to the source and drain of the simulated devices to account for extrinsic device resistance. The simulated electron concentration contours for three different channel lengths, shown in Fig. 2-6, clearly show the effect of carrier heating near the drain. The effect which is due to the the vertical diffusion

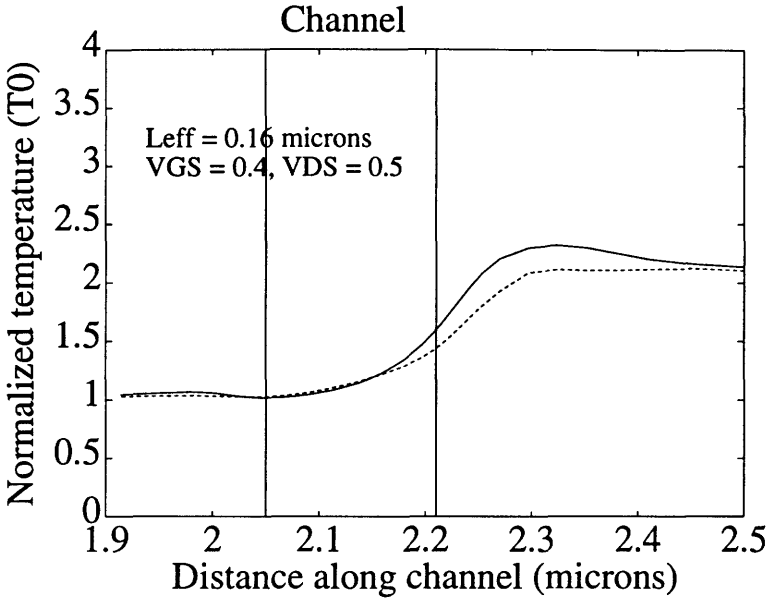

Figure 2-5: Results for discretization using (2-15) (dashed), and (2-27) (solid) but with a finer spatial discretization than in Fig. 2-4

of the carriers into the bulk is most pronounced for the shortest channel length. Contours of electron temperature for the shortest device are shown in Fig. 2-7.

2.4.1 Drain Current Calculation

The drain current predicted by the simulator and that actually measured for three different channel lengths is shown in Fig. 2-8. The predicted current is within 10% of the measured value for all three channel lengths and all the biases.

A question of some technological interest is whether the hydrodynamic model is needed for the prediction of drain currents and if so at what channel length. It has been suggested that due to velocity overshoot near the source the current in an ultra-short-channel may exceed the value predicted by the drift-diffusion model, which imposes velocity saturation. In the hydrodynamic model, of course, no such limitation is built-in and we should expect greater fidelity to experiment.

To answer this question we performed simulations using the hydrodynamic model described above and the drift-diffusion model using an electric field dependent mobility. In Fig. 2-9 we plot the computed small-signal transconductance of MOSFET's with different channel lengths using the drift-diffusion and the hydrodynamic models. The transconductance was calculated at a bias voltage of $V_{DS} = 2.0$ V and $V_{GS} = 1.2$ V. Clearly, for devices with channel lengths much longer than about 0.15 μm the difference between the two simulations is not significant. For shorter channel lengths, the simulations based on the hydrodynamic model predict a more rapid increase in transconductance than those based on the drift-diffusion model, The result obtained by the hydrodynamic model are quite similar to the experimen-

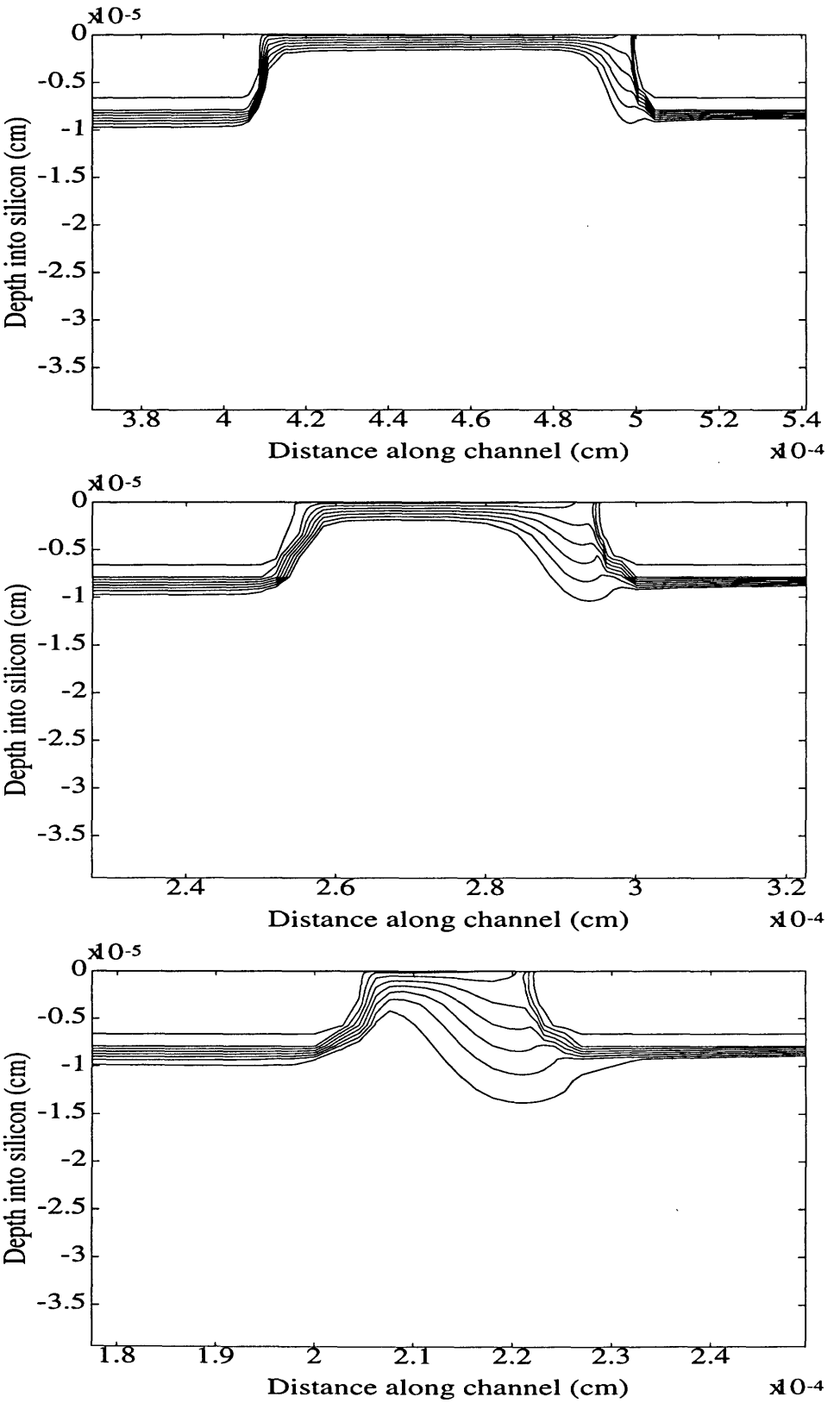

Figure 2-6: Simulated contours of electron concentration for three different channel lengths at the same bias conditions. Contours are separated by a factor of 3 change in the electron concentration.

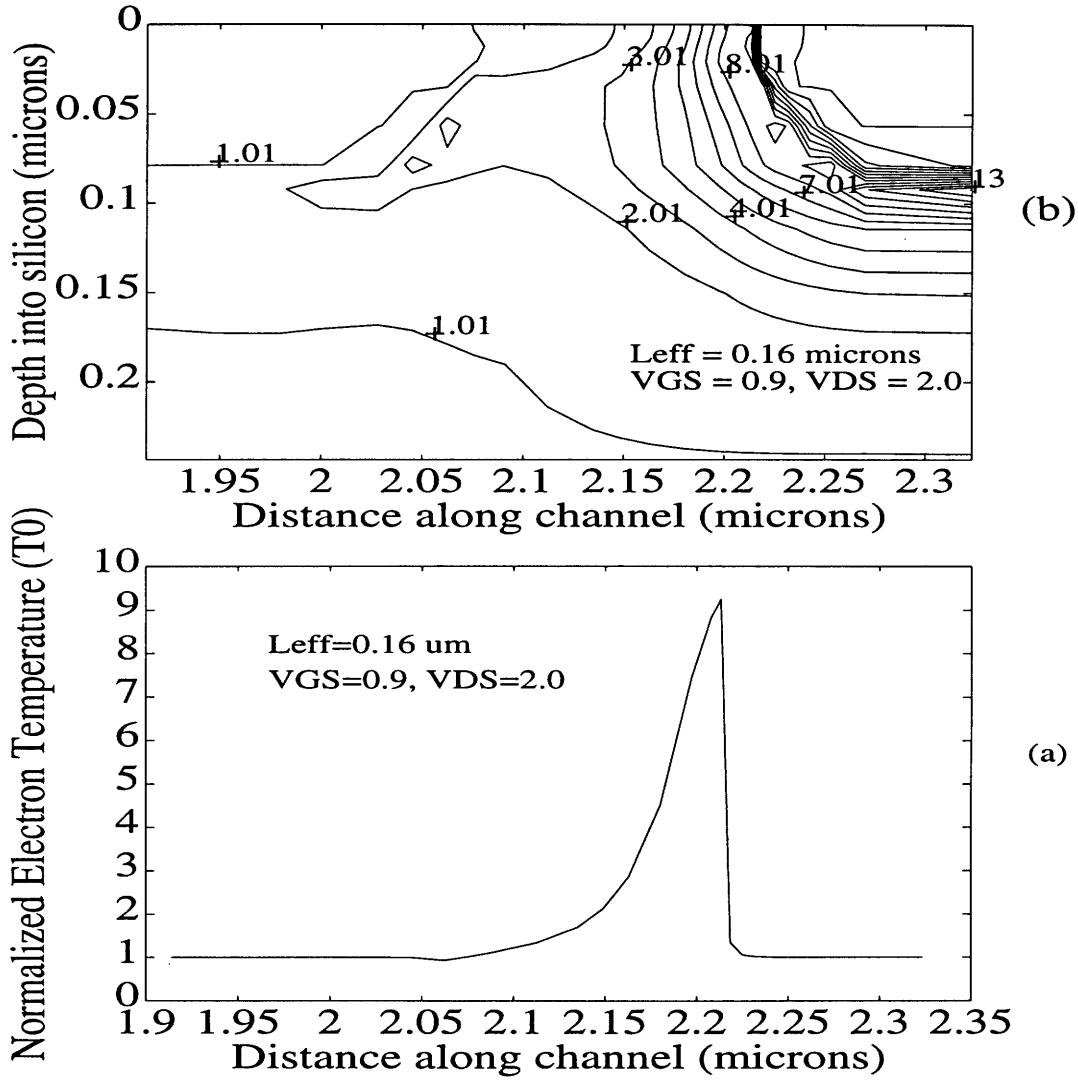

Figure 2-7: Simulated contours of electron temperature and a plot of the temperature at the silicon-oxide interface.

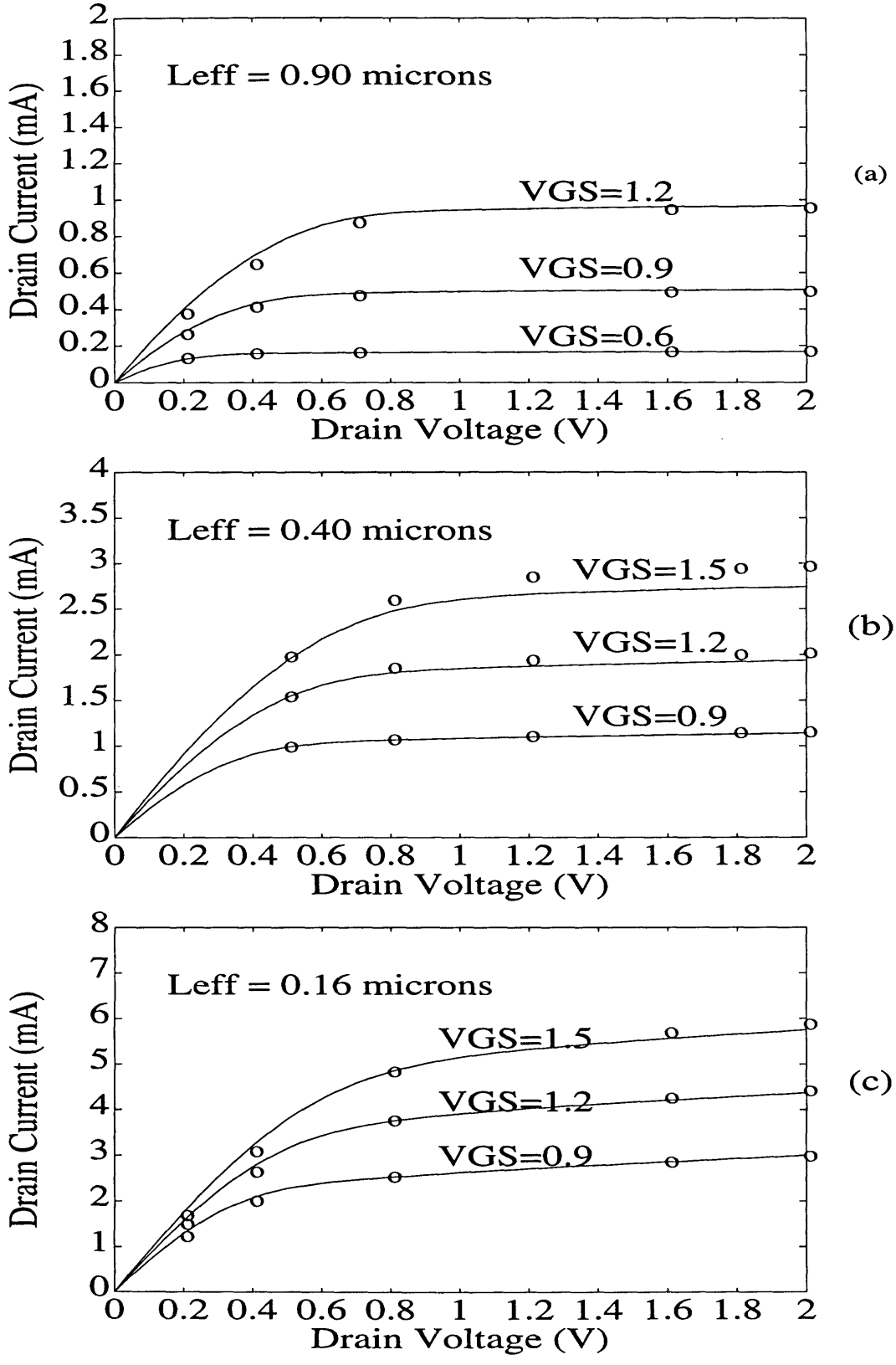

Figure 2-8: Experimental (solid) and simulated (o) drain current for three MOSFETs with channel lengths of (a) 0.90 μm (b) 0.40 μm (c) 0.16 μm.

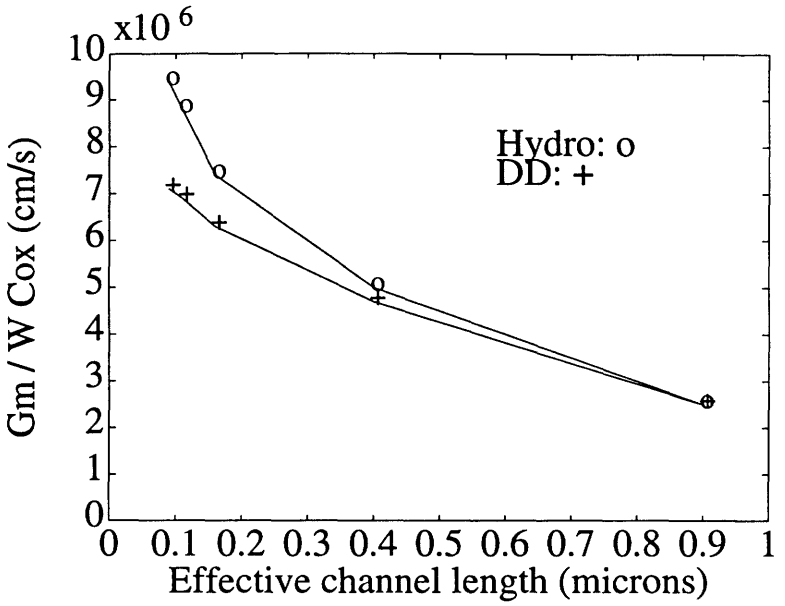

Figure 2-9: Transconductance at V_{GS} =1.2, V_{DS} = 2.0 using the hydrodynamic and drift-diffusion models for different channel lengths.

tal data reported in [34] although exact comparison is difficult due to the different device structures.

2.4.2 Substrate Current Calculation

As the substrate current in MOS transistors at high drain biases is primarily due to impact ionization, our approach for calculating the substrate current is to assume that it is proportional to the total number of electrons which have an energy above a threshold value. That is,

$$I_{sub} = C_{sub} W q \int_0^{L_x} \int_0^{L_y} dx\, dy\, n(x,y) \int_{\varepsilon_{Thresh}}^{\infty} d\varepsilon\, F(\varepsilon, T(x,y)) \tag{2.28}$$

where ε, ε_{Thresh} are the electron and threshold energies respectively; $F(\varepsilon, T)$ is the product of the electron energy distribution as a function of temperature and the density of states; L_x, L_y, W are the device length, height, and width respectively; and C_{sub} is a proportionality constant. Note that w, which is the average electron energy for the ensemble, is distinct from ε, which is a random variable for the energy of each electron in the ensemble.

For the two-dimensional simulations discussed in this paper, uniformity along the width is implicitly assumed, and therefore integration with respect to the width is replaced by multiplication.

Under high field conditions, it is well known that the actual electron energy distribution is substantially different from Maxwellian [35]. Recent Monte-Carlo studies [36], [37] suggest that the tail of the distribution function decays much faster than an exponential dependence. In [36], it is proposed that a more accurate model would be to use a cubic energy dependence in the exponent. This result was derived analytically using non-parabolic bands which, of

course, also changes the density of states. Thus, in our notation, this leads to

$$F(\varepsilon, T) = C_{dist}(T)\varepsilon^{1.25} \exp\left(-\chi \frac{\varepsilon^3}{T^{1.5}}\right) \tag{2.29}$$

where

$$C_{dist}(T) = \frac{1}{\int_0^\infty \varepsilon^{1.25} \exp\left(-\chi \frac{\varepsilon^3}{T^{1.5}}\right) d\varepsilon}. \tag{2.30}$$

The above substrate current model has only one free parameter, χ, in the distribution function, as we keep the threshold energy fixed at a value of 1.8 eV in (2.28) which is a reasonable value for the ionization energy. Changing χ is equivalent to scaling the energy axis and hence χ can be thought of as a contraction or dilation factor for the distribution function. The value of the proportionality constant, C_{sub}, is determined by equating the measured current and the calculated current at one bias point for the longer channel device. This constant which for our simulations was $8.10 \times 10^{11} s^{-1}$ is then used for all the other calculations.

To verify the accuracy of the above distribution function and choose an appropriate dilation factor, χ, we compared the experimentally measured impact ionization coefficient, α, in silicon with that obtained by our approach. The generation rate per unit volume for electrons can be written as

$$G = \alpha \; n \; |v_n| \tag{2.31}$$

where α is the impact ionization coefficient, n is the electron concentration and v_n is the electron velocity usually taken to be the saturation velocity. In our approach the generation rate per unit volume is

$$G = C_{sub} \; n \int_{\varepsilon_{Thresh}}^{\infty} d\varepsilon \, F(\varepsilon, T) \tag{2.32}$$

Hence the impact ionization coefficient is given by

$$\alpha = \frac{C_{sub}}{|v_n|} \int_{\varepsilon_{Thresh}}^{\infty} d\varepsilon \, F(\varepsilon, T) \tag{2.33}$$

The impact ionization rate thus obtained would be a function of the electron temperature; the measured data and models proposed for the impact ionization rate are given as function of electric field, and therefore it is necessary to convert the temperature dependence to a field dependence. To accomplish this, the relationship for the homogeneous case can be applied. Note that in our mobility model, the "low field mobility" is independent of the doping near the oxide interface, as surface effects dominate.

Using the expression for electric field versus temperature that can be derived for the homogeneous case, with $\mu_{n0} = 450 \, cm^2/(Vs)$, and the energy distribution function in (2.29) with a dilation factor of 1.0×10^5, where the energy is measured in eV and the temperature in Kelvin, (2.33) results in an impact ionization coefficient vs. inverse electric field curve as shown in Fig. 2-10. Experimental data from [38] is also shown in this figure. Although the

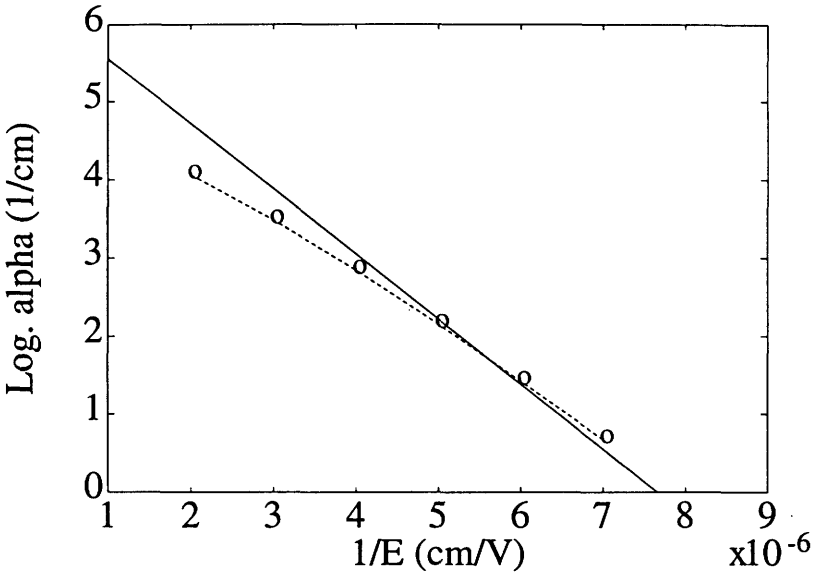

Figure 2-10: Comparison of the impact ionization coefficient, α, as a function of the inverse electric field between experimental data (solid) and that obtained from our model (2-33) (dash)

two do not match exactly, this is to be expected due to the large uncertainty in the measured data and the approximations in the model.

Fig. 2-11 shows the simulated and measured substrate currents as a function of the drain voltage at different gate biases for both devices using the cubic distribution function with the above parameters. Generation-recombination is ignored in our simulator so only a comparison with the hot carrier part of the substrate current is appropriate.

Fig. 2-12 shows the substrate current as a function of the gate voltage with the drain voltage as a parameter. The threshold voltage for this device is about 0.23 V, and the accuracy of the calculated substrate current seriously degrades for biases near the threshold voltage, although the peak substrate current is correctly predicted. This inaccuracy cannot be attributed to errors in the threshold voltage because a careful comparison of the simulated and measured drain currents even at low gate biases shows very little error. An explanation for this effect can be found if we consider that in the subthreshold regime with high drain voltage, the electric fields are extremely high (over 5×10^5 V/cm) in a very narrow region near the drain. Consequently, the carriers don't fully thermalize before entering the drain, and therefore the electron temperature in this region does not accurately reflect the tails of the distribution function. In this operating regime of the MOSFET, the energy-balance model is inadequate and only a method which correctly accounts for the high energy particles can be relied upon.

2.5 Summary

In this chapter we introduced the hydrodynamic model which is an extension of the classical drift-diffusion approach to modeling transport in semiconductor devices. In the

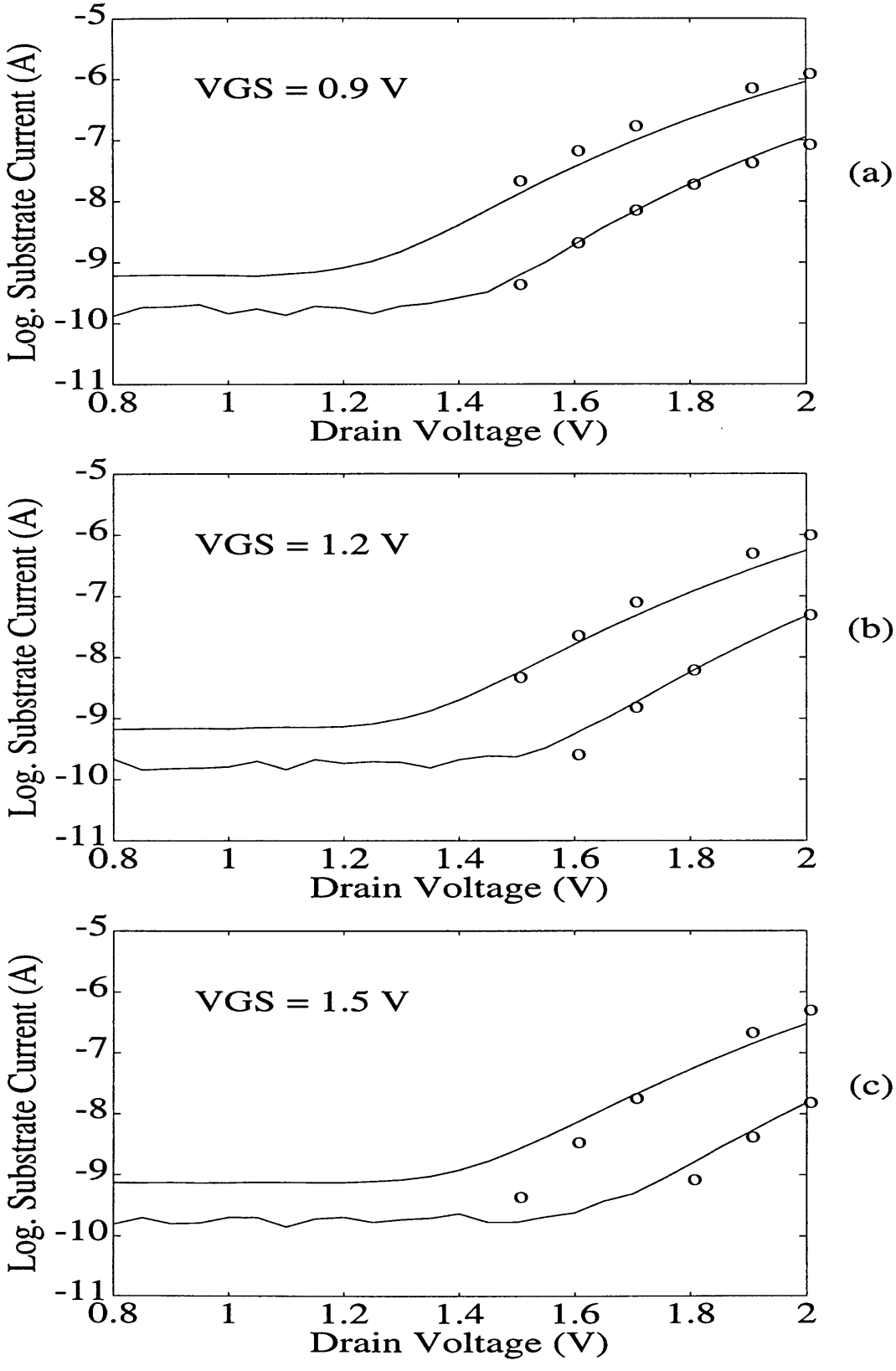

Figure 2-11: Simulated (o) and measured (solid) substrate current for the 0.16 μm (upper curve) and 0.40 μm (lower curve) devices using the cubic distribution function. (a) $V_{GS} = 0.9V$; (b) $V_{GS} = 1.2V$; (c) $V_{GS} = 1.5V$.

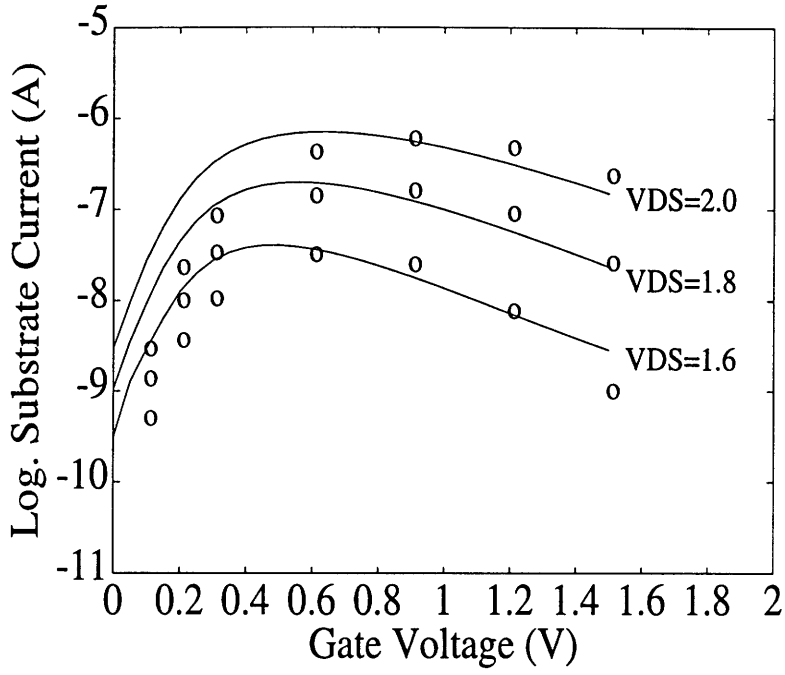

Figure 2-12: Simulated (o) and measured (solid) substrate current for a device with $L_{eff} = 0.20\,\mu$m as a function of the gate voltage for three different drain voltages.

hydrodynamic model, carrier energy momentum and particle number are all conserved. These conservation laws which lead to a set of coupled partial differential equations that form the hydrodynamic model were presented along with a set of physical parameters which were needed to successfully model realistic devices. Issues concerning the numerical techniques needed to actually solve the partial differential equations were then discussed. Finally we compared the simulated results with experimental data from advanced MOSFET's to test the model and the parameters used. It was shown that the drain currents obtained from the hydrodynamic model match the measured data quite well for a range of channel lengths and biases with one set of parameters, but the substrate current obtained by our technique while accurate for a large bias range produces large errors near threshold. We attribute this behavior of the simulation results from the hydrodynamic model to the use of the electron energy or temperature which is an ensemble average and therefore cannot in principle provide detailed information about the distribution of the hot carriers. If the distribution was roughly Maxwellian than this would be sufficient information but near the drain of a short channel MOSFET, for example, the distribution is far from a Maxwellian as two populations of hot and cold carriers exist simultaneously. Thus even if there is a considerable number of hot carriers with energy well above the impact ionization threshold, their effect on the electron temperature may be minimal if a large population of cold carriers at the lattice temperature is also present. And this, in fact is the situation in the drain region of a device at high fields. We, therefore, believe that only a method which provides direct information on the carrier energy distribution can be reliable and robust for this kind of simulation.

Chapter 3

Boltzmann Equation Solution– I

3.1 Introduction

The direct solution of the Boltzmann equation for modeling transport in semiconductors has been pursued for over two decades and many ingenious techniques have been developed for this purpose. These include an integral equation method [39], the Monte-Carlo method [40], a number of different basis function expansions [41] [42], and direct integration [43]. Of these techniques the one that has garnered the most attention is the Monte-Carlo method. The reason for this is twofold: first, it is relatively easy to implement (at least in its simpler manifestations), second, it is quite flexible and hence allows easy incorporation of details such as complex band structures and scattering mechanisms. The main drawback of the Monte-Carlo method is its computational expense especially when attempting to reduce the statistical noise in the low density tails of the distribution function.

As an alternative to the Monte-Carlo method for the solution of the Boltzmann equation, we shall consider the spherical harmonic expansion method in this chapter. In the next two sections some motivation for this approach is given and then a brief review of earlier work using this approach will be presented. Then our contribution in extending this technique to arbitrary order will be explained. Next the numerical difficulties encountered using this approach and the solutions found for the one-dimensional problem will be presented. Finally results using the arbitrary order approach for the test case of an n^+nn^+ diode will be presneted.

3.2 Spherical Harmonic Expansion

Due to the spherical symmetry of the band structure and the randomizing nature of most scattering processes we expect the distribution function to have some degree of spherical symmetry in momentum space. Also the equilibrium distribution, which is a Maxwellian, in the non-degenerate case, is spherically symmetric. Thus a basis function expansion of the distribution function should exploit this symmetry. Consequently a plausible choice for the basis functions is the family of surface spherical harmonics [44] [45] denoted by $Y_{lm}(\theta, \phi)$,

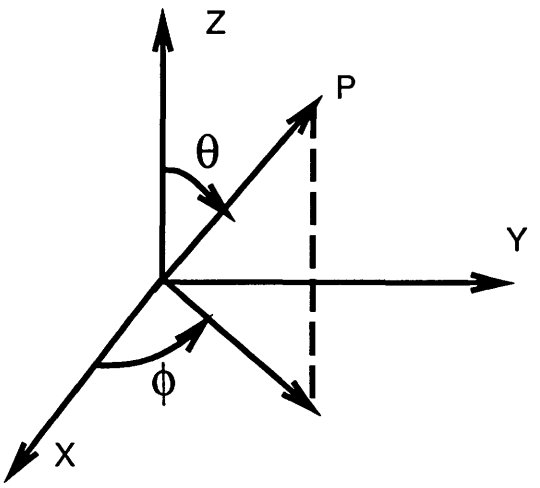

Figure 3-1: The coordinate system used in the spherical harmonic expansion.

l=0	$Y_{0,0}$	$\frac{1}{\sqrt{4\pi}}$
	$Y_{1,1}$	$\sqrt{\frac{3}{4}}\sin\theta\cos\phi$
l=1	$Y_{1,0}$	$\sqrt{\frac{3}{4}}\cos\theta$
	$Y_{1,-1}$	$\sqrt{\frac{3}{4}}\sin\theta\sin\phi$
	$Y_{2,2}$	$\frac{\sqrt{15}}{2\sqrt{4\pi}}\sin^2\theta\cos 2\phi$
	$Y_{2,1}$	$\frac{\sqrt{15}}{\sqrt{4\pi}}\sin\theta\cos\theta\cos\phi$
l=2	$Y_{2,0}$	$\frac{\sqrt{5}}{\sqrt{4\pi}}\left(\frac{3}{2}\cos^2\theta-\frac{1}{2}\right)$
	$Y_{2,-1}$	$\frac{\sqrt{15}}{\sqrt{4\pi}}\sin\theta\cos\theta\sin\phi$
	$Y_{2,-2}$	$\frac{\sqrt{15}}{2\sqrt{4\pi}}\sin^2\theta\sin 2\phi$

Table 3.1: Spherical harmonics up to second order (l=2).

which up to second order are given in Table 3.1. We therefore, express the distribution function as (see Fig. 3-1)

$$f(\mathbf{r},\mathbf{k}) = \sum_{lm} f_{lm}(\mathbf{r},k)\,Y_{lm}(\theta,\phi). \tag{3.1}$$

This expansion means that at each point, r, in real space the momentum space description of the distribution function is specified on each sphere of radius k by a weighted sum of surface spherical harmonics. The weight of each harmonic, Y_{lm}, is specified by f_{lm}. The essence of the problem is to find these weights on as many spheres in momentum space as desired. Of course, this choice for the expansion will be efficient only if a few spherical harmonics are needed to represent the momentum space distribution to the desired accuracy.

As an example of how a spherical harmonic expansion will model a typical distribution function, consider the displaced Maxwellian distribution which is commonly used as an

approximation to the true distribution for moderate electric fields. Fig. 3-2 shows a shifted Maxwellian and the Legendre polynomial expansion (in $\cos\theta$) for this distribution up to third order. The Legendre polynomials of $\cos\theta$ are just reduced spherical harmonics when there is rotational symmetry about the polar direction, i.e. there is no ϕ variation. Fig. 3-3 shows the projection of the shifted Maxwellian distribution onto the space of Legendre polynomials; only the first four Legendre polynomials for this distribution are shown in perspective. The coefficient for each Legendre polynomial is given by

$$f_n(k) = 2\pi \int P_n(\cos\theta)\, f_{DMW}(k,\theta)\, \sin\theta\, d\theta \tag{3.2}$$

where $P_n(\cos\theta)$ is the Legendre polynomial of order n and f_{DMW} is the drifted Maxwellian distribution given by:

$$\begin{aligned} f_{DMW} &= C_0 \exp\left[-(k_x^2 + k_y^2 + (k_z - k_D)^2\right] \\ &= C_0 \exp\left[-(k\sin\theta)^2 + (k\cos\theta - k_D)^2\right] \end{aligned} \tag{3.3}$$

where C_0 is a normalization constant and k_D is the shift in the Maxwellian from the origin.

From Fig. 3-2 it can be seen that the third order term in the expansion, $f_{3,0}$, is two orders of magnitude smaller than the first, $f_{0,0}$. This gives an indication of the level of accuracy attained by this approximation; obviously including higher orders in the expansion would reduce the error between the synthesized function and the original shifted Maxwellian. As the distribution becomes more displaced, the error in the Legendre polynomial expansion grows and higher order expansions are needed to maintain the same accuracy. For example in Fig. 3-4 a more shifted Maxwellian is expanded to third order again and as expected $f_{3,0}$ is now much larger in magnitude than in the previous example.

3.3 Boltzmann equation solution using spherical harmonics

In the previous section we showed how the spherical harmonics (or more specifically the Legendre polynomials) can be used to approximate a given distribution function. Of course our goal is to solve for the *unknown* distribution function that satisfies the Boltzmann equation with the appropriate boundary conditions. The standard way of doing this [42] [46] is to substitute the expansion (3.1) into the Boltzmann equation

$$\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f(\mathbf{r},\mathbf{k}) - \frac{q\mathcal{E}(\mathbf{r})}{\hbar} \cdot \nabla_{\mathbf{k}} f(\mathbf{r},\mathbf{k}) = \int S(\mathbf{k}',\mathbf{k}) f(\mathbf{r},\mathbf{k}')\, d^3\mathbf{k}' - f(\mathbf{r},\mathbf{k}) \int S(\mathbf{k},\mathbf{k}')\, d^3\mathbf{k}' \tag{3.4}$$

to generate a set of coupled partial differential equation in the expansion coefficients. This set of partial differential equations is then solved by first discretizing them to generate an algebraic problem, which is then solved to yield the desired set of coefficients f_{lm}.

For modeling transport in one real space dimension, z, we must consider a two dimensional momentum space in the polar coordinates, k and θ, where the angle is defined from the electric

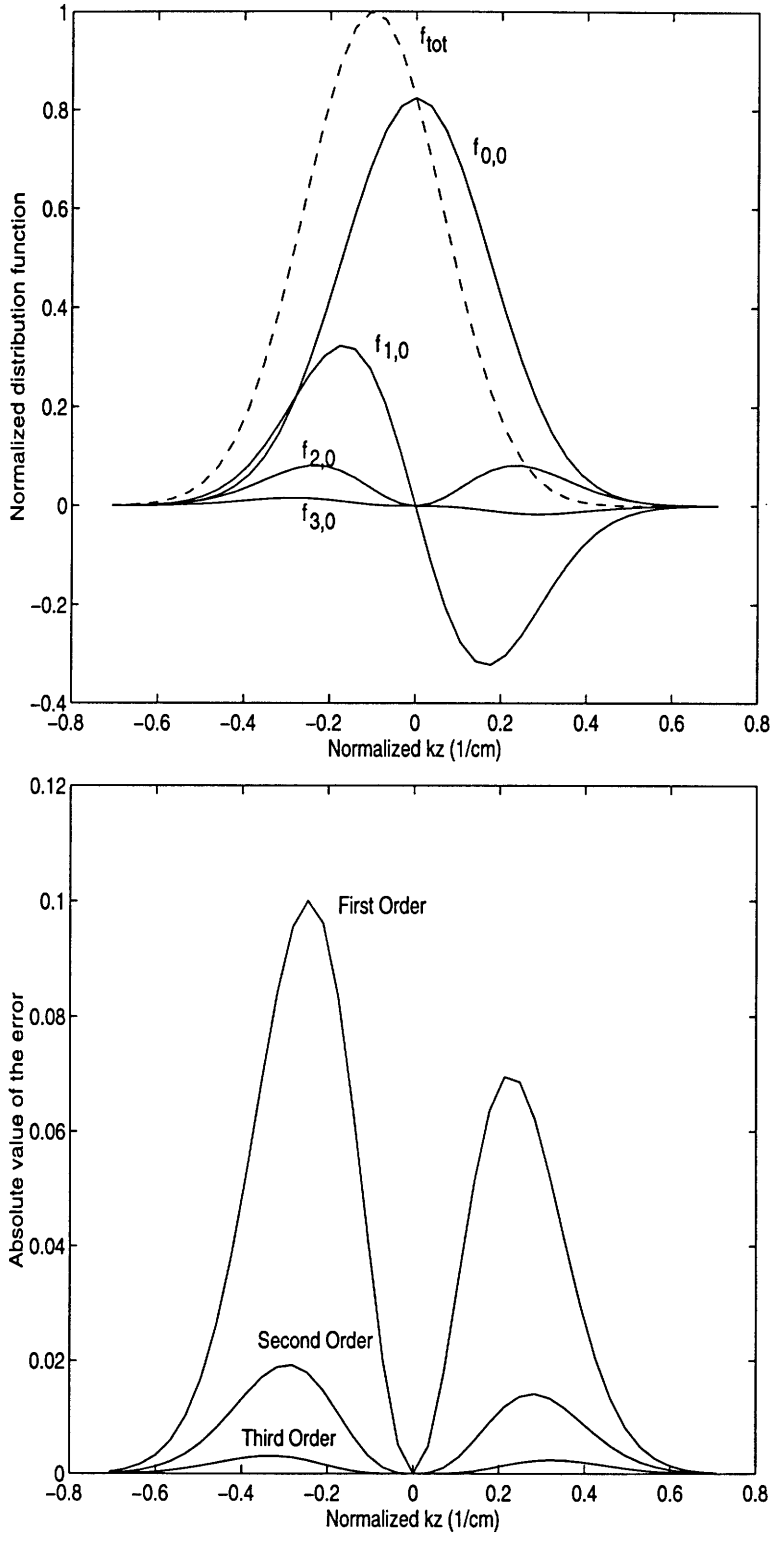

Figure 3-2: The top graph shows each of the lower order expansion components of the synthesized function, denoted by f_{tot}. The bottom graph shows the error between the shifted Maxwellian and the synthesized distribution for each order of approximation.

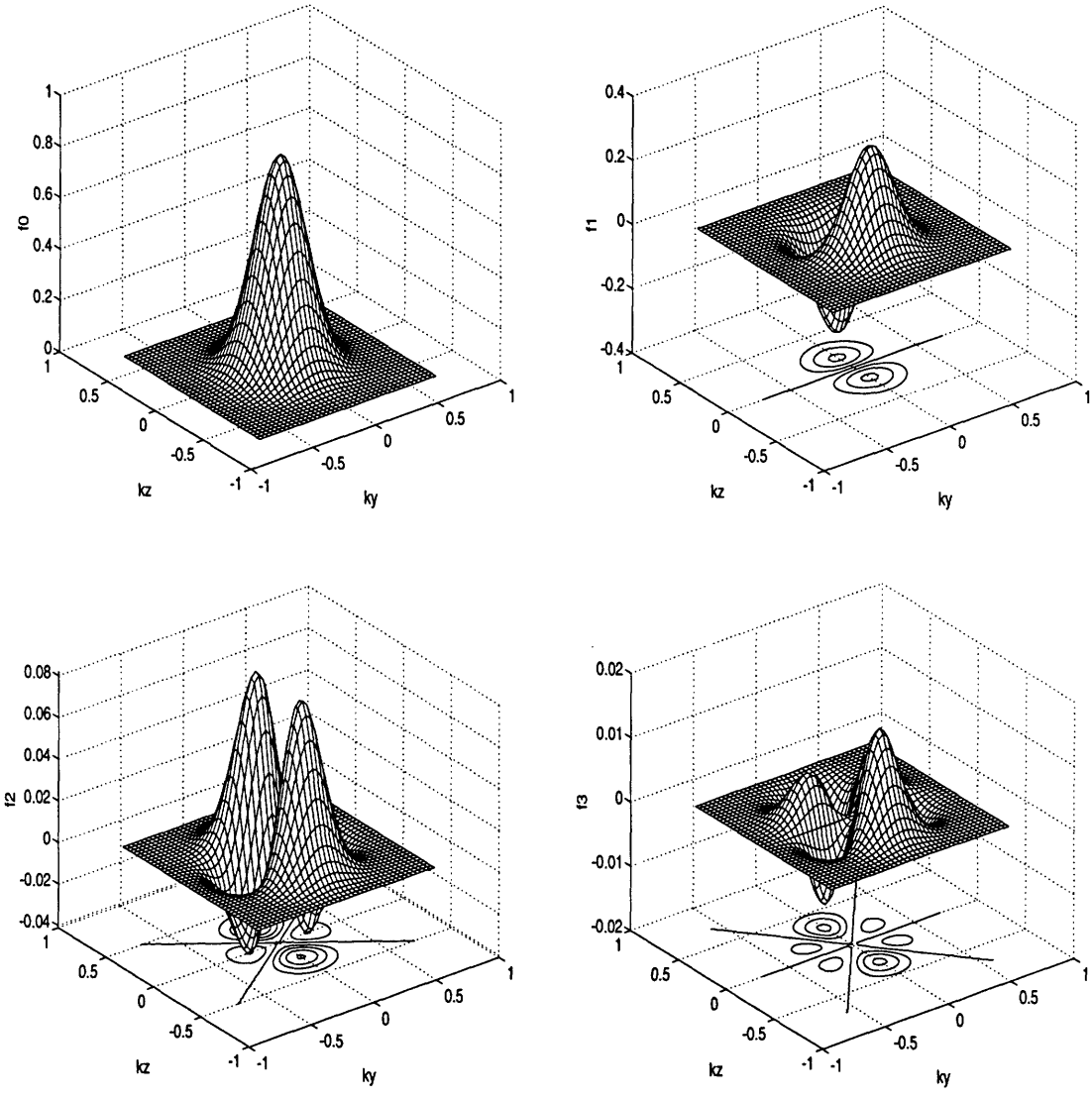

Figure 3-3: Zeroth, first, second and third order Legendre polynomials for shifted Maxwellian distribution.

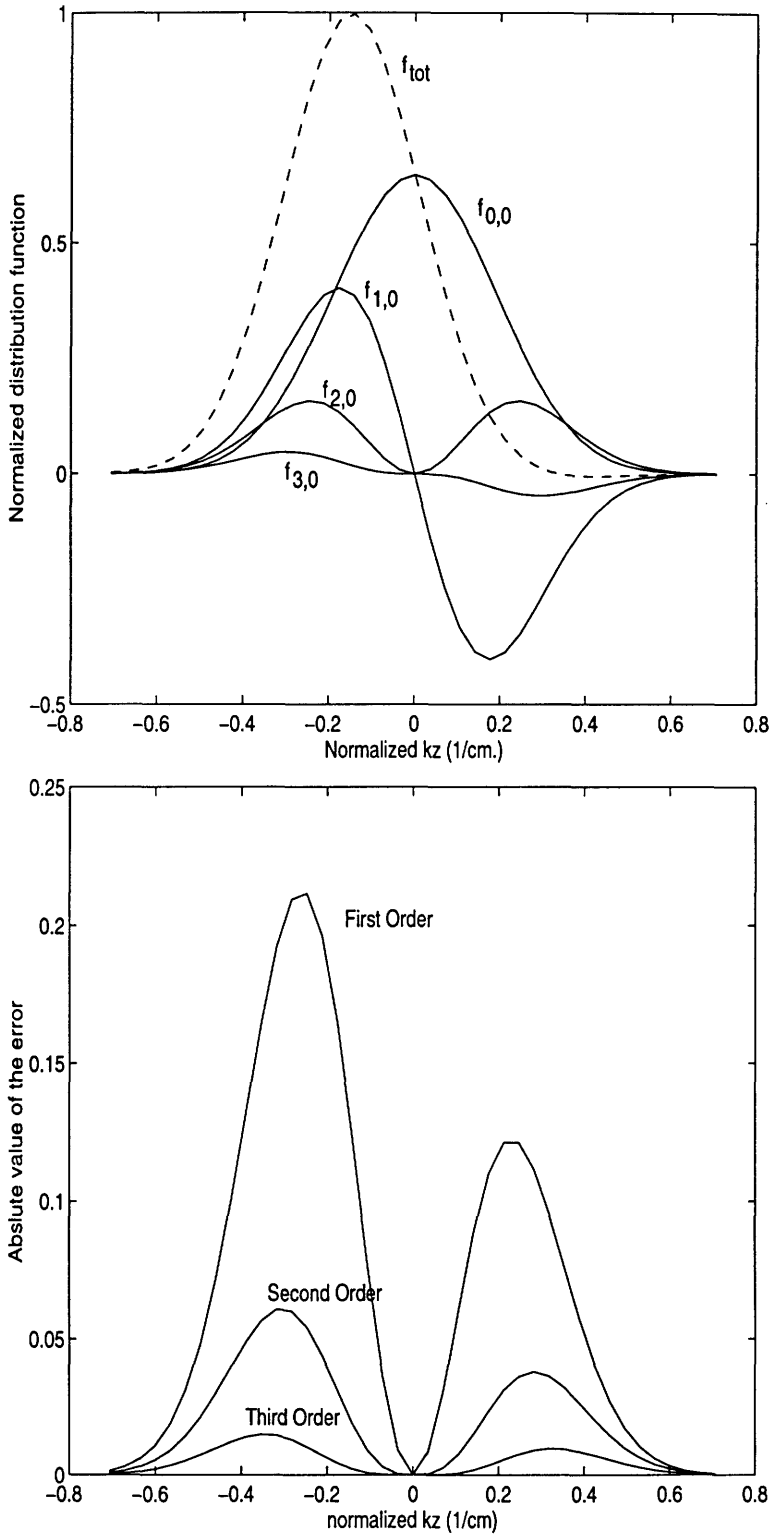

Figure 3-4: The top graph shows each of the lower order expansion components of the synthesized function, denoted by f_{tot}. The bottom graph shows the error between the shifted Maxwellian and the synthesized distribution for each order of approximation.

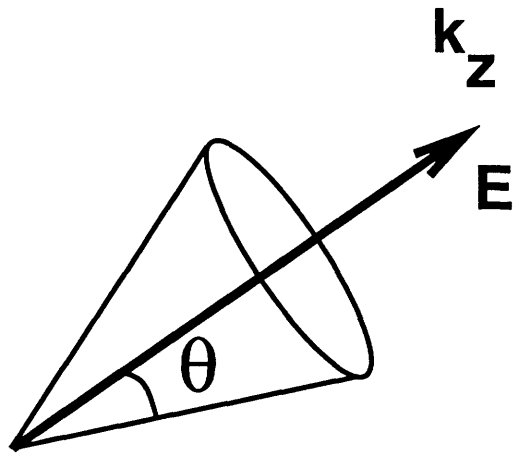

Figure 3-5: Coordinate system for the one dimensional real space problem.

field direction which is chosen to be the k_z axis, as shown in Fig. 3-5. Then the distribution function can be expanded in Legendre polynomials of the angle θ:

$$f(z,\mathrm{k}) = \sum f_n(z,k) P_n(\cos\theta). \tag{3.5}$$

Note that the Legendre polynomials in $\cos\theta$, $P_n(\cos\theta)$ are just the set of spherical harmonics $Y_{0,0}, Y_{1,0}, Y_{2,0}\ldots$ without any ϕ dependence and for this reason we shall drop the second index for the remainder of this section. We now use the above expansion for f in each term of the BTE.

3.3.1 The diffusion term: $\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f(\mathbf{r},\mathbf{k})$

The velocity $\mathrm{v}(\mathrm{k})$ is determined from the band structure: $\mathrm{v}(\mathrm{k}) = \frac{1}{\hbar}\frac{\partial E}{\partial \mathrm{k}}$, which is assumed to be spherically symmetric and described by a relation of the form: $\gamma(E) = \frac{\hbar^2 k^2}{2m^*} = \frac{\hbar^2 \mathrm{k}\cdot\mathrm{k}}{2m^*}$. This allows for a non-parabolic band structure by using an appropriate form for $\gamma(E)$. Of course, $\gamma(E) = E$ leads to the familiar parabolic band model. Thus, $\mathrm{v}(\mathrm{k})$ is given by

$$\mathrm{v}(\mathrm{k}) = \frac{1}{\hbar}\frac{\partial\gamma}{\partial \mathrm{k}}\frac{\partial E}{\partial\gamma} = \frac{\hbar \mathrm{k}}{m^*}\frac{\partial E}{\partial\gamma} = \frac{\hbar k}{m^*}\frac{\partial E}{\partial\gamma}\mathrm{i}_k = \left(\frac{2\gamma}{m^*}\right)^{\frac{1}{2}}\frac{\partial E}{\partial\gamma}\mathrm{i}_k = v(k)\mathrm{i}_k. \tag{3.6}$$

where i_k is the unit vector in the radial $\mathbf{k}$ direction. Note that the magnitude of the velocity is purely a function of the energy and its direction is always radial. Using the above expression for the velocity, we can write the diffusion term in one real space dimension as:

$$\mathrm{v}(\mathrm{k})\cdot\nabla_{\mathrm{r}} f(\mathrm{r},\mathrm{k}) = \mathrm{v}(\mathrm{k})\cdot\frac{\partial f}{\partial z}\mathrm{i}_z = v(k)\mathrm{i}_k\cdot\frac{\partial}{\partial z}\left[\sum f_n(z,k)P_n(\cos\theta)\right]\mathrm{i}_z. \tag{3.7}$$

The scalar product, above, can be evaluated by projecting the radial i_k vector onto the z-axis which evaluates to $\cos\theta$, whence, the diffusion term becomes:

$$\mathrm{v}(\mathrm{k})\cdot\nabla_{\mathrm{r}} f(\mathrm{r},\mathrm{k}) = v(k)\cos\theta\left[\sum\frac{\partial f_n(z,k)}{\partial z}P_n(\cos\theta)\right]. \tag{3.8}$$

If we now use the following identity for Legendre polynomials [44]:

$$xP_n(x) = \frac{(n+1)P_{n+1}(x) + nP_{n-1}(x)}{2n+1}, \tag{3.9}$$

we can rewrite the diffusion term as

$$v(k)\sum \frac{\partial f_n(z,k)}{\partial z}\left(\frac{(n+1)P_{n+1}(\cos\theta) + nP_{n-1}(\cos\theta)}{2n+1}\right). \tag{3.10}$$

3.3.2 The drift term: $\frac{q\mathcal{E}(\mathbf{r})}{\hbar}\cdot\nabla_{\mathbf{k}}f(\mathbf{r},\mathbf{k})$

As discussed in the previous section we are only considering variation in one real space direction conventionally chosen to be z, so we can replace $f(\mathrm{r},\mathrm{k})$ by $f(z,\mathrm{k})$. In spherical coordinates the gradient of any function $h(k,\theta)$ which has no ϕ dependence is:

$$\nabla_{\mathbf{k}}h(\mathrm{k}) = \frac{\partial h(k,\theta)}{\partial k}\mathbf{i}_k + \frac{1}{k}\frac{\partial h(k,\theta)}{\partial \theta}\mathbf{i}_\theta. \tag{3.11}$$

Hence the gradient of f with respect to the k vector can be explicitly written as:

$$\begin{aligned}
\nabla_{\mathbf{k}}f(\mathrm{r},\mathrm{k}) &= \nabla_{\mathbf{k}}\sum f_n(z,k)P_n(\cos\theta) \\
&= \sum \frac{\partial f_n(z,k)}{\partial k}P_n(\cos\theta)\mathbf{i}_k + \sum \frac{f_n(z,k)}{k}\frac{\partial P_n(\cos\theta)}{\partial\theta}\mathbf{i}_\theta \\
&= \sum \frac{\partial f_n(z,k)}{\partial k}P_n(\cos\theta)\mathbf{i}_k + \sum\left(\frac{f_n(z,k)}{k}\frac{\partial P_n(\cos\theta)}{\partial(\cos\theta)}(-\sin\theta)\right)\mathbf{i}_\theta.
\end{aligned} \tag{3.12}$$

where we have assumed that the distribution function has no ϕ dependence because of its symmetry about the z-directed electric field. We must now perform a scalar product of the gradient vector with the z-directed electric field which can be implemented by projecting the gradient onto the z-axis. For any vector of the form $f_r\mathbf{i}_\mathbf{r} + f_\theta\mathbf{i}_\theta$ the projection onto the z-axis is $f_r\cos\theta - f_\theta\sin\theta$. Therefore, the scalar product of the gradient with the electric field is:

$$\frac{q\mathcal{E}}{\hbar}\left[\sum \frac{\partial f_n(z,k)}{\partial k}P_n(\cos\theta)\cos\theta + \sum \frac{f_n(z,k)}{k}\frac{\partial P_n(\cos\theta)}{\partial(\cos\theta)}\sin^2\theta\right]. \tag{3.13}$$

To simplify the above expression we use the identity [44]:

$$(x^2-1)P_n'(x) = nxP_n(x) - nP_{n-1}(x) \tag{3.14}$$

along with (3.9) to obtain the relation:

$$(x^2-1)P_n'(x) = \frac{n(n+1)P_{n+1}(x) - n(n+1)P_{n-1}(x)}{2n+1}. \tag{3.15}$$

where $P_n'(x)$ is the derivative of $P_n(x)$ with respect to x. The drift term can then be written as:

$$\begin{aligned}
&\frac{q\mathcal{E}}{\hbar}\left[\sum \frac{\partial f_n(z,k)}{\partial k}\left(\frac{(n+1)P_{n+1}(x) + nP_{n-1}(x)}{2n+1}\right)\right. \\
&\left. - \sum \frac{f_n(z,k)}{k}\left(\frac{n(n+1)P_{n+1}(x) - n(n+1)P_{n-1}(x)}{2n+1}\right)\right]
\end{aligned} \tag{3.16}$$

where we have defined $x = \cos\theta$ for notational convenience. Rewriting the above equation to factor out the Legendre polynomials gives

$$\begin{aligned} \tfrac{q\mathcal{E}}{\hbar} \quad & \left[\sum \frac{1}{2n+1}\left(n\frac{\partial f_n(z,k)}{\partial k} + n(n+1)\frac{f_n(z,k)}{k}\right) P_{n-1}(x)\right. \\ + \quad & \left.\sum \frac{1}{2n+1}\left((n+1)\frac{\partial f_n(z,k)}{\partial k} - n(n+1)\frac{f_n(z,k)}{k}\right) P_{n+1}(x)\right] \end{aligned} \tag{3.17}$$

3.3.3 Complete BTE expansion in Legendre Polynomials

We now make a change of variable from the magnitude of the radial vector, **k**, to the energy, E and rewrite (3.17) in terms of the energy. To do this we need the following identities, which follow from (3.6):

$$\frac{1}{\hbar}\frac{\partial f_n(z,k)}{\partial k} = \frac{\partial f_n(z,E)}{\partial E}\frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{\partial f_n(z,E)}{\partial E}\frac{1}{\hbar}\frac{\partial E}{\partial \gamma}\frac{\partial \gamma}{\partial k} = \frac{\partial f_n(z,E)}{\partial E}v(E) \tag{3.18}$$

$$\frac{1}{v(E)}\left(\frac{1}{\hbar}\frac{f_n(z,k)}{k}\right) = f_n(z,E)\frac{\gamma'}{2\gamma}. \tag{3.19}$$

where $\gamma' = \frac{d\gamma}{dE}$. Under this change of variables, the diffusion term is unchanged as it did not have any explicit dependence on the momentum or energy vector:

$$v(E)\sum \frac{\partial f_n(z,E)}{\partial z}\left[\frac{(n+1)P_{n+1}(\cos\theta) + nP_{n-1}(\cos\theta)}{2n+1}\right], \tag{3.20}$$

but the drift term (3.17) now becomes:

$$\begin{aligned} q\mathcal{E}v(E) \quad & \left[\sum \frac{1}{2n+1}\left(n\frac{\partial f_n(z,E)}{\partial E} + \frac{n(n+1)\gamma'}{2\gamma}f_n(z,E)\right) P_{n-1}(x)\right. \\ + \quad & \left.\sum \frac{1}{2n+1}\left((n+1)\frac{\partial f_n(z,E)}{\partial E} - \frac{n(n+1)\gamma'}{2\gamma}f_n(z,E)\right) P_{n+1}(x)\right]. \end{aligned} \tag{3.21}$$

The sum of the above two expressions forms the left hand side of the Boltzmann equation, and if we write the above terms for the n^{th} order of the Legendre polynomial expansion we get:

$$\begin{aligned} \tfrac{1}{2n+1}v(E) \quad & \left[\left(n\frac{\partial f_{n-1}}{\partial z} + (n+1)\frac{\partial f_{n+1}}{\partial z}\right)\right. \\ - \quad & \left. q\mathcal{E}\left(n\frac{\partial f_{n-1}}{\partial E} + (n+1)\frac{\partial f_{n+1}}{\partial E} - \frac{n(n-1)\gamma'}{2\gamma}f_{n-1} + \frac{(n+1)(n+2)\gamma'}{2\gamma}f_{n+1}\right)\right] \end{aligned} \tag{3.22}$$

The two lowest order expansions are:

$n = 0$:

$$\frac{\partial f_1}{\partial z} - q\mathcal{E}\left(\frac{\partial f_1}{\partial E} + \frac{\gamma'}{\gamma}f_1\right) = \frac{1}{v(E)}\left(\frac{\partial f_0}{\partial t}\right)_c \tag{3.23}$$

$n = 1$:

$$\frac{\partial f_0}{\partial z} - 2\frac{\partial f_2}{\partial z} - q\mathcal{E}\left(\frac{\partial f_0}{\partial E} + 2\frac{\partial f_2}{\partial E} + \frac{3\gamma'}{\gamma}f_2\right) = \frac{3}{v(E)}\left(\frac{\partial f_1}{\partial t}\right)_c \tag{3.24}$$

where the right hand side is assumed to also have been expanded in spherical harmonics with the coefficients of the lowest order term being denoted by $\left(\frac{\partial f_0}{\partial t}\right)_c$, the next by $\left(\frac{\partial f_1}{\partial t}\right)_c$ etc. If we set all coefficients higher than the first to zero than the above equation simplifies to

$$\frac{\partial f_0}{\partial z} - q\mathcal{E}\left(\frac{\partial f_0}{\partial E}\right) = \frac{3}{v(E)}\left(\frac{\partial f_1}{\partial t}\right)_c. \tag{3.25}$$

The equations (3.23) and (3.25) are exactly the ones used in [46] [47], where the above equations are discretized after including the appropriate scattering mechanisms and then solved for the zeroth order, f_0 and first order, f_1 coefficients. The results obtained using this technique were satisfactory for the devices and biases studied in [46]. One important question that was not addressed in earlier work is whether first order expansions are sufficiently accurate for all realistic problems and if not then how can this technique be extended in a general way to higher orders. The obvious way to solve for higher order coefficients is to go back to the general set of equations and regenerate a larger set of PDE's for the higher order, and then discretize and implement them in a computer program. Clearly, this is not a desirable approach for each higher order. What is needed is a general method that automatically extends to arbitrary order. In the next section we present such a method.

3.4 Arbitrary Order Expansion

As illustrated in the previous section the technique for generating a set of partial differential equations which can be solved for the coefficients of the first n spherical harmonics consists in writing the distribution function as a sum of spherical harmonics and then using the orthogonality of the spherical harmonics to write n partial differential equations which couple the coefficients of all the n spherical harmonics. Once this set of equations is obtained it is discretized in energy and real space to produce an algebraic problem which can be solved using standard matrix techniques. One difficulty with this method is that if a higher order solution than originally used is needed then the set of differential equations must be revised and augmented, and this new set of partial differential equations has to be discretized and solved. An implementation that was based on the initial set of equations would have to modified to solve the new problem. Clearly a technique that accommodated itself to the solution of arbitrary order spherical harmonic expansion without human intervention would be superior. This can in fact be achieved by using a Galerkin method.

The basic idea is the same: the unknown solution is written as a sum of spherical harmonics whose coefficients have to be determined. But then instead of generating a partial differential equation for each harmonic we discretize the Boltzmann equation and multiply it by the conjugate harmonics and integrate over a sphere in momentum space. This produces a matrix equation for the coefficients of the expansion which accounts for the coupling between each spherical harmonic and all others. As the intermediate step of writing down a continuous set of differential equations is omitted, the problem can be solved for arbitrary size as the only change necessary is in the size of the matrix. This idea is developed formally below.

Substituting the spherical harmonic expansion into the Boltzmann Equation and multiplying by the conjugate spherical harmonic and integrating over a sphere in k-space yields:

$$\int Y^*_{l'm'}(\theta,\phi)\,\mathbf{v}(\mathbf{k})\cdot\frac{\partial}{\partial\mathbf{r}}\sum_{lm} f_{lm}(\mathbf{r},k)Y_{lm}(\theta,\phi)d\Omega - \int Y^*_{l'm'}(\theta,\phi)\frac{q\mathcal{E}(\mathbf{r})}{\hbar}\cdot\frac{\partial}{\partial\mathbf{k}}\sum_{lm} f_{lm}(\mathbf{r},k)Y_{lm}(\theta,\phi)d\Omega$$
$$= \int Y^*_{l'm'}(\theta,\phi)\;\mathcal{S}\left(\sum_{lm} f_{lm}(\mathbf{r},k)Y_{lm}(\theta,\phi)\right)d\Omega \tag{3.26}$$

where $\mathcal{S}$ denotes the scattering operator. In the following discussion we develop the Galerkin method in one real space dimension, conventionally chosen to be the z-axis, assuming spherical but not necessarily parabolic bands. The multidimensional real space problem is discussed in the next chapter.

3.4.1 The Diffusion Term

The first term in the above expansion can be written as:

$$\int Y^*_{l'm'}(\theta,\phi)\,v(k)\,\mathbf{i}_k\cdot\frac{\partial}{\partial z}\left[\sum_{lm} f_{lm}(z,k)Y_{lm}(\theta,\phi)\right]\mathbf{i}_z\,d\Omega \tag{3.27}$$

where $v(k)$ is given by (3.6). The scalar product above can be evaluated by remembering that the projection of the unit radial $\mathbf{i}_k$ vector onto the z-axis is $\cos\theta$. Thus after rearranging the terms we get:

$$v(k)\int Y^*_{l'm'}(\theta,\phi)\sum_{lm}\cos\theta\,\frac{\partial}{\partial z}\left[f_{lm}(z,k)\right]Y_{lm}(\theta,\phi)\,d\Omega. \tag{3.28}$$

If we now use a two point approximation to the space derivative:

$$\frac{\partial f_{lm}}{\partial z}\approx\frac{f^{i+1,j}_{lm}-f^{i,j}_{lm}}{\Delta z} \tag{3.29}$$

where the i index denotes discretization in space and the j index denotes discretization in energy, then the above expression at a point $(i,\,j)$ in the solution space can be written as:

$$v(k)\int\sum_{lm}\frac{f^{i+1,j}_{lm}-f^{i,j}_{lm}}{\Delta z}Y^*_{l'm'}(\theta,\phi)Y_{lm}(\theta,\phi)\cos\theta\,d\Omega \tag{3.30}$$

where the differential surface element, $d\Omega=\sin\theta\,d\theta\,d\phi$. Thus, if we let

$$G_{l'm';lm}=\int_0^{2\pi}\int_0^{\pi}Y^*_{l'm'}(\theta,\phi)Y_{lm}(\theta,\phi)\cos\theta\,\sin\theta\,d\theta\,d\phi\,, \tag{3.31}$$

then the diffusion term of the BTE can be written as

$$v^j\sum_{lm}\frac{f^{i+1,j}_{lm}-f^{i,j}_{lm}}{\Delta z}G_{l'm';lm} \tag{3.32}$$

where v^j is the magnitude of the velocity vector that corresponds to the discretized energy E^j. This expression can be written more compactly as:

$$\frac{v^j}{\Delta z} G \left(f^{i+1,j} - f^{i,j} \right) \tag{3.33}$$

where f is the vector of coefficients and G is the matrix defined by (3.31). One advantage of writing the diffusion term in this compact form is that it clearly shows the coupling between the different coefficients in the expansion. Note that G represents the coupling between harmonics at a given point in space whereas the approximation to the derivative in space is the coupling between harmonics at two different points in space.

3.4.2 The Drift Term

A similar derivation is performed for the drift term in the BTE. After substituting the spherical harmonic expansion for the distribution, performing the scalar product with the electric field, and multiplying by the conjugate harmonics, we can write:

$$- \int Y^*_{l'm'}(\theta,\phi) \frac{q\mathcal{E}(z)}{\hbar} \sum_{lm} \left[\frac{\partial f_{lm}(z,k)}{\partial k} Y_{lm}(\theta,\phi) \cos\theta - \frac{f_{lm}(z,k)}{k} \frac{\partial Y_{lm}(\theta,\phi)}{\partial \theta} \sin\theta \right] d\Omega. \tag{3.34}$$

If we now pull the electric field outside the integral as it is independent of k and write the expression as the sum of two integrals over Ω, we obtain

$$\frac{-q\mathcal{E}(z)}{\hbar} \int \sum_{lm} \left[\frac{\partial f_{lm}(z,k)}{\partial k} Y^*_{l'm'}(\theta,\phi) Y_{lm}(\theta,\phi) \cos\theta - \frac{f_{lm}(z,k)}{k} Y^*_{l'm'}(\theta,\phi) \frac{\partial Y_{lm}(\theta,\phi)}{\partial \theta} \sin\theta \right] d\Omega \tag{3.35}$$

We approximate the derivative with respect to the magnitude of the **k** vector or energy as a difference:

$$\frac{\partial}{\partial k} \left[f_{lm}(z,k) \right] = \frac{\partial}{\partial E} \left[f_{lm}(z,E) \right] \frac{\partial E}{\partial k} = \frac{\partial}{\partial E} \left[f_{lm}(z,E) \right] \hbar v(E) \approx \hbar v^j \left[\frac{f^{i,j+1}_{lm} - f^{i,j}_{lm}}{\Delta E} \right] \tag{3.36}$$

using (3.18) to change variables from the magnitude of the k vector to energy. With these substitutions the drift term is

$$\begin{aligned} -q\mathcal{E}^i \quad v^j \quad & \sum_{lm} \frac{f^{i,j+1}_{lm} - f^{i,j+1}_{lm}}{\Delta E} \int \int Y^*_{l'm'}(\theta,\phi) Y_{lm}(\theta,\phi) \cos\theta \, \sin\theta \, d\theta \, d\phi \\ +q\mathcal{E}^i \quad v^j \frac{\gamma' j}{2\gamma^j} \quad & \sum_{lm} f^{i,j}_{lm} \int \int Y^*_{l'm'}(\theta,\phi) \frac{\partial Y_{lm}(\theta,\phi)}{\partial \theta} \sin\theta \, \sin\theta \, d\theta \, d\phi \end{aligned} \tag{3.37}$$

where we have used (3.19) to express $\frac{f_{lm}(z,k)}{k}$ in a more convenient form. Also, γ^j and γ'^j refer to $\gamma(E)$ evaluated at the discretized energy E^j. Note that we have already encountered the first integral and it was denoted as $G_{l'm';lm}$. Analogously, the second integral term will be called $H_{l'm';lm}$, which is defined as:

$$H_{l'm';lm} = \int_0^{2\pi} \int_0^{\pi} Y^*_{l'm'}(\theta,\phi) \frac{\partial Y_{lm}(\theta,\phi)}{\partial \theta} \sin\theta \, \sin\theta \, d\theta \, d\phi \, . \tag{3.38}$$

The drift term can then be written more compactly as:

$$-q\mathcal{E}^i v^j \left[\frac{1}{\Delta E} G\left(f^{i,j+1} - f^{i,j}\right) - \frac{\gamma'^j}{2\gamma^j} H f^{i,j}\right] \tag{3.39}$$

where the matrix H was defined by (3.38).

3.4.3 The Complete Equation

Combining the drift and diffusion terms we can write the left hand side of the Boltzmann equation in the compact notation introduced above:

$$v^j \left[\frac{1}{\Delta z} G\left(f^{i+1,j} - f^{i,j}\right) - q\mathcal{E}^i \left(\frac{1}{\Delta E} G\left(f^{i,j+1} - f^{i,j}\right) - \frac{\gamma'^j}{2\gamma^j} H f^{i,j}\right)\right]. \tag{3.40}$$

The final task is to put the above expression in a matrix form. First we introduce some more notation to simplify the final form. Let

$$\begin{aligned} a^{i,j} &= \frac{q\mathcal{E}^i}{\Delta E} v^j \\ b^{i,j} &= \frac{q\mathcal{E}^i \gamma'^j}{2\gamma^j} v^j \\ c^{i,j} &= \frac{v^j}{\Delta z} \end{aligned} \tag{3.41}$$

and

$$\begin{aligned} W_1^{i,j} &= \left(-c^{i,j} + a^{i,j}\right) G + b^{i,j} H \\ W_2^{i,j} &= c^{i,j} G \\ W_3^{i,j} &= -a^{i,j} G \end{aligned} \tag{3.42}$$

Then the coefficient matrix can be written as:

$$\begin{bmatrix} \ddots & \ddots & & \ddots & & & \mathbf{0} \\ & W_1^{i,j} & W_2^{i,j} & & W_3^{i,j} & & \\ & & W_1^{i+1,j} & W_2^{i+1,j} & & W_3^{i+1,j} & \\ & & & \ddots & \ddots & & \ddots \\ & & & & & & \\ & & & & & & \\ & \mathbf{0} & & & & \mathbf{0} & \end{bmatrix} \begin{bmatrix} \vdots \\ f^{i,j} \\ f^{i+1,j} \\ \vdots \\ f^{i,j+1} \\ f^{i+1,j+1} \\ \vdots \end{bmatrix}$$

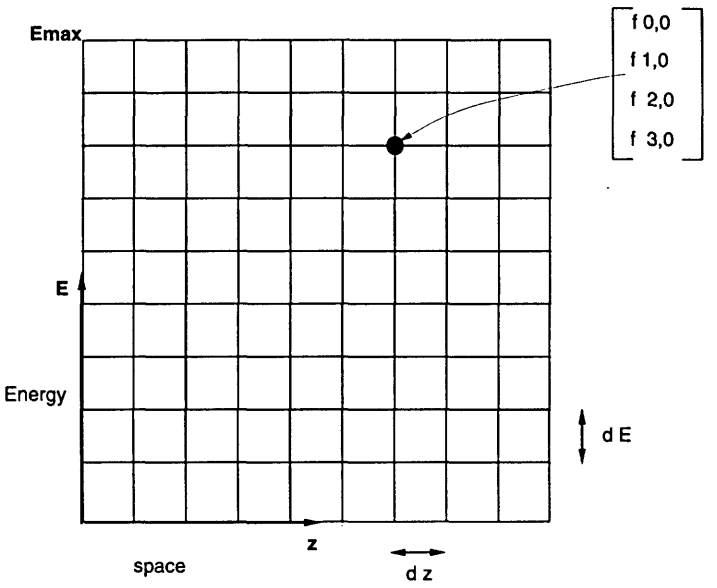

Figure 3-6: Mesh used for the one-dimensional real space problem.

The above matrix equation is the discrete equivalent of the Boltzmann equation for arbitrary order except for the scattering term which has not been included and the boundary conditions on the distribution function which will contribute to the right hand side. Note that the matrix is very sparse– the only off-diagonal non-zero blocks are the ones coupling neighbors in space (W_2) and energy (W_3). The particular location of the blocks depends on the ordering of the unknowns; the above pattern is for a row ordering based on the mesh in Fig. 3-6.

The order of the expansion enters only through the matrices G, H and the size of the vector f, otherwise the formulation is the same. G and H can be calculated *a priori* for whatever order is needed or they could be computed on the fly as they only involve trigonometric functions which can even be computed numerically if necessary. As an example, up to third order G and H are given below for the one-dimensional problem:

$$G = \begin{bmatrix} 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{15}} & 0 \\ 0 & \frac{2}{\sqrt{15}} & 0 & \frac{3}{\sqrt{35}} \\ 0 & 0 & \frac{3}{\sqrt{35}} & 0 \end{bmatrix} \qquad H = \begin{bmatrix} 0 & -\frac{2}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & -2\sqrt{\frac{3}{5}} & 0 \\ 0 & 2\sqrt{\frac{3}{5}} & 0 & -\frac{12}{\sqrt{35}} \\ 0 & 0 & \frac{6}{\sqrt{35}} & 0 \end{bmatrix}$$

Although we have assumed a particular form for the discrete approximation to the derivatives in space and energy, the development described above holds equally well for a different discrete approximation, the only change being that the non-zero blocks would be at a different position in the matrix. For example if one used a backward difference for the derivative in space instead of the forward difference used above, then instead of having a

non-zero block after the diagonal the non-zero block would be before it.

3.4.4 The Scattering Term

We will consider acoustic and optical phonon and ionized impurity scattering; and for all three cases we assume a single spherical band for simplicity. The details of the band structure and the scattering rates are given in section 3.6.

Acoustic Phonon Scattering

We will assume that acoustic phonon scattering is isotropic and completely elastic. It can therefore be written as

$$S_{ac}(\mathbf{k},\mathbf{k}') = c_{ac}\delta\left[E(\mathbf{k}') - E(\mathbf{k})\right] \tag{3.43}$$

where c_{ac} is the scattering rate. The net scattering term due to acoustic phonons is

$$\begin{aligned} & \int S_{ac}(\mathbf{k}',\mathbf{k}) f(\mathbf{k}') d^3\mathbf{k}' - f(\mathbf{k}) \int S_{ac}(\mathbf{k},\mathbf{k}') d^3\mathbf{k}' \qquad (3.44) \\ = \; & c_{ac} \int \delta\left[E(\mathbf{k}) - E(\mathbf{k}')\right] f(\mathbf{k}') d^3\mathbf{k}' - f(\mathbf{k}) c_{ac} \int \delta\left[E(\mathbf{k}') - E(\mathbf{k})\right] d^3\mathbf{k}' \\ = \; & c_{ac}\left[f_{0,0}Y_{0,0} - \sum_{lm} f_{lm}Y_{lm}(\theta,\phi)\right] g(E) \end{aligned}$$

where $g(E)$ is the density of states given below, which enters due to the change of variables from k to E

$$\frac{4\pi(2m^*)^{3/2}}{\hbar^3}\gamma^{1/2}\gamma'.$$

Using vector notation the above scattering term up to third order can be written as:

$$\begin{bmatrix} Y_{0,0} & Y_{1,0} & Y_{2,0} & Y_{3,0} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} f_{0,0} \\ f_{1,0} \\ f_{2,0} \\ f_{3,0} \end{bmatrix} c_{ac}g(E)$$

or more compactly

$$c_{ac}g(E)Y^T S_{ac} f$$

where S_{ac} is understood to be the matrix for the acoustic phonon scattering shown above and Y, f are the vectors of the spherical harmonics and coefficients respectively.

The last step is to generate the coefficient matrix by multiplying the conjugate harmonics and integrating over the unit sphere in k-space:

$$c_{ac}g(E)\int d\Omega\; Y^*Y^T S_{ac} f.$$

Optical Phonon Scattering

The optical phonon scattering rate is given by the expression

$$S_{op}(\mathbf{k},\mathbf{k}') = c_{op}\left[N_{op}\delta(E(\mathbf{k}') - E(\mathbf{k}) - \hbar\omega) + (N_{op}+1)\delta(E(\mathbf{k}') - E(\mathbf{k}) + \hbar\omega)\right] \tag{3.45}$$

where $\hbar\omega$ is the energy of the optical phonon and N_{op} the optical phonon number. The net scattering rate due to optical phonons is

$$\begin{aligned}
&\int S_{op}(\mathbf{k}',\mathbf{k})f(\mathbf{k}')d^3\mathbf{k}' - f(\mathbf{k})\int S_{op}(\mathbf{k},\mathbf{k}')d^3\mathbf{k}' \qquad (3.46)\\
=\ & c_{op}\int N_{op}\delta[E(\mathbf{k}) - E(\mathbf{k}') - \hbar\omega]f(\mathbf{k}')d^3\mathbf{k}' + c_{op}\int N_{op}^{+}\delta[E(\mathbf{k}) - E(\mathbf{k}') + \hbar\omega]f(\mathbf{k}')d^3\mathbf{k}'\\
-\ & f(\mathbf{k})c_{op}\int N_{op}^{+}\delta[E(\mathbf{k}') - E(\mathbf{k}) - \hbar\omega]d^3\mathbf{k}' - f(\mathbf{k})c_{op}\int N_{op}^{+}\delta[E(\mathbf{k}') - E(\mathbf{k}) + \hbar\omega]d^3\mathbf{k}'\\
=\ & c_{op}Y_{0,0}\left[N_{op}f_{0,0}(E-\hbar\omega)g^{-} + N_{op}^{+}f_{0,0}(E+\hbar\omega)g^{+} - N_{op}f_{0,0}(E)g^{+} - N_{op}^{+}f_{0,0}(E)g^{-}\right]\\
-\ & \sum_{lm\neq 0} f_{lm}Y_{lm}(\theta,\phi)c_{op}\left(N_{op}g^{+} + N_{op}^{+}g^{-}\right)
\end{aligned}$$

where $g^{-} = g(E-\hbar\omega)$, $g^{+} = g(E+\hbar\omega)$ are the density of states obtained from the band structure and $N_{op}^{+} = N_{op}+1$. In vector notation the above scattering term up to third order for one real space dimension is:

$$\begin{aligned}
c_{op}\begin{bmatrix} Y_{0,0} & Y_{1,0} & Y_{2,0} & Y_{3,0}\end{bmatrix}\Bigg(&\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{bmatrix}f_{0,0}(E+\hbar\omega)\\f_{1,0}(E+\hbar\omega)\\f_{2,0}(E+\hbar\omega)\\f_{3,0}(E+\hbar\omega)\end{bmatrix}(N_{op}^{+}g^{+}) \qquad (3.47)\\
&-\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}f_{0,0}(E)\\f_{1,0}(E)\\f_{2,0}(E)\\f_{3,0}(E)\end{bmatrix}(N_{op}g^{+} + N_{op}^{+}g^{-})\\
&+\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\begin{bmatrix}f_{0,0}(E-\hbar\omega)\\f_{1,0}(E-\hbar\omega)\\f_{2,0}(E-\hbar\omega)\\f_{3,0}(E-\hbar\omega)\end{bmatrix}(N_{op}g^{-})\Bigg).
\end{aligned}$$

Note that the scattering rate for the isotropic term depends on energies above and below the energy for which the coefficients are written. If space and energy were discretized and the energy step chosen such that $\hbar\omega$ was a multiple of the energy step then the above scattering rate would couple three different energies, E (with the index j), $E+\hbar\omega$ (with the index $j+l$) and $E-\hbar\omega$ (with the index $j-l$). Hence, the spherical harmonics vector would multiply a linear combination of f vectors for three different energies:

$$Y^{T}c_{op}\left[N_{op}^{+}g^{+}S_{op}^{+}f^{j+l} - \left(N_{op}g^{+} + N_{op}^{+}g^{-}\right)S_{op}f^{j} + N_{op}g^{-}S_{op}^{-}f^{j-l}\right]$$

where S_{op}, S_{op}^{-} and S_{op}^{+} are coupling matrices given above.

Ionized Impurity Scattering

The scattering operator for ionized impurity scattering in the Brooks-Herring model is:

$$S_{BH}(\mathbf{k},\mathbf{k}') = \frac{c_{BH}}{[1+\alpha(1-\cos\theta)]^2}\delta[E(\mathbf{k}')-E(\mathbf{k})] \tag{3.48}$$

where θ is the angle between the k and k′ vectors and $\alpha = \frac{2k^2}{\beta^2}$. β is the inverse of the Debye length and therefore depends on the doping concentration; c_{BH} is the scattering rate which depends linearly on the doping concentration. Before we write the scattering integral, we expand the above scattering rate in Legendre polynomials of $\cos\theta$ (as there is no ϕ dependence)

$$S_{BH}(\mathbf{k},\mathbf{k}') = \sum_n P_n(\cos\theta)\,\mathrm{sBH}_n\delta[E(\mathbf{k}')-E(\mathbf{k})] \tag{3.49}$$

where sBH_n are the coefficients of the Legendre polynomial expansion for the Brooks-Herring model which are given in Appendix C. Then the scattering integral can be written as

$$\begin{aligned} & \int S_{BH}(\mathbf{k}',\mathbf{k})f(\mathbf{k}')d^3\mathbf{k}' - f(\mathbf{k})\int S_{BH}(\mathbf{k},\mathbf{k}')d^3\mathbf{k}' \\ = & \left[\sum_{lm}\mathrm{sBH}_{lm}f_{lm}Y_{lm}(\theta,\phi) - \mathrm{sBH}_{0,0}\sum_{lm}f_{lm}Y_{lm}(\theta,\phi)\right]g(E) \\ = & \; g(E)Y^T S_{BH} f \end{aligned} \tag{3.50}$$

where S_{BH} is the net scattering matrix for the Brooks-Herring model, as shown below, up to third order

$$S_{BH} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathrm{sBH}_1 - \mathrm{sBH}_0 & 0 & 0 \\ 0 & 0 & \mathrm{sBH}_2 - \mathrm{sBH}_0 & 0 \\ 0 & 0 & 0 & \mathrm{sBH}_3 - \mathrm{sBH}_0 \end{bmatrix}. \tag{3.51}$$

For the first integral (the in-scattering term) we have used the Addition Theorem for spherical harmonics [44], whereas for the out-scattering integral the only non-zero term is from the lowest order harmonic which is isotropic. The next step is to multiply by the orthogonal spherical harmonics and integrate, which yields:

$$g(E)\int \left(Y^* Y^T S_{BH} f\right)\, d\Omega.$$

3.5 Numerics

A number of numerical issues arise in the Galerkin approach which need to be addressed differently from those that arise in the traditional approach using spherical harmonics. The difference exists primarily because the latter approach leads to a set of coupled partial differential equation whose discretization and subsequent solution is the main numerical focus

whereas in the approach presented here the partial differential equations are not explicitly formed but discretized earlier. Thus issues such as current conservation and the stability of the discretization need to be addressed in a more general way as they apply to the whole system rather than to each individual PDE as in the traditional approach.

3.5.1 Discretization

As noted in the previous section, in the arbitrary order method we discretize the differential operators in space and energy before we multiply by the orthogonal harmonics, which results in a matrix problem rather than a set of partial differential equations. The actual choice of the discrete approximation to the differential operator depends on obtaining a stable scheme and on the boundary conditions.

The time varying Boltzmann equation in one real-space dimension after expansion into spherical harmonics can be written as:

$$\frac{\partial f}{\partial t} + \frac{\partial (Gf)}{\partial z} - \mathcal{E}(z)\frac{\partial (Gf)}{\partial E} - \frac{\mathcal{E}(z)\gamma'}{2\gamma} Hf = \frac{1}{v(E)}\left(\frac{\partial f}{\partial t}\right)_c \tag{3.52}$$

This equation can be put into a discretized form as shown below, where we have used a box discretization with mesh spacing of Δz in space and ΔE in energy ($\Delta A = \Delta z \Delta E$):

$$\begin{aligned} \Delta A \frac{\partial f}{\partial t} + \Delta E\ G(f^{i+1/2,j} - f^{i-1/2,j}) \ &- \ \Delta z\ \mathcal{E}G(f^{i,j+1/2} - f^{i,j-1/2}) \\ &- \ \Delta A\ \frac{\mathcal{E}^i \gamma'^j}{2\gamma^j} H f^{i,j} = \frac{\Delta A}{v^j}\left(\frac{\partial f^{i,j}}{\partial t}\right)_c \end{aligned} \tag{3.53}$$

The terms $f^{i+1/2,j}, f^{i-1/2,j}, f^{i,j+1/2}, f^{i,j-1/2}$ represent the average 'flux' at the interface of the control volume as shown in Fig. 3-7. To completely specify the discretization method it is necessary to choose the 'average' flux at the interface. In the time dependent case the Boltzmann equation is a hyperbolic equation (a wave equation) and if time is discretized then at each time step the new values of the coefficients f can be computed either explicitly based on the values at the previous time steps or implicitly if the unknown values at the current time step are also used. If only a steady state solution is of interest then obviously an explicit time marching scheme is not possible and an implicit solution is required. But the stability issue remains the same: how should the flux be computed at the interfaces?

One approach is to everywhere use a simple two-point discretization for the flux for example one choice would be $f^{i+1/2,j} = f^{i,j}$ and $f^{i-1/2,j} = f^{i-1,j}$ and a similar expression for the flux in the direction of the energy axis. It must be remembered that the flux here is a vector quantity and therefore there exists the added freedom of choosing a different expression for each component of the flux vector. Hence, a large number of variations are possible even for a two point discretization scheme when two dimensions and a few orders (the size of the flux vector) are included. Some results from the one-dimensional wave equation can guide the search for the best discretization method. For example, it is known that in the one-dimensional

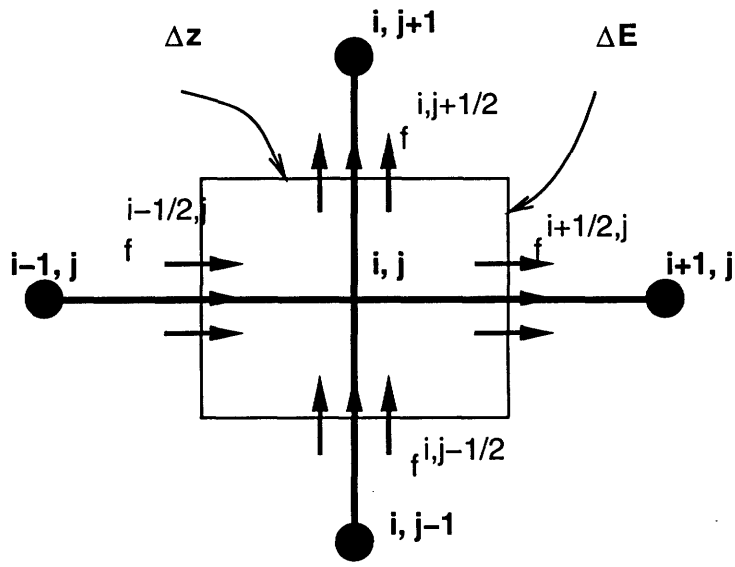

Figure 3-7: Control volume for energy-space discretization.

wave equation a symmetric discretization such that the net flux is given by $f^{i+1} - f^{i-1}$ is always unstable. This was also found to be true for the above problem. After some trial and error a first order discretization that was usually effective was found, which is denoted as a one-sided discretization in the following discussion. This scheme uses a forward one-sided, two-point approximation to the derivatives in space and energy of the coefficients of even harmonics $f_{0,0}$, $f_{2,0}$ etc. and a backward one-sided, two-point approximation in space and energy to the derivatives of the odd harmonics $f_{1,0}$, $f_{3,0}$ etc. This method is shown schematically in Fig. 3-8, and the results obtained up to third order using this method are shown in Fig. 3-9.

Although this discretization scheme is successful in most cases, under certain circumstances it can lead to instability, for example near the n^+n junction in the simulation of the n^+nn^+ diode [48]. An example of the instability is shown in Fig. 3-10. In this figure the coefficients are plotted as a function of space, but the oscillation are not in space, they are in energy in the region to the left of the n^+n junction. The salient fact about this junction is that this is the only part of the device where the electric field is positive for the bias conditions used in Fig. 3-10. We also know that the discretization shown in Fig. 3-8 is stable for a homogeneous field only if it is negative everywhere but not if it is positive. These two observations clearly imply that the sign of the field is the key to stable discretization. To better understand the stability issues we will use the insights from the one-dimensional wave equation and use it to guide our approach.

3.5.2 Upwinding Theory

It is well known that if the wave equation (in one-dimension)

$$U_t + a\,U_x = 0 \tag{3.54}$$

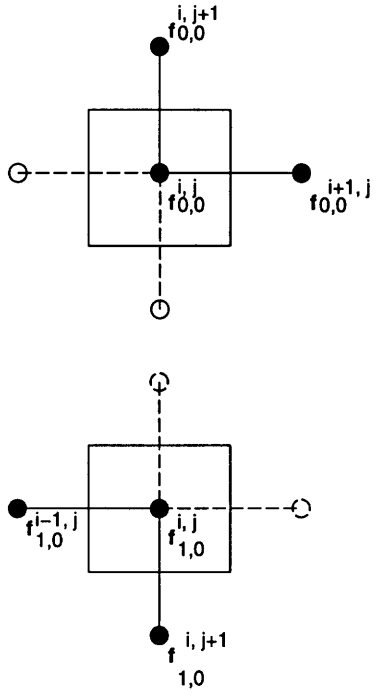

Figure 3-8: The one-sided discretization scheme, where the horizontal axis represents real space and the vertical direction energy. The index i corresponds to discretization in space and the index j to discretization in energy.

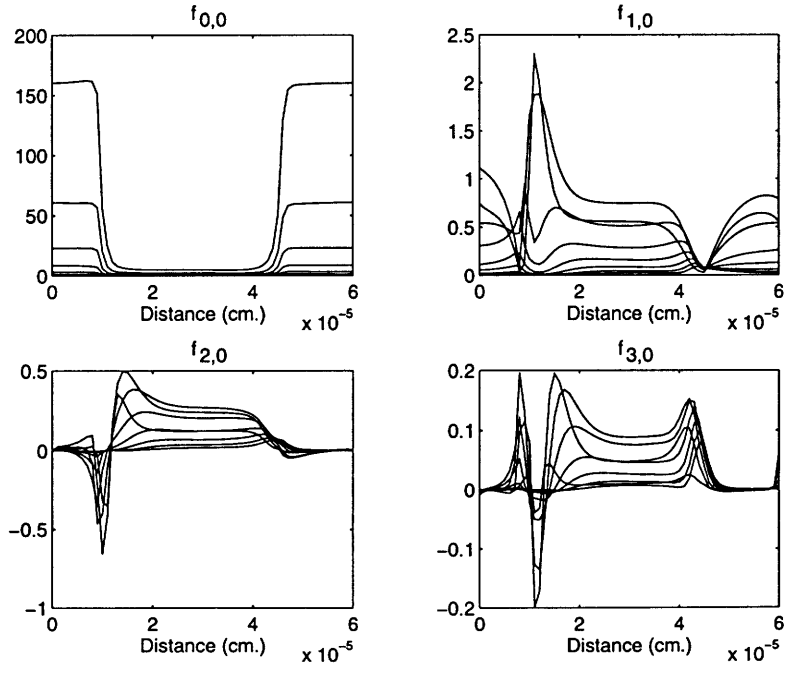

Figure 3-9: Stable results for the spherical harmonic coefficients (in arbitrary units) using the discretization shown in Fig. 3.8 for the simulation of an n^+nn^+ structure.

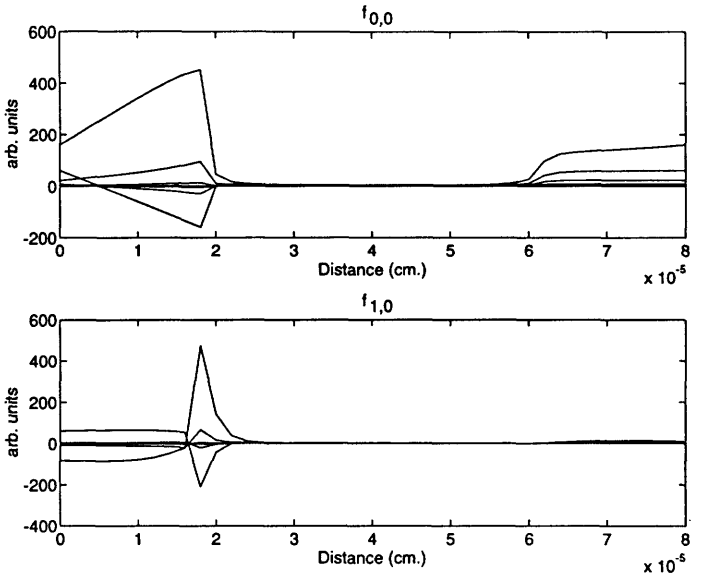

Figure 3-10: Unstable results obtained using the one-sided scheme.

is discretized using a forward Euler time marching scheme as shown below:

$$U_i^{k+1} - U_i^k = \frac{\Delta t}{\Delta x} a \left(U_{i+1/2}^k - U_{i-1/2}^k \right) \tag{3.55}$$

where k is the index for temporal discretization and i for spatial discretization, then for a stable discretization the spatial derivative must be done either in the forward or the backward direction depending on the sign of the velocity a [49]. This is know as upwind differencing. Thus, upwind differencing requires that

$$a > 0 \rightarrow \quad U_i^{k+1} - U_i^k = \tfrac{\Delta t}{\Delta x} a \left(U_{i+1}^k - U_i^k \right) \tag{3.56}$$

$$a < 0 \rightarrow \quad U_i^{k+1} - U_i^k = \tfrac{\Delta t}{\Delta x} a \left(U_i^k - U_{i-1}^k \right) \tag{3.57}$$

Note that in the control volume sense the above equation simply conserves the flux, that is if there is a net flux leaving a control volume in space then at the next time step the value of the function at the center of that control volume is reduced commensurately. Of course the change can be an increase or a decrease depending on the sign of the net flux.

In the Boltzmann equation, the energy derivative has a coefficient which changes sign depending on the electric field and therefore it is the energy derivative that must be 'winded'.

Now, up to first order the flux term in energy is given by Gf where

$$\mathrm{G} = \begin{bmatrix} 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & 0 \end{bmatrix} \qquad \mathrm{G\,f} = \begin{bmatrix} \frac{1}{\sqrt{3}} f_{1,0} \\ \frac{1}{\sqrt{3}} f_{0,0} \end{bmatrix}$$

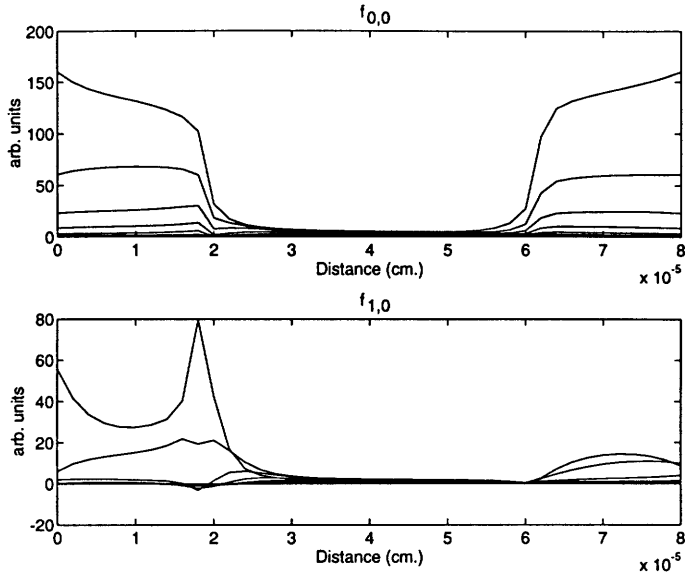

Figure 3-11: Stable results obtained using the upwinded scheme.

Therefore, the flux difference in energy has two possible expressions: if the electric field is negative it is

$$(-\mathcal{E}^i)\begin{bmatrix} f^j_{1,0} - f^{j-1}_{1,0} \\ f^{j+1}_{0,0} - f^j_{0,0} \end{bmatrix}$$

otherwise it is

$$\mathcal{E}^i\begin{bmatrix} f^{j+1}_{1,0} - f^j_{1,0} \\ f^j_{0,0} - f^{j-1}_{0,0} \end{bmatrix}$$

where we have suppressed the $\frac{1}{\sqrt{3}}$ factor for clarity. Note that the direction of the two point discretization for the even and odd harmonics are still opposite to each other as in the one-sided discretization but these directions flip sign depending on the electric field direction.

The results obtained using the upwinded discretization for the same structure and bias conditions as in Fig. 3-10 are shown in Fig. 3-11. Clearly the unstable oscillations introduced due to the discretization have been suppressed here. The advantage of using the more stable winded method is most clearly seen if $f_{0,0}$ is plotted as a function of energy in the left n^+ region for the two cases (Fig. 3-12).

An upwind scheme was also used in [50] where the BTE was solved by direct discretization of the distribution function in spherical coordinates but without using an expansion in spherical harmonics.

3.5.3 Boundary Conditions

The Boltzmann equation has only first order derivatives in space whereas physically we would like to specify the boundaries (Ohmic contacts for example) in space at both the left

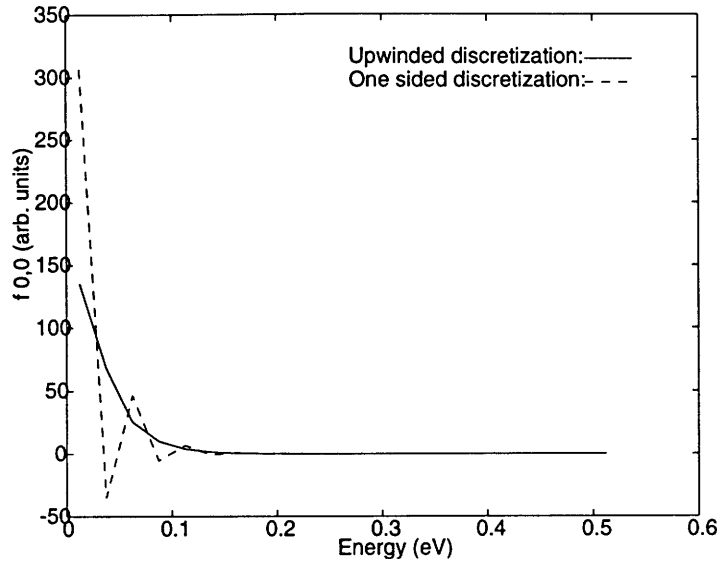

Figure 3-12: A comparison of the solution for $f_{0,0}$ obtained using the one-sided (dash) and upwinded (solid) discretization schemes.

and right end points. This seeming inconsistency can be resolved by using the fact that the Boltzmann equation expansion in spherical harmonics is really a set of equations. Thus, we specify the zeroth order harmonic at both edges but leave the first order coefficient unspecified everywhere. In general all even harmonics are fixed on both sides and all odd harmonics are left floating at the edges and their values calculated everywhere. We set the zeroth order harmonic to be the Maxwellian at the lattice temperature and the second and higher harmonics to zero. For low-order spherical harmonics these boundary conditions have a physical interpretation. Specifying the zeroth and second order harmonics, and solving for the first and third is equivalent to specifying the electron concentration and the electron temperature on both edges but solving for the electron current and the heat flux at those points.

In the energy direction we assume that all harmonics are zero beyond some maximum energy and also for the odd harmonics use the fact that the odd harmonics must be zero at zero energy. This is a necessary condition for the odd harmonics, otherwise there would be a discontinuity in the distribution at zero energy. This can be demonstrated for the first order harmonic term $f_{1,0}(k)\cos\theta$ by considering the distribution along the k_z axis— for $k_z > 0$, $\cos\theta = 1$ but for $k_z < 0, \cos\theta = -1$. Therefore if $f_{1,0}$ had a non-zero value in the neighborhood of $k_z = 0$ then the distribution would have a discontinuity at $k_z = 0$.

3.5.4 Current Conservation

In the previous section we discussed issues regarding the stability of the discretization, in this section we will be concerned with guaranteeing current conservation with the chosen discretization method.

The equation for the zeroth harmonic including acoustic and optical phonon and ionized impurity scattering is:

$$\frac{\partial f_{1,0}}{\partial z} - q\mathcal{E}\left(\frac{\partial f_{1,0}}{\partial E} + \frac{\gamma'}{\gamma} f_{1,0}\right) = S[f_{0,0}(E), f_{0,0}(E-\Delta E), f_{0,0}(E+\Delta E)] \tag{3.58}$$

where the scattering operator is given by

$$S[,,] = \frac{\sqrt{3}}{v(E)} c_{op}\left[g^{+}\left(N_{op}^{+} f_{0,0}(E+\Delta E) - N_{op} f_{0,0}(E)\right) + g^{-}\left(N_{op} f_{0,0}(E-\Delta E) - N_{op}^{+} f_{0,0}(E)\right)\right] \tag{3.59}$$

and $g^{+} = g(E+\Delta E)$, $g^{-} = g(E-\Delta E)$ are the density of states at the appropriate energies. N_{op} is the number of optical phonons which is given by the Bose-Einstein distribution, $N_{op} = \left[e^{\hbar\omega_{op}/k_B T} - 1\right]^{-1}$ and $N_{op}^{+} = N_{op} + 1$.

In one dimension the current is only in the z-direction and is given as:

$$\begin{aligned} J_z = q\int f_{tot}(\mathrm{k}) v_z(\mathrm{k})\, d^3\mathrm{k} &= q\int f_{tot}(\mathrm{k}) v(\mathrm{k}) \cos\theta\, d^3\mathrm{k} \\ &\propto \int f_{1,0}(k) v(k) d^3\mathrm{k} \\ &\propto \int f_{1,0}(E) v(E) g(E) dE \\ &\propto \int f_{1,0}(E)\gamma(E) dE. \end{aligned} \tag{3.60}$$

The only component of the distribution function that contributes to the current in the z-direction is the $f_{1,0}$ coefficient. This of course follows from the orthogonality property of the harmonics. Using this property we can write the equation for the zeroth harmonic essentially as a current conservation equation as follows:

$$\begin{aligned} \frac{\partial f_{1,0}}{\partial z} - q\mathcal{E}\left(\frac{\partial f_{1,0}}{\partial E} + \frac{\gamma'}{\gamma} f_{1,0}\right) &= S[f_{0,0}(E), f_{0,0}(E-\Delta E), f_{0,0}(E+\Delta E)] \\ \gamma\frac{\partial f_{1,0}}{\partial z} - q\mathcal{E}\left[\gamma\frac{\partial f_{1,0}}{\partial E} + \gamma' f_{1,0}\right] &= \gamma\, S[f_{0,0}(E), f_{0,0}(E-\Delta E), f_{0,0}(E+\Delta E)] \\ \int \frac{\partial(\gamma f_{1,0})}{\partial z} dE - q\mathcal{E}\int\left[\frac{\partial(\gamma f_{1,0})}{\partial E}\right] dE &= \int \gamma(E)\, S[f_{0,0}(E), f_{0,0}(E-\Delta E), f_{0,0}(E+\Delta E)] dE \\ \int \frac{\partial(\gamma f_{1,0})}{\partial z} dE &= q\mathcal{E}\left[\gamma(E) f_{1,0}\right]_0^{\infty} \\ &+ \int \gamma(E)\, S[f_{0,0}(E), f_{0,0}(E-\Delta E), f_{0,0}(E+\Delta E)] dE \end{aligned} \tag{3.61}$$

In the last equation above, the left hand side is proportional to the spatial derivative of the current, which follows from (3.61), and therefore must be zero if current is to be conserved. Using the boundary condition that $\gamma(E) f_{1,0} \to 0$ as $E \to \infty$, which follows from the assumption that the total distribution approaches zero "faster" than any polynomial function of energy, the first term is zero. Hence, the integral of the scattering term over energy must reduce to zero for current to be conserved. From the scattering operator given in (3.59), it can be easily

shown that this condition is satisfied:

$$
\begin{aligned}
&\int \gamma(E)\ S[f_{0,0}(E), f_{0,0}(E-\Delta E), f_{0,0}(E+\Delta E)]dE \qquad (3.62)\\
&\propto \int \frac{\gamma(E)}{v(E)}\left[g^{+}\left(N_{op}^{+}f_{0,0}(E^{+}) - N_{op}f_{0,0}(E)\right) + g^{-}\left(N_{op}f_{0,0}(E^{-}) - N_{op}^{+}f_{0,0}(E)\right)\right]dE\\
&\propto \int \gamma^{1/2}\gamma'\left[g^{+}\left(N_{op}^{+}f_{0,0}(E^{+}) - N_{op}f_{0,0}(E)\right) + g^{-}\left(N_{op}f_{0,0}(E^{-}) - N_{op}^{+}f_{0,0}(E)\right)\right]dE\\
&\propto \int \gamma^{1/2}\left[g^{+}\left(N_{op}^{+}f_{0,0}(\gamma^{+}) - N_{op}f_{0,0}(\gamma)\right) + g^{-}\left(N_{op}f_{0,0}(\gamma^{-}) - N_{op}^{+}f_{0,0}(\gamma)\right)\right]d\gamma\\
&\propto \int \sqrt{\gamma}\sqrt{\gamma+\Delta\gamma}\left(N_{op}^{+}f_{0,0}(\gamma^{+}) - N_{op}f_{0,0}(\gamma)\right) + \sqrt{\gamma}\sqrt{\gamma-\Delta\gamma}\left(N_{op}f_{0,0}(\gamma^{-}) - N_{op}^{+}f_{0,0}(\gamma)\right)d\gamma\\
&= 0
\end{aligned}
$$

where we made a change of variable from E to γ, and the last identity was obtained by a change of variables ($\gamma \rightarrow \gamma + \Delta\gamma$) in the second term of the integrand.

Thus, current conservation is inherent in the continuous case and the discretized equation must preserve this property, which is achieved by using the discretization discussed in the previous section. Note that performing a higher order expansion leaves the above current conservation equation unchanged as the zeroth order equation only depends upon the first order coefficient and not any higher ones.

3.5.5 Self-consistent solution with Poisson's equation

In all the discussion up to now we have assumed that the electric field was known or given independently. Of course, in general the electric potential or field has to be solved self-consistently with the distribution function. Thus, along with the Boltzmann equation we need to discretize and solve Poisson's equation. Using the box-discretization method mentioned earlier it is fairly straightforward to set up the discretized Poisson's equation. The only complication comes about from the fact that in the self-consistent formulation the system of equations becomes nonlinear: the electric field multiplies the energy derivative of the distribution function coefficients. Hence, this nonlinear system must be solved. A standard way of solving nonlinear systems is Newton's method, which is known to converge quadratically [51]. Therefore if a good initial guess is known, after typically five to eight iterations, the solution is found. The size of the matrix problem increases only slightly as there is only one extra unknown at each real space point (there may be 50 coefficients in energy at each point) but the cost in the number of iterations is of course significant. Also the sparsity of the matrix is reduced slightly because there are many non-zero entries associated with the

extra unknown as shown below for a typical Jacobian matrix:

$$\begin{bmatrix} \ddots & \ddots & & \ddots & & & 0 \\ & W_1^{i,j} & W_2^{i,j} & & W_3^{i,j} & & \\ & & W_1^{i+1,j} & W_2^{i+1,j} & & W_3^{i+1,j} & \\ & & \ddots & \ddots & & & \ddots \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ P & \dots & P & 1 & -2 & 1 & 0 \\ & P & \dots & 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} \vdots \\ \delta f^{i,j} \\ \delta f^{i+1,j} \\ \vdots \\ \delta f^{i,j+1} \\ \delta f^{i+1,j+1} \\ \vdots \\ \delta \psi^{i,j} \\ \delta \psi^{i+1,j} \\ \vdots \end{bmatrix}$$

where the P matrix blocks are the coefficients in the discrete approximation to the integral of f_{lm} over energy which gives the electron concentration at that point in real space. Using this method, we have found no difficulty in converging from the zero potential inital guess.

3.6 The Physical Model

For the simulation results presented in this and subsequent chapters the following scattering mechanisms and band structure are used.

3.6.1 The Band Model

A single parabolic band with an effective mass of 0.26 times the electron mass is assumed:

$$E = \frac{\hbar^2 k^2}{2m^*}$$

3.6.2 Acoustic Phonon Scattering

The scattering probability for an electron starting from an initial momentum state k to the state k′ under acoustic phonon scattering is [13]:

$$P(\mathbf{k}, \mathbf{k}') = \frac{2\pi k_B T_0}{\hbar V u_l^2 \rho} \mathcal{E}_1^2 \delta \left[E(\mathbf{k}') - E(\mathbf{k}) \right] \tag{3.63}$$

where V is the volume, u_l^2 is the sound speed in silicon, ρ is the density of silicon, $\mathcal{E}_1$ is the deformation potential for the lattice. As we are working in energy space we need to convert

the above scattering probability from momentum space to energy space which is done by integrating over the three dimensional k space after appropriate scaling:

$$\frac{V}{(2\pi)^3}4\pi\int P(E(\mathrm{k}),E(\mathrm{k}'))k'^2dk' = \frac{V}{2\pi^2}\frac{\sqrt{2}m^{*3/2}}{\hbar^3}\int P(E,E')E'^{1/2}dE'. \tag{3.64}$$

Hence the acoustic scattering rate becomes

$$P(E,E') = \frac{\sqrt{2}m^{*3/2}k_BT_0}{\pi\hbar^4u_l^2\rho}\mathcal{E}_1^2\delta\left[E-E'\right]E^{1/2}. \tag{3.65}$$

The values of the parameters used are:

m^*	0.26 m_0
u_l	9.00 $\times 10^3$ ms^{-1}
ρ	2.33 $\times 10^3$ kgm^{-3}
$\mathcal{E}_1$	9.00 eV

3.6.3 Optical Phonon Scattering

For optical phonons the scattering rate is [13]:

$$P(\mathrm{k},\mathrm{k}') = \frac{\pi(D_tK)^2}{V\rho\omega_{op}}\{N_{op};N_{op}+1\}\delta\left[E(\mathrm{k}')-E(\mathrm{k})\mp\hbar\omega\right] \tag{3.66}$$

where D_tK is the coupling constant, ω_{op} is the frequency of the optical phonon and N_{op} is the number of optical phonons using Bose statistics. Note that N_{op} is associate with $-\hbar\omega$ and $N_{op}+1$ is associated with the $+\hbar\omega$ for the cases of absorption and emission of optical phonons.

The above probability can be written in energy space to yield:

$$P(E,E') = \frac{(D_tK)^2}{\rho\omega_{op}}\frac{m^{*3/2}}{\sqrt{2}\pi\hbar^3}\{N_{op};N_{op}+1\}E^{1/2}\delta\left[E'-E\mp\hbar\omega\right] \tag{3.67}$$

The parameter values used in the above equation are:

D_tK	5.0×10^8 eV cm^{-1}
$\hbar\omega_{op}$	50 meV

3.6.4 Ionized Impurity Scattering

For ionized impurity scattering we use the Brooks-Herring model [13] which is based on assuming a shielded potential (the Yukawa potential) for the ionized impurity. The resulting scattering probability is given as:

$$S(\mathrm{k},\mathrm{k}') = \frac{Z^2q^4}{\epsilon_{Si}^2\hbar}N_i\frac{1}{\left[\beta^2+2k^2(1-\cos\theta)\right]^2}\delta[E'-E] \tag{3.68}$$

where Z is the ordinality of the impurity charge, N_i is the impurity concentration, β is the Debye length and θ is the angle between the k and k' vectors, which are assumed to have the same magnitude k as the scattering is considered to be elastic.

3.7 Simulation Results

In this section we present results using the arbitrary order expansion method described in the previous sections of this chapter. We will first present results for the homogeneous problem with no spatial variation and compare those results with bulk Monte-Carlo data which is a test of the implementation of our scattering terms. Then results for one-dimensional structures such as uniform doping and n^+nn^+ diode are presented. All the results presented are using a self-consistent scheme and were obtained using a direct sparse solver on the sparse matrix generated by the Galerkin method. For the one-dimensional problem the size of the sparse matrix problem is approximately 10,000 unknowns and therefore using a direct sparse method rather than an iterative method is not a major handicap. For a much larger problem or for two real space dimensions direct sparse solvers become less efficient than conjugate gradient type iterative methods if a good preconditioner is available, and an iterative method would then be the matrix solver of choice.

3.7.1 Homogeneous problem

Here we set all spatial derivatives to zero and then solve for the coefficients. Using results up to first order we can match the distribution function obtained from Monte-Carlo simulations [52] if the same scattering mechanisms and band structure are used. In Fig. 3-13 the electron mean velocity calculated using the spherical harmonic approach and the Monte-Carlo method is plotted as a function of the applied electric field along with experimental data. The graph clearly demonstrates that the results from the spherical harmonic method and Monte-Carlo are essentially the same except for the noise in the Monte-Carlo data. The experimental data is qualitatively similar but the saturation velocity is about 20% lower because of the simplistic band and scattering model used in the simulation.

3.7.2 Resistor

If we simulate a uniformly doped finite region of silicon (a resistor) with an applied bias, we expect to see a uniform field in the bulk and some carrier heating if the field is large enough. Fig. 3-14 and 3-15 show the first four coefficients and the macroscopic parameters obtained from such a simulation with an average field of 1×10^4 V/cm. There are a number of noteworthy features even in this simplest inhomogeneous problem.

1. The magnitude of the coefficients in the uniform part of the resistor decreases monotonically as the order is increased.

2. The electron temperature in the bulk of the resistor is above the lattice temperature as evinced by the compression of the lines of $f_{0,0}$ at the different energies at the center as compared to the edges. The magnitude of $f_{0,0}$ decreases faster at the edges than at the

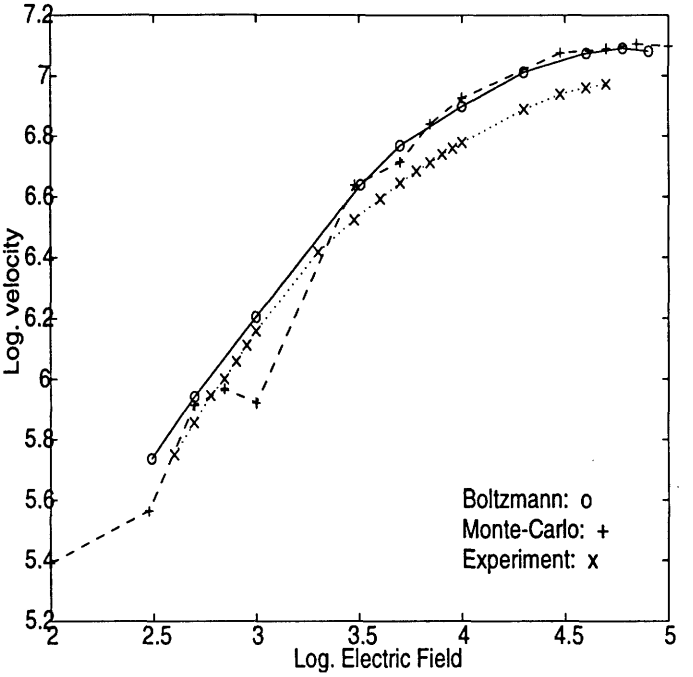

Figure 3-13: Mean electron velocity in the bulk for undoped silicon as a function of the electric field calculated using the spherical harmonic expansion method and a Monte-Carlo method along with experimental data.

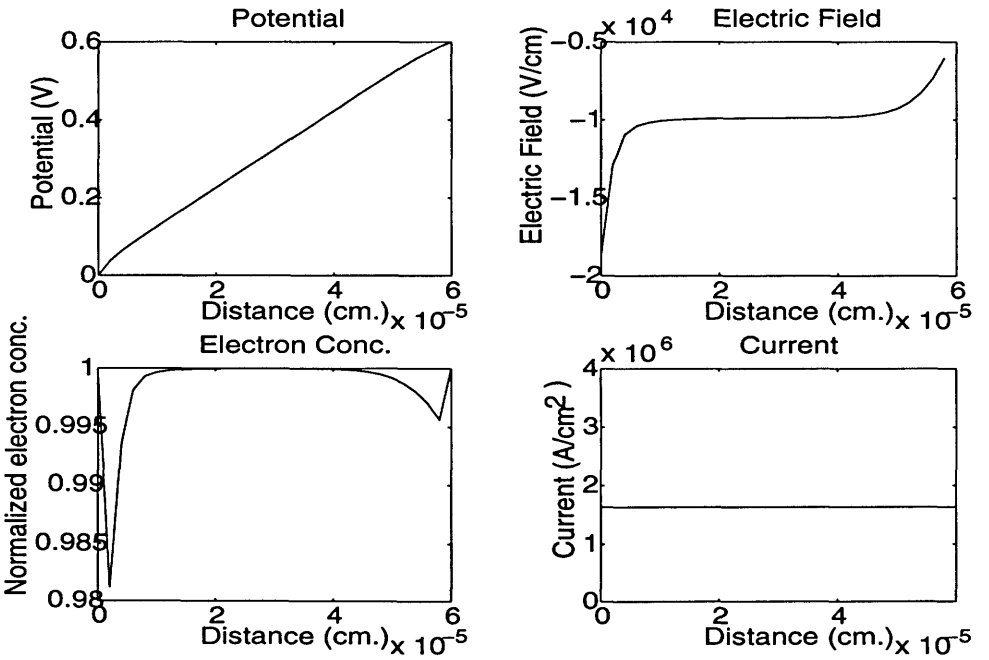

Figure 3-14: The potential, electric field, normalized electron concentration and the normalized current for a resistor obtained from the Boltzmann equation solution up to third order.

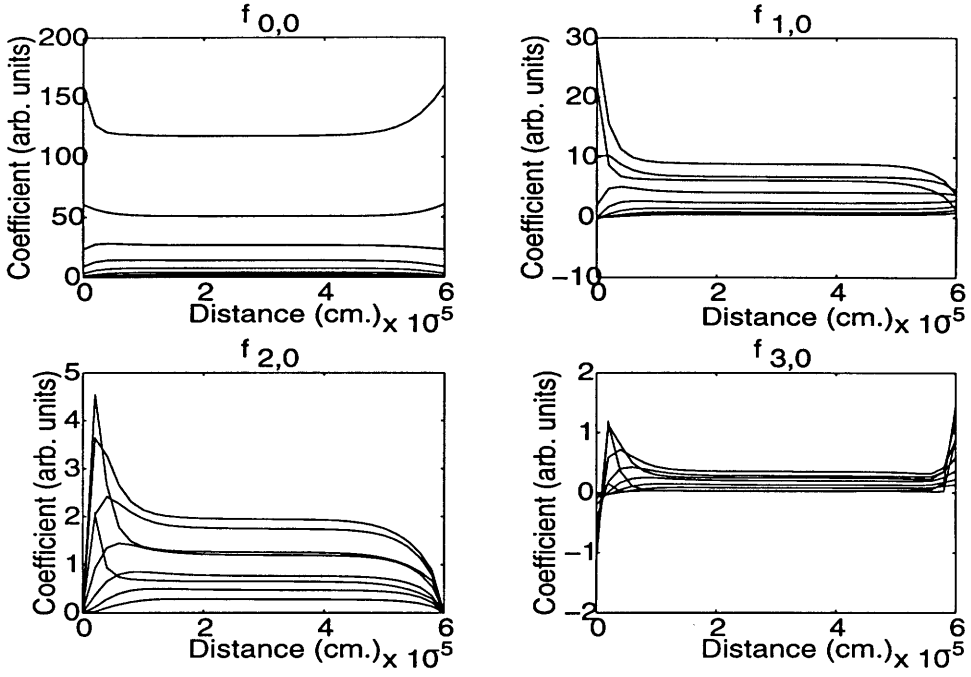

Figure 3-15: Zero to third order coefficients as a function of position for different electron energies (25meV spacing in energy) for the 0.6 μm resistor shown in the previous figure.

center as a function of energy which is equivalent to having a lower electron temperature at the edges than the center.

3. There is a small charge buildup near both the left and right contacts. This occurs because at the contacts we have assumed that the isotropic part of the distribution ($f_{0,0}$) is a Maxwellian at the lattice temperature, whereas in the center of the resistor the electron gas is at an elevated temperature. Thus, there would be an excess thermal current if the field were uniform all the way up to the contacts. Hence, to maintain a constant current the electric field deviates from its value in the bulk to counteract these thermal currents.

4. The current is constant as was predicted in section 3.5, using the discretization methods described earlier.

3.7.3 Diode

Fig. 3-16 shows the computed spherical harmonics coefficients for an n^+nn^+ diode as a function of position at different energy values (separated by 25 meV). The electron concentration and the current obtained from the calculated distribution function along with the potential and electric field obtained by the self-consistent solution of Poisson's equation for the same device are shown Fig. 3-17.

A comparison between the electric field and electron temperature obtained using the hydrodynamic model [28] and the solution of the Boltzmann equation for this particular

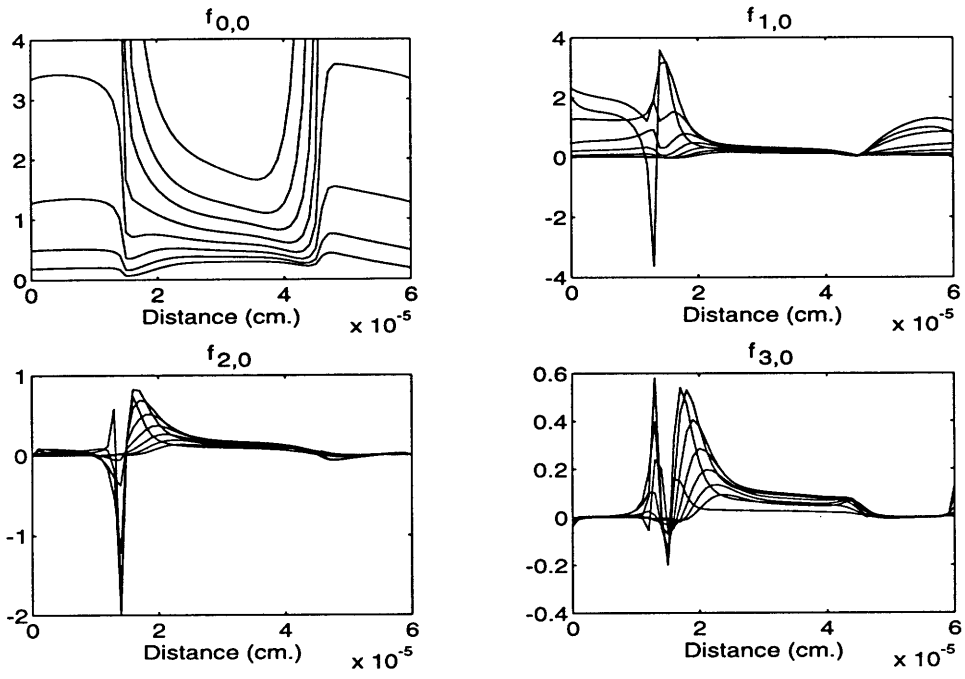

Figure 3-16: Spherical harmonic coefficients for a $0.6\,\mu$m n^+nn^+ diode with a doping of $2 \times 10^{18} cm^{-3}$ and $1 \times 10^{17} cm^{-3}$ in the n^+ and n regions and a bias of $0.8V$.

structure is shown in Fig.3-18 and Fig.3-19. Although the field and temperature obtained using the two approaches are not identical because the mobility model used in the hydrodynamic corresponds to slightly different scattering parameters than those used in the Boltzmann solver, the results are quite comparable.

In Fig.3-20 the distribution functions obtained using different orders of spherical harmonics are compared at two points along the diode. The impact of including higher orders is obvious as it tends to produce a more streamed distribution than if only the first two orders were considered. The electron temperatures obtained in the two cases quite close which demonstrates the lack of sensitivity to the details of the distribution in this macroscopic parameter. The coefficients of the distribution as a function of the energy are shown in Fig. 3-21 at two positions along the diode. Note that the distribution at the peak electric field point ($z = 0.44\,\mu$m) shows the mixing of two carrier populations: hot carriers from the source and cold carriers from the drain. This can be adduced from the two slopes that are evident in the isotropic part of the distribution ($f_{0,0}$): the larger slope (lower temperature) up to 0.1eV is due to cold carriers from the drain and the smaller slope (higher temperature) beyond 0.1eV is due to hot carriers from the source. This example also demonstrates the difficulty any averaging scheme, such as a moments method, would have in correctly estimating the hot carrier population. For instance, the average temperature at $z = 0.44\,\mu$m is close to the lattice temperature (as shown in Fig. 3-18) as it is dominated by the cold electrons from the drain even though there is very substantial number of hot carriers present, as seen in Fig. 3-21.

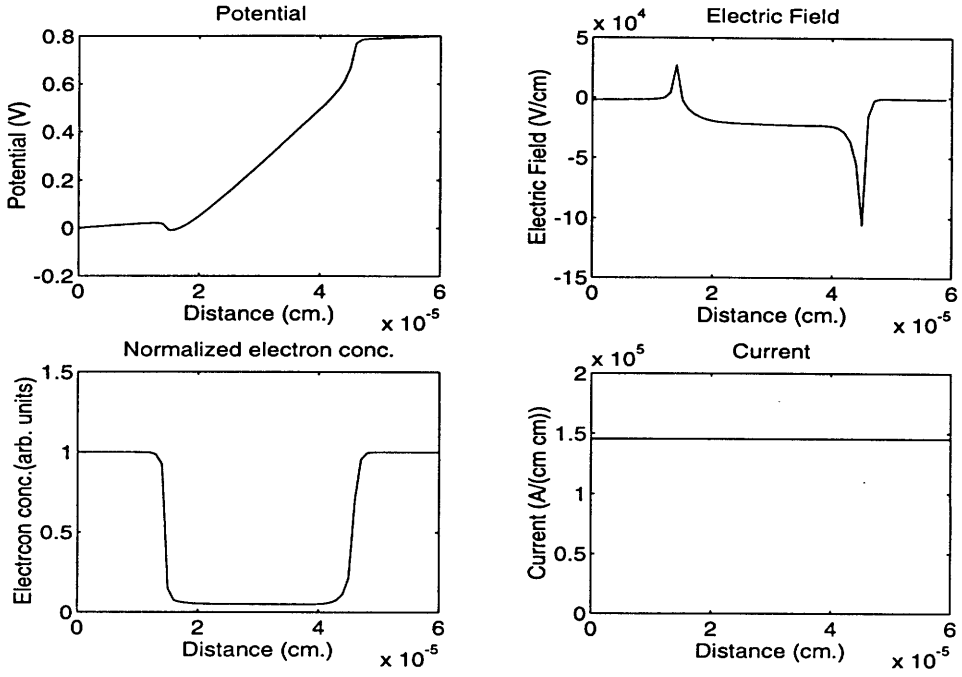

Figure 3-17: Average quantities for a 0.6 μm n^+nn^+ diode with a doping of $2 \times 10^{18} cm^{-3}$ and $1 \times 10^{17} cm^{-3}$ in the n^+ and n regions and a bias of $0.8V$.

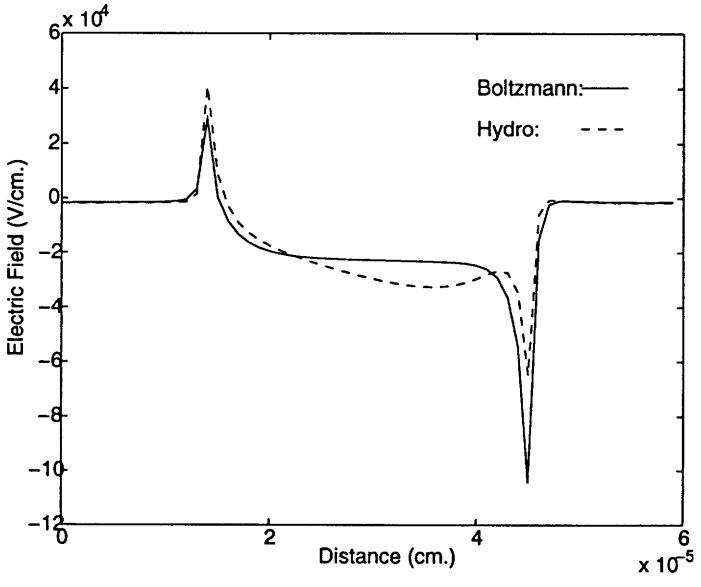

Figure 3-18: A comparison of the electric field computed from the solution of the Boltzmann and Poisson's equations with that obtained from the hydrodynamic model for a 0.6 μm n^+nn^+ diode with a doping of $2 \times 10^{18} cm^{-3}$ and $1 \times 10^{17} cm^{-3}$ in the n^+ and n regions and a bias of $0.8V$.

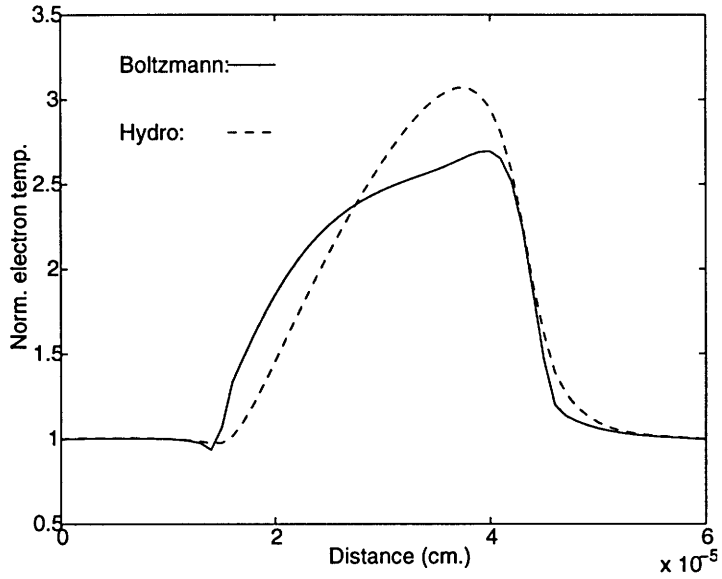

Figure 3-19: A comparison of the electron temperature computed from the solution of the Boltzmann and Poisson's equations that obtained from the hydrodynamic model for a 0.6 μm n^+nn^+ diode with a doping of $2 \times 10^{18} cm^{-3}$ and $1 \times 10^{17} cm^{-3}$ in the n^+ and n regions and a bias of $0.8V$.

3.8 Summary

In this chapter we have presented a Galerkin method that allows the use of an arbitrary order spherical harmonic expansion for the solution of the Boltzmann equation. Earlier work which was restricted to low orders was discussed and then the arbitrary order scheme was developed. Simulation results, up to third order, based on the Galerkin method for one dimensional structures were presented. From the simulations results it was demonstrated that spherical harmonic coefficients beyond the the first can be significant and neglecting them may not be appropriate in certain cases. Also it was shown that under high and rapidly varying fields the calculated distribution function can differ significantly from a displaced Maxwellian or other simple form and therefore any macroscopic method will perform poorly at estimating the details of the distribution function. This result casts doubts on the ability of a macroscopic model to accurately predict substrate and gate currents without resorting to an excessive number of fitting parameters.

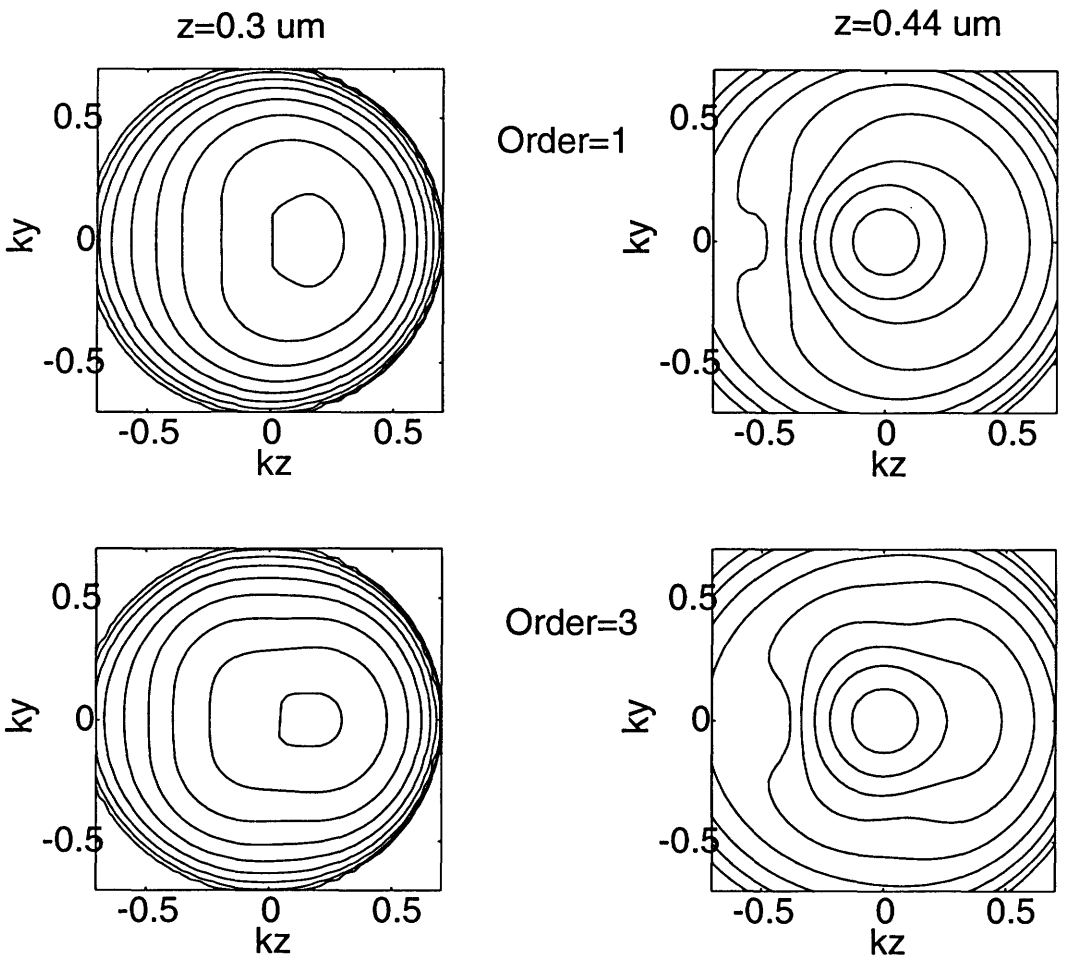

Figure 3-20: Contours of constant distribution function (separated by 3x) over a normalized k_z, k_y plane for the device of Fig. 1 up to the first and up to third order harmonic expansions at z=0.44 μm and z = 0.30 μm.

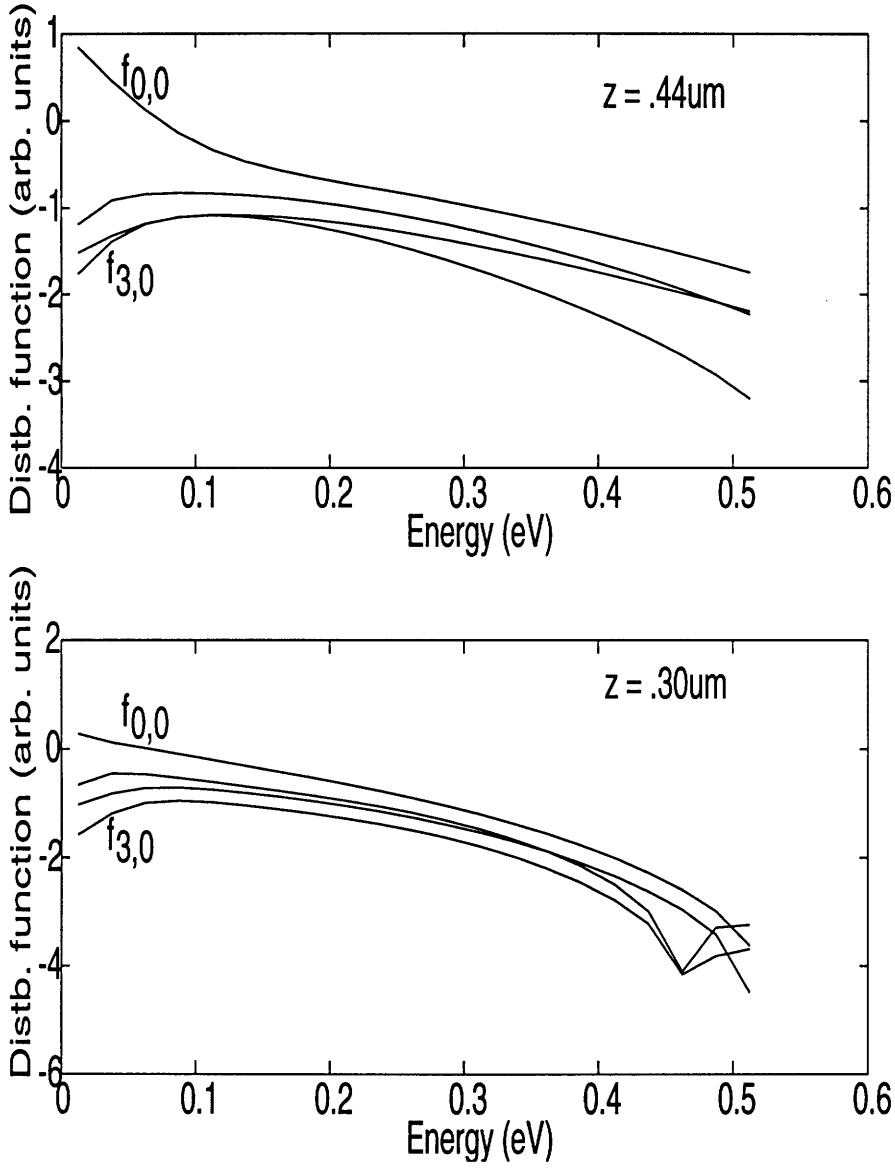

Figure 3-21: Coefficients of spherical harmonics up to third order for the 0.6 μm n^+nn^+ diode at 0.8V shown in the preceding figures at z=0.3 μm and z = 0.44 μm.

Chapter 4

Boltzmann Equation Solution– II

4.1 Introduction

The main goal of this chapter is to extend the analysis of the previous chapter to multiple real space dimensions. It may seem at first glance that this would entail minimal theoretical development and primarily involve computational issues. Of course computational issues become more significant as the complexity of the problem increases but some of the underlying theory needs to be revisited when one attempts to extend the solution of the Boltzmann equation to two or more real space dimensions. This is necessitated by the fact that in one dimension the electric field direction can always be chosen as the polar direction from which the spherical angles are defined. For higher dimensions this requirement can still be met but at the cost of rotating the coordinate system to keep the polar direction aligned with the changing electric field direction which itself is not known *a priori*. Alternately one could use a fixed coordinate system everywhere for the momentum space expansion, but the consequences of this choice too need to be explored and understood. We shall therefore in the next section begin by discussing the relative merits of these two fundamental choices. In section 3, the Boltzmann equation solution will be reformulated using the rotated coordinate system while section 4 will discuss the numerical implementation issues. Simulation results using this approach will be presented in section 5 and finally a summary section will recapitulate the findings of this chapter.

4.2 Coordinate System Choice

Just as in the one-dimensional case, we begin by first expanding the distribution function in spherical harmonics in momentum space:

$$f(\mathbf{r},\mathbf{k}) = \sum_{lm} f_{lm}(\mathbf{r},k) Y_{lm}(\theta,\phi) \tag{4.1}$$

where the $Y_{lm}(\theta,\phi)$ are spherical harmonics. One key distinction between the one-dimensional case and the higher dimensional problem is that in the former the electric field can always be

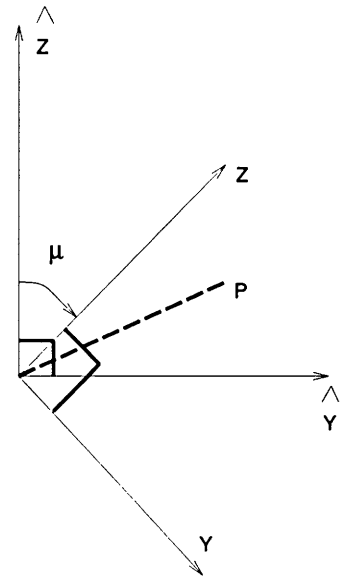

Figure 4-1: Rotated coordinate system

chosen to be one of the axes, and in our convention we chose it to be the z-axis. Thus the angle θ that is used in the spherical harmonics is uniquely determined by this choice. In higher dimensions the electric field, of course, is not always aligned to any fixed set of coordinate axes. There is, therefore, no privileged direction to use as the polar direction for the spherical coordinate system and we can arbitrarily choose any direction to be the fixed pole direction from which the angles θ and ϕ are measured. The disadvantage of this approach is that we would not be exploiting the approximate symmetry of the distribution function about the electric field direction. As the electric field varies from the fixed polar direction, the distribution function will not have much symmetry about the polar axis and all the harmonic components would be non-zero and in fact their magnitudes for a given order will be commensurate.

An alternate method is to guarantee that even in two dimensions the polar direction is always aligned with the electric field, but this can only be achieved at the cost of having a coordinate system (in k-space) that is rotated at each real space point from a fixed coordinate system. Thus, in two dimensions, if the electric field is assumed to always lie in the yz plane, say, then we need to rotate the coordinate axes about the x-axis. Hence at each point in real space, the z and y axes must be rotated by an angle μ from some "global" $(k, \hat{\theta}, \hat{\phi})$ coordinate system to form a "local" coordinate system (k, θ, ϕ) as shown in Fig. 4-1.

The advantage of imposing this requirement is that now we can expect the distribution function to have some degree of symmetry about the polar direction, which implies that the magnitude of the spherical harmonics that account for asymmetry will be small (for each order of expansion). And in fact we would need to do higher order expansions only for the spherical harmonics aligned with the polar direction as the asymmetric contribution could be neglected. For example, for expansions up to third order this approach would reduce the number of unknown coefficients at each real space point from 10 to 5. Clearly the approach using a rotated

coordinate system would reduce the computational cost compared to a fixed coordinate system method but the loss in accuracy (as we neglect the small "off-diagonal" harmonics) should be small if this method is to be useful. One way of estimating the effectiveness of this technique is to consider a known distribution function whose asymmetry about the z-axis can be varied, and calculate the spherical harmonic coefficients for such a series of functions. If the function has perfect symmetry about the z-axis then all "off-diagonal" harmonics which involve $\sin\phi$ or $\cos\phi$ would be zero, but as the degree of symmetry is reduced, the off-diagonal coefficients would become significant if the coordinate system is kept fixed.

The results of such an experiment are shown in Figs. 4-2 and 4-3. For simplicity and some physical realism we chose a displaced Maxwellian as the known distribution function (as was done in the previous chapter), with the direction of displacement from the z-axis being varied to generate a family of functions. This was done for two values of the magnitude of the displacement, a smaller one for Fig. 4-2 and larger one for Fig. 4-3. It is evident from the figures that for a displacement vector at an angle of 45° from the z-axis essentially all the harmonics are non-zero, whereas at the lowest angle considered (5°) the off-diagonal harmonics are a factor of four or five times smaller than the diagonal entries at each order of expansion.[1] Thus neglecting the off-diagonal entries in those cases should reduce the accuracy of the calculation only minimally. Therefore, as long as the distribution function has a high degree of symmetry about the field direction then the rotation of the coordinate system to align with the field should allow us to neglect the off-diagonal harmonics without incurring a high penalty in accuracy.

To take advantage of the rotated coordinate system approach we first need to establish a relationship between the spherical harmonics in the two cases. The relationship between the local variables, rotated by an angle μ, and the global variables, up to first order, can be derived from elementary geometry and is given by,

$$\begin{aligned} \cos\hat{\theta} &= \cos\theta\cos\mu - \sin\theta\sin\phi\sin\mu \\ \sin\hat{\theta}\sin\hat{\phi} &= \cos\theta\sin\mu + \sin\theta\sin\phi\cos\mu \\ \sin\hat{\theta}\cos\hat{\phi} &= \sin\theta\cos\phi \end{aligned} \tag{4.2}$$

which can be written more compactly in matrix form as:

$$\begin{bmatrix} \sin\hat{\theta}\cos\hat{\phi} \\ \sin\hat{\theta}\sin\hat{\phi} \\ \cos\hat{\theta} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\mu & \sin\mu \\ 0 & -\sin\mu & \cos\mu \end{bmatrix} \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}. \tag{4.3}$$

In the above equation the global harmonics are written as a linear combinations of the local harmonics, but in fact the converse is more useful. This can be done easily, as the matrix above

[1] Note that the coefficients of the harmonics $Y_{1,1}, Y_{2,1}, Y_{2,-2}$ in these expansions are always zero because of the two-dimensional nature of the problem. This will be proven more generally in appendix D.

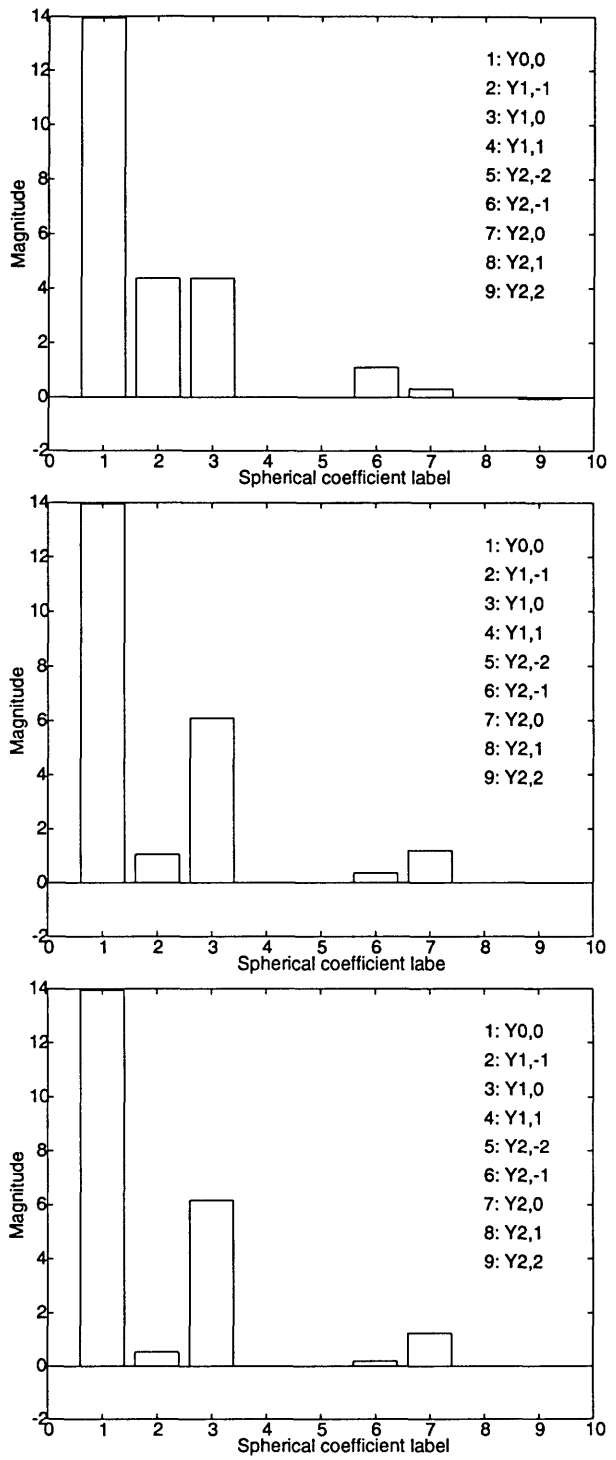

Figure 4-2: Coefficients of the spherical harmonics up to second order for a Maxwellian distribution displaced from the origin along directions 45°, 10°, and 5° from the origin.

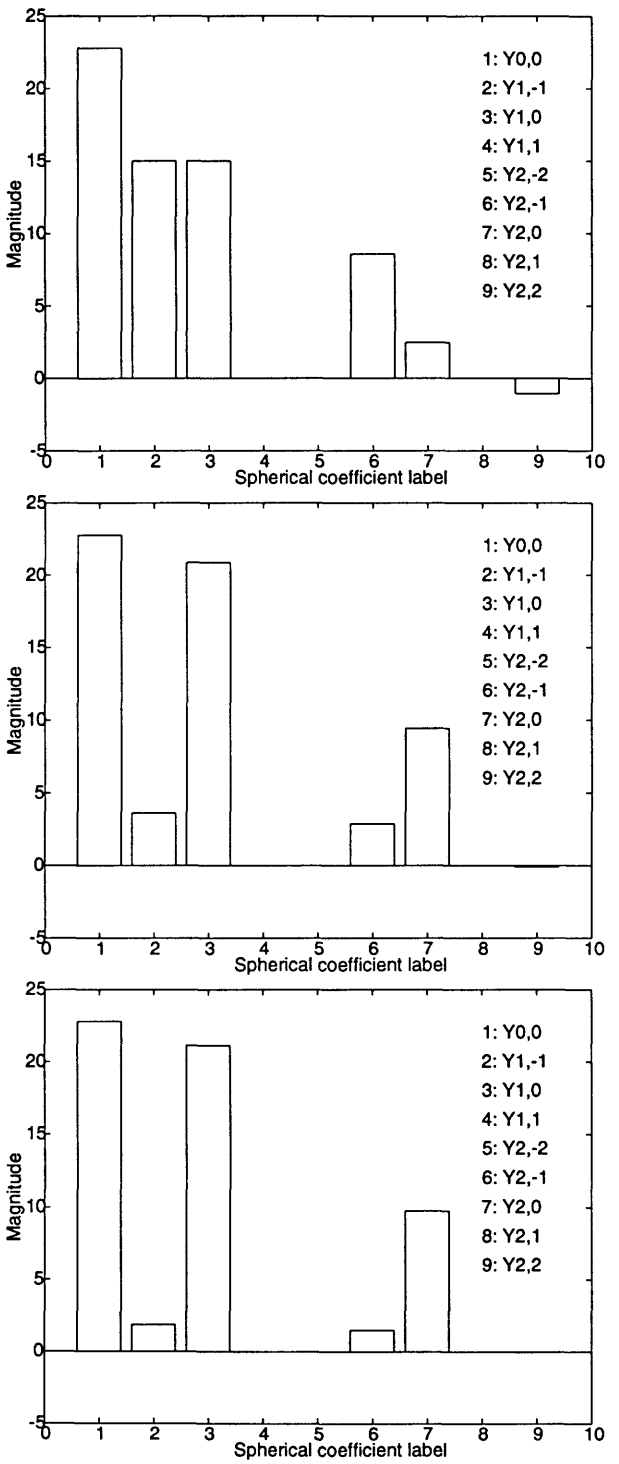

Figure 4-3: Coefficients of the spherical harmonics up to second order for a Maxwellian distribution displaced from the origin along directions 45°, 10°, and 5° from the origin.

is invertible:

$$\begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\mu & -\sin\mu \\ 0 & \sin\mu & \cos\mu \end{bmatrix} \begin{bmatrix} \sin\hat{\theta}\cos\hat{\phi} \\ \sin\hat{\theta}\sin\hat{\phi} \\ \cos\hat{\theta} \end{bmatrix}. \tag{4.4}$$

Here, the rotated harmonics are written as a linear combination of the "global" harmonics, which in matrix terms amounts to multiplication by a matrix. If we consider the expansion of the distribution function in spherical harmonics as a linear combination of the basis functions (spherical harmonics) then we can write the expansion as an inner product:

$$\sum_{lm} f_{lm}(\mathrm{r},k)Y_{lm}(\theta,\phi) = \begin{bmatrix} Y_{0,0} & Y_{1,-1} & Y_{1,0} & Y_{1,1} & \dots \end{bmatrix} \begin{bmatrix} f_{0,0} \\ f_{1,-1} \\ f_{1,0} \\ f_{1,1} \\ \vdots \end{bmatrix} = Y^T f_{\text{local}}. \tag{4.5}$$

If we wish to write the above expansion in global spherical harmonics, that is in terms of $\hat{Y}_{0,0}$, $\hat{Y}_{1,-1}$ etc., then we use the relationship already established between the rotated and unrotated harmonics:

$$\begin{aligned} f &= \sum_{lm} f_{lm}(\mathrm{r},k)Y_{lm}(\theta,\phi) = Y^T f_{\text{local}} \\ &= \begin{bmatrix} \hat{Y}_{0,0} & \hat{Y}_{1,-1} & \hat{Y}_{1,0} & \hat{Y}_{1,1} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\mu & \sin\mu & 0 \\ 0 & -\sin\mu & \cos\mu & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_{0,0} \\ f_{1,-1} \\ f_{1,0} \\ f_{1,1} \end{bmatrix} \\ &= \hat{Y}^T R f_{\text{local}}. \end{aligned} \tag{4.6}$$

Note that the rotation matrix, R, reduces to an identity matrix when $\mu = 0$, i.e. when there is no rotation, which is of course a necessary condition. It can also be shown that this can be done for higher order harmonics, though the rotation matrix is more complex.

4.3 Formulation Using A Rotated Coordinate System

The approach we will follow for the solution of the two-dimensional problem is summarized below:

1. Expand the distribution function in local coordinates.
2. Write the drift, diffusion and the scattering terms of the BTE in local coordinates.
3. Convert to global coordinates by multiplying by the rotation matrix.
4. Discretize in space and energy.

5. Multiply by conjugate harmonics and integrate over a unit sphere in k-space to form the coefficient matrix equation.

6. Solve the matrix problem to obtain the coefficients at all points in real space.

This approach only differs in Step 3 from that followed in the one-dimensional problem. In the following sections we will expand upon the above steps for each of the drift, diffusion and scattering terms in the BTE.

4.3.1 The Diffusion Term

The diffusion term of the Boltzmann equation is

$$\mathbf{v}(\mathbf{k}) \cdot \nabla_r f(\mathbf{r}, \mathbf{k}),$$

which after substituting the spherical harmonic expansion in local coordinates becomes

$$\mathbf{v}(\mathbf{k}) \cdot \nabla_r \sum_{lm} f_{lm}(\mathbf{r}, k) Y_{lm}(\theta, \phi; \mu) = \mathbf{v}(\mathbf{k}) \cdot \nabla_r (Y^T f).$$

If we assume spherical bands then $\mathbf{v}(\mathbf{k})$ is in the radial k direction and its projection onto the z-axis is $v(k)\cos\theta$ and onto the y-axis is $v(k)\sin\theta\sin\phi$. Hence, without any rotation of the coordinate system, the diffusion term would be

$$v(k)\left(Y^T \frac{\partial f}{\partial z}\cos\theta + Y^T \frac{\partial f}{\partial y}\sin\theta\sin\phi\right)$$

assuming no variation in the x-direction.

To generate the coefficient matrix we multiply this term by the orthogonal complement of the spherical harmonics and integrate over the unit sphere in k-space:

$$\begin{aligned} & v(k)\int d\Omega\ Y^* \left(Y^T \frac{\partial f}{\partial z}\cos\theta + Y^T \frac{\partial f}{\partial y}\sin\theta\sin\phi\right) \\ = \ & v(k)\int d\Omega\ Y^* Y^T \cos\theta \frac{\partial f}{\partial z} + v(k)\int d\Omega\ Y^* Y^T \sin\theta\sin\phi \frac{\partial f}{\partial y}. \end{aligned} \tag{4.7}$$

Note that Y^*Y^T is an $n \times n$ matrix where n is the order of the spherical harmonic expansion. The first few terms of the coefficient matrix are shown below:

$$\begin{bmatrix} \int d\Omega\ Y^*_{0,0} Y_{0,0}\cos\theta & \int d\Omega\ Y^*_{0,0} Y_{1,0}\cos\theta & \cdots \\ \int d\Omega\ Y^*_{1,0} Y_{0,0}\cos\theta & \int d\Omega\ Y^*_{1,0} Y_{1,0}\cos\theta & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

The above expansion would be correct if there were no rotations, but in fact we need to convert the local spherical harmonics into global harmonics before we do the dot product with

the velocity. As described in the previous section this is just equivalent to a multiplication by the rotation matrix. Thus, the expansion in global coordinates is:

$$v(k)\left(\hat{Y}^T\frac{\partial(Rf)}{\partial z}\cos\theta+\hat{Y}^T\frac{\partial(Rf)}{\partial y}\sin\theta\sin\phi\right).$$

The coefficient matrix can be generated by multiplication by orthogonal harmonics and integration over the unit sphere in k-space:

$$v(k)\left(\int d\Omega\,\hat{Y}^*\hat{Y}^T\cos\theta\right)\frac{\partial(Rf)}{\partial z}+v(k)\left(\int d\Omega\,\hat{Y}^*\hat{Y}^T\sin\theta\sin\phi\right)\frac{\partial(Rf)}{\partial y} \tag{4.8}$$

It should be remembered that the rotation matrix, R, in the above expression is a function of real space and in taking spatial derivatives we should be cognizant of this fact.

Finally the derivatives in space and energy must be discretized to generate a matrix problem for the coefficients in space and energy for any desired order. The discretized version of the diffusion term is:

$$v(E)\left[\frac{1}{\Delta z}\mathcal{G}_z(R^{i+1}f^{i+1}-R^if^i)+\frac{1}{\Delta y}\mathcal{G}_y(R^{j+1}f^{j+1}-R^jf^j)\right]$$

where the script variables represent matrices (defined below) which form the non-zero blocks in the sparse coefficient matrix and i and j are the indices for the discretized z and y variables.

$$\mathcal{G}_z = \int d\Omega\;\hat{Y}^*\hat{Y}^T\cos\theta \tag{4.9}$$

$$\mathcal{G}_y = \int d\Omega\;\hat{Y}^*\hat{Y}^T\sin\theta\sin\phi \tag{4.10}$$

In the above discretization we approximated the spatial derivatives for the product Rf rather than expand it out because the former approach assures current conservation whereas the discretization of the expanded terms would entail some error in current conservation.

4.3.2 The Drift Term

The drift term after expansion in local spherical harmonics is

$$\frac{q\mathcal{E}}{\hbar}\cdot\nabla_k\sum_{lm}f_{lm}(\mathbf{r},k)Y_{lm}(\theta,\phi;\mu),$$

which can be written in the more compact vector notation after expanding in global harmonics as

$$\frac{q\mathcal{E}}{\hbar}\cdot\nabla_k(Y^Tf)=\frac{q\mathcal{E}}{\hbar}\cdot\nabla_{\hat{k}}(\hat{Y}^TRf).$$

The gradient in momentum space can be written explicitly in spherical coordinates:

$$\frac{\partial}{\partial\hat{k}}\left(\hat{Y}^TRf\right)\mathbf{i}_\mathbf{k}+\frac{1}{\hat{k}}\frac{\partial}{\partial\hat{\theta}}\left(\hat{Y}^TRf\right)\mathbf{i}_\theta+\frac{1}{\hat{k}\sin\hat{\theta}}\frac{\partial}{\partial\hat{\phi}}\left(\hat{Y}^TRf\right)\mathbf{i}_\phi, \tag{4.11}$$

and taking the dot product of the gradient with the electric field yields:

$$\frac{q\mathcal{E}_z}{\hbar}\left[\cos\theta\frac{\partial}{\partial k}\left(Y^T Rf\right) - \frac{1}{k}\sin\theta\frac{\partial}{\partial\theta}\left(Y^T Rf\right)\right] \tag{4.12}$$
$$+ \frac{q\mathcal{E}_y}{\hbar}\left[\sin\theta\sin\phi\frac{\partial}{\partial k}\left(Y^T Rf\right) + \frac{1}{k}\cos\theta\sin\phi\frac{\partial}{\partial\theta}\left(Y^T Rf\right) + \cos\phi\frac{1}{k\sin\theta}\frac{\partial}{\partial\phi}\left(Y^T Rf\right)\right]$$

where we have taken off our hats as a courtesy to the reader. The last step in the derivation of the coefficient matrix is to multiply by the conjugate harmonics and integrate over the unit sphere. For the sake of brevity the detailed steps will be shown only for the terms multiplied by $\mathcal{E}_z$.

If we now make a change of variable from the magnitude of the k vector to energy by using the relation, $v(k) = \frac{1}{\hbar}\frac{\partial E}{\partial k}$, as was done in the one-dimensional case, we obtain:

$$v(E)q\mathcal{E}_z\left[\left(Y^T R\cos\theta\right)\frac{\partial f(z,E)}{\partial E} - \frac{1}{2E}\left(\frac{\partial Y^T}{\partial\theta}R\sin\theta\right)f\right].$$

To generate the coefficient matrix, we multiply the above expression by the conjugate harmonics and integrate over the unit sphere in **k**-space:

$$v(E)q\mathcal{E}_z\left[\left(\int d\Omega\; Y^* Y^T R\cos\theta\right)\frac{\partial f(z,E)}{\partial E} - \frac{1}{2E}\left(\int d\Omega\; Y^*\frac{\partial Y^T}{\partial\theta}R\sin\theta\right)f\right].$$

Upon discretization this can be written as:

$$v(E)q\mathcal{E}_z\left[\frac{1}{\Delta E}\mathcal{G}_E^z R(f^{l+1} - f^l) - \frac{1}{2E}\mathcal{H}_E^z Rf\right]$$

where l is the index for the discretized energy variable, $\mathcal{G}_E^z$ and $\mathcal{H}_E^z$ are defined below:

$$\mathcal{G}_E^z = \int d\Omega\; Y^* Y^T\cos\theta \tag{4.13}$$

$$\mathcal{H}_E^z = \int d\Omega\; Y^*\frac{\partial Y^T}{\partial\theta}\sin\theta. \tag{4.14}$$

Finally, it should be remembered that the z component of the electric field $\mathcal{E}_z$ depends on the angle of rotation through: $\mathcal{E}_z = |\mathcal{E}|\cos\mu$. For the $\mathcal{E}_y$ term we get:

$$v(E)q\mathcal{E}_y\left[\frac{1}{\Delta E}\mathcal{G}_E^y R(f^{l+1} - f^l) - \frac{1}{2E}\mathcal{H}_E^y Rf\right]$$

where $\mathcal{G}_E^y$ and $\mathcal{H}_E^y$ are defined below:

$$\mathcal{G}_E^y = \int d\Omega\; Y^* Y^T\sin\theta\sin\phi \tag{4.15}$$

$$\mathcal{H}_E^y = -\left(\int d\Omega\; Y^*\frac{\partial Y^T}{\partial\theta}\cos\theta\sin\phi + \int d\Omega\; Y^*\frac{\partial Y^T}{\partial\phi}\frac{1}{\sin\theta}\cos\phi\right). \tag{4.16}$$

Although there is a $\sin\theta$ in the denominator in the expression for $\mathcal{H}_E^y$, that should not be a source of concern becuase it is automatically removed when multiplied by $d\Omega = \sin\theta d\theta d\phi$.

4.3.3 The Complete Equation

Thus the complete left hand side of the Boltzmann equation using the compact notation introduced in the previous section is:

$$\begin{aligned} v(E) \quad & \left[\mathcal{G}_z\left(\frac{R^{i+1}f^{i+1}-R^{i}f^{i}}{\Delta z}\right)+\mathcal{G}_y\left(\frac{R^{j+1}f^{j+1}-R^{j}f^{j}}{\Delta y}\right)\right. \\ & -q\mathcal{E}_z\left(\mathcal{G}_E^z R\frac{\left(f^{l+1}-f^{l}\right)}{\Delta E}-\frac{1}{2E}\mathcal{H}_E^z Rf\right) \\ & \left.-q\mathcal{E}_y\left(\mathcal{G}_E^y R\frac{\left(f^{l+1}-f^{l}\right)}{\Delta E}-\frac{1}{2E}\mathcal{H}_E^y Rf\right)\right]=\mathcal{S}(Rf). \end{aligned} \tag{4.17}$$

4.3.4 The Scattering Term

We will consider acoustic and optical phonon and ionized impurity scattering; and for all three cases assume a single spherical band for simplicity, as was done in the one-dimensional case. Because the formulation of the scattering term is essentially independent of the spatial variation, the results obtained in two real space dimensions are almost the same as in one dimension with one exception which arises because of our use of a rotated coordinate system. As has been demonstrated in the case of the drift and diffusion terms, in each case we need to multiply the distribution function by the rotation matrix and this also holds true in the case of the scattering operator. Thus we shall simply state the final result for each type of scattering for an expansion to first order without an elaborate derivation. The generalization to higher orders is obvious from the form of the scattering matrices.

Acoustic Phonon Scattering

After including the effect of rotation, the acoustic phonon scattering term is

$$c_{ac}g(E)Y^T S_{ac}\,R\,f$$

where S_{ac} is given by:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{4.18}$$

To generate the coefficient matrix by we multiply by the conjugate harmonics and integrate over the unit sphere in k-space:

$$\int d\Omega\; Y^* Y^T S_{ac}\,R\,f.$$

Optical Phonon Scattering

The scattering term for optical phonons after including the effects of rotation is

$$Y^T c_{op} \left[- \left(N_{op} g^+ + N_{op}^+ g^- \right) S_{op} R f^k + N_{op} g^- S_{op}^- R f^{k-l} + N_{op}^+ g^+ S_{op}^+ R f^{k+l} \right]$$

where S_{op}, S_{op}^- and S_{op}^+ are coupling matrices given below:

$$S_{op} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad S_{op}^- = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad S_{op}^+ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The coefficient matrix can then be generated directly.

Ionized Impurity Scattering

The scattering term for ionized impurity scattering is:

$$\int d\Omega \; Y^* Y^T \, S_{BH} \, Rf$$

where S_{BH} is the net scattering matrix for the Brooks-Herring model, as shown below, up to first order

$$S_{BH} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -\mathrm{sBH}_0 & 0 & 0 \\ 0 & 0 & \mathrm{sBH}_1 - \mathrm{sBH}_0 & 0 \\ 0 & 0 & 0 & -\mathrm{sBH}_0 \end{bmatrix} \tag{4.19}$$

and the expansion coefficients, $\mathrm{sBH}_0, \mathrm{sBH}_1 \ldots$ are given in Appendix C.

4.4 Numerics

4.4.1 Symmetry of the Two Dimensional Problem

In one real space dimension we argued that because of symmetry of the distribution function around the electric field direction the distribution could be represented without using the ϕ dependent harmonics. In other words only the so called Legendre polynomials $Y_{0,0}, Y_{1,0}, Y_{2,0} \ldots$ were needed. In two space dimensions this condition does not hold; but even now there is symmetry as the distribution function has no variation in the k_x direction, when the electric field lies in the yz plane. This implies that the distribution function will have symmetry about the $k_x = 0$ plane. Due to this symmetry only half of the spherical harmonic coefficients are non-zero, more specifically all coefficients of harmonics of the form $\cos m\phi$ where m is odd are zero and all coefficients of harmonics of the form $\sin m\phi$ where m is even are zero. This result is proven in Appendix D.

4.4.2 Boundary conditions

In section 3 of this chapter we developed a method to generate a matrix for the coefficients of the spherical harmonic expansion of the Boltzmann equation using a rotated momentum space coordinate system. One aspect of the problem which was not discussed is the nature of the boundary conditions that must be imposed to obtain a physically reasonable solution. To help do this we begin by writing explicitly the expansion to first order under a rotated formulation:

$$\frac{\partial f_{1,y}}{\partial y} + \frac{\partial f_{1,z}}{\partial z} - q\mathcal{E}_y\left[\frac{\partial f_{1,y}}{\partial E} + \frac{f_{1,y}}{E}\right] - q\mathcal{E}_z\left[\frac{\partial f_{1,z}}{\partial E} + \frac{f_{1,z}}{E}\right] = \frac{\sqrt{3}}{v(E)}\left(\frac{\partial f_{0,0}}{\partial t}\right)_c \quad (4.20)$$

$$\frac{\partial f_{0,0}}{\partial y} - q\mathcal{E}_y\frac{\partial f_{0,0}}{\partial E} = -\frac{\sqrt{3}}{v(E)\tau(E)}f_{1,y} \quad (4.21)$$

$$\frac{\partial f_{0,0}}{\partial z} - q\mathcal{E}_z\frac{\partial f_{0,0}}{\partial E} = -\frac{\sqrt{3}}{v(E)\tau(E)}f_{1,z} \quad (4.22)$$

where the $\tau(E)$ is the scattering rate which includes both acoustic and optical phonon scattering; the scattering rate in the first equation cannot be written simply as a function of the local energy as it couples higher and lower energies as was discussed in the one-dimensional case. The expressions $f_{1,z}$ and $f_{1,y}$ are composite spherical harmonics coefficients arising because of the rotation and are defined by:

$$f_{1,z} = \cos\mu f_{1,0} - \sin\mu f_{1,-1} \quad (4.23)$$

$$f_{1,y} = \sin\mu f_{1,0} + \cos\mu f_{1,-1} \quad (4.24)$$

A number of observations can be made about the system of equations (4.20)-(4.24). The first observation is that if $\mu = 0$ the system becomes completely symmetric in $f_{1,-1}$ and $f_{1,0}$ the spherical harmonics corresponding to variation in the y and z axis. Second, the current in the y and z direction is given by the integral over energy of the right hand sides of (4.21) and (4.22) respectively. This also means that the zeroth order equation, (4.20), is the current conservation equation similar to the case in one-dimension.

In general for a first order partial differential equation system in three variables and two dimensions we would need to specify the values of all three unknown functions at two of the boundaries as illustrated in Fig. 4-4(a). If we specify one of the functions, say, $f_{0,0}$ at each of the two boundaries in z and y then we need to specify only one of the other functions at that boundary as shown in Fig. 4-4(b). For our particular set of PDE's this condition is relaxed even more because when $f_{0,0}$ is specified all along the the left boundary, say, then its derivative in y is also known along that boundary and therefore the left hand side of (4.21) is known completely which just gives an algebraic equation for $f_{1,-1}$ and $f_{1,0}$, which is equivalent to the other boundary condition that needed to be specified there. Thus specifiying $f_{0,0}$ at both the z boundaries is a sufficient boundary condition for the left and right edges. A similar argument holds for the boundary conditions at the top and bottom boundaries. Hence, if we specify $f_{0,0}$ at all four boundaries we should be able to solve the above system of equations.

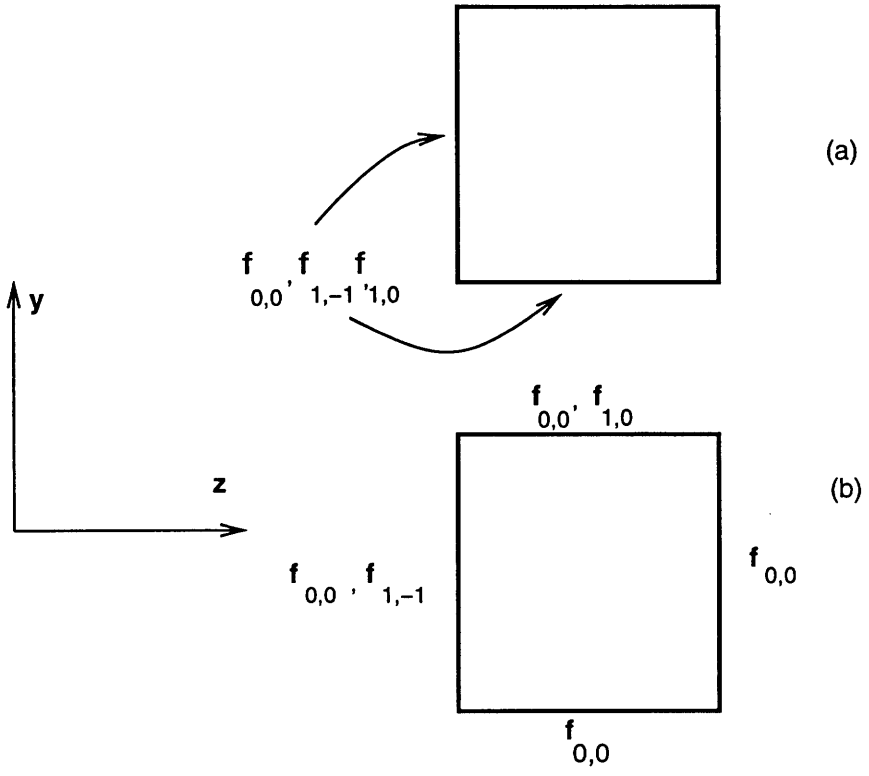

Figure 4-4: Boundary conditions for the two-dimensional solution.

Note that the number of boundary conditions imposed by this argument is two less than the number mentioned earlier and this reduction arises from the special structure of the equations in our system: there is only one spatial derivative in each of (4.21) and (4.22). Physically this reduction arises because the current flow in each direction is dependent on the field and the derivative of the spherical harmonic coefficient in that direction only.

This new set of boundary conditions is almost where we want to be but not quite, because it still requires the specification of the isotropic part of the distribution function $f_{0,0}$, at all four boundaries. This is approximately equivalent to requiring Ohmic contacts on all sides, which is not always the physical situation. To be able to simulate insulating and reflecting boundary conditions, we need a way of specifying zero normal current boundary conditions. Fortunately this too can be done in a fairly straightforward manner by utilizing the fact that the integral of the right hand side of (4.21) and (4.22) is the current in the y and z directions respectively. Thus instead of specifiying $f_{0,0}$ at the boundary for all energy, we can specify the integrand of the current to be zero and let $f_{0,0}$ be an unknown to be solved for. This will of course ensure that the normal current is zero and still provide the requisite number of boundary conditions. Finally it should be mentioned that the boundary conditions in the energy direction are the same as in the one-dimensional case i.e. beyond some maximum energy all coefficients are assumed to be zero and the odd harmonic coefficients are assumed to be zero at the origin in energy.

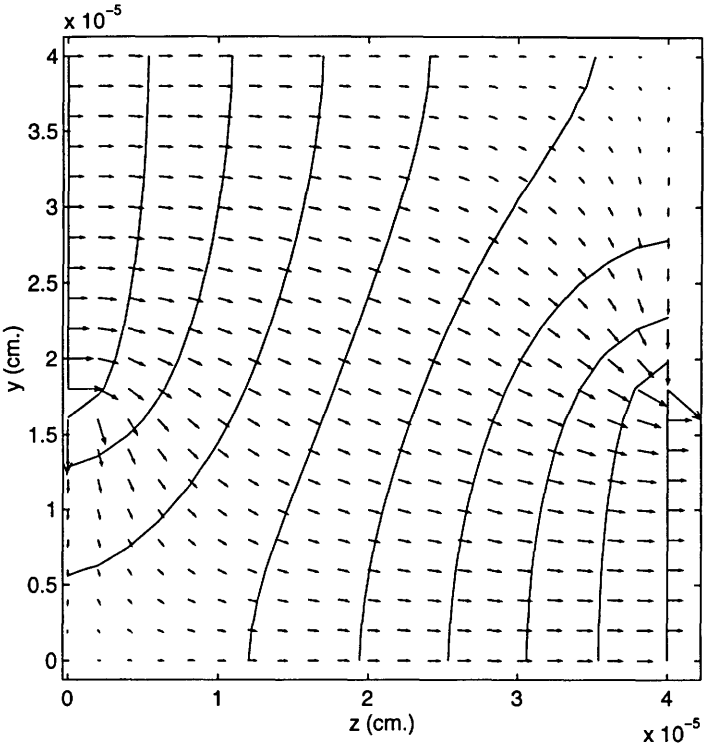

Figure 4-5: Current flow and equipotentials in a resistive structure which has contacts at the bottom right and top left sides.

An example of the implementation of this boundary condition approach is shown in Fig. 4-5, for a purely resistive structure in which the right and left sides are half contact and half insulator. Thus the current must either be normal or tangential to the boundaries and this condition is clearly met using this discretization as shown in Fig. 4-6.

4.4.3 Self-consistent solution with Poisson's equation

In principle we can obtain a self-consistent solution to the Boltzmann equation by solving both the Boltzmann and Poisson equations simultaneously as was done in the one dimensional case. But in two dimensions we adopt a different strategy: we use an iterative method that alternatingly solves the Boltzmann equation and Poisson's equation and uses the solution of one as an input to the other. The reason for this choice was due to the complexity of calculating the Jacobian matrix that is needed in the coupled solution of the non-linear Boltzmann/Poisson system. The Jacobian is significantly harder to calculate in two dimensions than in one dimension because the electric field and thus the rotation matrices are functions of the potential and need to be correctly treated whereas in one dimension this issue did not arise.

Formally the discretized Boltzmann equation can be written as:

$$\mathcal{B}(f_{lm}^{i,j,l}, \psi^{i,j}) = 0 \tag{4.25}$$

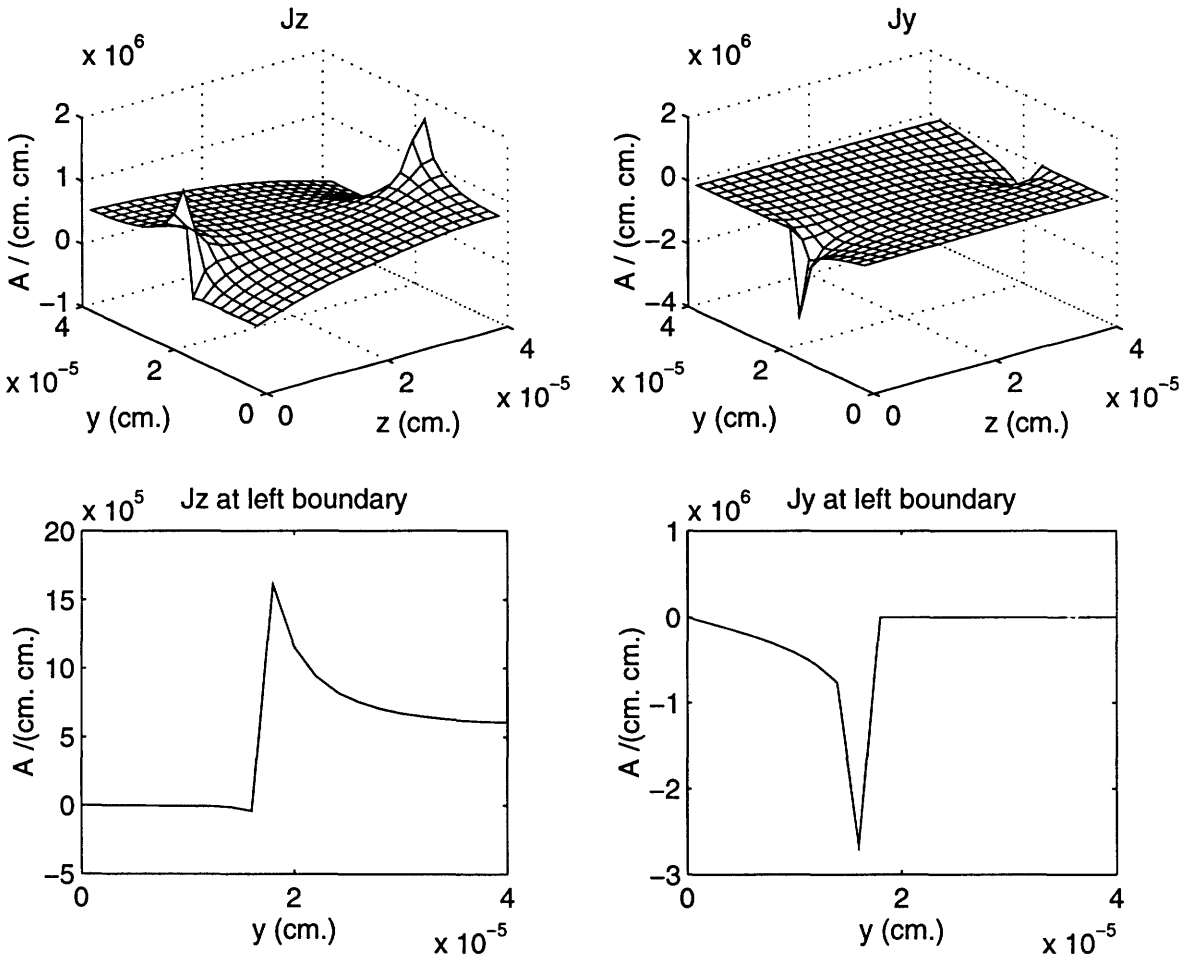

Figure 4-6: Perspective plots of the current for the resistive structure (top) and the currents in J_z and J_y directions at the left edge (bottom).

and the discretized Poisson's equation as:

$$\mathcal{P}(n^{i,j}, \psi^{i,j}) = 0 \tag{4.26}$$

where $n^{i,j}$ is the electron concentration calculated by integrating the distribution function over energy or performing the discrete approximation to the integral for the discretized distribution function, $n^{i,j} = \sum_l f_{0,0}^{i,j,l} \Delta E$. Thus the iteration proceeds by first guessing the potential everywhere and then solving the Boltzmann equation (4.25) to obtain the distribution function which is then used to calculate the electron concentration, which in turn is used to solve the Poisson equation (4.26). The last step of the iteration is to use the potential calculated from Poisson's equation as the new corrected value of the potential to be used in the solution of the Boltzmann equation. One important detail that must be mentioned here is that the iteration (or relaxation) only converges when we solve the non-linear Poisson equation

$$\nabla^2 \psi = -\frac{q}{\epsilon_{Si}} \left(p + n_C e^{\psi / V_{th}} + N_D - N_A \right) \tag{4.27}$$

which is the equation obtained by assuming no current flow, where $n_C = n \exp[-\psi^0 / V_{th}]$ and V_{th} is the thermal voltage and ψ^0 is the initial guess of the potential. The above equation yields the correct potential in equilibrium when there exists non-uniform doping in the device. An example of the iteration process is shown in Fig. 4-7, where a bias is applied on the right contact and the solution converges to the correct answer in about ten iterations.

It is of course possible to use a drift-diffusion or hydrodynamic simulation to get the approximate potential solution and use that as the initial guess and then use the non-linear Poisson's equation to refine the solution. This would reduce somewhat the number of relaxation iterations needed. Finally, it should be mentioned that the above relaxation procedure may fail to converge in some cases, specially under high current conditions, as is the case when a relaxation method is used for the solution of the drift-diffusion and Poisson's equation [53].

4.4.4 Solution of the Matrix Equation

When solving the Boltzmann equation in one real space dimension we used a direct sparse matrix solution method to solve for the coefficients. This method is viable in the one-dimensional case where the problem is equivalent to a two-dimensional problem in space and energy and the number of unknowns is usually less than 10,000 for typical problems (for example assume 50 real space points and 50 energy space points and a third order solution: $10,000 = 50 \times 50 \times 4$). In two real space dimensions the number of unknowns is higher by an order of magnitude. Also the bandwidth of the sparse matrix is larger now by an order of magnitude, which also tends to slow down the sparse solver more than what would be expected from an increase in problem size alone. Hence, we chose to explore alternatives to the direct solver and the most attractive approaches are conjugate gradient methods such as GMRES, CGS, BiCG etc. [54] [55]. Rather than discuss the theoretical justifications of these

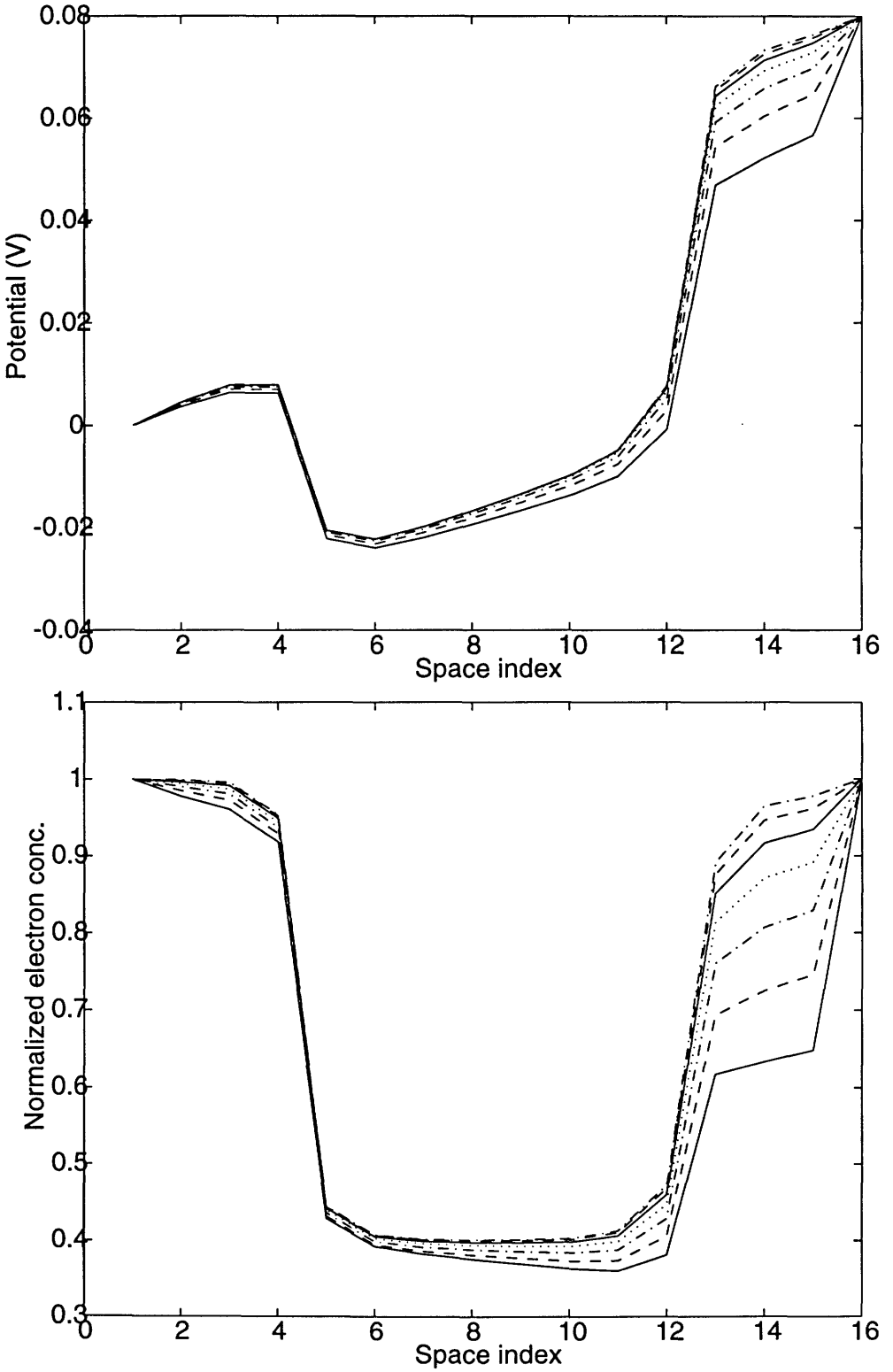

Figure 4-7: The normalized potential (top) and electron concentration (bottom) across the channel region of a two-dimensional device after each Gummel iteration.

techniques or their relative merits which are extensively discussed in the numerical analysis literature. Our approach will be purely empirical and treat the iterative solver as a black box which provides a fast matrix solution [56].

Obviously the one topic where we do have to wet our feet in numerical analysis is in obtaining an effective pre-conditioner for whichever iterative solver we use. The pre-conditioner which is an approximation to the inverse of the matrix but is much cheaper to factor is an input to the iterative solvers and is used by them to improve the convergence rate. Otherwise the iteration is only guaranteed to terminate in N iterations (for GMRES) where N is the order of the sparse matrix. If we order the coefficient matrix such that the lowest index corresponds to discretization in energy, the next index is discretization in z and the last in y then the coefficient matrix can be considered to be a banded matrix of blocks of coefficients as shown in Fig. 4-8. Each block represents the the coupling in energy of all the coefficients at a given point in space. The first band are the block coefficients of the neighboring points in z and the outer band are the blocks of the neighbors in y. We use the block diagonals as our choice for the pre-conditioner because they are relatively easy to factor and the block diagonal matrix would in fact give the correct solution if there was no current flow, that is if there was no coupling in space. The size of each block is the number of unknowns in energy times the number of coefficients so it is quite moderate and can be factored very efficiently.

Using this pre-conditioner we obtained fairly good convergence with both CGS and GMRES. Both methods converged in about 150 iterations (to a tolerance of 1.0×10^{-7}). As is known in the literature [57], CGS tends to be more efficient computationally but has erratic convergence. This was also our experience, as CGS would occasionally not converge at all. For fairly moderate sized problems of approximately 20,000 unknowns CGS would converge in a about three minutes on a SPARCStation-10, whereas a direct solve takes about three times as long.

4.5 Simulation results in two dimensions

4.5.1 Uniform Doping

Here we present results obtained using the rotated coordinate system approach discussed in this chapter. The first question that must be addressed is whether this technique works at all. A simple test is to see the effect of the rotation on a simple resistive structure, which is shown in Fig. 4-10. If we plot the coefficients as a function of real space for this problem (see Fig. 4-9) the effectiveness of the rotation method can be seen. Notice that the $f_{1,-1}$ coefficient is zero almost everywhere while that is not the case with $f_{0,0}$. This is true despite the fact that there is substantial current flow in both directions as shown in Fig. 4-10. The angle of rotation μ can be seen from Fig. 4-11. The higher order coefficients up to third order when only the diagonal coefficient that is $f_{2,0}$ and $f_{3,0}$ are retained are shown in Fig. 4-12. The rotation matrix for this

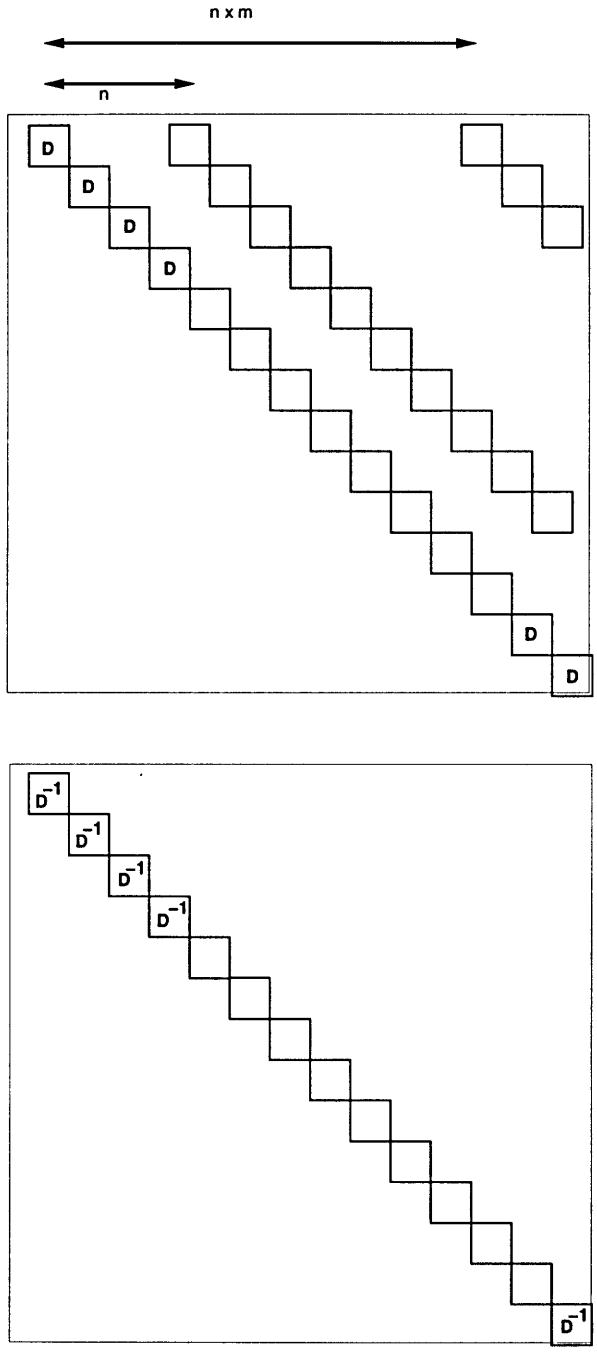

Figure 4-8: (a) The block structure of the coefficient matrix for the two dimensional real space problem. n is the number of mesh lines in the discretization in the z and m is the number of mesh lines in the discretization in the y dimension. (b) The pre-conditioner obtained by inverting only the diagonal blocks.

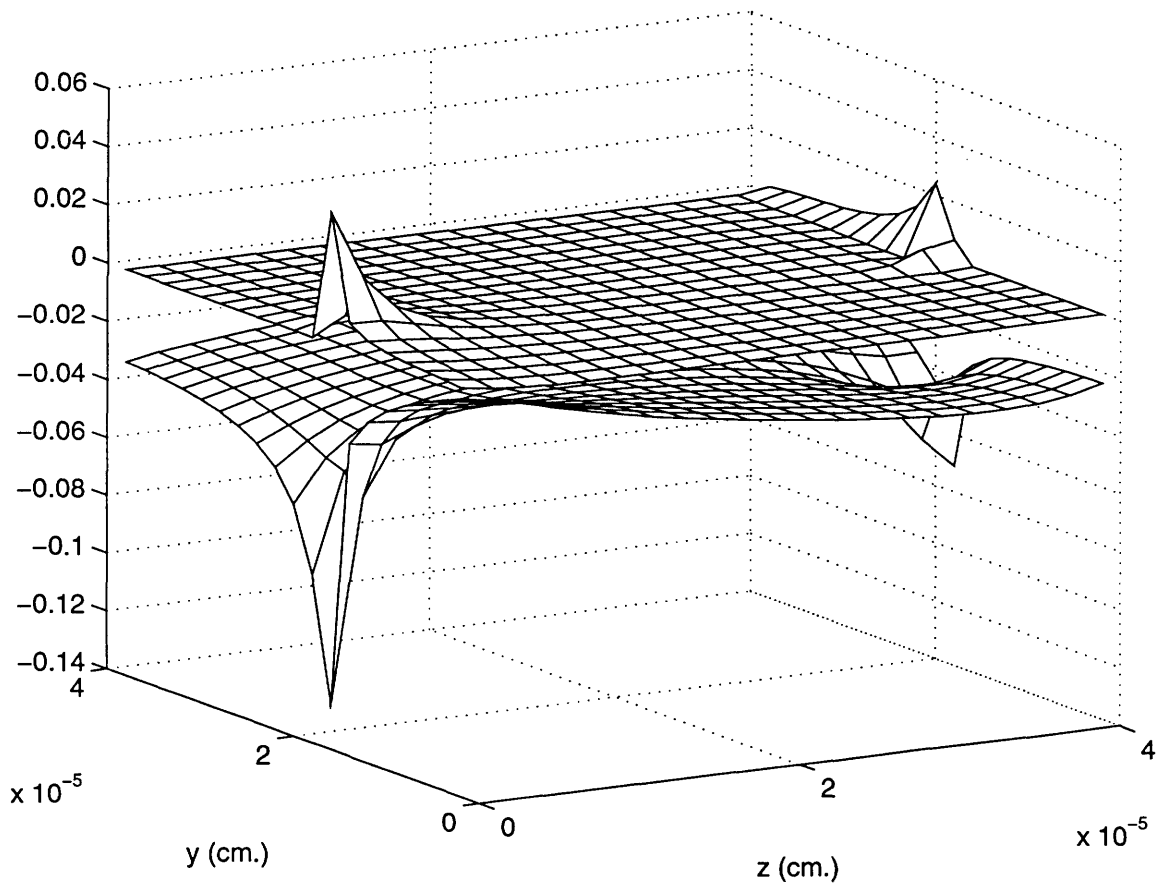

Figure 4-9: $f_{1,0}$ and $f_{1,-1}$ as a function of real space at an energy of 25 meV for the structure shown in Fig. 4.10.

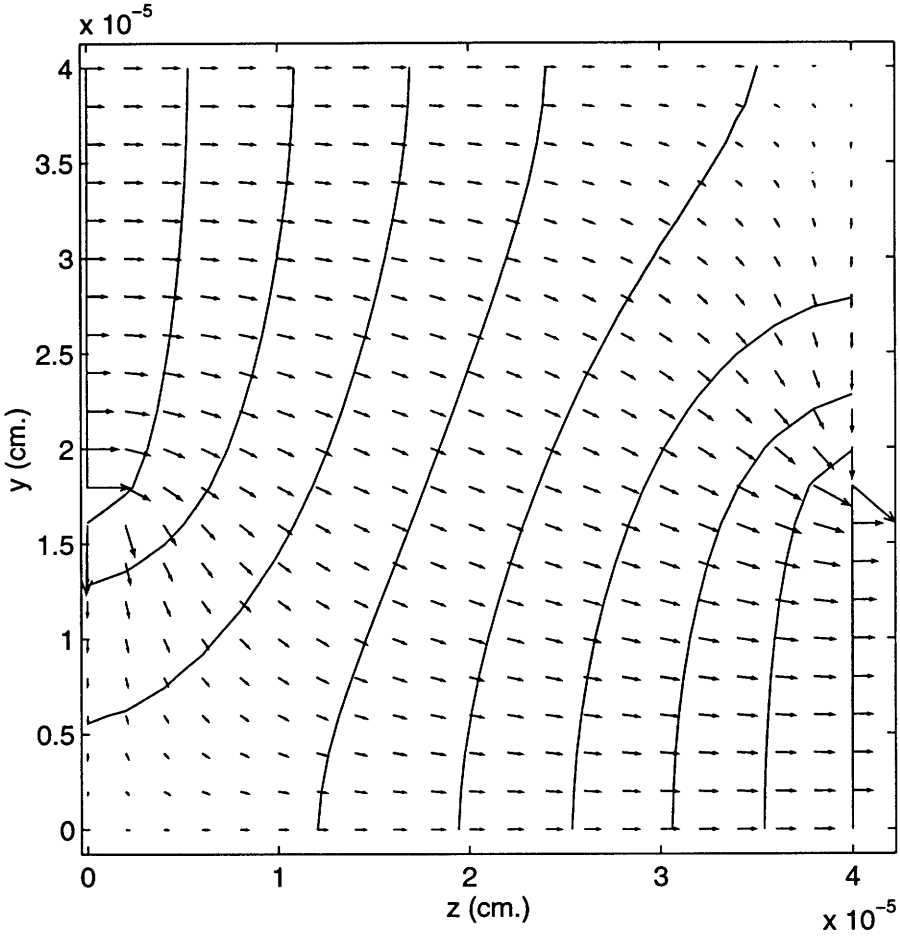

Figure 4-10: Current flow in a resistive structure which has contacts at the bottom right and top left sides.

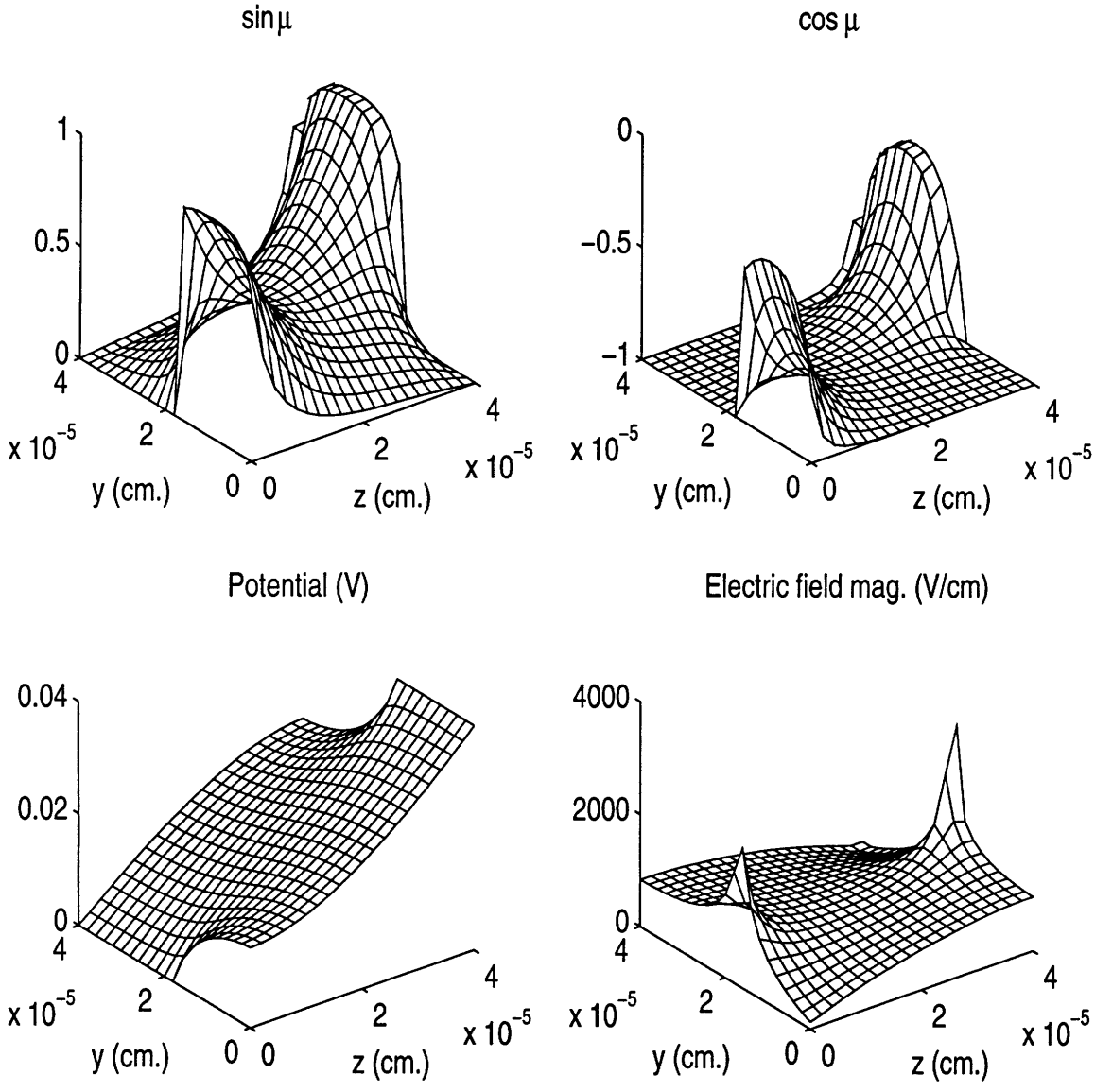

Figure 4-11: The sine and cosine of the rotation angles, the potential and electric field for the resistive structure shown in Fig. 4.10.

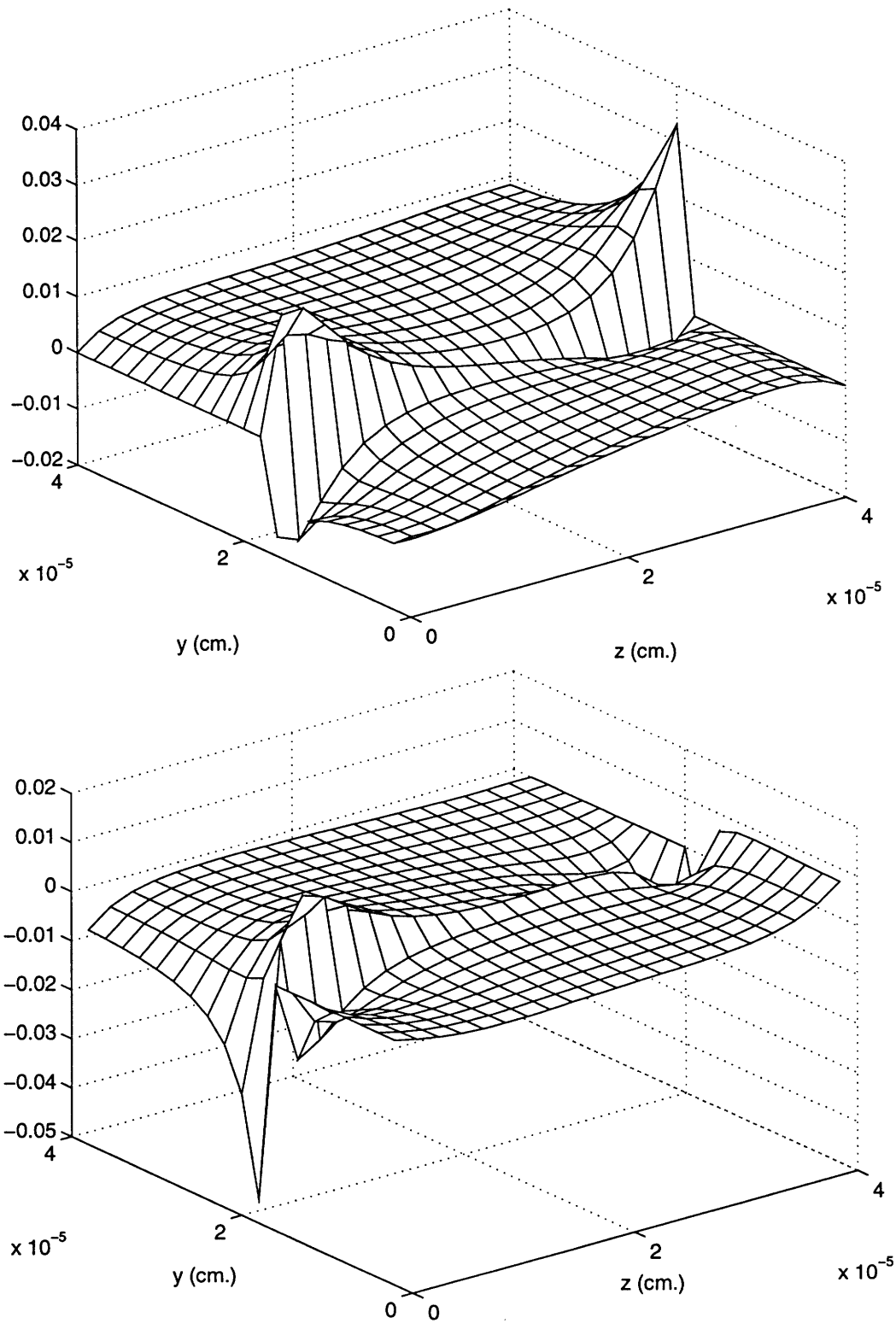

Figure 4-12: The higher order coefficients $f_{2,0}$ (top) and $f_{3,0}$ (bottom) for the homogeneous doping example.

third order system is given below:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & c & s & 0 & 0 \\ 0 & -s & c & 0 & 0 \\ 0 & 0 & 0 & c^2 - \frac{1}{2}s^2 & 0 \\ 0 & 0 & 0 & 0 & c^3 - \frac{3}{2}cs^2 \end{bmatrix}$$

where $c = \cos\mu$ and $s = \sin\mu$ and the vector of spherical harmonics is:

$$[Y_{0,0}\ Y_{1,0}\ Y_{1,-1}\ Y_{2,0}\ Y_{3,0}].$$

4.5.2 Inhomogeneous Doping

For the next simulation we used a structure which has a doping profile which mimics some of the features of a channel of a MOSFET. The doping profile, electron concentration and the potential and field for this structure are shown in Fig. 4-13. Fig. 4-14 shows the current and the equipotentials for the above structure. The isotropic part of the distribution function for the above simulation is shown in Fig. 4-15 for a number of different energies. Note that the distribution function has a heated region near the drain which has a much higher population of carriers at elevated energies than in the rest of the device.

4.6 Summary

In this chapter we extended the Galerkin method approach to the solution of the Boltzmann equation in two real space dimensions using a spherical harmonic expansion in momentum space. The key theoretical results are, first, the use of a rotated momentum space coordinate frame so as to "diagonalize" the problem and reduce the number of spherical harmonics that must be simulated to a given level of accuracy. Second, the implementation of physically reasonable boundary conditions for the rotated coordinate problem. Third, the use of a discretization that maintains current conservation and stability. Using the above results we demonstrated the effectiveness of the rotated coordinate frame approach and the correctness of the implementation for structures with both uniform and non-uniform doping profiles. To obtain an efficient solution to the large matrix problem that results in two dimensions we used an iterative approach with a choice of pre-conditioner that improves the convergence of the iteration. This results was then used in a Gummel type scheme to obtain a self-consistent solution with Poisson's equation.

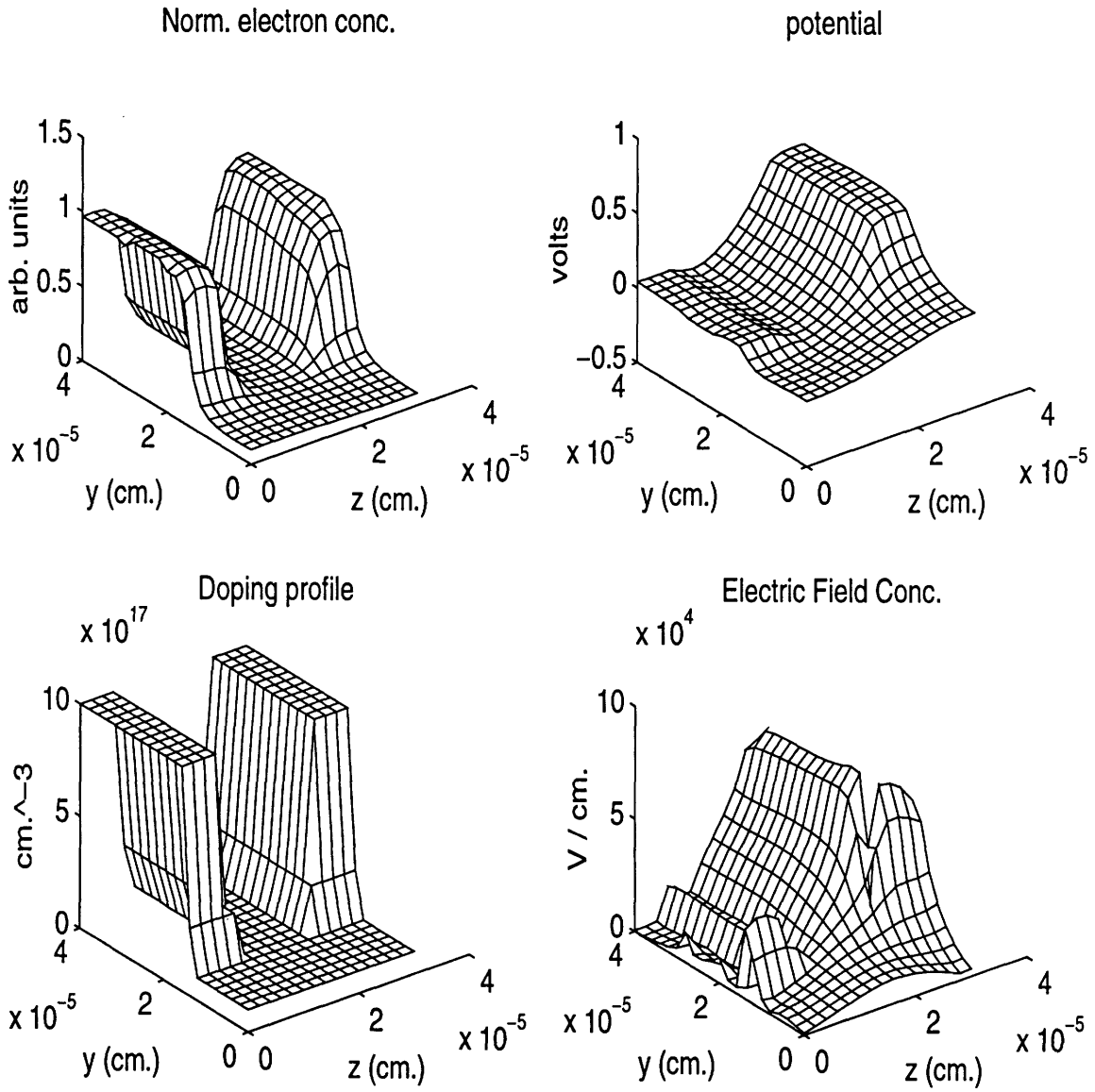

Figure 4-13: The doping profile, electron concentration, potential and electric field for a two dimensional structure with non-uniform doping.

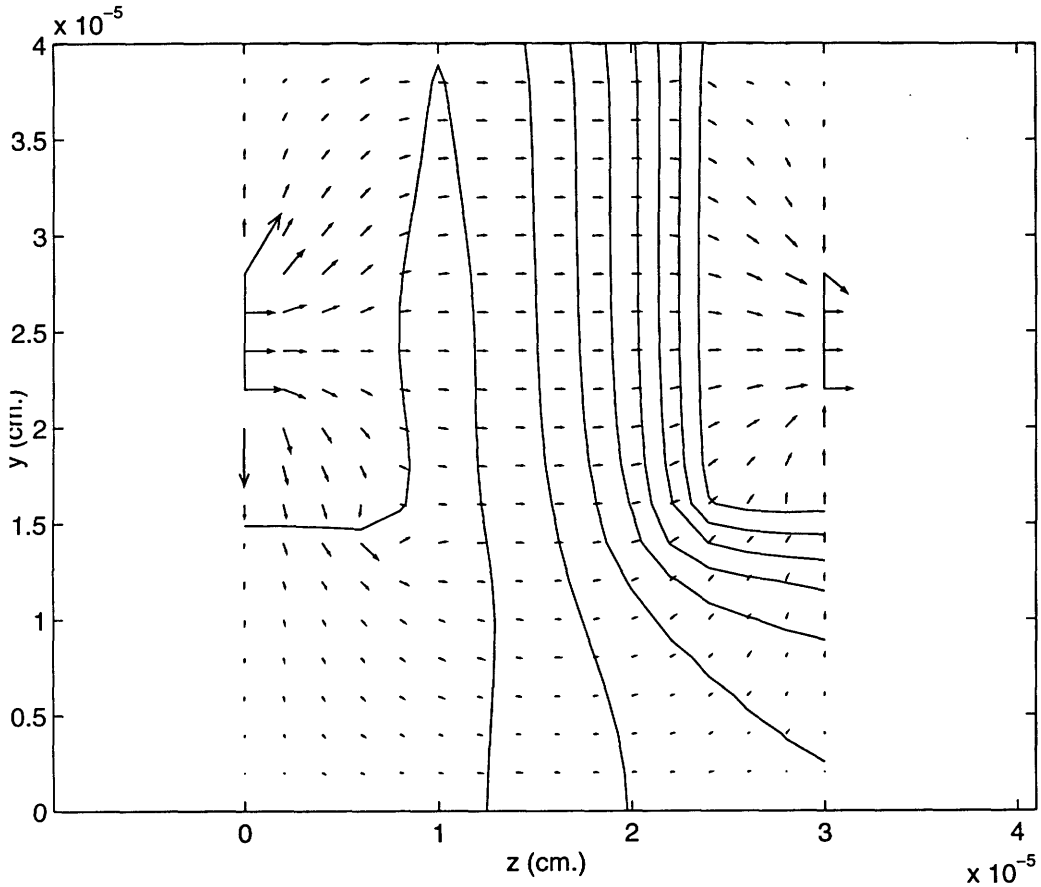

Figure 4-14: Electric current flow lines and equipotentials for the structure shown in Fig. 4.13.

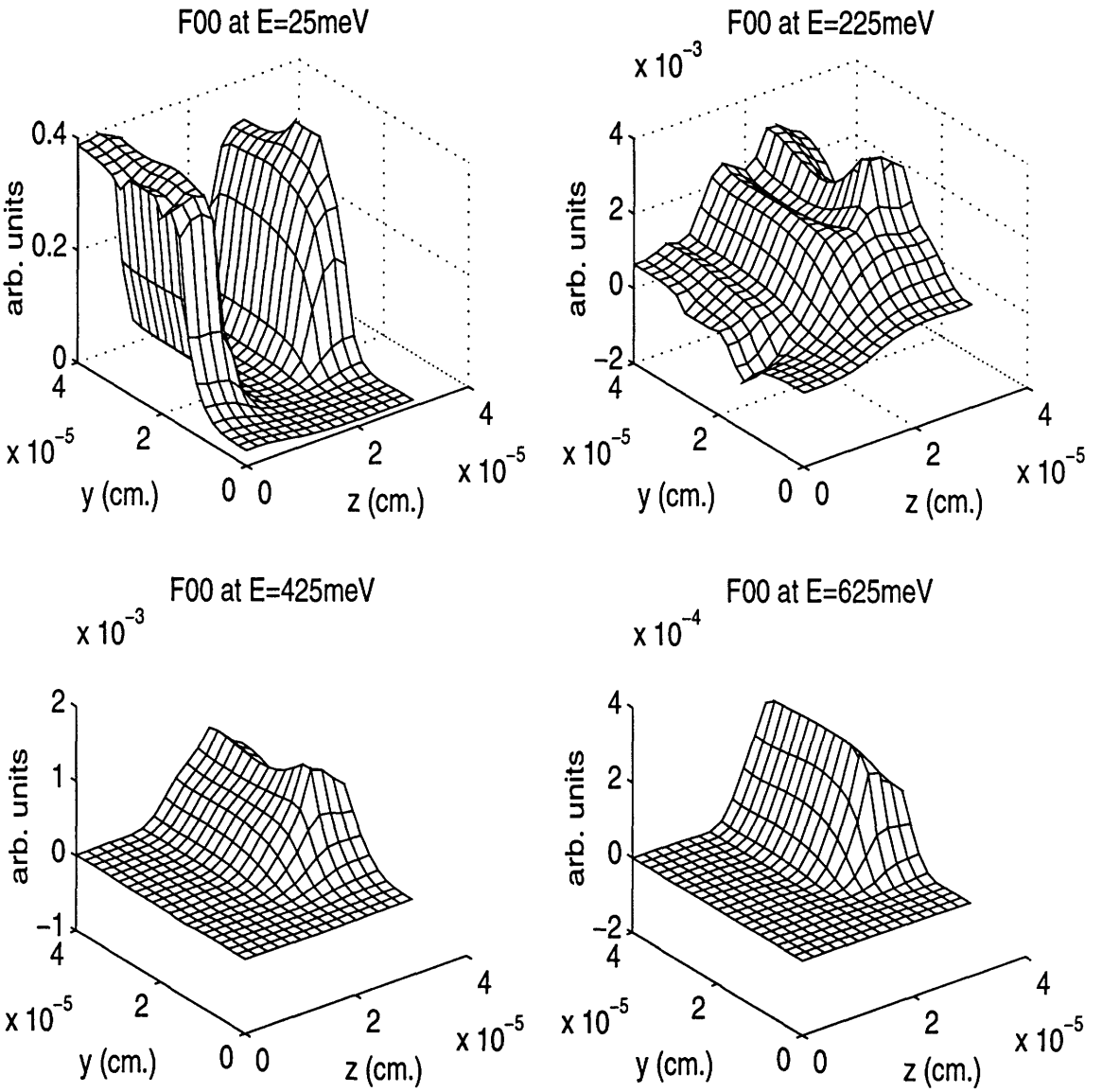

Figure 4-15: The isotropic part of the distribution at energies of 25, 225, 425 and 625 meV for the two-dimensional structure shown in Fig. 4.13.

Chapter 5

Conclusion

5.1 Conclusion

The work described in this dissertation is an assay or attempt to solve the problem of determining the distribution of hot carriers in semiconductor devices. This knowledge is useful because it is an essential input in determining the reliability of highly scaled devices, but it is also becoming central to the problem of determining even the macroscopic properties, as the mean carrier energy becomes significantly larger than the thermal energy. In other words most carriers will be hot carriers in these devices when compared with the lattice temperature. Thus the ability to obtain the distribution function of the carriers efficiently and with sufficient accuracy will become an even more important task than it is today.

We use the semiclassical model as a description of the behavior of charge carriers in semiconductors. This description which leads to the Boltzmann equation is essentially classical because the electrons are treated as classical particles whose position and momentum can be defined simultaneously, except for the use of a quantum-mechanical band structure and scattering rates.

In this thesis two quite different approaches were used to solve the Boltzmann equation to obtain the distribution function. In the first approach the distribution function was not explicitly calculated from the Boltzmann equation rather the moments of the distribution function were computed by generating a set of moment balance equations from the Boltzmann equation. These moment balance equations are physically equivalent to the equations for conservation of particle number, momentum and energy. From these averages over the distribution function, estimates of the hot carriers population can be obtained. We used a method that assumes an explicit analytic form for the distribution function which is more appropriate under high field conditions than the more commonly used guess of a Maxwellian or a displaced Maxwellian distribution. This distribution function has a cubic dependence on energy in the exponent. Using this "ansatz" for the distribution we calculated the number of hot carriers (electrons for the NMOS case) and from that the substrate current by integrating the distribution above a threshold energy which represents the energy required for impact

ionization. Although the substrate current obtained using this approach was fairly accurate when compared with the experimental data for a range of biases and channel lengths, it had a large error when the transistor was biased near threshold at a high drain voltage. This bias condition produces the highest fields near the drain of the transistor as it corresponds to deep saturation. As this electric field is large and rapidly changing near the drain we believe that the error in the simulated substrate current and the underlying hot carrier population arises from the use of an average energy or temperature in the distribution. Also the cubic energy dependence may not be accurate in this field range. The way to avoid these errors therefore is to explicitly calculate the distribution function rather than ensemble averages over it.

As the carrier distribution is a function of a five dimensional phase space in two real space dimensions, a method which uses a direct discretization of phase space would be prohibitively expensive computationally. Thus the most common way to solve the Boltzmann equation to obtain the distribution function is to use a Monte-Carlo method. In the work presented here we used a less well-established but potentially more efficient method which is based on the idea of using a basis function expansion to describe the distribution function in momentum space while treating the real space variation using a conventional finite difference discretization. As the distribution function has some degree of spherical symmetry, a natural choice for the basis functions for the momentum space is the set of spherical harmonic functions. Although this approach has been used in the past, the major contribution of the work presented here is to extend this approach by using a Galerkin method which allows the use of arbitrary order expansions in both one and two real space dimensions. Results in earlier work had been restricted to only the first order expansion. The significance of the ability to do higher order expansions arises from the fact that under high and rapidly changing fields the distribution function can be quite anisotropic and only with the use of higher orders can it be accurately represented by the spherical harmonic functions in momentum space. In two real space dimensions we also formulated and implemented a rotated coordinate approach in which the polar direction for the spherical harmonics is aligned with the electric field at every real space point. In the rotated formulation the two dimensional problem has almost the same symmetry as the one-dimensional problem and therefore the magnitude of the off-diagonal harmonics (those that depend on the ϕ coordinate) is small and they can be neglected. This of course reduces the computational cost by a factor of two or three as compared to the case where a fixed polar direction is used. The issues pertaining to current conservation and boundary conditions in the two-dimensional case were also addressed in this work.

Because we used a simplified parabolic band model in the spherical harmonic expansion approach we did not directly compare the substrate current obtained using this method with experimental data, although comparisons were done with the hydrodynamic model in one-dimension. Other researchers have used advanced band structures and more realistic scattering mechanisms with a low order spherical harmonic expansion [58]. Thus the work presented here can easily utilize those results and in effect the two approaches complement each other.

5.2 Future Work

To make the spherical harmonic expansion approach to the solution of the Boltzmann equation a real competitor with the Monte-Carlo method, further work must be done in a number of different areas. These areas may be divided into issues that relate to numerical efficiency and those that relate to physics.

Perhaps the most important numerical enhancement that needs to be made to the spherical harmonic approach is the use of a higher order discretization scheme. The two-point discretization method used in this thesis requires the use of a large number of mesh lines in energy and space. The problem is most severe when a large doping gradient is encountered, in which case a fine mesh spacing must be used to maintain an accurate approximation to the electron concentration which may vary many orders of magnitude over a distance of only a few hundred angstroms. Ideally, an exponentially fitted discretization such as the Scharfetter-Gummel method that is used in the drift-diffusion and hydrodynamic models could be used here also. The main difficulty with using such a method here is that it has to be applied to a vector of unknowns, the spherical harmonic coefficients, rather than just to a scalar, the electron concentration. An alternate to the Scharfetter-Gummel discretization would be the reformulation of the Boltzmann equation using the logarithm of the distribution function, which is roughly equivalent to the use of quasi-Fermi levels instead of the electron and hole concentrations in the drift-diffusion equations and would also lead to the compression of the dynamic range of the variable used [53]. Our early attempts at using this approach were unfortunately not successful, because the use of a logarithm renders the scattering term non-linear in the new variables and leads to numerical instability when the discretized non-linear problem is solved, nevertheless it is likely that there are other numerical techniques that would ameliorate this difficulty.

Another way of improving the efficiency of the solution would be to use a better preconditioner for the iterative solvers. Although the block diagonal preconditioner works reasonably well, perhaps including some part of the spatial variation could further improve the convergence rate. Also the use of a full Newton iteration rather than a relaxation may require fewer iteration steps especially at higher biases.

Including more sophisticated physical models would obviously make the simulation results more accurate and reliable. Although we used a single parabolic band, it is quite straightforward to use a non-parabolic model. The only change in the formulation would be in the value of the electron velocity as a function of energy and the scattering integrals in k-space. For any spherically symmetric band we could just use a table lookup to determine the electron velocity and the density of states for the integrals in k-space. A more elaborate modification would be to include higher order energy bands. This has been demonstrated for the first order solution in the work of other research groups [47]. Basically in this case a number of separate Boltzmann problems are solved for each band and the coupling between them occurs only at the band edges in energy.

Bibliography

[1] E. Takeda and N. Suzuki. An empirical model for device degradation due to hot-carrier degradation. *Electron Device Lett.*, 4:111–113, 1983.

[2] C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Cho, and K. W. Terril. Hot-electron-induced MOSFET degradation— Model, monitor and improvement. *IEEE Trans. Electron Devices*, ED-32:375–385, 1985.

[3] W. Weber, C. Werner, and G. Dorda. Degradation of N-MOS transistors after pulsed stress. *Electron Device Lett.*, EDL-5:518, 1984.

[4] K. R. Hoffman, C. Werner, W. Weber, and G. Dorda. Hot electron and hole emission effects in short n-channel MOSFET's. *IEEE Trans. Electron Devices*, ED-32:691–699, 1985.

[5] P. Bellens, P. Heremans, G. Groeseneken, and H. A. Maes. Hot carrier effect in n-channel mos transistors under alternate stress conditions. *Electron Device Lett.*, 9:232–234, 1988.

[6] W. Weber. Dynamic stress experiments for understanding hot carrier degradation phenomena. *IEEE Trans. Electron Devices*, ED-35:1476–1485, 1988.

[7] B. Doyle, M. Bourcerie, J. C. Marchetaux, and A. Boudou. Interface state creation and charge trapping in the medium-to-high gate voltage range ($v_d/2 \leq v_g \leq v_d$) during hot-carrier stressing of N-MOS transistors. *IEEE Trans. Electron Devices*, 37:744–754, 1990.

[8] S. P. Joshi, R. Lahri, and C. Lage. Poly emitter bipolar hot carrier induced h_{fe} degradation in BiCMOS transistors. In *IEDM Tech. Digest*, pages 182–185, 1987.

[9] D. Quon, P. K. Gopi, G. J. Sonek, and G. P. Li. Hot carrier induced bipolar transistor degradation due to base dopant compensation by hydrogen: theory and experiment. *IEEE Trans. Electron Devices*, 41:1824–1830, 1994.

[10] B. K. Ridley. *Quantum Processes in Semiconductors*. Oxford University Press, Oxford, 1993.

[11] M. V. Fischetti and S. E. Laux. Monte Carlo analysis of electron transport in small semiconductor devices including band structure and space-charge effects. *Physical Review B*, 38(14):9723–9744, 1988.

[12] V. M. Agostinelli, T. J. Bordelon, X. Wang, K. Hasnat, C-F. Yeap, D. B. Lemersal, A. F. Tasch, and C. M. Maziar. Two-dimensional energy-dependent models for the simulation of substrate current in submicron MOSFET's. *IEEE Trans. Electron Devices*, 41:1784–1795, 1994.

[13] C. Jacoboni and L Reggiani. The Monte-Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Reviews of Modern Physics*, 55(3):645–704, 1983.

[14] J. W. Slotboom, G. Streutker, M. J. v. Dort, P. H. Woerlee, A. Pruijmboom, and D. J. Gravesteijn. Non-local impact ionization in silicon devices. In *IEDM Tech. Digest*, pages 127–130, 1991.

[15] S. Selberherr, A. Schütz, and H. Pötzl. MINIMOS–A two-dimensional MOS transistor analyzer. *IEEE Trans. Electron Devices*, ED-27:1770–1780, 1980.

[16] M. R. Pinto, C. S. Rafferty, H. R. Yeager, and R. W. Dutton. PISCES-IIB. Technical Report, Stanford Electronics Lab., Stanford University, 1985.

[17] E. M. Buturla, P. E. Cottrell, B. M. Grossman, and K. A. Salsburg. Finite-Element analysis of semiconductor devices: The FIELDAY Program. *IBM J. Res. Develop.*, 25:218, 1981.

[18] T. Thurgate and N. Chan. An impact ionization model for two-dimensional device simulation. *IEEE Trans. Electron Devices*, ED-32:400–404, 1985.

[19] W. Hänsch and M. Miura-Mattausch. The hot-electron problem in small semiconductor devices. *J. Appl. Phys.*, 60:650–656, 1986.

[20] A. K. Henning, N. N. Chan, J. T. Watt, and J. D. Plummer. Substrate current at cryogenic temperatures: measurment and two-dimensional model for CMOS technology. *IEEE Trans. Electron Devices*, ED-34:64–73, 1987.

[21] B. Meinerzhagen and W. L. Engl. The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon submicrometer MOS transistors. *IEEE Trans. Electron Devices*, ED-35:689–697, 1988.

[22] K. Bløtekjaer. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electron Devices*, ED-17:38–47, 1970.

[23] R. K. Cook and J. Frey. Two-dimensional numerical simulation of energy transport effects in Si and GaAs MESFET's. *IEEE Trans. Electron Devices*, ED-29:970–977, 1982.

[24] H. Hänsch and S. Selberherr. MINIMOS 3: A MOS simulator that includes energy balance. *IEEE Trans. Electron Devices*, ED-34:1074–1078, 1987.

[25] M. Rudan and F. Odeh. Multi-dimensional discretization scheme for the hydrodynamic model of semiconductor devices. *COMPEL*, 5:149–183, 1986.

[26] A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan, and G. Baccarani. A new discretization strategy of the semiconductor equations comprising momentum and energy balance. *IEEE Trans. Computer-Aided Design*, 7:231–242, 1988.

[27] F. J. Blatt. *Physics of electronic conduction in solids*. McGraw-Hill, New York, 1968.

[28] K. Rahmat, J. White, and D. A. Antoniadis. Computation of drain and substrate currents in ultra-short-channel NMOSFET's using the hydrodynamic model. *IEEE Trans. Computer-Aided Design*, 12:817–824, 1993.

[29] G. Baccarani and M. R. Wordeman. An investigation of steady-state velocity overshoot effects in Si and GaAs devices. *Solid State Electron.*, 28:407–416, 1985.

[30] D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon read diode oscillator. *IEEE Trans. Electron Devices*, ED-16:64–77, 1969.

[31] A. Gnudi and F. Odeh. Engineering and numerics of the hydrodynamic model. Invited Talk SIAM meeting, Chicago, JULY 1990.

[32] T. W. Tang. Extension of the Scharfetter-Gummel algorithm to the energy balance equation. *IEEE Trans. Electron Devices*, ED-31:1912–1914, 1984.

[33] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith. Reduction of channel hot-electron-generated substrate current in sub-150-nm channel length Si MOSFET's. *Electron Device Lett.*, 9:497–499, 1988.

[34] G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, S. Rishton, and E. Ganin. High transconductance and velocity overshoot in NMOS devices at the 0.1-μm gate-length level. *Electron Device Lett.*, 9:464–466, 1988.

[35] J. M. Higman, K. Hess, C. G. Hwang, and R.W. Dutton. Coupled Monte-Carlo drift diffusion analysis of hot-electron effects in MOSFETS. *IEEE Trans. Electron Devices*, ED-36:930–937, 1989.

[36] D. Cassi and B. Riccò. An analytical model of the energy distribution of hot electrons. *IEEE Trans. Electron Devices*, ED-37:1514–1521, 1990.

[37] N. Goldsman, L. Hendrickson, and J. Frey. Reconciliation of a hot-electron distribution function with the lucky electron-exponential model in silicon. *Electron Device Lett.*, 11:472–474, 1990.

[38] J. W. Slotboom, G. Streutker, G. J. T. Davids, and P. B. Hartog. Surface impact ionization in silicon devices. In *IEDM Tech. Digest*, pages 494–497, 1987.

[39] H. D. Rees. Numerical solution of electron motion in solids. *J. Phys. C: Solid State Phys.*, 5:641–656, 1972.

[40] R. W. Hockney, R. A. Warriner, and M. Reiser. Two-dimensional particle models in semiconductor-device analysis. *Electron. Lett*, 10:484–486, 1974.

[41] Yannick L. Le Coz. *Semiconductor device simulation: a spectral method for solution of the Boltzmann Transport Equation*. PhD thesis, Massachusetts Institute of Technology, 1988.

[42] E. M. Conwell and M. O. Vassell. High-field transport in n-type GaAs. *Physical Review*, 166(3):797–821, 1968.

[43] T. Iizuka and M. Fukuma. Carrier transport simulator for silicon based on carrier distribution function evolutions. *Solid State Electron.*, 33:27–34, 1990.

[44] G. Sansone. *Orthogonal Functions*. Interscience Publishers, New York, 1959.

[45] Thomas M. MacRobert. *Spherical harmonics: an elementary treatise on harmonic functions with applications*. Methuen, London, 1947.

[46] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh. Multidimensional spherical harmonics expansion of Boltzmann Equation for transport in semiconductors. *Appl. Math. Lett*, 5(3):85, 1992.

[47] A. Gnudi, D. Ventura, and G. Baccarani. One-dimensional simulation of a bipolar transistor by means of spherial harmonics expansion of the boltzmann equation. In *Proc. of SISDEP Conf.*, page 205, 1991.

[48] K. Rahmat, J. White, and D. A. Antoniadis. A Galerkin method for the arbitrary order expansion in momentum space of the Boltzmann equation using spherical harmonics. In *Proc. of NUPAD V*, pages 133–136, 1994.

[49] R. D. Richtmeyer and K. W. Morton. *Difference Methods for Initial-Value Problems*. Interscience Publishers, New York, 1967.

[50] E. Fatemi and F. Odeh. Upwind finite difference solution of Boltzmann equation applied to electron transport in semiconductor devices. *Journal of Computational Phys.*, 108:209–217, 1993.

[51] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, New York, 1970.

[52] A. Henning. Private communication.

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

[54] Y. Saad and M. H. Schultz. GMRES: a generalized minimum residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comp.*, 7:856–869, 1986.

[55] P. Sonneveld. CGS, a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM J. Sci. Statist. Comp.*, 10:36–52, 1989.

[56] Ricard Barrett *et al. Templates for the solution of linear systems : building blocks for iterative methods*. SIAM, Philadelphia, 1994.

[57] Noël M. Nachtigal, Satish C. Reddy, and Lloyd N. Trefethen. How fast are nonsysmmetric matrix iterations? *SIAM J. Matrix Anal. Appl*, 13:778–795, 1992.

[58] A. Gnudi, D. Ventura, and G. Baccarani. Modeling impact ionization in a BJT by means of a spherical harmonic expansion of the Boltzmann Transport Equation. *IEEE Trans. Computer-Aided Design*, 12(11):1706–1713, 1994.

[59] Jarvis Jacobs. Modeling the effects of the Si/SiO_2 interface proximity and transverse field on carrier mobility in MOSFET's, SM Thesis 1988, MIT.

Appendix A

Derivation of the Hydrodynamic Model

We shall here provide a detailed derivation of the hydrodynamic model starting from the steady state Boltzmann equation[1]:

$$\mathbf{u} \cdot \nabla_x f - \frac{q}{m^*} \mathcal{E} \cdot \nabla_u f = \left(\frac{\partial f}{\partial t} \right)_c \tag{A.1}$$

which is the same as 2.1 except for a change of variables to velocity space, **u** rather than k-space for convenience.

We begin by defining the average quantities, electron density n; momentum density **P**; and energy density W; as follows,

$$\begin{aligned} n &= \int f \, d^3u \\ \mathbf{P} &= m \int \mathbf{u} f \, d^3u \\ W &= \frac{m^*}{2} \int \mathbf{u} \cdot \mathbf{u} \, f \, d^3u \end{aligned} \tag{A.2}$$

Similarly we define the average flow quantities, electron flux density $\hat{n}$; momentum flux density $\hat{P}$; and energy flux density **S** as follows:

$$\begin{aligned} \hat{n}_i &= \int u_i f \, d^3u \\ \hat{P}_{ij} &= m^* \int u_i u_j f \, d^3u \\ S_i &= \frac{m^*}{2} \int u_i (\mathbf{u} \cdot \mathbf{u}) \, f \, d^3u \end{aligned} \tag{A.3}$$

Note that the flow quantities are in general tensors; electron and energy flow are first order tensors (vectors) whereas momentum flow is a second order tensor.

[1]This section uses the notation of an unpublished report by M. Reichelt.

Armed with these definitions and the following integral properties, we begin by taking moments of the BTE and generate conservation laws for electron number, momentum and energy.

$$\int \frac{\partial f}{\partial u_i}\, d^3u = 0 \tag{A.4}$$

$$\int u_j \frac{\partial f}{\partial u_i}\, d^3u = -n\delta_{ij} \tag{A.5}$$

$$\int u_j u_j \frac{\partial f}{\partial u_i}\, d^3u = -\frac{2}{m^*} P_j \delta_{ij} \tag{A.6}$$

The above integral relations require the assumption that the distribution function, f, decays to zero faster than a second order polynomial at infinite velocity which is physically justifiable. We shall also be interchanging integrals and derivatives rather cavalierly in the following sections under the assumption that the distribution function is well behaved.

A.1 Zeroth Moment

To compute the first conservation law we multiply the BTE by unity and integrate each term over velocity space. The first term therefore becomes:

$$\begin{aligned} \int \mathbf{u} \cdot \nabla_x f \; d^3u &= \int \sum_i \left[u_i \frac{\partial f}{\partial x_i}\, d^3u \right] \\ &= \sum_i \frac{\partial}{\partial x_i} \left(\int u_i f \; d^3u \right) \\ &= \nabla_x \cdot \int \mathbf{u} f \; d^3u \\ &= \nabla_x \cdot \hat{\mathbf{n}} \end{aligned} \tag{A.7}$$

For the second term we write out the scalar product explicitly to yield:

$$\frac{q}{m^*} \mathcal{E} \cdot \nabla_u = \frac{q}{m^*} \sum_i \mathcal{E}_i \frac{\partial f}{\partial u_i} \tag{A.8}$$

Then upon integrating over velocity space we get:

$$\int \frac{q}{m^*} \sum_i \left[\mathcal{E}_i \frac{\partial f}{\partial u_i}\, d^3u \right] = \frac{q}{m^*} \sum_i \left[\mathcal{E}_i \int \frac{\partial f}{\partial u_i}\, d^3u \right] = 0 \tag{A.9}$$

where we have used the relation A.4. Finally the right hand side or the scattering term yields:

$$\int \left(\frac{\partial f}{\partial t} \right)_c d^3u = \left(\frac{\partial n}{\partial t} \right)_c \tag{A.10}$$

Thus the equation for the zeroth moment is

$$\nabla_x \cdot \hat{\mathbf{n}} = \left(\frac{\partial n}{\partial t} \right)_c \tag{A.11}$$

which of course is just the current conservation equation if we remember the definition of $\hat{\mathbf{n}} = \frac{\mathbf{P}}{m^*}$ as the electron flux.

A.2 First Moment

To obtain the equations for the first moment we need to multiply the BTE by $\mathbf{u}$ and integrate over velocity space. As $\mathbf{u}$ is a vector we shall first do this for the j^{th} component of the velocity and then generalize to the vector case.

For the first term of the BTE we get:

$$\int u_j \sum_i u_i \frac{\partial f}{\partial x_i}\, d^3u = \sum_i \int u_j u_i \frac{\partial f}{\partial x_i}\, d^3u = \frac{1}{m^*} \sum_i \frac{\partial}{\partial x_i} \hat{P}_{ij} = \frac{1}{m^*} (\nabla_x \hat{\mathbf{P}})_j \tag{A.12}$$

For the second term we obtain:

$$\begin{aligned} \frac{q}{m^*} \int u_j \sum_i \left[\mathcal{E}_i \frac{\partial f}{\partial u_i} \right] d^3u &= \frac{q}{m^*} \sum_i \left[\int u_j \mathcal{E}_i \frac{\partial f}{\partial u_i}\, d^3u \right] \\ &= \frac{q}{m^*} \sum_i \left[\mathcal{E}_i \int u_j \frac{\partial f}{\partial u_i}\, d^3u \right] \\ &= -\frac{q}{m^*} \mathcal{E}_j n \end{aligned} \tag{A.13}$$

where we have used the relation A.5 in the last step. The scattering term yields:

$$\int u_j \left(\frac{\partial f}{\partial t} \right)_c d^3u = \frac{\partial}{\partial t} \left(\int u_j f\, d^3u \right)_c = \left(\frac{\partial \hat{n}_j}{\partial t} \right)_c d^3u \tag{A.14}$$

The complete first moment equation (after multiplying by m^*) is:

$$(\nabla_x \hat{\mathbf{P}})_j + \frac{q}{m^*} n \mathcal{E}_j = \left(\frac{\partial P_j}{\partial t} \right)_c \tag{A.15}$$

A.3 Second Moment

We now multiply each term of the BTE by $\frac{1}{2}\mathbf{u} \cdot \mathbf{u} = \frac{u^2}{2}$. For the first term we obtain:

$$\begin{aligned} \int \frac{u^2}{2} \sum_i \left[u_i \frac{\partial f}{\partial x_i}\, d^3u \right] &= \sum_i \left[\int \frac{u^2}{2} u_i \frac{\partial f}{\partial x_i}\, d^3u \right] \\ &= \frac{1}{2} \sum_i \left[\frac{\partial}{\partial x_i} \left(\int u_i u^2 f\, d^3u \right) \right] \\ &= \frac{1}{m^*} \sum_i \frac{\partial}{\partial x_i} S_i \\ &= \frac{1}{m^*} \nabla_x \cdot \mathbf{S} \end{aligned} \tag{A.16}$$

For the second term we get:

$$\begin{aligned} \frac{q}{m^*} \int \frac{u^2}{2} \sum_i \mathcal{E}_i \frac{\partial f}{\partial u_i}\, d^3u &= \frac{q}{2m^*} \sum_i \left[\mathcal{E}_i \int u^2 \frac{\partial f}{\partial u_i}\, d^3u \right] \\ &= -\frac{q}{m^{*2}} \sum_i \mathcal{E}_i P_i \\ &= -\frac{q}{m^{*2}} \mathcal{E} \cdot P \end{aligned} \tag{A.17}$$

where we have used A.6 at the second step in the derivation. For the scattering term we get:

$$\int \frac{u^2}{2}\left(\frac{\partial f}{\partial t}\right)_c d^3u = \frac{\partial}{\partial t}\left(\int \frac{u^2}{2} f\right)_c d^3u = \frac{1}{m^*}\left(\frac{\partial W}{\partial t}\right)_c \tag{A.18}$$

Putting all three terms together (and after multiplication by m^*) we obtain the equation for the second moment:

$$\nabla_x \cdot \mathbf{S} + \frac{q}{m^*}\mathcal{E} \cdot P = \left(\frac{\partial W}{\partial t}\right)_c \tag{A.19}$$

A.4 Simplification

Summarizing the results of the previous sections, we can write the first three moment equations as:

$$\frac{1}{m^*}\nabla_x \cdot \mathbf{P} = \left(\frac{\partial n}{\partial t}\right)_c \tag{A.20}$$

$$(\nabla_x \hat{\mathbf{P}})_j + \frac{q}{m^*} n\mathcal{E}_j = \left(\frac{\partial P_j}{\partial t}\right)_c \tag{A.21}$$

$$\nabla_x \cdot \mathbf{S} + \frac{q}{m^*}\mathcal{E} \cdot P = \left(\frac{\partial W}{\partial t}\right)_c \tag{A.22}$$

We shall next write the above equations in the more convenient and intuitive macroscopic variables, electron concentration n, electron mean velocity $\mathbf{v}$, and the electron temperature $\mathbf{T}$, which are defined below. The electron concentration has already been defined in A.2. The mean velocity is given by

$$\mathbf{v} = \frac{\int \mathbf{u} f \; d^3u}{\int f \; d^3u} = \frac{\frac{1}{m^*}\mathbf{P}}{n}. \tag{A.23}$$

Before we can define the electron temperature, we need to define a variable η, which denotes the variation of the actual electron velocity from the mean:

$$\eta = \mathbf{u} - \mathbf{v}. \tag{A.24}$$

The random velocity from the mean, η has the important property that

$$\int \eta_i f \; d^3u = \int (u_i - v_i) f \; d^3u = \int u_i f \; d^3u - v_i \int f \; d^3u = nv_i - v_i n = 0. \tag{A.25}$$

Using the random velocity η we define the temperature:

$$nk_B \hat{T}_{ij} = m^* \int \eta_i \eta_j \; d^3u. \tag{A.26}$$

Similarly we define a heat flow density $\mathbf{Q}$ as:

$$Q_i = \frac{m^*}{2} \int \eta_i \eta^2 f \; d^3u \tag{A.27}$$

A.4.1 Zeroth Moment Equation

In terms of the above macroscopic variables, we have for the zeroth order equation:

$$\nabla_x(n\mathbf{v}) = \left(\frac{\partial n}{\partial t}\right)_c \tag{A.28}$$

A.4.2 First Moment Equation

We must first write the momentum density, $\mathbf{P}$, in terms of the mean velocity $\mathbf{v}$, which follows almost directly from the definition of the mean velocity:

$$\mathbf{P} = m^* n\mathbf{v}. \tag{A.29}$$

The momentum flow density must also be rewritten:

$$\hat{P}_{ij} = m^* \int u_i u_j f \; d^3u = m^* \int (v_i + \eta_i)(v_j + \eta_j) f \; d^3u = m^* n v_i v_j + n k_B \hat{T}_{ij} \tag{A.30}$$

where we have used the property A.25. Using the above expression for momentum flow density we can express the divergence of the flow density as follows:

$$\begin{aligned}
(\nabla_x \hat{P})_j &= \sum_i \left[\frac{\partial}{\partial x_i}\hat{P}_{ij}\right] \\
&= \sum_i \left[\frac{\partial}{\partial x_i}(m^* n v_i v_j + n k_B \hat{T}_{ij})\right] \\
&= m^* \left[\sum_i v_j \frac{\partial}{\partial x_i}(n v_i) + \sum_i v_i \frac{\partial}{\partial x_i}(n v_j) + \sum_i \frac{\partial}{\partial x_i}\left(n k_B \hat{T}_{ij}\right)\right] \\
&= m^* v_j \nabla_x(n\mathbf{v}) + m^* n\mathbf{v} \cdot \nabla_x(v_j) + \nabla_x(n k_B \hat{T}_{ij}).
\end{aligned} \tag{A.31}$$

Putting the pieces together for the first moment equation we can write:

$$m^* v_j \nabla_x(n\mathbf{v}) + m^* n\mathbf{v} \cdot \nabla_x(v_j) + (\nabla_x n k_B \hat{T}_{ij}) + \frac{q}{m^*} n\mathcal{E}_j = \left(\frac{\partial P_j}{\partial t}\right)_c \tag{A.32}$$

A.4.3 Second Moment Equation

As usual we rewrite the energy flow density in terms of the macroscopic variables:

$$\begin{aligned}
S_i &= \frac{m^*}{2} \int u_i u^2 f \; d^3u \\
&= \frac{m^*}{2} \int (v_i + \eta_i)(v^2 + 2\mathbf{v}\cdot\eta + \eta^2) f \; d^3u \\
&= \frac{m^*}{2} v_i \left(v^2 \int f \; d^3u + \int \eta^2 f \; d^3u\right) + \left(m^* \mathbf{v} \cdot \int \eta_i \eta f \; d^3u\right) + \frac{m^*}{2} \int \eta_i \eta^2 f \; d^3u \\
&= w n v_i + (n k_B \hat{\mathbf{T}}\mathbf{v})_i + Q_i
\end{aligned} \tag{A.33}$$

where the electron energy is given by

$$w = \frac{1}{2} m^* v^2 + \frac{1}{2} \mathrm{Tr}(k_B \hat{\mathbf{T}}). \tag{A.34}$$

Thus the conservation equation for the second moment is:

$$\nabla_x \cdot (wn\mathbf{v} + (nk_B\hat{\mathbf{T}}\mathbf{v}) + \mathbf{Q}) + qn\mathbf{v} \cdot \mathcal{E} = \left(\frac{\partial W}{\partial t}\right)_c \tag{A.35}$$

or more compactly:

$$\nabla_x \cdot \mathbf{S} + qn\mathbf{v} \cdot \mathcal{E} = \left(\frac{\partial W}{\partial t}\right)_c \tag{A.36}$$

where $\mathbf{S} = (wn\mathbf{v} + (nk_B\hat{\mathbf{T}}\mathbf{v}) + \mathbf{Q})$

A.4.4 Closure and thermal conductivity

The problem with the above set, specially the energy equation, is that it requires knowledge of the next higher moment i.e. the third moment. This is evident in the definition of $Q_i = \frac{m^*}{2}\int \eta_i\eta^2 f\ d^3u$. This is a problem in general with moment methods– each conservation law involves moments lower and higher than the one for which the equation is being written. The lower moments cause no problems as they will be solved as well but obviously if the moment equations are truncated at some level the higher moments are unavailable. Thus, in the jargon, closure needs to be achieved. Two possible solutions suggest themselves: one is to just leave out the higher moment under the presumption that it is negligible. The other is to rely on some physical intuition or knowledge to estimate the highest moment. For the hydrodynamic equations both approaches have been used, but perhaps the more plausible is to use the physical argument that the heat flow is related to the temperature gradient and thus assume that

$$\mathbf{Q} = -\kappa\nabla_x T \tag{A.37}$$

while at the same time assuming that the temperature is a scalar. In the above equation κ is the thermal conductivity which is proportional to the electron concentration. Expressions for it can be calculated depending on what scattering mechanisms are dominant. With these assumptions the energy flux is

$$\mathbf{S} = -\kappa\nabla_x T + wn\mathbf{v} + nk_BT\mathbf{v} \tag{A.38}$$

where $w = \frac{1}{2}m^*v^2 + \frac{3}{2}k_BT$.

The only other simplification needed to put the above equations in the standard form is to invoke a relaxation time approximation to account for the scattering terms which were left undefined. This assumption allows us to write the scattering terms in terms of two parameters: the momentum relaxation time τ_n and the energy relaxation time τ_w:

$$\left(\frac{\partial \mathbf{v}}{\partial t}\right)_c = -\frac{\mathbf{v}}{\tau_\mathbf{n}} \tag{A.39}$$

$$\left(\frac{\partial w}{\partial t}\right)_c = -\frac{w - w_0}{\tau_w} \tag{A.40}$$

Any change in electron concentration itself is, of course, due to net generation, therefore:

$$\left(\frac{\partial \mathbf{n}}{\partial t}\right)_c = -U \tag{A.41}$$

With some manipulations we can now put the equations in their standard form:

$$\nabla \cdot \mathbf{J} = qU \tag{A.42}$$

$$\mathbf{J} - \frac{\tau_n}{q}(\mathbf{J} \cdot \nabla)\left(\frac{\mathbf{J}}{n}\right) = q\mu n\mathcal{E} + q\mu n\nabla\left(\frac{k_B T}{q}\right) + qD_n\nabla_n \tag{A.43}$$

$$\nabla_x \cdot \mathbf{S} - \mathbf{J} \cdot \mathcal{E} = -Uw - n\frac{w - w_0}{\tau_w} \tag{A.44}$$

where we have used the following definition for the electron mobility

$$\mu = \frac{q\tau_n}{m^*} \tag{A.45}$$

and the generalized Einstein relation:

$$\frac{D}{\mu} = \frac{k_B T}{q} \tag{A.46}$$

Appendix B

Parameters used in the Hydrodynamic Model

The expression for electron mobility which includes the effects of the scattering of the electrons by the lattice and ionized impurities is [53] :

$$\mu_{LI} = 88\left(\frac{T_0}{300}\right)^{-0.57} + \frac{1252\left(\frac{T_0}{300}\right)^{-2.33}}{1+\frac{N}{1.432\times10^{17}}\left(\frac{T_0}{300}\right)^{-0.57}} \tag{B.1}$$

where N is the total doping concentration in cm^{-3}, and T_0 is the lattice temperature.

The effects of surface scattering are included by using an expression that depends on both the distance from the surface and the lateral electric field [59]

$$\mu_{LIS} = \frac{\mu_{LISmin} + (\mu_{LI} - \mu_{LISmin})(1-f)}{1+f\left(\frac{E_t}{EREF}\right)^{LISexp}} \tag{B.2}$$

where f is given by

$$f = \frac{2.0e^{-1\times10^{12}d^2}}{1.0+e^{-2\times10^{12}d^2}} \tag{B.3}$$

where d is the distance from the silicon–oxide interface in cm. and E_t is the transverse electric field in V/cm. The constants in the above expression are $\mu_{LISmin} = 638\,cm^2/(Vs)$; $EREF = 5.0 \times 10^5$ V/cm; $LISexp = 1.30$.

The distance function, f, has a value of one at the surface and its value approaches zero deeper in the bulk. Consequently, near the surface the mobility is dominated by surface effects whereas deeper into the bulk, the surface effects become negligible.

To include the high field effects in the drift-diffusion simulations we used the electric field dependence given by equation 4.1–56 in [53]:

$$\mu_{LISE} = \frac{2\,\mu_{LIS}}{1+\sqrt{1+\left(\frac{2\mu_{LIS}E}{v_{sat}}\right)^2}} \tag{B.4}$$

Appendix C

Spherical Harmonic Coefficients for Ionized Impurity Scattering

The Brooks-Herring model for scattering due to ionized impurities can be written as:

$$S(\mathbf{k},\mathbf{k}') = \frac{c_{BH}}{[1+a(1-\cos\theta)]^2}\delta[E(\mathbf{k})-E(\mathbf{k}')] \tag{C.1}$$

where c_{BH} is the scattering rate, $a = 2\frac{k^2}{\beta^2}$ depends on the magnitude of the k vector and the Debye length β. We shall below give the spherical harmonic expansion coefficients for the above expression up to third order after normalizing by c_{BH}.

$$S_{0,0} = \frac{\sqrt{2}}{1+2a}$$

$$S_{1,0} = \sqrt{\frac{3}{2}}\left(\frac{2(1+\frac{1}{a})}{1+2a} - \frac{1}{a^2}\log(1+2a)\right)$$

$$S_{2,0} = 3\sqrt{\frac{5}{2}}\frac{1}{a^2} + \sqrt{10}\left(1+\frac{3}{a}+\frac{3}{2a^2}\right)\frac{1}{1+2a} - \frac{3\sqrt{10}}{2a^2}\left(1+\frac{1}{a}\right)\log(1+2a)$$

$$S_{3,0} = \frac{5\sqrt{14}}{a^2}\left(1+\frac{1}{a}\right) + \sqrt{\frac{7}{2}}\left(1+\frac{6}{a}+\frac{15}{2a^2}+\frac{5}{2a^3}\right)\frac{2}{1+2a} - \sqrt{\frac{7}{2}}\frac{1}{a}\left(\frac{6}{a}+\frac{15}{a^2}+\frac{15}{a^3}\right)\log(1+2a)$$

Appendix D

Symmetry of the Two-dimensional Distribution

A function $g(\theta, \phi)$ that has symmetry about the $k_x = 0$ plane as shown in Fig. D-1 has the following property:

$$\begin{aligned} g(\theta, \phi) &= g(\theta, \pi - \phi) & 0 \leq \phi \leq \pi \\ g(\theta, \phi) &= g(\theta, -\phi - \pi) & -\pi \leq \phi \leq 0 \end{aligned} \tag{D.1}$$

which can be written more symmetrically as:

$$\begin{aligned} g(\theta, \frac{\pi}{2} + d) &= g(\theta, \frac{\pi}{2} - d) & 0 \leq d \leq \frac{\pi}{2} \\ g(\theta, -\frac{\pi}{2} + d) &= g(\theta, -\frac{\pi}{2} - d) & -\frac{\pi}{2} \leq d \leq 0 \end{aligned} \tag{D.2}$$

We now evaluate the spherical harmonic coefficients, g_{lm} for the function $g(\theta, \phi)$ which has the symmetry property described above:

$$g_{lm} = \int d\theta \int d\phi \; Y_{lm}(\theta, \phi) g(\theta, \phi). \tag{D.3}$$

But as we are primarily interested in exploiting the symmetry in ϕ, we shall focus on the integral over ϕ which will be considered to be the inner integral in the double integral above. Thus, writing out only the ϕ dependent part of both the $\cos\phi$ and $\sin\phi$ spherical harmonics:

$$\begin{aligned} I_c &= \int_{-\pi}^{\pi} d\phi \, (\cos m\phi) g(\theta, \phi) \\ I_s &= \int_{-\pi}^{\pi} d\phi \, (\sin m\phi) g(\theta, \phi) \end{aligned} \tag{D.4}$$

Consider the I_c integral first:

$$\begin{aligned} I_c &= \int_{0}^{\pi} d\phi \, (\cos m\phi) g(\theta, \phi) + \int_{-\pi}^{0} d\phi \, (\cos m\phi) g(\theta, \phi) \\ I_c &= I_c^+ + I_c^- \end{aligned} \tag{D.5}$$

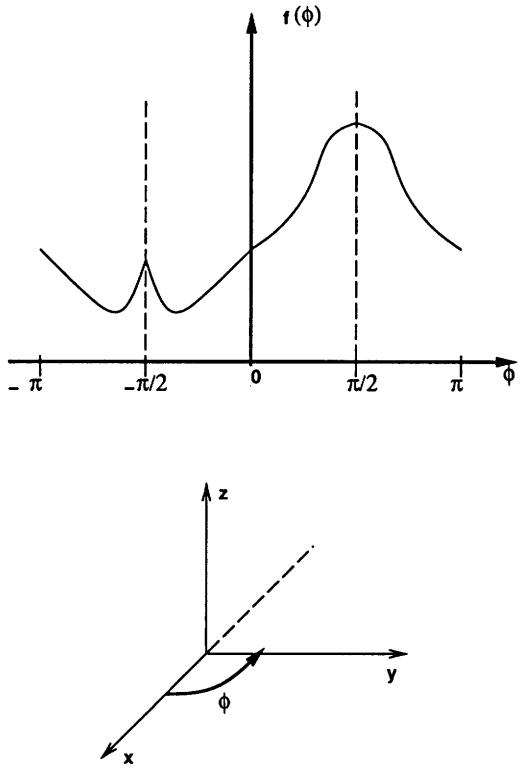

Figure D-1: Example of a distribution function having symmetry about $\phi = \pi/2$ and $\phi = -\pi/2$

If we now focus on the integral over the positive domain only, I_c^+, we can write:

$$\begin{aligned} I_c^+ &= \int_0^{\pi} d\phi\,(\cos m\phi) g(\theta,\phi) \\ &= \int_0^{\frac{\pi}{2}} d\phi\,(\cos m\phi) g(\theta,\phi) + \int_{\frac{\pi}{2}}^{\pi} d\phi\,(\cos m\phi) g(\theta,\phi) \\ &= \int_0^{\frac{\pi}{2}} d\phi\,(\cos m\phi) g(\theta,\phi) - \int_{\frac{\pi}{2}}^{0} d\phi'\,[\cos m(\phi'-\pi)]\, g(\theta,\phi') \end{aligned} \tag{D.6}$$

where we have made a change of variables $\phi' = \pi - \phi$ and used the symmetry property (D.2). The above expression can be simplified as follows:

$$\begin{aligned} I_c^+ &= \int_0^{\frac{\pi}{2}} d\phi\,(\cos m\phi) g(\theta,\phi) + \int_0^{\frac{\pi}{2}} d\phi' \cos m\phi' (-1)^m g(\theta,\phi') \\ &= 2\int_0^{\frac{\pi}{2}} d\phi\,(\cos m\phi) g(\theta,\phi) \qquad \text{m even} \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad \text{m odd} \end{aligned} \tag{D.7}$$

To complete the argument we need to evaluate the integral over the negative range of ϕ, I_c^-:

$$\begin{aligned} I_c^- &= \int_{-\pi}^{0} d\phi\,(\cos m\phi) g(\theta,\phi) \\ &= \int_{-\pi}^{-\frac{\pi}{2}} d\phi\,(\cos m\phi) g(\theta,\phi) + \int_{-\frac{\pi}{2}}^{0} d\phi\,(\cos m\phi) g(\theta,\phi) \end{aligned} \tag{D.8}$$

$$= \int_{-\pi}^{-\frac{\pi}{2}} d\phi\,(\cos m\phi)g(\theta,\phi) - \int_{-\frac{\pi}{2}}^{-\pi} d\phi'\,[\cos m(-\pi-\phi')]\,g(\theta,\phi')$$

where we have made a change of variables $\phi' = -\pi - \phi$ and used the symmetry property (D.2). The above expression can be simplified as follows:

$$\begin{aligned} I_c^- &= \int_{-\pi}^{-\frac{\pi}{2}} d\phi\,(\cos m\phi)g(\theta,\phi) + \int_{-\pi}^{-\frac{\pi}{2}} d\phi'\cos m\phi'(-1)^m g(\theta,\phi') \qquad \text{(D.9)} \\ &= 2\int_{-\pi}^{-\frac{\pi}{2}} d\phi\,(\cos m\phi)g(\theta,\phi) \qquad \text{m even} \\ &= 0 \qquad \text{m odd} \end{aligned}$$

Upon combining the results for the integrals I_c^+ and I_c^- we get the condition that the coefficients of all spherical harmonics containing expressions of the form $\cos m\phi$ where m is odd are zero. The analogous condition for the harmonics containing $\sin m\phi$ is that the coefficients for these harmonics are zero if m is even, although the derivation for this is omitted here. Thus the coefficients of the harmonics $Y_{1,1}, Y_{2,1}, Y_{3,1}\ldots$ and $Y_{2,-2}, Y_{3,-2}\ldots$ will be zero for any two dimensional real space function.