PARTICLE-BASED FULL-BAND SIMULATION OF CHARGE TRANSPORT IN SEMICONDUCTORS.

By

JULIEN BRANLARD

Submitted in partial fulfillment of the requirement for the degree of Electrical and Computer Engineering in the Graduate College of Ecole Nationale Supérieure de l’Électronique et de ses Applications, Cergy, France.

Approved

Illinois Institute of Technology
Chicago, Illinois
June 2002
ABSTRACT

As the minimum feature size shifts into the sub-0.1 μm regime, there is a higher demand for accurate and fast computer simulation program. The Ensemble Monte Carlo (EMC) method has a long history of success in simulating carrier transport in semiconductor devices. Unfortunately, this technique can be computationally intensive, limiting its application, particularly when a full band representation of the electronic structure is implemented. The Cellular Monte Carlo (CMC) method was developed to reduce this computational burden. Although faster, this technique can require an impractical amount of memory. A new hybrid EMC/CMC has been implemented, to optimize the tradeoff between memory and speed. In this work a general overview of charge transport in semiconductors is presented. Simulation issues related to the hybrid EMC/CMC modeling tool are described and finally, results obtained with this hybrid implementation are given and confronted with published data.
Acknowledgements

I first would like to thank my advisor Dr. Marco Saraniti who has always been next door to answer my questions, guide my initiatives and debug my code. I would like to thank him also for his patience on the days my brain didn’t respond, his moral support and his good wine. Many thanks also, to Dr. Shela Aboud, (congratulations!) who was here to help me when Marco couldn’t take it any more. I would also like to thank my friends and my family, who contributed to this work, sometimes on purpose, sometimes not, by encouraging me or mocking me, giving me the will to go ahead. To them and to all the others, I thank you all!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>................................................................. vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>........................................................................... ix</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>........................................................................... 1</td>
</tr>
<tr>
<td>CHAPTER I</td>
<td>Basic notions and definitions ........................................................................ 3</td>
</tr>
<tr>
<td>1- Basic crystallographic definitions</td>
<td>........................................................................ 3</td>
</tr>
<tr>
<td>a) The Bravais lattice</td>
<td>........................................................................ 3</td>
</tr>
<tr>
<td>b) The Wigner-Seitz primitive cell</td>
<td>........................................................................ 4</td>
</tr>
<tr>
<td>c) The reciprocal lattice</td>
<td>........................................................................ 4</td>
</tr>
<tr>
<td>d) The first Brillouin zone</td>
<td>........................................................................ 5</td>
</tr>
<tr>
<td>2- Bonding forces in solids</td>
<td>........................................................................ 6</td>
</tr>
<tr>
<td>a) Ionic bonding: insulators</td>
<td>........................................................................ 6</td>
</tr>
<tr>
<td>b) Metallic bonding: metals</td>
<td>........................................................................ 6</td>
</tr>
<tr>
<td>c) Covalent bonding: semiconductors</td>
<td>........................................................................ 7</td>
</tr>
<tr>
<td>3- Band structure</td>
<td>.................................................................................. 7</td>
</tr>
<tr>
<td>a) Definition</td>
<td>.................................................................................. 7</td>
</tr>
<tr>
<td>b) Direct and indirect transitions</td>
<td>........................................................................ 9</td>
</tr>
<tr>
<td>4- Effective mass</td>
<td>................................................................................. 10</td>
</tr>
<tr>
<td>5- Drift of carriers in electric fields</td>
<td>......................................................................... 11</td>
</tr>
<tr>
<td>a) Carriers in an electric field</td>
<td>......................................................................... 11</td>
</tr>
<tr>
<td>b) Conductivity and mobility</td>
<td>............................................................................. 13</td>
</tr>
<tr>
<td>CHAPTER II</td>
<td>Physical models of charge transport in semiconductors ........................................ 14</td>
</tr>
<tr>
<td>1- Band structure</td>
<td>.................................................................................. 14</td>
</tr>
<tr>
<td>a) The Brillouin Zone, the irreducible wedge</td>
<td>.................................................................... 14</td>
</tr>
<tr>
<td>b) Approximations of the conduction band</td>
<td>...................................................................... 16</td>
</tr>
<tr>
<td>c) Approximations of the valence band</td>
<td>...................................................................... 17</td>
</tr>
<tr>
<td>d) Full band representation</td>
<td>............................................................................. 18</td>
</tr>
<tr>
<td>e) 3D representation</td>
<td>............................................................................. 21</td>
</tr>
</tbody>
</table>
2- The Boltzmann equation (semiclassical theory of charge transport) .......... 22
   a) The Boltzmann transport equation ........................................... 22
   b) Change of particle number due to carriers displacement ............... 23
   c) Change of particle number due to scattering ............................... 24
   d) Validity of the BTE ................................................................... 25

3- Scattering mechanisms ................................................................... 26
   a) Phonon scattering ..................................................................... 26
   b) Impurity scattering ................................................................... 26
   c) Carrier-carrier scattering .......................................................... 27

CHAPTER III Particle-based approach for charge transport simulation .... 28

1- The self-consistent particle-based method ......................................... 28
   a) Particle-based methods ............................................................. 28
   b) Self-consistent method .............................................................. 30
   c) Poisson’s equation .................................................................... 30

2- The Monte Carlo method ................................................................. 33
   a) Historical approach .................................................................. 33
   b) The MC approach ................................................................. 34

3- The Cellular Automaton method ...................................................... 35
   a) Description ............................................................................... 35
   b) CA for semiconductor simulation, the CMC method .................... 35

4- Scattering tables ............................................................................. 36
   a) Scattering tables in the EMC program ....................................... 36
   b) Scattering tables in the CMC program ....................................... 38
   c) The hybrid EMC/CMC method, speed up and performances ....... 39

CHAPTER IV Velocity Calculation ....................................................... 40

1- Operators, the Bloch theorem ............................................................ 40
   a) Operators .................................................................................. 40
   b) The Bloch theorem ............................................................... 41

2- Computation of the velocity .............................................................. 42

3- Two numerical approaches ............................................................. 43
## CHAPTER V  Results and discussion

1- Simulation characteristics  

2- Comparison with referenced data  
   a) Velocity – field characteristics 
   b) Energy – field characteristics 

## CONCLUSION, PROJECTS AND FUTURE WORK

## APPENDIX A  Computing the first derivative of the energy

## REFERENCES
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.1.1</td>
<td>Unit cell for three types of cubic lattice structures (sc, bcc and fcc)</td>
<td>3</td>
</tr>
<tr>
<td>I.1.2</td>
<td>A simple cubic three-dimensional Bravais lattice</td>
<td>4</td>
</tr>
<tr>
<td>I.1.3</td>
<td>The Wigner-Seitz cell for a two-dimensional Bravais lattice</td>
<td>4</td>
</tr>
<tr>
<td>I.1.4</td>
<td>Brillouin Zone for zincblende and diamond lattices</td>
<td>5</td>
</tr>
<tr>
<td>I.2.1</td>
<td>Ionic bonding in NaCl and covalent bonding in Si</td>
<td>6</td>
</tr>
<tr>
<td>I.2.2</td>
<td>The diamond lattice structure and its unit cell</td>
<td>7</td>
</tr>
<tr>
<td>I.3.1</td>
<td>The simplified band structure of a crystal</td>
<td>8</td>
</tr>
<tr>
<td>I.3.2</td>
<td>Typical band structures at 0K</td>
<td>9</td>
</tr>
<tr>
<td>I.3.3</td>
<td>Direct and indirect electron transitions</td>
<td>10</td>
</tr>
<tr>
<td>I.4.1</td>
<td>Effective mass of electrons and holes</td>
<td>11</td>
</tr>
<tr>
<td>II.1.1</td>
<td>Sampling region for the calculation of the band structure</td>
<td>14</td>
</tr>
<tr>
<td>II.1.2</td>
<td>The GaAs band structure</td>
<td>17</td>
</tr>
<tr>
<td>II.1.3</td>
<td>Full band structure representation for common semiconductor materials</td>
<td>20</td>
</tr>
<tr>
<td>II.1.4</td>
<td>Typical constant-energy surfaces for electrons and holes in cubic semiconductors</td>
<td>21</td>
</tr>
<tr>
<td>II.1.5</td>
<td>Constant energy surfaces at 0.1 eV and 0.4 eV for the first conduction band of Si</td>
<td>22</td>
</tr>
<tr>
<td>II.2.1</td>
<td>Cube in real and momentum space used for calculating particle balance in the Boltzmann equation</td>
<td>23</td>
</tr>
<tr>
<td>II.2.2</td>
<td>Cubes in k-space used for calculating balance in the Boltzmann equation</td>
<td>24</td>
</tr>
<tr>
<td>III.1.1</td>
<td>Flow chart of particle-based simulators</td>
<td>29</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>III.1.2</td>
<td>Comparison between the convergence behavior of the multi-grid and Successive Over Relaxation Poisson solvers, as applied to a HEMT</td>
<td>32</td>
</tr>
<tr>
<td>III.4.1</td>
<td>An example of the hybrid implementation of the EMC and the CMC methods and a comparison of the EMC/CMC performance</td>
<td>39</td>
</tr>
<tr>
<td>IV.3.1</td>
<td>Conduction band velocity for Si when moving along the &lt;111&gt; axis</td>
<td>44</td>
</tr>
<tr>
<td>IV.3.2</td>
<td>Conduction band velocity for Si when moving along the &lt;100&gt; axis</td>
<td>44</td>
</tr>
<tr>
<td>V.1.1</td>
<td>Energy versus field and velocity versus field data curve for InP from DAMOCLES</td>
<td>45</td>
</tr>
<tr>
<td>V.1.2</td>
<td>Average energy versus electric field for Si, GaAs and InP</td>
<td>47</td>
</tr>
<tr>
<td>V.1.3</td>
<td>Drift velocity versus electric field for Si, GaAs and InP</td>
<td>48</td>
</tr>
<tr>
<td>V.2.1</td>
<td>Drift velocity versus electric field for Si</td>
<td>49</td>
</tr>
<tr>
<td>V.2.2</td>
<td>Drift velocity versus electric field for GaAs and InP</td>
<td>50</td>
</tr>
<tr>
<td>V.2.3</td>
<td>Energy versus electric field for Si, GaAs and InP</td>
<td>51</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.1</td>
<td>Band gaps of common semiconductor materials</td>
</tr>
<tr>
<td>II.1</td>
<td>Element of the Point Group $T_d$</td>
</tr>
<tr>
<td>II.2</td>
<td>Local pseudopotential form factors for common semiconductor materials</td>
</tr>
<tr>
<td>II.3</td>
<td>Non-local pseudopotential form factors for common semiconductor materials</td>
</tr>
<tr>
<td>II.4</td>
<td>Spin orbit parameters for common semiconductors</td>
</tr>
<tr>
<td>II.5</td>
<td>Effective mass, $m^*/m_0$ for Germanium and Silicon</td>
</tr>
</tbody>
</table>
INTRODUCTION

During the past 25 years, the semiconductor industry has made great progress, implementing continually smaller devices. At the same time, the demand for accurate and fast computer simulation programs for designing, testing and optimizing new devices keeps increasing. A reliable computer model of semiconductor devices can be an important design tool. In many circumstances, new device characteristics can be determined faster and cheaper using computer simulations, than by traditional experimental techniques.

The advantages of computer simulation can be described as follows:

1- Parameters. The parameters used in any computer simulation are precisely known. In contrast, parameters such as doping profiles are only approximately known in laboratory experiments.

2- Parameter range. The range of parameters accessible to laboratory measurements is limited. In computer simulations, however, a much wider range of parameters may be used.

3- Speed. Computer simulation is much faster than laboratory measurements, particularly when a device requires manufacturing processes, such as mask making, material deposition, etc…

In spite of the advantages of computer simulation listed above, one must keep in mind that the simulation is only as good as the computer model used. This fact implies that sophisticated models have to be used in the theoretical analysis, in order to obtain accurate results. The simulation must deal with non-local phenomena, various scale-lengths, non-linear behavior, and quantities that differ by many orders of magnitude. All these aspects make the development of an efficient and robust simulator challenging.

Numerical simulation methods have been developed along with the improvement of computer systems. More specifically, the particle-based approach [Hoc88] consists of the simulation of the motion of one or more electrons inside the crystal, subject to the action of external forces due to applied electric field and scattering mechanisms. By statistically solving charge transport, using a significant portion of the population of charge carriers, a deep insight of the semiconductor device can be obtained. Among the stochastic methods developed for charge transport simulation, the most popular are the Monte Carlo (MC) simulation [Jac83, Fis88, Hoc88] and the Cellular Automaton (CA) [Kom92, Sar98]. The MC approach is now considered a well-established technique for semiconductor device simulation. However, it is also the most costly of all computational approaches. For this reason, the CA method has been developed. It is faster and
physically equivalent to the MC method. Both of these methods are computationally demanding algorithms, and require fast CPUs and large amounts of memory.

In chapter I, basic semiconductor definitions and notions used in subsequent chapters are presented. Chapter II describes the physical models of charge transport in semiconductor. A detailed overview of the band structure, which determines the electrical characteristics of semiconductors, is also given in this chapter. Two approaches for transport simulation (EMC and CMC) are presented and compared in chapter III. Results obtained with an hybrid implementation EMC/CMC, observations and comparison with published data are given in chapter IV and V. Finally, a brief presentation of new projects and future work will conclude this report.
CHAPTER I

BASIC NOTIONS AND DEFINITIONS

In this chapter, we present some of the fundamental semiconductor definitions and notions that will be used and discussed in detail in later chapters. After a brief introduction to solids and crystals, semiconductors will be studied. In particular, basic notion such as the band structure and the effective mass of a carrier will be discussed.

1- Basic crystallographic definitions

a) The Bravais lattice

A crystal is a particular example of a solid, in which atoms are arranged in a periodic fashion. That is, there is some basic arrangement of atoms, which is repeated throughout the entire solid. Figure I.1.1 presents a few simple lattice structures.

![Unit cell for three types of cubic lattice structures](image)

*Figure I.1.1: Unit cell for three types of cubic lattice structures (sc, bcc and fcc).*

The Bravais lattice [Ash81] is a fundamental concept in the description of any crystalline solid. It specifies the periodic array in which the repeated units of the crystal are arranged. The Bravais lattice summarizes only the geometry of the underlying periodic structure, regardless of what the actual units may be. A three dimensional Bravais lattice consists of all points with position vector \( \mathbf{R} \) of the form,

\[
\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3 ,
\]

where \( \mathbf{a}_1 , \mathbf{a}_2 \) and \( \mathbf{a}_3 \) are any three vectors not in the same plane, and \( n_1 , n_2 \) and \( n_3 \) are any integer values. The entire lattice can be generated with the vectors \( \mathbf{a}_i \) which are called primitive vectors [Ash81]. Thus, the point \( \sum n_i \mathbf{a}_i \) in the lattice is reached by moving \( n_i \) steps of length \( \mathbf{a}_i \) in the direction of \( \mathbf{a}_i \), for \( i = 1,2 \) and 3. Figure I.1.2 presents a simple example of a Bravais lattice.
b) The Wigner-Seitz primitive cell

A unit cell that just fills space without overlapping when translated through all lattice is called a primitive cell [Ash81]. One can always choose a primitive cell with the full symmetry of the Bravais lattice. By far the most common such choice is the Wigner-Seitz cell. The Wigner-Seitz cell about a lattice point is the region of space that is closer to that point than to any other lattice point.

In two dimensions, the Wigner-Seitz cell about a lattice point can be constructed by drawing lines connecting the point to all others in the lattice, bisecting each line, as shown in Fig. I.1.3.

Figure I.1.3: The Wigner-Seitz cell for a two dimensional Bravais lattice.
The six sides of the cell bisects the lines joining the central points to its six nearest neighboring points.

c) The reciprocal lattice

For a given set of point $\mathbf{R}$ constituting a Bravais lattice, we can choose a special value of the wave vector $\mathbf{k}$, such that the plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ will have the periodicity of the Bravais lattice. The set of all wave vectors $\mathbf{K}$ that yield plane waves with the periodicity of a given Bravais lattice is known as its reciprocal lattice [Ash81]. Analytically, $\mathbf{K}$ belongs to the reciprocal lattice of a Bravais lattice of points $\mathbf{R}$, provided the relation,

$$e^{i\mathbf{K} \cdot \mathbf{R}} = 1. \quad (1.2)$$

Note that the reciprocal of a Bravais lattice is a Bravais lattice and that the reciprocal of the reciprocal lattice is the original lattice itself.
d) The first Brillouin zone

The first Brillouin zone [Ash81] is the Wigner-Seitz primitive cell of the reciprocal lattice. Let $\mathbf{a}_1, \mathbf{a}_2$ and $\mathbf{a}_3$ be a set of primitive vectors for the direct Bravais lattice $\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$. Then the reciprocal lattice can be generated by the three primitive vectors, [Ash81],

$$
\begin{align*}
\mathbf{b}_1 &= 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \\
\mathbf{b}_2 &= 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \\
\mathbf{b}_3 &= 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}.
\end{align*}
$$

Any vector $\mathbf{k} \in \mathbf{K}$ of the reciprocal lattice can now be written as a linear combination of these three vectors,

$$
\mathbf{k} = k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 + k_3 \mathbf{b}_3,
$$

and since the reciprocal lattice is defined by $e^{i\mathbf{k} \cdot \mathbf{R}} = 1$ for $\mathbf{k} \in \mathbf{K}$,

$$
\mathbf{k} \cdot \mathbf{R} = 2\pi (k_1 n_1 + k_2 n_2 + k_3 n_3).
$$

For $e^{i\mathbf{k} \cdot \mathbf{R}} = 1$ to be true for any $\mathbf{R}$, $\mathbf{k} \cdot \mathbf{R}$ must be equal to an integer multiple of 2. This requires that the coefficients $k_i$ be integers. It follows that for a given lattice constant $a$, the dimension of the first Brillouin zone is given in units of $2\pi / a$. The Wigner-Seitz cell for the body-centered cubic Bravais lattice is a truncated octahedron. The surrounding cube is a conventional body-centered cubic cell with a lattice point at its center and on each vertex. The hexagonal faces are regular and bisect the line joining the points on the vertices. The square faces bisects the lines joining the points in the center of each of the six neighboring cubic cells (not drawn). The Brillouin zone for the diamond lattice (Si, Ge) and zincblende (GaAs, InP) is shown in Fig. II.3.1. The most important symmetry points and lines are also indicated for a crystal with lattice constant $a$.

![Brillouin Zone for zincblende and diamond lattices](Ash81)

The center of the zone $\Gamma=2\pi/a \ (0,0,0)$

The $\langle 111 \rangle$ axis ($\Delta$) and its intersection with the zone edge $L=2\pi/a \ (1/2,1/2,1/2)$

The $\langle 100 \rangle$ axis ($\Delta$) and its intersection with the zone edge $X=2\pi/a \ (1,0,0)$

The $\langle 110 \rangle$ axis ($\Sigma$) and its intersection with the zone edge $K=2\pi/a \ (3/4,3/4,0)$
2- Bonding forces in solids

The interactions of electrons in neighboring atoms of a solid are responsible for holding a crystal together. Three types of bonding in solids can be defined; they are ionic, metallic and covalent bonding.

a) Ionic bonding: insulators

In ionic bonding, all electrons are tightly bound to the atoms, so there are no free electrons available to participate in current flow. For example, in the case of NaCl, each Na atom is surrounded by six nearest Cl atoms and vice versa. In the lattice, each Na atom gives up its outer electron to a Cl atom so that the crystal is made up of ions, as shown in Fig. I.2.1a), which presents the NaCl structure. The Na⁺ ion has a positive charge, having lost an electron, and the Cl⁻ ion has a negative charge, having gained an electron. Each Na⁺ ion exerts an electrostatic attractive force upon its six Cl⁻ neighbors and vice versa. These Columbic forces pull the lattice together until a balance is reached with repulsive forces.

![NaCl structure](image)

*Figure I.2.1: Ionic bonding in NaCl a) and covalent bonding in Si b.) [Str00]*

b) Metallic bonding: metals

In a metal, the electron in the outer orbital of each atom is contributed to the crystal, so that the solid is made up of ions with closed shells immersed in a sea of free electrons. The forces holding the lattice together arise from an interaction between the positive ion cores and the surrounding free electrons. The electrons at the periphery of the metal are free to move about the crystal under the influence of external fields.
c) **Covalent bonding: semiconductors**

Covalent bonding is exhibited by the diamond lattice semiconductors, which is shown in Fig. I.2.2. In these crystals, each atom shares its four valence electrons with its four nearest neighbors. The bonding forces arise from the interaction between the shared electrons, and each electron pair constitutes a covalent bond. The electron doesn’t belong to any particular atom, but to the bond. The two electrons are indistinguishable except that they must have opposite spin to satisfy the Pauli exclusion principle, which states that no two electrons in a given interacting system may have the same quantum state. Although there are no free electrons at 0K, electrons can be thermally or optically excited out of a covalent bond and thereby become free to participate in conduction. Compound semiconductors such as GaAs have mixed bonding, in which both ionic and covalent bonding forces participate.

![Diamond lattice structure](image)

*Figure I.2.2: The diamond lattice structure a) and its unit cell b).*

A diamond lattice is obtained by placing atoms at a \((1/4,1/4,1/4)\) translated position from each atom of the fcc structure. [Str00, Ash81]

3- **Band structure**

a) **Definition**

Electrons in atoms are restricted to sets of discrete energy levels. In a similar fashion, electrons in crystals are restricted to certain energies and are not allowed at other energies. These restricted energies can be seen as energy bands that represent the so-called band structure of a crystal or a semiconductor and are represented as a function of the electron energy \(E\) with respect to its wave vector \(\mathbf{k}\). The equation relating \(E\) to \(\mathbf{k}\) is called the dispersion relation and changes with the medium. In vacuum, the electron momentum is \(\mathbf{p} = mv\). In many ways, \(\hbar\mathbf{k}\) is a natural extension of \(\mathbf{p}\) to the case of periodic structures and is known as the crystal momentum [Ash81] to emphasize this similarity, although \(\mathbf{k}\) is actually the wave vector of the electron.
In vacuum, the electron energy is given by,

$$E = \frac{1}{2} m v^2 = \frac{1}{2} \frac{p^2}{m} = \frac{\hbar^2}{2m} k^2.$$  \hspace{1cm} (1.6)

Thus the free electron energy is parabolic with wave vector $k$. In a crystal, the dispersion relation is more complicated but the parabolic shape of the band structure can still be observed at low energies. A simplified example of a crystal band structure is shown in Fig. I.3.1. A certain number of bands may be completely filled with electrons, all others remaining empty. The highest occupied energy level at 0K is called the valence band; the lowest unoccupied energy level is called the conduction band. The difference between the top of the valence band and the bottom of the conduction band is known as the band gap and is characterized by its energy $E_g$.

![Figure I.3.1: The simplified band structure of a crystal.][Str00]

Only the bands that are almost empty or full of electrons are involved in charge transport phenomena. Thus, it is the band structure of a semiconductor that is responsible for its electrical characteristics. In order for electrons to experience acceleration in an applied electric field, they must be able to move into new energy states, that is to say there must be empty states available for the electrons to move into. For instance, Si at 0K has a valence band completely filled with electrons and an empty conduction band. There can be no charge transport within the valence band, since no empty states are available. The same applies for the conduction band. Thus, Si at 0K behaves exactly as an insulator. In fact, all semiconductor materials at 0K have the same structure as insulators. The difference lies in the size of the band gap $E_g$, which is much smaller in semiconductors than in insulators. (Si $E_g = 1.1$ eV, Diamond $E_g = 5$ eV). The relatively small band gap of semiconductors allows for the excitation of electrons from the lower valence band to the upper conduction band by thermal or optical energy, which increases the number of available states for conduction. The band gaps of some common semiconductors are listed in Table I.1 at two different temperatures.
Table I.1
Band gaps of common semiconductor materials. [Sze81]

<table>
<thead>
<tr>
<th></th>
<th>Si</th>
<th>Ge</th>
<th>GaAs</th>
<th>InP</th>
<th>GaN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_g$ [eV] 77K</td>
<td>1.21</td>
<td>0.744</td>
<td>1.51</td>
<td>1.42</td>
<td>-</td>
</tr>
<tr>
<td>$E_g$ [eV] 300K</td>
<td>1.12</td>
<td>0.664</td>
<td>1.42</td>
<td>1.34</td>
<td>3.39</td>
</tr>
</tbody>
</table>

In the case of metals, the conduction and the valence bands overlap or are partially filled. Thus electrons can move freely under the influence of an external field. Figure I.3.2 shows the difference between insulators, conductors and semiconductors, in terms of band structure.

![Figure I.3.2: Typical band structures at 0K.](image)

b) Direct and indirect transitions

The band structure of GaAs has a minimum in the conduction band and a maximum in the valence band for the same momentum value ($k=0$). On the other hand, the valence band of Si is maximal at a different value of $k$ than the conduction band minimum. Thus, an electron making a transition from the conduction band to the valence band in Si requires some change in energy and in $k$, whereas only a change in energy is required for GaAs. This defines two classes of semiconductor energy bands, indirect and direct transitions, which are illustrated in Fig. I.3.3. In a direct transition, an electron in the conduction band can fall to an empty state in the valence band, giving off the energy difference $E_g$ as a photon of light. This process is more complicated for indirect transitions, because it also involves a change in $k$. Since phonons have no non-relativistic mass, their momentum is negligible and they cannot account for the required change of momentum. An electron going from the conduction band to the valence band may instead go first through some defect state $E_t$ within the band gap. In this type of transition, the energy is generally given up as heat to the lattice rather than as an emitted photon.
As seen in the previous sections, the periodicity of a crystal lattice induces a periodic potential that interacts with electrons. As a result, the “wave-particle” motion of electrons in a crystal cannot be expected to be the same as in free space. Thus, in applying the usual equations of electrodynamics to charge carriers in a solid, we must use altered values of the particle mass. In doing so, most of the influences of the lattice are taken into consideration, so that electrons and holes can be treated as “almost free” carriers in most computations. For this purpose, we introduce the effective mass \( m^* \) of an electron that takes into account the shape of the energy bands in three-dimensional \( k \)-space, taking appropriate averages over the various energy bands. In vacuum we have from Eq. 1.6,

\[
\frac{d^2E}{dk^2} = \frac{\hbar^2}{m}.
\]

Although electrons in solids are not free, most energy bands are close to parabolic at their minima (for conduction bands) or maxima, (for valence bands) and we define the effective mass of an electron in a crystal as

\[
m^* = \frac{\hbar^2}{d^2E/dk^2}.
\]

From this definition, we see that the curvature of the band determines the electron mass. The curvature \( d^2E/dk^2 \) is positive at the conduction band minimum but is negative at the valence band maximum (see Fig. I.4.1). Electrons near the top of the valence band have a negative effective mass. Valence band electrons with negative charge and negative mass move in an electric field in the same directions as holes with positives charges and positive mass. We can fully account for charge transport in the valence band by considering hole motion.
5- Drift of carriers in electric fields

Knowledge of carrier concentration in a solid (n electrons and p holes per cubic centimeters) is necessary for calculating current flow in the presence of electric or magnetic fields. In addition to these concentration values, the collisions of the charge carriers with the lattice and with the impurities must be taken into account. These processes will affect the ease with which electrons and holes can flow through the crystal, that is, their mobility within the solid. Collisions and scattering processes depend on temperature, which affect the thermal motion of the lattice atoms, and the velocity of the carriers.

a) Carriers in an electric field

The charge carriers in a solid are in constant motion, even at thermal equilibrium. At room temperature, the thermal motion of an individual electron can be visualized as random scattering from lattice impurities, other electrons, and defects of the lattice structure. Since the scattering is random, there is no net motion of the group of electrons over any period of time. For an individual electron, the probability of returning to its starting point after some period of time \( t \) is negligibly small. However, if a large number of electrons is considered, (typically \( 10^{16} \) cm\(^{-3} \) in a n-type semiconductor), there will be no preferred direction of motion for the group of electrons and no net current flow. If an electric field \( E_x \) is applied in the x-direction, each electron experiences a net force \( (-qE_x) \) from the field. This force may be insufficient to alter appreciably the random path of an individual electron; the effect when averaged over all electrons however is a net motion of the group in the negative x-direction. If \( p_x \) is the x-component of the total momentum of the group, the force of the field on the electrons is [Str00],

\[
\frac{dp_x}{dt}_{\text{field}} = -nqE_x. \tag{1.9}
\]

The previous equation indicates a continuous acceleration of the electrons in the negative x-direction. This is not actually the case, because in a crystal at room temperature, this
acceleration is balanced by the deceleration due to scattering. Thus, the net rate of change of momentum must be equal to zero in steady state, when taking into account the collision processes.

For a group of $N_0$ electrons at time $t=0$, we define $N(t)$ the number of electrons that have not undergone a collision by time $t$. The rate of decrease in $N(t)$ at any time $t$ is proportional to the number left unscattered at $t$ and has the following expression, [Str00],

$$N(t) = N_0 e^{-t/\tau},$$  \hspace{1cm} (1.10)

where $\tau$ represents the mean time between scattering events, called the mean free time. The probability that an electron has a collision in the time interval $dt$ is $dt / \tau$. Thus the differential change in $p_x$ due to collision in time $dt$ is

$$dp_x = -p_x \frac{dt}{\tau}. \hspace{1cm} (1.11)$$

The rate of change of $p_x$ due to the decelerating effect of collisions is

$$\left. \frac{dp_x}{dt} \right|_{\text{collisions}} = -p_x \frac{1}{\tau}. \hspace{1cm} (1.12)$$

The sum of the acceleration due to the electric field and deceleration due to the scattering is equal to zero,

$$\left. \frac{dp_x}{dt} \right|_{\text{collision}} + \left. \frac{dp_x}{dt} \right|_{\text{field}} = 0, \text{ thus } -\frac{p_x}{\tau} - nqE_x = 0. \hspace{1cm} (1.13)$$

We then define the average momentum per electron and the average net velocity,

$$\langle p_x \rangle = \frac{p_x}{n} = -q\bar{E}_x, \hspace{1cm} (1.14)$$

$$\langle v_x \rangle = \frac{\langle p_x \rangle}{m_n} = -\frac{q\bar{E}_x}{m_n}, \hspace{1cm} (1.15)$$

where the angular brackets indicate an average over the entire group of electrons. Equation 1.15 shows a constant net velocity in the negative $x$-direction, as expected. The average net velocity defines the net drift of an average electron in response to the electric field.
b) Conductivity and mobility

The current density resulting from the net drift is the number of electrons crossing a unit area per unit time, \( n \langle v_x \rangle \), multiplied by the charge of the electron, \(-q\),

\[
J_x = -qn \langle v_x \rangle = \frac{q^2 n}{m^*_n} E_x, \tag{1.16}
\]

where \( m^*_n \) is the effective mass of electrons. The electron mobility \( \mu_n \) is introduced, which describes the ease with which electrons drift in the material. It is a very important quantity characterizing semiconductor materials and is defined as \([Str00]\),

\[
\mu_n = \frac{q\bar{v}_x}{m^*_n}. \tag{1.17}
\]

The mobility can also be expressed as the average particle drift per unit velocity per unit field,

\[
\mu_n = -\frac{\langle v_x \rangle}{E_x}, \quad \text{and thus } J_x = qn \mu_n E_x. \tag{1.18}
\]

To account for the hole current, we replace \( n \) by \( p \), \( q \) by \(-q\) and \( \mu_n \) by \( \mu_p = +\langle v_x \rangle/E_x \),

\[
J_x = q(n \mu_n + p \mu_p) E_x = \sigma E_x, \tag{1.19}
\]

where \( \sigma \) is the conductivity of the semiconductor, and values of \( \mu_n \) and \( \mu_p \) can be found in tables for all common semiconductor materials.

The electron mobility is a crucial value for the purpose of this work. Although simulations are performed in the microscopic scale, the mobility accounts for the macroscopic behavior of semiconductors, such as current and material resistivity, and allows us to compare our results with published experimental data.
CHAPTER II

PHYSICAL MODELS OF CHARGE TRANSPORT IN SEMICONDUCTORS

In this chapter, we present in detail the transport models used in semiconductor simulations. We first give a detailed description of the electronic structure of a semiconductor. Then, we introduce the Boltzmann transport equation and discuss how its solution governs the semiconductor conduction. Finally, we give a brief explanation of the scattering mechanisms involved in the charge transport simulations.

1. Band Structure

a) The Brillouin zone and the irreducible wedge

The band structure of a semiconductor is represented by the dispersion relation $E=E(k)$, where $k$ is related to the momentum of the electron by the relation $p = \hbar k$. Thus, the band structure diagram is also called the energy-momentum diagram. It is usually obtained by solving the Schrödinger equation of a one-electron problem. Due to the periodicity of the lattice, one can show that the energy $E(k)$ is periodic in the reciprocal lattice. Therefore, it is sufficient to use only $k$ in a primitive cell of the reciprocal lattice, conventionally, the first Brillouin zone.

![Figure II.1.1: Sampling region for the calculation of the band structure. The region is 1/48 of the total Brillouin zone and is called the irreducible wedge.](image)

In actual calculations, it can be costly to compute the dispersion relation $E(k)$ for all points of the Brillouin zone. However, due to its high symmetry properties, it is only necessary to do computations in a small portion of the Brillouin zone, and then apply symmetry operations to obtain $E(k)$ everywhere. The minimal volume in which $E(k)$ is computed is called the irreducible wedge [Hes99], and is shown in Fig.II.1.1 for the
diamond structure. It corresponds to 1/48 of the total volume of the Brillouin zone and is defined by,

\[ 0 \leq k_z^* \leq k_y^* \leq k_x^* \leq 1, \quad \text{and} \quad k_z^* + k_y^* + k_x^* \leq \frac{3}{2}, \tag{2.1} \]

where the \( k^* \) are the momentum coordinates normalized by \( 2\pi/a \). (for example \( k_z^* = \frac{k_z}{2\pi/a} \))

<table>
<thead>
<tr>
<th>Table II.1</th>
<th>Element of the Point Group ( T_d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q_1(x_1x_2x_3) )</td>
<td>( Q_2(x_1\bar{x}_2\bar{x}_3) )</td>
</tr>
<tr>
<td>( Q_5(x_2x_3x_1) )</td>
<td>( Q_6(\bar{x}_2x_3\bar{x}_1) )</td>
</tr>
<tr>
<td>( Q_9(x_3x_1x_2) )</td>
<td>( Q_{10}(\bar{x}_3\bar{x}_1x_2) )</td>
</tr>
<tr>
<td>( Q_{13}(\bar{x}_1x_1\bar{x}_2) )</td>
<td>( Q_{14}(\bar{x}_1\bar{x}_1\bar{x}_2) )</td>
</tr>
<tr>
<td>( Q_{17}(x_2\bar{x}_1\bar{x}_3) )</td>
<td>( Q_{18}(\bar{x}_2x_1\bar{x}_3) )</td>
</tr>
<tr>
<td>( Q_{21}(x_3\bar{x}_1x_2) )</td>
<td>( Q_{22}(\bar{x}_3\bar{x}_1x_2) )</td>
</tr>
</tbody>
</table>

The set of symmetry operations (rotations, reflections, etc…) used to project the energy found in the irreducible wedge to the entire BZ, transform a crystal lattice into itself and is called the \textit{point group} of the lattice [Mor69]. Forty-eight such operations transform a cube into itself. Table II.1 shows 24 operations that form the subgroup \( T_d \). In this table, the operation \( Q_3(\bar{x}_1x_2\bar{x}_3) \), for example, means \( Q_3 f(x_1,x_2,x_3) = f(-x_1,x_2,-x_3) \) for any function of the coordinates \((x_1,x_2,x_3)\). The 24 remaining operations can be derived from these by applying the inversion \( Q_0 f(\bar{x}_1,\bar{x}_2,\bar{x}_3) \).

One of the main requirements to correctly model charge transport in a semiconductor is an accurate representation of the band structure of the crystal. The accuracy of the band structure model is crucial because it governs the calculation of the classical motion of particles and their quantum-mechanical scattering rate. There is a variety of numerical techniques that have been developed to compute the band structure of semiconductors. Among those methods, the three most frequently used are the pseudopotential method [Phi58], the orthogonalized plane-wave method [All55, Her55], and the \( k \cdot p \) method [Kan56]. The method used in this work is the pseudopotential approach and is discussed in the following sections.
b) Approximations of the conduction band:

There are three levels of approximations, the \textit{parabolic}, the \textit{non-parabolic}, and the \textit{full band} representation of the dispersion relation of a semiconductor. In the region around the minimum of the conduction band, which is taken to occur around \( k=0 \) to simplify the following discussion, the dispersion relation, \( E(k) \) can be approximated by a quadratic function of \( k \). This approximation defines the parabolic band structure as,

\[
E(k) = \sum_i \sum_j \frac{\hbar^2}{2} \left( \frac{1}{m_{i,j}} \right) k_i k_j
\]  (2.2)

where \( \left( \frac{1}{m_{i,j}} \right) = \frac{1}{\hbar^2} \left( \frac{\partial^2 E}{\partial k_i \partial k_j} \right) \) is the inverse effective mass tensor, and \( k \) is measured from the center of the valley. In the simplest analytic model, \( E(k) \) takes the following parabolic form,

\[
E(k) = \frac{\hbar^2 k^2}{2m}.
\]  (2.3)

This representation is only valid for low energies, near the valley center and for values of \( k \) farther from the minimum of the conduction, the parabolic approximation doesn’t hold any longer. The non-parabolic approximation modifies the previous expression by allowing the energy, \( E(k) \), to satisfy the following equation,

\[
E(1+\alpha E) = \frac{\hbar^2 k^2}{2m} = \gamma(k),
\]  (2.4)

which results in the following non-parabolic solution,

\[
E(k) = \frac{-1 + \sqrt{1 + 4\alpha \gamma(k)}}{2\alpha},
\]  (2.5)

where \( \gamma(k) \) is the parabolic approximation of \( E(k) \), (from Eq.2.3), and \( \alpha \) is a non-parabolicity parameter that can be obtained by the \( k \cdot p \) theory.

Although recent simulation approaches use approximations like the one described above, it is not always sufficient for accurate numerical simulations. An analytical approximation of the band structure may only hold under low electric field conditions. To investigate the high-field behavior of semiconductors, a full band representation of the energy dispersion is often necessary. The need for full band representation and the example of several semiconductor materials will be presented and discussed later in this chapter.
c) Approximations of the valence band

In most cubic semiconductors, the valence bands are very similar one to another. The two upper most bands are referred as the heavy-hole and light-hole bands, and are generally degenerated about \( k = 0 \), and the lower one as the split off or spin orbit band, as shown in Fig. II.1.2.

The compound GaAs structure is a direct material, with a band gap of 1.43 eV at room temperature. The conduction band minimum (\( k = 0 \)) is called \( \Gamma \). There are also two higher-lying indirect minima in the GaAs conduction band, L (lower) and X (upper). The heavy-hole and the light-hole valence bands are degenerated. The bottom spin-orbit valence band is lower by about 0.34 eV owing to the spin-orbit splitting.

**Figure II.1.2: The GaAs band structure.**

The spherical parabolic expression of Eq.2.3 is appropriate for the split off valence band, while for the two degenerate valence bands, the energy around the maximum is given by a more elaborate expression that account for the warped equienergetic surfaces, but does not allow the definition of an effective mass tensor,

\[
E(k) = ak^2 [1 \mp g(\vartheta, \phi)],
\]

where \( \mp \) refers to heavy and light holes, respectively, \( \vartheta \) and \( \phi \) are the polar and azimuthal angles of \( k \) with respect the crystallographic axis. The function \( g(\vartheta, \phi) \) contains the angular dependence of the two degenerate bands, given by [Ott75],

\[
g(\vartheta, \phi) = \left[ b^2 + c^2 (\sin^4 \vartheta \cos^4 \phi + \sin^2 \vartheta \sin^2 \phi \cos^2 \vartheta) \right]^{1/2},
\]

where,

\[
a = \frac{\hbar^2}{2m_0} |A|, \quad b = \frac{|B|}{|A|}, \quad c = \frac{|C|}{|A|},
\]

and A, B, C are the inverse valence band parameters, given in [Dre55].
d) Full band representation

Any solid can be thought of as a set of tightly bound spherical ion-cores sitting in the electron system formed by the valence, or conduction electrons. These outer electrons are responsible for all ordinary physical and chemical properties of the solid. One would ideally like to describe the interaction of an outer electron with the ion and its surrounding cloud of electrons in such terms that one can calculate the whole range of solid state properties. Computing the real ionic potential turns out to be an impossible task and the theory of the pseudopotential approximation has shown to be an efficient approach to determine the full band structure representation. In this method, the real ionic potential is replaced by a pseudopotential that matches the real potential at the boundaries and that interpolates it in the interstitial regions away from the ion-core. Therefore, we can obtain the band structure once we are given an appropriate model potential which is easier to use in this calculation than the real ion potential.

The pseudopotential concept was introduced by J.C. Phillips in 1958 [Phi58]. The Empirical Pseudopotential Method (EPM) initially was based on a simplified local approximation [Coh70]. The word “empirical” refers to the method chosen to determine the form of this model pseudopotential and the logarithmic derivatives at the core / interstitial-region interface. In this approach, experiments have played a prominent role in determining the theoretical parameters involved in EPM calculations. The term “local” expresses the fact that this pseudopotential is only a function of position. The local approximation has been proven to be sufficient to explain experimental data available for semiconductors. However, if the calculation is extended to the valence bands and if we compare the results to the experiment, discrepancies can arise [Pol73, Gro72]. In addition, band topologies and critical optical point symmetries as calculated by a local EPM for Ge and GaAs have been found to be in error. Owing to the nature of these discrepancies, it was speculated that purely local pseudopotential techniques could not yield satisfactory results, and that an energy dependent and non-local pseudopotential should be considered.

The pseudopotential of an ion of index $\alpha$ (anion or cation) is usually described as [Che76],

$$V_p^\alpha (r) = V_L^\alpha (r) + V_{NL}^\alpha (r, E), \quad (2.9)$$

where the local term $V_L$ is a simple function of position, and the non-local term $V_{NL}$ is expressed as,

$$V_{NL} (r, E) = \sum_{l=0}^{\infty} A_l^\alpha (E) f_l^\alpha (r) \Pi_l, \quad (2.10)$$

where $A_l^\alpha (E)$ is an energy-dependent well depth, $f_l^\alpha (r)$ is a function simulating the effect of core states with $l$ symmetry, and $\Pi_l$ is a projection operator of the $l^{th}$ angular momentum component. The pseudopotentials computation involves only a small number of parameters needed in the local and non-local approximations.
The local parameters used for the calculation of the band structure of Ge, Si, GaAs, InP and GaN are listed in Table II.2, and the non-local parameters are listed in Table II.3.

**Table II.2 [Pot81, Hes91]  
Local pseudopotential form factors for common semiconductor materials (in Ry)**

<table>
<thead>
<tr>
<th></th>
<th>$a$ (Å)</th>
<th>$V_d(3)$</th>
<th>$V_d(4)$</th>
<th>$V_d(8)$</th>
<th>$V_d(11)$</th>
<th>$V_d(3)$</th>
<th>$V_d(4)$</th>
<th>$V_d(8)$</th>
<th>$V_d(11)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>5.43</td>
<td>-0.257</td>
<td>-0.203</td>
<td>-0.040</td>
<td>0.033</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ge</td>
<td>5.65</td>
<td>-0.236</td>
<td>-0.160</td>
<td>0.019</td>
<td>0.056</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GaAs</td>
<td>5.64</td>
<td>-0.214</td>
<td>-0.157</td>
<td>0.014</td>
<td>0.067</td>
<td>0.055</td>
<td>0.038</td>
<td>0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>InP</td>
<td>5.86</td>
<td>-0.235</td>
<td>-0.176</td>
<td>0.000</td>
<td>0.053</td>
<td>0.080</td>
<td>0.060</td>
<td>0.033</td>
<td>0.030</td>
</tr>
<tr>
<td>GaN</td>
<td>4.50</td>
<td>-0.320</td>
<td>0.000</td>
<td>0.030</td>
<td>0.060</td>
<td>0.250</td>
<td>0.210</td>
<td>0.000</td>
<td>0.020</td>
</tr>
</tbody>
</table>

**Table II.3 [Pot81, Hes91]  
Non-local pseudopotential form factors for common semiconductor materials**

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_c$</th>
<th>$\alpha_d$</th>
<th>$\beta$</th>
<th>$R_s$ (Å)</th>
<th>$R_d$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>0.550</td>
<td>-</td>
<td>0.000</td>
<td>1.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Ge</td>
<td>0.000</td>
<td>-</td>
<td>0.295</td>
<td>0.000</td>
<td>0.00</td>
</tr>
<tr>
<td>GaAs</td>
<td>0.000</td>
<td>0.000</td>
<td>0.625</td>
<td>0.125</td>
<td>1.06</td>
</tr>
<tr>
<td>InP</td>
<td>0.300</td>
<td>0.000</td>
<td>0.350</td>
<td>0.550</td>
<td>1.06</td>
</tr>
<tr>
<td>GaN</td>
<td>0.000</td>
<td>0.000</td>
<td>2.4273</td>
<td>2.4273</td>
<td>1.06</td>
</tr>
</tbody>
</table>

For heavier elements, the spin-orbit interaction becomes rather large. In particular, the structure and splitting of degeneracies at the top of the valence band at the zone center ($\Gamma$ point) is strongly affected by this interaction. An example is CdTe where the energy bands split by nearly 1 eV at the valence-band maximum. Thus, since we are also interested in the transport properties of holes, particularly at low energies, it is necessary to include an additional term in order to obtain the correct effective masses, non parabolicity, and warping of the valence bands. The spin-orbit parameters are given in Table II.4. Figure II.1.3 shows the results of the EPM calculations obtained with the parameters mentioned above for several common semiconductor materials, in the irreducible wedge. The plots for Ge, GaAS and InP are calculated with non-local spin-orbit interaction and the local interaction has been used for the Si and GaN calculation.

**Table II.4 [Pot81, Hes91]  
Spin orbit parameters for common semiconductor materials:**

<table>
<thead>
<tr>
<th></th>
<th>$\zeta_{\text{anion}}$</th>
<th>$\zeta_{\text{cation}}$</th>
<th>$\alpha$</th>
<th>$\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>4.60</td>
<td>-</td>
<td>1.00</td>
<td>0.000157</td>
</tr>
<tr>
<td>Ge</td>
<td>5.34</td>
<td>-</td>
<td>1.00</td>
<td>0.000930</td>
</tr>
<tr>
<td>GaAs</td>
<td>5.34</td>
<td>5.34</td>
<td>0.334</td>
<td>0.000572</td>
</tr>
<tr>
<td>InP</td>
<td>4.60</td>
<td>4.95</td>
<td>3.599</td>
<td>0.001700</td>
</tr>
<tr>
<td>GaN</td>
<td>5.34</td>
<td>5.34</td>
<td>0.334</td>
<td>0.001150</td>
</tr>
</tbody>
</table>
Figure II.1.3: Full band structure representation for common semiconductor materials: a) Silicon, b) Germanium, c) Gallium Arsenate d) Indium Phosphate e) Gallium Nitrate
e) 3D representation:

The band structure diagrams in Figure II.1.3 are obtained by plotting allowed values of the energy versus the propagation constant \( k \) (crystal momentum. Since the full relationship between \( E \) and \( k \) is a complex surface, it is often helpful to visualize it in three dimensions. In the case of the parabolic approximation, the energy has the form of Eq. 2.14 assuming the effective mass is a scalar quantity and has the same value in all three directions (spherical case). However, if the effective mass is a tensor, then the equi-energetic surface may be elliptical,

\[
E(k) = \frac{\hbar^2}{2} \left[ \frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z} \right],
\]

where \( k_i \) and \( k_t \) are the longitudinal and the transverse components of \( k \) respectively, and \( 1/m_l \) and \( 1/m_t \) are the longitudinal and the transverse components of the inverse effective mass tensor respectively. The effective masses of Ge and Si are shown in Table II.5.

| Effective mass, \( m^*/m_0 \) for Germanium, and Silicon. [Sze81] |
|-----------------|-----------------|-----------------|-----------------|
| Electrons       | Holes           |
| Ge              | Ge              | Ge              | Ge              |
| \( m_i^* \) = 1.64 | \( m_i^* \) = 0.082 | \( m_{ih}^* \) = 0.44 | \( m_{ih}^* \) = 0.28 |
| Si              | Si              | Si              | Si              |
| \( m_i^* \) = 0.98 | \( m_i^* \) = 0.19 | \( m_{ih}^* \) = 0.16 | \( m_{ih}^* \) = 0.49 |

As seen in the previous section, the parabolic approximation is multiplied by a warping factor \( g(\vartheta, \phi) \) for the case of the valence band. The different shapes of the constant-energy surfaces for the three cases considered above are shown in Fig. II.1.4.

![Diagram](image)

\[ E(k) = \frac{\hbar^2 k^2}{2m} [1 - g(\vartheta, \phi)] \]

a) Spherical  b) Ellipsoidal  c) Warped

*Figure II.1.4: Typical constant-energy surfaces for electrons and holes in cubic semiconductors. [Jac89]*
Although many simulation approaches use, at least partially, approximations like the ones presented in Fig. II.1.4, the analytical energy band approach often fails under high electric field conditions. Figures II.1.5 a) and b) show a 3-D contour plot of the constant energy surfaces for the first conduction band of silicon to illustrate the deviation from the analytical model. For an energy contour level of 0.1 eV, the equi-energetic surfaces are ellipsoids and can still be reasonably represented with the analytical description. However, for a contour level set at 0.4 eV, the constant energy surfaces are no longer ellipsoids and the analytical model fails. This explains why the full band representation is crucial for accurate high field simulations.

Figure II.1.5: Constant energy surfaces at a) 0.1 eV and b) 0.4 eV for the first conduction band of Si.

2- The Boltzmann Equation (semiclassical theory of charge transport)

a) The Boltzmann transport equation

The equation that describes semiconductor transport phenomena in semiclassical terms is the Boltzmann Transport Equation (BTE) [Rei65],

\[
\frac{\partial f}{\partial t} = - \mathbf{v} \cdot \nabla f - k \nabla_{\mathbf{v}} f + \frac{\partial f}{\partial t}_{\text{collisions}},
\]  

(2.12)
where the effective distribution function $f(k,r,t)$ is a function of the position of the electron $r$, its momentum $k$ and time, and the velocity $v$ is defined by $v = \frac{\partial r}{\partial t}$. Once the time evolution of $f$ (\(\frac{\partial f}{\partial t}\)) is known, all physical quantities, such as the drift velocity of carriers and their mean energy can be obtained as functions of the electric field, the lattice temperature, and the carrier concentration gradient (i.e. the change of doping in the semiconductor with respect to space). In this equation, the displacement of charge carriers follows the classical laws of Newtonian mechanics. However, the collision term is derived from the quantum theory of scattering. Thus, the charge transport theory used in the present work is said to be a semi-classical approach.

b) Change of particle number due to carriers displacement

An approach for calculating particle balance in the Boltzmann equation is to consider the seven dimensional phase space $(k:3, r:3, t:1)$ as an elementary cube at position $r$ in the real space and an elementary cube at position $k$ in the momentum space, as shown in Fig.II.2.1.

![Figure II.2.1: Cube in real a) and momentum b) space used for calculating particle balance in the Boltzmann equation.](image)

One first calculates the numbers of electrons, entering the cube in the left $dydz$ plane, and the number of electrons leaving the corresponding plane to the right of the cube in an interval of time $dt$. Since the distance traveled in the $x$-direction by electrons with velocity $v$ is $v_x dt$, one has

\[
\text{Incoming: } f(k, r, t)dkdzyv_x dt,
\]

\[
\text{Outcoming: } f(k, r(x+dx, y, z), t)dkdzyv_x dt.
\]

The particle gain in the $x$-dimensions is therefore

\[
-v_x[f(k, r(x+dx, y, z), t)dydzdk - f(k, r, t)dydzdkdt = -v_x\frac{\partial f}{\partial x} dydzdk dt.
\]
The particle gain is obtained in a similar manner in the \( y \)- and \( z \)-direction. In three dimensions, the net particle gain is \(-\mathbf{v}.\nabla f d\mathbf{k} d\mathbf{r} / dt\), where the velocity \( \mathbf{v} \) is given by

\[
\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{1}{\hbar} \nabla_k E(\mathbf{k}),
\]

where \( E(\mathbf{k}) \) is the dispersion relation and \( \nabla_k = \left( \frac{\partial}{\partial k_x}, \frac{\partial}{\partial k_y}, \frac{\partial}{\partial k_z} \right) \).

In a similar manner, one can obtain the change of the number of electrons at \( \mathbf{k} \) in momentum space, replacing \( dx \) by \( dk_x \), and \( v_x = \frac{dx}{dt} \) by \( \frac{dk_x}{dt} \), so that

\[
-\frac{dk}{dt} \nabla_k f d\mathbf{k} d\mathbf{r} / dt,
\]

where \( \frac{dk}{dt} = e\mathbf{E} \) and \( \mathbf{E} \) is the electric field and \( e \) is the electron charge.

c) Change of particle number due to scattering

Electrons can be scattered by phonons, lattice impurities, etc… which can change their wave vectors from \( \mathbf{k} \) to \( \mathbf{k}' \) (for a given position \( \mathbf{r} \) in real space). Figure II.2.2 shows two infinitesimal volumes in \( \mathbf{k} \) space to illustrate the scattering events.

\[\text{Figure II.2.2: Cubes in } k\text{-space used for calculating balance in the Boltzmann equation.}\]
The outgoing electrons, that is the electrons scattering from state \( k \) to state \( k' \) are given by

\[
out = -\sum_{k'} S(k, k') f(k, r, t) dkdrt,
\]

where \( S(k, k') \) is the scattering probability. The effective distribution function \( f(k, r, t) \) allows for the fact that an electron has to be in state \( k \) in order to be scattered out of it. The incoming electrons, that is the electrons coming into state \( k \) form some other state \( k' \) are given by

\[
in = \sum_{k'} S(k', k) f(k', r, t) dkdrt.
\]

The Boltzmann equation is obtained by balancing the particle numbers and the change in \( f \) given by the net change of incoming and outgoing particles. One has, from Eq. 2.12, 2.18 and 2.19,

\[
\frac{\partial f(k, r, t)}{\partial t} = -v \cdot \nabla_r f(k, r, t) - \frac{1}{\hbar} F_0 \cdot \nabla_k f(k, r, t) + \sum_{k'} \left[ f(k', r, t) S(k', k) - f(k, r, t) S(k, k') \right],
\]

where \( F_0 \) is the force (\( = -eE \) for an electric field \( E \)). Equating the third term in Eq. 2.20 with the collision term \( \frac{\partial f}{\partial t}_{\text{collision}} \) results in the BTE given in Eq. 2.12.

Finding solutions to the Boltzmann equation is not a trivial task. Even in the case of linear solutions, with simple scattering mechanisms, approximations are needed. From the analytical point of view, transport phenomena in nonlinear regimes would be completely described by solving the Boltzmann equation. This is however a very complex mathematical problem whose analytical solution still remains unfound today. The introduction of numerical techniques to solve the BTE has been a big step forward. Among these methods, particle-based approaches are extremely popular, and will be discussed in more detail in the following chapter.

d) Validity of the BTE

To ensure the validity of the semi-classical transport theory, the momentum uncertainty must be much smaller than the average momentum of the carrier population and at the same time, the position uncertainty must be much smaller than the mean free path.
\[ l = v \tau = (p/m) \tau \] where \( \tau \) is the mean time between two collisions. These two conditions, \((\Delta p \ll p \text{ and } \Delta x \ll l)\) must be confronted [Jac89] with Heisenberg uncertainty principle,

\[ \hbar = \Delta p \Delta x \ll pl = \frac{p^2}{m} \tau \sim k_B T \tau, \quad (2.22) \]

where \( k_B \) is the Boltzmann constant and \( T \) the absolute temperature. Therefore, the semi-classical approach holds when

\[ \tau \gg \frac{\hbar}{k_B T} \sim 10^{-13} - 10^{-14} \text{sec}. \quad (2.23) \]

Collisions occur with a frequency that is lower than the one expressed above, therefore, the requirement concerning the momentum and position uncertainty is usually fulfilled.

### 3- Scattering mechanisms

In the semi-classical approach, the dynamics of the electron interaction are assumed to be independent of the applied field, and the collisions are assumed to be instantaneous. The electronic transitions of interest for charge transport in semiconductors can be classified as *intravalley*, when the initial and final states lie in the same valley, or *intervalley*, when they lie in different valleys. The most important sources of scattering are phonons, crystal impurities and other carriers, and are described below.

#### a) Phonon scattering

Scattering due to the lattice vibrations, the phonons, often dominates the mobility of semiconductors, at least around room temperature, and is treated with the perturbation theory. Two types of phonon scattering can be distinguished, *acoustic* and *optical* scattering, depending on whether the electron scatters with an acoustic or an optical phonon respectively. Acoustic phonons are related to the acoustic (i.e. in phase) mode of vibration of the lattice; optical phonons are related to the optical mode (i.e. out of phase) mode of vibration of the lattice. We assume that the lattice vibration is a linear combination of these two modes.

#### b) Impurity scattering

Ionized and neutral impurities in the crystal can also be a source of scattering. The effect of ionized impurity scattering decreases as the temperature of the crystal rises. This fact is accompanied by the increasing importance of phonon scattering under high
temperature. Furthermore, at any lattice temperature, as the field is increased, the effect of Coulomb scattering decreases when electrons are heated by the field. Neutral impurities have a very small cross-section at normal concentrations. Thus, their influence on transport processes is minimal and will be neglected in this work.

c) Carrier-carrier scattering

Among the scattering mechanisms of interest in transport theory, carrier-carrier interaction is the most difficult to treat, because in this case, the scattering probability depends on the distribution of electrons with respect to time, momentum and space, which is unknown and make this relation non-linear. In a process where an electron is scattered by another electron, (electron-electron scattering), the total momentum of the electron gas is not changed. Thus, electron-electron scattering has little influence on the mobility. However, since it is always combined with another scattering mechanism, which it may enhance, it can have quite an important influence. We account for the long-range carrier-carrier interaction in the self-consistent solution of the Poisson equation. The short-range Coulombic interactions are included in the screening of other scattering mechanisms.
CHAPTER III

PARTICLE-BASED APPROACH FOR TRANSPORT SIMULATION

In this chapter, we present the self-consistent particle-based method and its application to semiconductor simulation. Then, we will introduce the particular method used in this work along with the similarities and differences with traditional Monte Carlo approaches.

1- The self-consistent particle-based method

a) Particle-based methods

As for all differential equations, solving the Boltzmann transport equation requires initial values and boundary conditions. Computer simulation may be regarded as the theoretical exercise of numerically solving an initial-value, boundary-value problem. At time $t=0$, the initial state of the system is specified in some finite region. The main part of the calculation is the time-step cycle, in which the state of the system is incremented forwards in time by a small time-step, $\delta t$. Although the amount of data that can be handled by computers is large, it is finite. Much of the effort of computational scientists, and of this work, is devoted to obtaining good simulation models of the physical systems with the available computer resources, which frequently means simulating only a representative domain. Methods of discretization used in obtaining simulation models include finite-difference methods and finite-element methods.

Particle-based simulation methods [Hoc88] are applied to any system that is modeled as a set of separate elements, each of them singularly considerable. Thus, the model of particle-based simulation can be formalized as describing a set of particles interacting via fields. In semiconductor simulations, typical attributes of a carrier are its charge, mass, position, momentum and energy.

As previously mentioned, in particle-based simulations, the time evolution of a simulated population is followed through phase space, subject to its environment. The simulated population corresponds to a significant portion of charges carriers, typically holes and electrons, and the simulated environment is a representation of the semiconductor material, and its interface with the external world. All the elements of the simulated population are tracked in the phase space to obtain the distribution function, which is the statistical solution to the problem. The phase space is generally represented as a continuous six-dimensional space, where each particle carries a charge and a state $(r,k)$ that changes dynamically due to the interaction with the material, and external perturbations. The main algorithm is generally implemented in a cycle over the carrier population. During the cycle, the position and momenta of the particles are continuously tracked and updated accordingly.
Figure III.1.1 presents the algorithm of a typical particle-based semiconductor simulation program. After an initialization step, where material characteristics, boundary conditions, external forces, etc… are initialized, the carrier distribution is computed as a function of the semiconductor domain. This carrier distribution in position space defines an electric field that is computed by solving Poisson’s equation. Within the carrier dynamics portion of the algorithm, the new field is used to update the carrier position in real space and its momentum. The simulated population is then allowed to undergo scattering events. During this period of time, $\delta t$, the electric field is kept constant. This process is then repeated for the next time step until the total simulated time is reached. In other words, within the adopted scheme, the field is computed at a given instant of the simulated time, and taken as constant in the course of a time step $\delta t$. During $\delta t$, the dynamics of the carriers under the effect of the field are calculated. Since the carrier distribution in phase space changes under the effect of the electric field, and due to the interactions with the material, a new field is computed at the end of $\delta t$, according to the new distribution. Therefore, the length of such an interval is crucial [Hoc88] for the correct application of the method. The computer time required by traditional finite difference Poisson solvers is generally comparable to the time needed to simulate the dynamics in the Monte Carlo approach. Hence, the development of fast and highly efficient Poisson solvers is a key issue for particle-based simulators [Sar96].
b) Self-consistent method

Three main particle models are defined [Hoc88]: the particle-particle (PP) model, in which the force on each particle is the sum of the forces due to other particles; the particle-mesh (PM) model, which obtains the force by averaging on a discrete grid the contribution of all particles; and the particle-particle-particle-mesh (P³M) where a PP model is used to compute the short-range components of the force, and a PM model is used to determine the long-range components. The most popular model used for semiconductor simulations is the particle-mesh method. In this approach, the charge distribution due to carriers (electrons and holes) is mapped onto a discrete grid over which Poisson’s equation is solved. The resulting electric field is in turn responsible for the new displacement of the carriers, and consequently for a new charge distribution, which is again used to update the field distribution. This iterative sequence of updating the solution of Poisson’s equation and the carrier displacement is repeated until convergence is reached, and forms the so-called self-consistent scheme.

c) Poisson’s equation

The common aspect of all particle-based simulations is the combined self-consistent solution of Boltzmann’s transport equation and Poisson’s equation. The former is solved statistically via a stochastic procedure that tracks the phase-space distribution of the charge carriers in time. At discrete time steps, the resulting charge density is used as input to the Poisson solver, returning the electrostatic potential that accelerates the carriers. At each step, the Poisson solver can access the previously computed potential as an initial guess. Due to the nature of the system, it is necessary to frequently update the fields, typically on the time scale of a femtosecond [Hoc71].

Poisson’s equation is derived for Maxwell’s equation and describes the behavior of the electrostatic potential from the carrier density,

\[ \nabla \cdot (\varepsilon \nabla \Phi) = -\rho, \]  

(3.1)

where \( \Phi \) is the potential, \( \rho \) the charge density and \( \varepsilon \) the dielectric constant, assumed to be step-wise constant. In the general two-dimensional case, this equation becomes, for a homogeneous dielectric constant,

\[ \nabla^2 \Phi = f(x, y). \]  

(3.2)

By employing finite difference on a set of grid points denoted by \( \Omega_n \), this differential equation transforms into an algebraic matrix equation of the form,

\[ Au = f, \]  

(3.3)

where the vector \( u \) is the unknown potential, the matrix \( A \) represents the Laplacian operator \( \nabla^2 \) and \( f \) is the forcing function.
Among methods to solve differential equations, **iterative methods** are well adapted to the sequential nature of the simulation algorithm. They build a sequence of approximations to the required solution, starting from an initial guess, in such a way that convergence to that solution is achieved. The current potential values can be generally used as a good initial guess for the next required solution. In such a framework, a sequence of approximations \( v_0, v_1, \ldots, v_n, \ldots \) is constructed to achieve convergence to the solution \( u \) [You71], where \( v \) is the approximation to \( u \) at the \( i \)th iteration. Since the exact solution \( u \) is unknown, one may define the residual

\[
r^i = f - Av^i
\]  

(3.4)

as a computable measure of the deviation of \( v^i \) from \( u \). The **algebraic error** \( e^i \) of the approximation \( v^i \) is defined by

\[
e^i = u - v^i.
\]  

(3.5)

Subtracting Eq.3.4 from Eq.3.3 and rearranging terms, it is easily seen that \( e^i \) obeys the so-called residual equation,

\[
Ae^i = r^i.
\]  

(3.6)

Iterative methods can be interpreted as applying a relaxation operator to \( v^i \) yielding a more accurate approximation \( v^{i+1} \) by reducing the error \( e^i \). The sequence of approximations \( v_0, v_1, \ldots, v_n, \ldots \) is said to be “relaxed” to the solution \( u \).

There exist various numerical iterative methods that generate this approximation sequence, with different convergence characteristics. Among them, one can cite **Jacobi’s iteration**, [Dah74], **Gauß-Seidel’s method** [You71] and the **Successive Over Relaxation** (SOR) method [Dah74]. Despite the good results obtained with these methods, their rate of convergence is highly related to the number of grid points and is unacceptably low when applied to realistic device simulation, and are therefore inefficient Poisson solver methods.

The method employed here is derived from the SOR iterative method and is called the **multi-grid method**. The basic idea is to employ different length scales to efficiently reduce the approximation error given by Eq.3.5. Specifically, one solves exactly the residual equation, Eq.3.4, on a grid \( \Omega_{n-1} \) that is coarser than \( \Omega_n \). The resulting values of \( e^i \) is an approximation used to correct the previous approximation \( v^i \) that has been determined on the original grid \( \Omega_n \), as described by Eq.3.7,

\[
v^{i+1} = v^i - e^i.
\]  

(3.7)

The simplest version of the multi-grid algorithm is the so-called two-grid iteration, which employs only two grid levels. In the \( i \)th iteration, this procedure starts from the approximation \( v' \) of \( u \) and consists of the following five steps [Sar96],
1. Smooth $v^i$ on the grid $\Omega_n$ by applying some suitable relaxation scheme, called pre-smoothing.

2. Compute the residual from Eq.3.4 and transfer it to the coarser grid $\Omega_{n-1}$. The step is called restriction.

3. Solve exactly Eq.3.6 on the grid $\Omega_{n-1}$.

4. Interpolate the resulting $e^i$ to the finer grid $\Omega_n$, this step is called prolongation. Subsequently, calculate $v^{i+1}$ from Eq.3.7.

5. Smooth $v^{i+1}$ on the grid $\Omega_n$ by applying some relaxation method. This step is called post-smoothing.

It is possible to extend the two-grid algorithm to a sequence of grids that are increasingly coarser. This is achieved by recursively applying the complete algorithm (step 1 through 5) at step 3. The recursion stops when the coarsest grid $\Omega_0$ is reached. At that grid level, Eq.3.6 is solved exactly. Since this grid usually only contains a few points, this can easily be done. This multi-case algorithm defines one complete multi-grid iteration. The procedure is then repeated until the required convergence threshold is reached.

The multi-grid method shows good performance and is implemented as the Poisson solver in this work. It allows a faster simulation of different families of semiconductor devices with complex geometries and boundary conditions.

![Figure III.1.2: Comparison between the convergence behavior of the multi-grid and Successive Over-Relaxation Poisson solvers, as applied to a HEMT.](image)

The computer time, in seconds, is measured every iteration in the multi-grid (circles) and every 100 iterations in the SOR (squares). The convergence threshold is the maximum allowed relative difference between two successive iterations.

Figure III.1.2 presents a comparison between the classical SOR method and the multi-grid algorithm for a High Electron Mobility Transistor (HEMT) [Sar96]. In this application, the multi-grid scheme is seen to be about 30 times faster than the SOR.
Applied to a MOSFET, the multi-grid method converges about 20 times faster than the SOR method [Sar96]. Typically, the solution of Poisson’s equation by the multi-grid method requires about 1 second with a $10^{-6}$ threshold, less than half the time needed for a Monte Carlo cycle of 1 femtosecond when 150,000 particles are used.

2- The Monte Carlo method

The Monte Carlo (MC) methods belong to the most sophisticated, but at the same time the most costly, of all simulators [Zan93]. For this reason, the MC method still remains restricted to university and laboratory research, and has not yet become a common modeling tool for device design. Therefore, it is highly desirable to develop a simulator of comparable physical content and accuracy as the MC, but faster, and capable of parallel computing. The Cellular Automaton (CA) [Kom92, Sar98] was developed to meet these requirements. Very good agreement is obtained between the present CA and previously reported results from the MC simulation of the energy and velocity-field characteristics in Si, GaAs and InP.

a) Historical approach

The name of the Monte Carlo method comes from the gambling casinos in the city of Monte Carlo in Monaco. The mathematical techniques used in this method are, in fact, based on the selection of random numbers [Shr66]. In principle, the MC method can be considered as a very general mathematical tool for the solution of a great variety of problems. In the case of semiconductor device simulation, carrier-transport problems are described to some extent of accuracy, by well-defined transport equations. The MC method applied to these equations may yield more information than that obtainable from the direct solution of these equations. For instance, the MC or CA solution of the Boltzmann Transport Equation (BTE) not only gives the distribution function that verifies the equation, but also yields information on the fluctuations that are lost in the BTE itself.

The application of MC techniques to semiconductor simulation was first introduced by Kurosawa at the Semiconductor Conference held in Kyoto in 1966 [Kur66]. Since then, the method has been greatly improved [Jac83] and is widely used to obtain results for various situations in practically all semiconductor materials. Among the most significant developments of the MC techniques are the work of Price [Pri68], with the introduction of the self-scattering scheme and the extension of the method to many-particle simulations [Leb71], to harmonic fields [Pri73] and to degenerate statistics [Bos76, Lug85].
b) The MC approach:

Within the MC approach, the position of each electron and hole is tracked, continuously, both in position and momentum space. The physics of the system is defined by the band structure of the semiconductor material, and by the related scattering mechanisms.

Before the simulation, a setup procedure computes the scattering rates for each interaction or physical mechanism included in the model, and builds lookup tables. The geometrical constraints, such as interfaces, boundary conditions and doping concentrations are also defined in this initial procedure. The real simulative part of the algorithm is a sequence of push-and-scatter actions on the carrier population. The total simulated time is discretized into homogeneous steps $\Delta t$, and Poission’s equation is solved at the beginning of each interval. The electric field is then obtained by derivation of the electrostatic potential and the equations of motion are solved for each carrier. Successively, a stochastic procedure selects the scattering mechanism as a function of their tabulated probability, and computes the new position of the carriers in momentum space. The push-and-scatter sequence is repeated over the entire population of simulated carriers until the end of the time step $\Delta t$. At this point, the dynamics of the system are stopped, and the new charge distribution is used to again solve Poisson’s equation. The described procedure is the so-called self-consistent interaction, which is repeated over the entire simulation time.

An important point to be stressed is related to the time duration of the push-and-scatter procedure, which takes place within the $\Delta t$ time window between two consecutive updates of the Poisson’s equation. Two main approaches are adopted for the duration of the push-and-scatter process. The first one is the synchronous ensemble method used in the present code, which applies a prefixed time step $\Delta t$ for the numerical integration of the equations of motion during the push-and-scatter process. This way allows an easy inclusion of transient phenomena with the inclusion of carrier-carrier interactions [Fis88]. The synchronous nature of the particle population also allows a good control of the computational load due to the particles themselves and can easily be shared between different processes and/or processors. A second, more traditional, approach relies on stochastically generated $\Delta t$ [Hoc88]. These time intervals differ from each other; the simulated time of each carrier can be different within the same $\Delta t$ between two consecutive solutions of Poisson’s equation. This solution allows slightly faster code than the first one with a fixed $\Delta t$. However, due to its asynchronous attribute, the resulting programs are less prone to parallelization and debugging.

The most widely used particle-based method in semiconductor simulation is the Ensemble Monte Carlo (EMC). Whereas the Monte Carlo approach focuses on single-electron bulk simulation (assuming the bulk is ergodic), the EMC approach considers an ensemble of charge carriers, allowing for transients and devices simulations. The use of the full band structure for EMC was pioneered by K. Hess and coworkers at the University of Illinois Urbana-Champaign, to investigate high field electron transport in GaAs and Si [Hes91]. A drawback of this method is the time required for the calculation of the final state after scattering. Each time a scattering event occurs, the range of
energies compatible with this scattering is computed and a search over the entire Brillouin Zone is performed, to identify the destination cells. A random choice among these cells determines the final state of the carrier. This calculation is extremely demanding in terms of computational resources and it will be discussed in more detail in the following sections. To overcome the computational limitations due to the final state selection within the EMC formalism, a method inspired from the Cellular Automaton (CA) was been developed at the Technische Universität München by P. Vogl and coworkers [Kom92]

3- The Cellular Automaton method

a) Description

A cellular automaton is a discrete dynamical system that evolves in discrete time steps [Wol83]. Each cell of the lattice is characterized by a finite number of Boolean states. In the context of semiconductor simulations [Kom92], the Boolean states are classified as being occupied or empty. In each time step, a set of transition rules updates the state occupancies, synchronously, on all lattice cells. The important point is that these rules are local in the sense that they only act on the occupancies of a given cell and its nearest (possibly second nearest) neighbors. For this reason, the CA constitutes one of the very few algorithms for dynamical systems that can optimally utilize massively parallel processors.

There are two crucial factors that set cellular automaton apart from standard finite-difference methods for solving partial-differential equations. First, it is demonstrated [Fri86] that one can construct a set of extremely simple microscopic Boolean evolution rules that asymptotically yield distribution equations by macroscopic averaging. Thus, it is possible to replace highly non-linear partial differential equations with a large set of individual trivial problems that involve only logical and therefore, numerically stable operations and can be solved in parallel. This is one reason for the numerical efficiency of the CA. The second factor is the reduction of all physical variables to Boolean values. Microscopic transport equations contain scattering processes that depend on many physical parameters. Often, the solution is insensitive to some of these parameters and it is not necessary to store all of them with high precision.

b) CA for semiconductors simulation, the CMC method:

Within the CA formalism, every final state is tabulated for every initial state based on the appropriate conservation laws. A detailed insight is given in the next section. In this way, the choice of final state is reduced to the generation of a single random number during runtime, greatly reducing the CPU time necessary for scattering. Although the time required to determine the final state after scattering is reduced within the CA framework, the amount of memory (computer RAM) required for storing all the states can be very high, particularly in the full band description. A full band Cellular Monte Carlo (CMC) (renamed from the original CA [Kom92]) was recently [Sar02] developed using an
inhomogeneous discretization of the BZ to reduce the amount of computer storage needed to adequately conserve energy and momentum. To further improve the energy resolution, the hybrid method was developed, which implements the traditional EMC in particular regions of the BZ to reduce the amount of space required for the CMC and where a high accuracy is needed. The program in this work is this hybrid CMC/EMC implementation, coupled with an efficient multigrid Poisson solver.

4- Scattering Tables

To account for scattering mechanisms, scattering tables are built and extensively used in the simulation. The primary difference between the EMC and the CMC can be seen by looking in detail at the generation of the scattering tables in both cases. These tables store the probability of a carrier in a given state \( k \) to scatter to another state \( k' \). For each scattering mechanism, i.e. for every type of scattering defined in Chapter II, we define \( P_j(k, k') \) as the probability of scattering from state \( k \) to state \( k' \) with scattering mechanism \( j \). By summing these probabilities over all \( j \), we get the probability for a carrier to scatter out of state \( k \) to state \( k' \), regardless of the scattering mechanism,

\[
P(k, k') = \sum_j P_j(k, k').
\] (3.8)

Another approach is to sum these probabilities over all final states \( k' \). This yields the probability of a carrier to scatter out of state \( k \), for a particular scattering mechanism,

\[
P_j(k) = \sum_{k'} P_j(k, k').
\] (3.9)

In the next paragraphs, we have an insight of the generation of the scattering tables in the EMC and in the CMC. The following is a short comparison in terms of the performance and speed-up of the two algorithms.

a) Scattering tables in the EMC program

In the case of the EMC, for every \( k \) vector, a table is built, such that each of its elements is the sum of all \( P_j(k) \) for \( j \leq i \) (where again \( P_j(k) \) is the probability to scatter from state \( k \) with scattering mechanism \( j \)). This means that for a given state \( k \), the first element of the corresponding table is the probability \( P_1(k) \) to scatter out of \( k \) with scattering mechanism 1, the second element is the probability \( P_{1+2}(k) = P_1(k) + P_2(k) \) to scatter out of state \( k \) with scattering mechanism 1 or 2, etc... Assuming there are \( N \) scattering mechanisms, the last element of the table for a given \( k \) is

\[
P_{1+2+...+N}(k) = \sum_{j=1}^N P_j(k).
\] (3.10)
Once the scattering table is built up for every possible state $k_i$, each probability is normalized to the maximum total probability $P_{\text{max}}(k)$ defined by

$$P_{\text{max}}(k) = \max_i \left[ P_{1+2+\ldots+N}(k_i) \right].$$  

During the simulation, a first random number is chosen between 0 and 1. If this number is greater than $P_{1+2+\ldots+j}(k)/P_{\text{max}}(k)$, we assume the corresponding scattering mechanism doesn’t affect the carrier; its final momentum state at the end of the time step $\Delta t$ remains unchanged. However, if it is smaller, a second random number is chosen to determine which scattering mechanism is going to affect the carrier. Then, we know the original state $k$, the energy of the carrier $E(k)$, the type of scattering involved and, in case of phonon scattering, if it is an emission or absorption, and what type of phonon the carrier scatters with. With this information, one is able to determine precisely what final state $k'$ the carrier scatters to. The following is a description of the algorithm used for this purpose, (in the case of phonon scattering),

1. A carrier $(k, E(k))$ is scattering with a phonon that has a maximum and a minimum energy $\varepsilon_{\text{max}}$ and $\varepsilon_{\text{min}}$.

2. Find all $N$ cells $k_i'$ in the Brillouin Zone that have an energy in the range $[E(k) + \varepsilon_{\text{min}}, E(k) + \varepsilon_{\text{max}}]$, or $[E(k) - \varepsilon_{\text{max}}, E(k) - \varepsilon_{\text{min}}]$ in the case of an absorption, or emission, respectively.

3. For all $N$ cells in that energy range, compute $P(k, k_i')$, where $k_i'$ is the new state after scattering.

4. Build the partial sums $P_{1+2+\ldots+i}(k, k_i') = \sum_{j=1}^{i} P(k, k_j')$ and normalize them

$$P_{1+2+\ldots+i}(k, k_i') = P_{1+2+\ldots+i}(k, k_i') / P_{1+2+\ldots+N}(k, k_N').$$

5. Pick a random number in $[0,1]$ to determine which $k_i'$ cell is selected.

6. Define the wave vector of the phonon $q = k_i' - k$ as the difference between the momentum $k_i'$ and the original one. The energy of the carrier is then defined as $E_{\text{final}} = E(k) + \varepsilon(q)$, or $E_{\text{final}} = E(k) - \varepsilon(q)$ in the case of absorption, or emission, respectively.

7. Find the new momentum $k_{\text{final}}'$ in the cell centered around $k'$ that minimizes $|E(k_{\text{final}}') - E_{\text{final}}|$.

This process is repeated for every $k$ vector, at each time step $\Delta t$. It involves heavy computation and is a bottleneck in terms of simulation time. The idea of the CMC is to
compute the transition tables beforehand, save them in memory and keep them accessible throughout the simulation. That means one has to compute the scattering probabilities for every scattering mechanism, to and from each possible \( \mathbf{k} \) vector. Typically, this takes approximately two days to compute. However, these probabilities are then available to the program and don’t need to be recomputed at each time step. This saves a lot of time and can reduce the simulation time by 50 times, as will be shown later.

\textbf{b) Scattering tables in the CMC program}

The procedure to build the scattering tables in the CMC is simpler than the one described above. For a given state vector \( \mathbf{k} \), one computes the probability to scatter from state \( \mathbf{k} \) to a state \( \mathbf{k}_i' \) by any of the \( N \) scattering mechanisms,

\[
P(\mathbf{k}, \mathbf{k}_i') = \sum_{j=1}^{N} P_j(\mathbf{k}, \mathbf{k}_i'),
\]

where \( P_j(\mathbf{k}, \mathbf{k}_i') \) is the probability to scatter from \( \mathbf{k} \) to \( \mathbf{k}_i' \) by scattering mechanism \( j \). Unlike in the EMC formalism, only the sums \( P(\mathbf{k}, \mathbf{k}_i') \) are stored in the CMC approach. This saves memory but also loses the information about the type of scattering involved. The first element of the table corresponding to a given state \( \mathbf{k} \) is the probability \( P(\mathbf{k}, \mathbf{k}_1') \) to scatter with any phonon from \( \mathbf{k} \) to \( \mathbf{k}_1' \). The second element of the table is the probability \( P(\mathbf{k}, \mathbf{k}_2') + P(\mathbf{k}, \mathbf{k}_1') \) to scatter with any phonon from \( \mathbf{k} \) to \( \mathbf{k}_2' \) or to \( \mathbf{k}_1' \), etc…. The last element of the table for a given \( \mathbf{k} \) is then the sum of all probabilities to scatter from \( \mathbf{k} \) to any other \( \mathbf{k}_i' \),

\[
P(\mathbf{k}) = \sum_{\mathbf{k}'} P(\mathbf{k}, \mathbf{k}').
\]

As in the EMC method, these probabilities are then normalized to the maximum total probability \( P_{\text{max}}(\mathbf{k}) = \max_{i} [P(\mathbf{k}_i')] \) and a random number is picked to determine whether the carrier at state \( \mathbf{k} \) is going to scatter or not. If the carrier happens to scatter, a second random number is generated to determine which \( \mathbf{k}_i' \) the carrier scatters to. This means that at the end of the time step \( \Delta t \), the carrier has moved in the \( \mathbf{k} \)-space from the grid point of coordinates \( \mathbf{k} = (k_x, k_y, k_z) \) to the new grid point \( \mathbf{k}_i' = (k_{x1}', k_{y1}', k_{z1}') \) and the carrier is placed at the center of the cell of coordinate \( \mathbf{k}_i' \). The energy tuning performed in the ECM formalism, (step 1 through 7 of the algorithm described previously) cannot be applied here because the knowledge of the scattering mechanism involved is lost and therefore, the maximum and minimum energy \( \epsilon_{\text{max}} \) and \( \epsilon_{\text{min}} \) of the phonon involved are unknown. This approximation is acceptable if the size of the grid is small. It follows that the finer the mesh, the smaller the error.
c) **The hybrid EMC/CMC method, speed up and performances**

As seen earlier, the complex models used to represent carrier transport can limit the applicability of the full band EMC method due to the intensive computational resources required. The CMC approach was developed to reduce this high computational demand. However, the CMC algorithm requires a huge amount of memory, typically, for adequate energy conservation, the requirement can be as high as 2.5 GB of computer RAM. A hybrid EMC/CMC method is used in this work to optimize the trade-off between the memory consuming CMC and the slower EMC. In the hybrid implementation, the CMC is used in the most active areas of momentum space, while the EMC is used elsewhere. The EMC scattering tables are then used in the region of the BZ where the scattering rate is low. These regions include the high-energy portions of the BZ where the number of carriers is low, and near the minimum of the conduction and maximum of the valence band, where the scattering probability is small. The implementation of these EMC tables increases the amount of memory available for the CMC, allowing for a finer discretization mesh.

The program related to this work is coded so that both methods, the EMC and the CMC, are available. The user can choose to use either one in a given energy range. To optimize the simulation time and the accuracy of the results, the EMC is generally used where fewer scatterings take place, and the CMC in the regions with a high density of state. Figure III.4.1a) gives an example of the hybrid use of both methods depending on the energy range. Figure III.4.1b) shows the performance comparison between the EMC and the CMC by looking at the time per iteration for each program. It is shown that at low field the speed-up is limited since the scattering rate is low in these areas. However, at high fields, the CMC method is proved to be up to 50 times faster than the EMC method.

![Figure III.4.1: a) An example of the hybrid implementation of EMC and CMC. b) EMC/CMC performance comparison. The arrows indicate the scale to consider.](image-url)
CHAPTER IV
COMPUTING THE VELOCITY

In this chapter, we describe the computation of the velocity of the electrons. After an analytical presentation of the velocity computation, two different numerical approaches will be considered. In this work, both methods are implemented and their agreement checked.

1- Operators – The Bloch Theorem

a) Operators

The operator that corresponds to the observable linear momentum is called the momentum operator and is given by

\[ \hat{p} = -i\hbar \nabla. \] (4.1)

In the case of a particle constrained to move in one dimension (x-direction for example), the momentum has only one component \( p_x \), and the momentum operator is

\[ \hat{p}_x = -i\hbar \frac{\partial}{\partial x}, \] (4.2)

where the corresponding eigenvalue equation for this operator is

\[ -i\hbar \frac{\partial}{\partial x} \psi = p_x \psi. \] (4.3)

The values \( p_x \) represent the possible values that measurement of the x-component of the momentum will yield. The eigenfunction \( \psi(x) \), corresponding to a specific value of the momentum \( p_x \), is such that \( |\psi(x)|^2 \, dx \) is the probability of finding the particle with momentum \( p_x \) in the interval \( x, x+dx \).

The velocity operator can be derived from the momentum operator. From \( \mathbf{p} = m \mathbf{v} \), we get \( \mathbf{v} = \mathbf{p} / m \) and thus,

\[ \hat{\mathbf{v}} = \frac{\hat{\mathbf{p}}}{m} = -i\hbar \frac{\nabla}{m} = \frac{\hbar}{mi} \nabla, \] (4.4)

where \( m \) is the mass of the electron. For each physical observable \( C \), we define its expectation value as

\[ < C > = \int dr \, \psi^* \hat{C} \psi, \] (4.5)

where \( \hat{C} \) is the corresponding operator with eigenvectors \( \psi \) and their conjugates \( \psi^* \).
The expectation value for the velocity operator is

\[ < \hat{v} > = \int dr \; \psi^* \hat{v} \psi . \quad (4.6) \]

The operator corresponding to the energy is the Hamiltonian, with momentum \( \mathbf{p} \) replaced by its operator counterpart, \( \hat{\mathbf{p}} \). For a single particle of mass \( m \), in a potential field \( V(r) \), the energy operator is

\[ \hat{H} = \frac{\hat{\mathbf{p}}^2}{2m} + V(r) = -\frac{\hbar^2}{2m} \nabla^2 + V(r) . \quad (4.7) \]

The eigenvalue equation for \( \hat{H} \) is called the time-independent Schrödinger equation,

\[ \hat{H} \psi(r) = E\psi(r) , \quad (4.8) \]

and yields the possible energies \( E \) that the particle may have.

---

**b) The Bloch Theorem**

The Bloch theorem states that the eigenstates of the Hamiltonian operator, (given in Eq.4.7), can be chosen to have the form of a plane wave times a function \( u_{nk}(r) \) that has the periodicity of the Bravais lattice,

\[ \psi_{nk}(r) = e^{ik \cdot r} u_{nk}(r) , \quad (4.9) \]

where the crystal momentum \( k \) is fixed and \( u_{nk}(r + R) = u_{nk}(r) \) for all position vector \( R \) in the Bravais lattice. Substituting this into the Schrödinger equation, we find that \( u_{nk}(r) \) is determined by the solution to the eigenvalue problem,

\[ H_k u_{nk}(r) = \left( \frac{\hbar^2}{2m} \left( \frac{1}{i} \nabla + k \right)^2 + V(r) \right) u_{nk}(r) = E_k(r) u_{nk}(r) , \quad (4.10) \]

with the periodic boundary condition, \( u_{nk}(r) = u_{nk}(R + r) \), where \( R \) is the primitive vector of the Bravais lattice. Equation 4.8 is an eigenvalue problem restricted to a single primitive cell of the crystal. Because the eigenvalue problem is set in a fixed finite volume, we expect to find an infinite family of solutions with discretely spaced eigenvalues, which are commonly labeled with the band index \( n \). For a given \( n \), (for a given energy band), the eigenstates and eigenvalues are periodic functions of \( k \) in the reciprocal lattice and give the wave function and the electron energy respectively. In Eq.4.10, \( k \) appears only as a parameter in the Hamiltonian \( H_k \). We therefore expect each of the energy levels, for a given \( k \), to vary continuously as \( k \) varies and thus defines the energy level of an electron as a family of continuous functions \( E(k) \), defining the band structure.
2- Computation of the velocity

The expression of the velocity can be obtained by looking at a solution of the Schrödinger equation for a given $k+q$, (with $q$ small with respect to $k$). The Hamiltonian for $k+q$ is

$$\hat{H}_{k+q} = \hat{H}_k + \frac{\hbar^2}{m} q \left( \frac{1}{i} \hat{\nabla} + k \right) + \mathcal{O}(q^2) = \hat{H}_k + \hat{H}^\prime + \mathcal{O}(q^2),$$

(4.11)

which has a corresponding eigenvalue, $E_n(k+q)$. We use perturbation theory [Lib92] applied to the equation above to treat this problem, with a second order expansion of the eigenvalues and the eigenvectors. The Schrödinger equation becomes

$$\hat{H}_n(r) = E_n \psi_n(r) \text{ with } \hat{H}_{k+q} = \hat{H}_k + \hat{H}^\prime + \mathcal{O}(q^2) \text{ and } \begin{cases} \psi_n = \psi_n^{(0)} + \psi_n^{(1)} + \psi_n^{(2)} + \ldots \\ E_n = E_n^{(0)} + E_n^{(1)} + E_n^{(2)} + \ldots \end{cases}$$

First order perturbation theory states that, the eigenvalues have the following form,

$$E_n^{(1)} = \int d^nr \psi_n^{(0)*} \hat{H}^\prime \psi_n^{(0)},$$

(4.12)

where the integral over $r$ is the expectation value of $\hat{H}^\prime$, taken to the fist order, and substituting the expression of $\hat{H}^\prime$ from Eq.4.11 into Eq.4.12 we get

$$E_n^{(1)} = \int d^nr \psi_n^{(0)*} \left[ \frac{\hbar^2}{m} q \left( \frac{1}{i} \hat{\nabla} + k \right) \right] \psi_n^{(0)}.$$

(4.13)

Now, looking at the Taylor expansion of the energy about $q$,

$$E_n(k+q) = E_n(k) + \sum_i \frac{\partial E_n}{\partial k_i} q_i + \mathcal{O}(q^2) = E_n^{(0)} + E_n^{(1)} + \mathcal{O}(q^2),$$

(4.14)

and comparing Eq.4.13 and Eq.4.14 we have

$$\sum_i \frac{\partial E_n}{\partial k_i} q_i = \int d^nr \psi_n^{(0)*} \left[ \frac{\hbar^2}{m} q \left( \frac{1}{i} \hat{\nabla} + k \right) \right] \psi_n^{(0)},$$

(4.15)

or

$$\nabla_k E_n = \int d^nr \psi_n^{(0)*} \left[ \frac{\hbar^2}{m} \left( \frac{1}{i} \hat{\nabla} + k \right) \right] \psi_n^{(0)}.$$

(4.16)
If we express Eq.4.16 in terms of the Bloch functions from Eq.4.13, it can be written,

\[
\nabla_k E_n = \int dr \begin{pmatrix} \frac{h^2}{m} \left( \frac{1}{i} \nabla + k \right) \end{pmatrix} u_n
\]

(4.17)

or

\[
\nabla_k E_n = h \int dr \psi_n^* \frac{\hbar}{mi} \nabla \psi_n.
\]

(4.18)

Using the fact that \( \frac{\hbar}{mi} \nabla \) is the velocity operator, and \( \int dr \psi_n^* \hat{\psi} \psi_n = \langle \hat{\psi} \rangle \) is the expectation value for the velocity operator, we finally obtain

\[
v = \frac{1}{\hbar} \nabla_k E_n,
\]

(4.19)

which means that the velocity of the electrons can be obtained by deriving their energy with respect to the momentum \( k \).

### 3- Two numerical approaches

One numerical approach is directly derived from Eq.4.19 and consists in taking the discrete derivative of the energy found in every cell of the mesh. For each grid point, the difference between the energy of neighboring cells is computed. Dividing this difference by the reduced Plank constant \( \hbar \), yields the velocity of the electron (Eq.4.19). Since the problem has to be solved in three dimensions, the numerical derivative is taken in each direction (Vx, Vy and Vz). For a fine grid, the discrete derivation gives a good approximation of the velocity. However, this method is less reliable close to the boundaries of the BZ.

Another approach is to compute the velocity directly using the eigenvectors of the Schrödinger equation (Eq.4.8) with the expectation value of the velocity operator, (Eq.4.6), and the discrete version of this equation is

\[
v = \frac{\hbar^2}{m_i} \sum_i \left( \text{Re}^2[\psi_i] + \text{Im}^2[\psi_i] \right) \Delta k,
\]

(4.20)

where the eigenvectors \( \psi_i \) are known, and can be extracted from the Hamiltonian matrix. This second solution has been implemented in the course of this work. The corresponding code is given in Appendix A. The results obtained are comparable with the ones obtained using the discrete derivation of the energy. Figure IV.3.1 and Fig. IV.3.2 show the velocities computed with both methods, along different axis in k-space. In these figures Vx, Vy and Vz correspond to the velocity computed directly from the eigenvectors,
whereas Dx, Dy and Dz correspond to the velocities computed with the discrete method. In k-space, when moving along <111> axis, k_x, k_y and k_z vary from \(-2\pi/a\) to \(2\pi/a\). When moving along the <100> axis, only k_x varies from \(-2\pi/a\) to \(2\pi/a\) whereas k_y and k_z are zero.

**Figure IV.3.1:** Conduction band velocity for Si when moving along the <111> axis.

**Figure IV.3.2:** Conduction band velocity for Si when moving along the <100> axis.

As seen in these figures, an excellent agreement is obtained. In order to appreciate the differences between these two methods and eventually show the contribution of the later one, the main program needs to be slightly modified to allow the user to choose either one of the methods. At the time of this work, this change had not been implemented. More accurate results are expected, particularly close to the boundaries of the BZ.
CHAPTER V

RESULTS - DISCUSSION

In this chapter, we present the simulation results and comparison with data obtained with an established full-band MC code called DAMOCLES [Fis96b]. Although generally good agreement is obtained for a wide range of material, there are still some discrepancies between the simulation results and the experimental data. Explanations are given and possible solutions are discussed.

1- Simulations characteristics

First, we take a look at the data in literature to have a good sense of the expected range of values. A commonly used way of presenting data is the energy-field and velocity-field curve, and are given in Fig. V.1.1 for InP. From the energy-field curve, (Fig. V.1.1 a), we can see that the energy of electrons ranges from 0.038eV to 1.5eV when the applied electric field changes from 10 V/cm to 10^6 V/cm respectively. For the interest of our simulation, this means it is only necessary to model carriers with an average energy within this range. From the velocity-field curve, (Fig. V.1.1 b), we see that the velocity range is between 10^5 cm/s and 10^8 cm/s and there is an interesting behavior when the electric field is around 10^4 V/cm for this material. These types of features helps us in choosing the measurement points.

Figure V.1.1: Energy versus field a) and velocity versus field b) data curve for InP from DAMOCLES. [Fis98b]
Simulations were run in momentum space for both electrons and holes to calibrate the steady state bulk properties using the hybrid EMC/CMC. Typically, 20,000 carriers, (10,000 electrons and 10,000 holes) were simulated for a total duration of 5 ps and a time step of 0.2-0.4 fs. The total time required for one field value is approximately 2 hours with the CMC algorithm. The steady state drift velocity as a function of the electric field applied in the <100> direction, as well as the ensemble energy as a function of the field were studied. Figure V.1.2 presents the energy-field curve for different semiconductor materials obtained for electrons and holes and Fig. V.1.3 presents the velocity-field curves.
ENERGY- FIELD curves:

Figure V.1.2: Average energy versus electric field for Si, GaAs and InP.
VELOCITY- FIELD curves:

Figure V.1.3: Drift velocity versus electric field for Si, GaAs and InP.
2- Comparison with referenced data

In this work, data is taken from M. V. Fischetti [Fis96b] as the reference because they are widely accepted results. Fischetti uses analytic approximations in the conduction band of III-V compounds for energies below 0.3 eV whereas in this work, a complete full-band representation of the $E(k)$ dispersion relation for all materials has been implemented. We present here our hybrid EMC/CMC simulation results and compare them to the data from M. V. Fischetti.

a) Velocity - field characteristics

Figure V.2.1 shows the hybrid simulation results of the velocity-field characteristics of Si compared to M. V. Fischetti’s results.

For electrons, the velocity at high fields is lower than Fischetti’s results. There are two possible reasons to this velocity undershot at high electric fields. The first one could be related to the relatively coarse grids used around the L (0.5,0.5,0.5) valley. This coarse grid can cause some spurious diffusion in the final state selection. From the band structures (Chapter II) we can see that the velocity in the L valley is lower than that in the X valley. The excess population in the L valley could account for this difference. Another reason could be the isotropic impact ionization model used in this work. The approximations inherent to this model make the average velocity lower at high fields. Possible solutions to address this problem are: use a finer grid around L points, provided enough memory is available, or use a more sophisticated model for impact ionization.

**Figure V.2.1: Drift velocity versus electric field for Si.**
Figure V.2.2 presents the same curves as Fig. V.2.1 for other common semiconductor materials and their comparison with M. V. Fischetti’s results.

![Graphs showing drift velocity versus electric field for GaAs and InP.](image)

*Figure V.2.2: Drift velocity versus electric field for GaAs and InP.*

We can see that a good agreement is obtained but the undershot observed for Si seems to be observed for other materials. This shows the limits of the model used in the simulations. To achieve better results, a finer grid should be used, along with a more detailed model, especially for the impact ionizations phenomena. For InP, due to limitation of published data, only the results for electrons are shown.
b) Energy - field characteristics

We present here, (Fig. V.2.3), the energy-field curves for the same materials as above, and the comparison with Fischetti’s results.

*Figure V.2.3: Energy versus electric field for Si, GaAs and InP.*
For the energy field curves, a very good agreement with published data is obtain for certain materials (InP or GaAs for example). For other materials, the results are not as satisfactory (Si for example). We can see in the case of silicon that the energy found for holes and electron is higher than Fischetti’s results throughout the electric field range considered. A probable reason is the too coarse grid used in these simulations. The approximation of the CMC when computing the final state after scattering might yield energy values higher than the ones expected.
CONCLUSION, PROJECTS, FUTURE WORK:

In this document, general results of the hybrid EMC/CMC simulations of steady-state carrier transport in several zincblende semiconductor materials, by using a full-band, particle-based approach have been presented. As seen before [Sar98, Sar99], this hybrid method shows very good agreement with experimental published data and also with results obtained with classical Monte Carlo simulations [Fis98].

Work has been done to implement a computation of the carrier velocity, using a different approach that should yield higher precision when integrated with the main program, especially in regions close to the boundaries. One of the objectives for future work is to reduce the memory allocated for the scattering tables involved in the simulation of a given number of particles. To achieve this, we will be working on two areas. The first is to change the storage format for the probabilities generated throughout the algorithm of the scattering tables. So far, the probabilities are stored as floating point numbers, which have a 7 digit precision. The idea is to code and normalize them as short integers, with a 4 digits precision. Since a short integer is coded in 2 bytes instead of 4 bytes for a floating-point number, the first benefit of this change is a reduction of 25% of the total size of the scattering tables. Another advantage of this change will be a gain in speed in the process of selecting the scattering process involved. Once the probabilities are stored as short integers, only a short integer random number needs to be picked, where a floating point random number was previously required. This is interesting because generating a random short integer is much faster than generating a random floating-point number. Reducing the size of the scattering table would reduce the required memory and allow for the use of a finer grid, which could address the energy conservation problems mentioned in Chapter III and V. However, one must be careful that the loss of precision of the random number doesn’t significantly affect the validity of the results.

Another means of reducing the size of the scattering tables would be to change the approach of storing the address of the k states. Presently, the address of the final k’ state that the carrier could scatter to is coded as an absolute value. Due to the large number of particles simulated, the number of possible k’ states is large and must be stored as a double floating point number. In scattering phenomena, the energy exchanged by particles is finite. It is thus inaccurate to consider that from a given position in the k-pace, a carrier can possibly scatter to any other k’ position. This new position in the k-space is limited to a certain range determined by the maximum energy involved in carrier scattering phenomena. The idea is then to code the possible destination k’ in relative instead of absolute addressing by using long instead of double floating point numbers. This would also save 2 bytes per address and reduce, by an additional 25%, the total size of the scattering table. One has to be careful however that by increasing the number of simulated cells, this relative addressing will not fail. Here too, the idea is to leave to the user the opportunity to choose the addressing preference.
Along with this work aiming to reduce the amount of memory required for simulations, the research group at the Illinois Institute of Technology is also going to gain benefits from the new-coming cluster of PCs with faster CPUs and increased RAM. This hardware enhancement will also allow us to simulate a finer mesh and improve the accuracy of our results, by exploring some effects yet left aside for computational cost reasons. This faster cluster should also allow us to simulate more semiconductor devices and in particular 3D-devices showing a 3D doping profile and geometry. The journey only begins!
**APPENDIX A**

## COMPUTING THE FIRST DERIVATIVE OF THE ENERGY

in /dynamics/band.c

```c
static DERIV *Ene1Deriv(Kpseudo *kp,
    Predata *pd,
    Complex *evec,
    Gpoint *krec,
    int numbands,
    int hamdim,
    DZMAT **dnloc)
{
    THEDOUBLE hs2mo = ACHBAR/MASEL * ACHBAR/2.0;
    static DERIV *der0=NULL;
    THEDOUBLE coeff,gpk;
    THEDOUBLE rvectorij,ivectorij,rvectorkj,ivectorkj;
    THEDOUBLE dnlocrik,dnlociik;
    THEDOUBLE xon1,xon2,xon3,xon4;
    THEDOUBLE der_nl[3];
    THEDOUBLE derx,dery,derz;
    static FLAG firstime=YES;
    int nbands=0;
    int i,j,k,id,i_sp;
    int ndim=2*hamdim;

    if((kp==NULL)&&(pd==NULL)) {
        if(firstime==YES){
            WERRS("Ene1Deriv: requested deallocation of non existing data.");
        } else {
            FREE(der0);
            firstime=YES;
            return(NULL);
        }
    }

    if((pd->inithamdata.mode) == SPIN_ORBIT) {
        ndim = 2*hamdim;
        nbands=numbands*2;
    } else {
        ndim = hamdim;
        nbands=numbands;
    }

    if(firstime==YES) {
        der0=(DERIV*)(calloc(nbands,sizeof(DERIV)));
        firstime=NO;
    }

    for(j=0;j<nbands;j++) {
        der=der0+j;
        der->dx=der->dy=der->dz=0.0;
        derx=dery=derz=0.0;
        for(i=0;i<ndim;i++) {
            rvectorij=(*(evec+j*ndim+i)).re;
            ivectorij=(*(evec+j*ndim+i)).im;
            coeff = hs2mo*(rvectorij*rvectorij + ivectorij*ivectorij);
            der_nl[0]=der_nl[1]=der_nl[2]=0.0;
            for(id=0;id<3;id++) {
                rvectorijprime=*(evec+j*ndim+i+id*ndim).
                rvectorijprime+=*(evec+j*ndim+i+id*ndim+1).
                coeff = hs2mo*(rvectorij*rvectorijprime + ivectorij*ivectorijprime);
                der_nl[id]=der_nl[id]-coeff;
            }
        }
    }
}
```

This function computes the first derivative of the energy from the eigenvectors given by the eigenproblem function.

**author:** julien branlard - julien@neumann.ece.iit.edu

**creation date:** Saturday, February 23, 2002
rvectorkj = (*{evec+j*ndim+k}).re;
ivectorkj = (*{evec+j*ndim+k}).im;
dnlocrik = dnloc[id]->me[i][k].re;
dnlociik = dnloc[id]->me[i][k].im;
xon1 += rvectorij * rvectorkj * dnlocrik;
xon2 += rvectorij * ivectorkj * dnlociik;
xon3 += ivectorij * ivectorkj * dnlocrik;
xon4 += ivectorij * rvectorkj * dnlociik;
}
der_nl[id] += xon1 + xon2 + xon3 + xon4;
}
if (i >= hamdim) i_sp = i - hamdim;
else
    i_sp = i;
gpk = (kp->kx) + (krec+i_sp)->x;
derx += 2 * coeff * gpk + der_nl[0] * CAREL;
gpk = (kp->ky) + (krec+i_sp)->y;
dery += 2 * coeff * gpk + der_nl[1] * CAREL;
gpk = (kp->kz) + (krec+i_sp)->z;
derz += 2 * coeff * gpk + der_nl[2] * CAREL;
}
der->dx = (MYFLOAT) derx;
der->dy = (MYFLOAT) dery;
der->dz = (MYFLOAT) derz;
}
return (der0);
REFERENCES


[Fis96b] Fischetti, M.V., Sano, N., Laux, S.E. and Natori, K., “Full-band structure theory of high-field transport and impact ionization of electrons and holes in Ge, Si, and GaAs”, Processings of the 1966 International Conference of semiconductor processes and devices (SISPAD’ 96), Tokyo, Japan, 1966


