INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313-761-4700  800-521-0600
A tight-binding approach to resonant tunneling diode simulation

Boykin, Timothy Bruce, Ph.D.
Stanford University, 1992

Copyright ©1992 by Boykin, Timothy Bruce. All rights reserved.
A TIGHT-BINDING APPROACH TO RESONANT TUNNELING DIODE SIMULATION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Timothy Bruce Boykin
June 1992
I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James S. Harris, Jr. (Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

R. Fabian W. Pease

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Walter A. Harrison (Applied Physics)

Approved for the University Committee on Graduate Studies:

iii
ABSTRACT

We have developed a simulation for resonant tunneling diodes (RTDs) based on an empirical tight-binding model. Using the tight-binding approach allows a more careful treatment of the heterointerfaces than is possible within the more common envelope-function model. We are therefore able to take into account the effects of multiple valleys and bands in a physically meaningful way. This is particularly important for devices with barriers made of indirect bandgap materials such as AlAs.

In order to carry out these calculations, we have developed the first numerically stable transfer matrix method. With this improved method, we are able to transfer across large distances (> 1000 Å) and have for the first time included space-charge regions in a tight-binding RTD simulation.

We examine the transmission coefficients calculated with this model and discuss their features. In particular, we discuss the importance of the AlAs X-minima in devices with AlAs barriers. We also demonstrate how transverse effective-mass differences can shift the $k_{||} \neq 0$ transmission graph relative to that for $k_{||} = 0$. Finally, we present current density calculations and compare them to those made with the envelope-function model.
ACKNOWLEDGEMENTS

I owe many people thanks for the advice, friendship, and help they have extended to me during my time at Stanford. First among them is my advisor, Prof. James Harris. This research is at the present time somewhat unusual for the field of electrical engineering and Prof. Harris' patience, particularly in the middle stages of the project when progress became difficult, along with his willingness to grant me the necessary independence are most appreciated. I understand that supporting this research was something of a gamble and I am grateful to Prof. Harris for taking the risk; I think that many other engineers would not have done so.

I also owe many thanks to the other members of my committee. I appreciate Prof. Mark Horowitz's service as chair of the Orals committee. I am very grateful to Prof. Walter Harrison for sharing his insights on and deep understanding of condensed-matter physics with me. I also thank him for his cheerful and careful reading of and helpful comments on my preprints. My first project as a Harris group member involved collaboration with Prof. Pease and his group and I would like to thank him for his help there in addition to his service on the committee. Finally, I must thank Profs. Harris, Harrison, and Pease again for serving as references for me; I couldn't have gotten a job without their help!

Prof. Harris has assembled a superb group of graduate students with whom I have been privileged to work. Several among them deserve special mention. Ed Wolak, a former Harris group member, introduced me to the problems of heterointerface scattering and helped get me started in this area. Paul van der Wagt has been my frequent co-conspirator. There are very few people who can grow your wafer, tell you what you're doing wrong processing it, and simultaneously discuss (at a very high level) the most esoteric aspects of condensed-matter theory; Paul can do all that and more! It has been great working with him. When I first joined the Harris group Douglas Bone and I worked with Prof. Pease and Ken Williams of the Solid State Lab on a mobile ultrahigh vacuum system. Although my research eventually went in a rather different direction, I value both
the knowledge gained and the experience itself; Douglas, Ken, and Prof. Pease were most congenial collaborators. I will miss discussing physics, the Stanford administration, and a variety of other topics with Ken Shepard. It has also been a pleasure working with current and former students Susan Lord, Bob Lodenkamper, Bardia Pezeshki, Peng Cheng, Damian Costa, Kevin Lear, Byung Gook Park, Winston Fu, and Jae Hoon Kim. I would be remiss if I failed to mention the excellent administrative support provided by Gail Chun-Creech, Susie Burns, and Sherrie Harvey.

Finally, I thank my family (Mom, Dad, Melissa, Mike, Michael Brooks, Grandmother Boykin, Grandmother & Papa Owens, Uncle Brooks, and all the rest) and friends for their love and support; I couldn’t have done it without them!
CHAPTER 3: TIGHT-BINDING MODELS

3.1 UNDERLYING APPROXIMATIONS

3.2 OPERATORS, ORBITALS, AND SYMMETRIES

3.2.1 Rotations and Translations of the Coordinate System
3.2.2 Behavior of the State Vector Under Rotations and Translations
3.2.3 Symmetries of the Zincblende Crystal
3.2.4 Löwdin Orbitals
3.2.5 Matrix Elements and the Nearest-Neighbor Model

3.3 PERFECT CRYSTALS: BULK TIGHT-BINDING HAMILTONIANS

3.4 INTERFACE SYSTEMS AND THE TRANSFER MATRIX

3.4.1 Tight-Binding Hamiltonian for the Interface System and the Single-Monolayer Transfer Matrix
3.4.2 Eigenvalues, Eigenvectors, and Properties of the Single-Monolayer Transfer Matrix in the Nearest-Neighbor Model
3.4.3 The Usual Transfer Matrix Approach: An Example Calculation of a Tunneling Problem
3.4.4 Removal of the Numerical Instability from and Boundary Conditions for the Interface Tight-Binding Hamiltonian

3.5 RESONANT TUNNELING DIODES

3.5.1 Specification of the Non-Bulk Transfer Matrices
3.5.2 The Transmission Coefficient and the Current Density
3.5.3 Simulation Results: Current-Voltage Characteristics
3.5.4 Simulation Results: Transmission Coefficients

3.6 SINGLE-BARRIER DIODES

3.7 THE SINGLE-INTEGRAL APPROXIMATION

3.8 CRITICISMS OF THE PRESENT TIGHT-BINDING MODEL

3.9 CONCLUSIONS

NOTES
CHAPTER 4: CONCLUSIONS

4.1 SUMMARY

4.2 SUGGESTIONS FOR FUTURE WORK

4.2.1 Improvements Within the Context of the Tight-Binding Model
4.2.2 Other Improvements

NOTES

APPENDIX

A.1 BULK TIGHT-BINDING HAMILTONIAN FOR THE sp^3s* MODEL
A.2 THE TIGHT-BINDING METHOD APPLIED TO SUPERLATTICES

NOTES
# FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Schematic conduction-band diagram of an RTD under zero bias</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Schematic conduction-band diagram of an RTD at resonant bias</td>
<td>6</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematic conduction-band diagram of an RTD at valley bias</td>
<td>7</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic RTD $I-V$ characteristic</td>
<td>7</td>
</tr>
<tr>
<td>2.5</td>
<td>Conduction-band diagram for an RTD as calculated from Poisson's equation</td>
<td>16</td>
</tr>
<tr>
<td>2.6</td>
<td>$J-V$ curve for a 10 monolayer single-barrier RTD</td>
<td>18</td>
</tr>
<tr>
<td>2.7</td>
<td>$I-V$ curve for a 10 monolayer single-barrier RTD, including series resistance</td>
<td>18</td>
</tr>
<tr>
<td>2.8</td>
<td>Schematic diagram of an MBE growth chamber</td>
<td>21</td>
</tr>
<tr>
<td>2.9</td>
<td>RTD fabrication</td>
<td>23</td>
</tr>
<tr>
<td>3.1</td>
<td>The two lowest conduction bands of GaAs and AlAs from $\Gamma$ to $X$</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>GaAs layers oriented in the [001] direction</td>
<td>47</td>
</tr>
<tr>
<td>3.3</td>
<td>Layers diagram for the RTD tunneling problem</td>
<td>57</td>
</tr>
<tr>
<td>3.4</td>
<td>$J-V$ curve for a triple-barrier RTD</td>
<td>70</td>
</tr>
<tr>
<td>3.5</td>
<td>Low-energy transmission coefficient for the triple-barrier RTD of Figure 3.4</td>
<td>73</td>
</tr>
<tr>
<td>3.6</td>
<td>Electron probability densities for the triple-barrier RTD of Figure 3.4</td>
<td>74</td>
</tr>
<tr>
<td>3.7</td>
<td>The complex bandstructures of GaAs and AlAs for $k_{</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>High-energy transmission coefficient for the triple-barrier RTD of Figure 3.4</td>
<td>78</td>
</tr>
<tr>
<td>3.9</td>
<td>Low-energy transmission coefficient for a 5 monolayer AlAs barrier</td>
<td>85</td>
</tr>
<tr>
<td>3.10</td>
<td>Low-energy transmission coefficient for a 10 monolayer AlAs barrier</td>
<td>85</td>
</tr>
<tr>
<td>3.11</td>
<td>Low-energy transmission coefficient for a 20 monolayer AlAs barrier</td>
<td>86</td>
</tr>
<tr>
<td>3.12</td>
<td>Low-energy transmission coefficient for a 50 monolayer AlAs barrier</td>
<td>86</td>
</tr>
<tr>
<td>3.13</td>
<td>High-energy transmission coefficient for a 5 monolayer AlAs barrier</td>
<td>91</td>
</tr>
</tbody>
</table>
3.14 High-energy transmission coefficient for a 10 monolayer AlAs barrier 91
3.15 High-energy transmission coefficient for a 20 monolayer AlAs barrier 92
3.16 Comparison of $k_\parallel = 0$ and $k_\parallel \neq 0$ transmission coefficients for a 20 monolayer AlAs barrier 92
3.17 Comparison of $k_\parallel = 0$ and $k_\parallel \neq 0$ transmission coefficients for a 6/18/6 GaAs/AlAs double-barrier RTD 95

xiii
# TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Implant-Isolated RTD Process</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Same-Site and Nearest-Neighbor Matrix Elements</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Tight-Binding Parameters</td>
<td>40</td>
</tr>
<tr>
<td>3.3</td>
<td>Energies, Minima, and Effective Masses</td>
<td>40</td>
</tr>
<tr>
<td>3.4</td>
<td>Blocks of the Interface Tight-Binding Hamiltonian</td>
<td>48</td>
</tr>
<tr>
<td>3.5</td>
<td>Wavefunction Decompositions for Figure 3.5</td>
<td>79</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Modern electronic devices continue to decrease in size with the passing of each year. Crystal growth techniques such as molecular beam epitaxy (MBE) reliably produce films as thin as one monolayer (~ 3 Å). Devices with such fine dimensions, which are often referred to as nanostructures and include, for example, the resonant tunneling diode (RTD), can only be understood in quantum mechanical terms, unlike the more familiar bipolar junction transistor (BJT), heterojunction bipolar junction transistor (HBT), and metal-oxide-semiconductor field-effect transistor (MOSFET), which are still essentially classical in their behavior. If the present trends in shrinking device dimensions continue -- and indications are that they will -- then future devices will be governed increasingly by quantum, rather than classical, mechanics. It is therefore essential to understand in as much detail as possible the operation of those quantum mechanical electronic devices we have at our disposal today.

The RTD is an excellent device for studying the physics of nanostructures. Because the electrons and/or holes are confined in one dimension only, this greatly simplifies the analysis and modeling. Thanks to the fine control available with MBE and the excellent interface quality of MBE-grown materials, RTDs usually exhibit very strong quantum effects at room temperature. Indeed, it is rarely necessary to cool an RTD below 77 K to find a wealth of physics!

Somewhat surprisingly, however, today there are still no really accurate RTD models. While it is rather simple to predict many qualitative aspects of RTD behavior, achieving quantitative correctness is an entirely different matter. The reason is no doubt due to the myriad of processes taking place in an RTD. The tunneling electrons do not behave in the same simple manner as the tunneling free-particles so familiar from first-year quantum mechanics texts. The electrons also interact with lattice vibrations and one another. In addition, even though MBE crystal quality is excellent, it is not perfect, and
there are defects, impurities and rough interfaces which provide scattering centers for the electrons. If we are to fully understand the physics of RTDs and the other quantum devices which will succeed them, advanced RTD models are essential. The complexity of the interactions taking place in an RTD at this point in time precludes addressing all of the "nonidealities" (i.e., any behavior which deviates from the simple free-particle outlook), so in order to produce the most accurate predictions we must carefully choose those we shall study. The two most important are doubtless the electrons' markedly non-free behavior, even in a perfect bulk material, and the interaction of the electrons with lattice vibrations. Either one alone is a large problem in and of itself; in this work we shall address only the former.

Chapter 2 of this dissertation presents the background material necessary for the advanced RTD model presented in Chapter 3. Chapter 2 begins with a discussion of the principles of RTD operation. Following this, we briefly state and examine the envelope-function techniques used in most RTD models. We conclude with a short discussion of the growth and fabrication of the devices themselves.

Chapter 3 presents the advanced RTD simulation, based on an empirical tight-binding model. With this a model we describe the electron wavefunction in terms of a localized-orbital basis, so it is naturally appropriate for treating interface systems. Tight-binding techniques result in electron wavefunctions having the proper $E(k)$ dispersion relations throughout all of $k$-space (not only in the first Brillouin zone). While tight-binding models are familiar to condensed matter physicists, they have only very recently begun to be employed by electronic device engineers. Therefore, in Chapter 3 we first state or briefly develop the techniques necessary for tight-binding models in general. Next, we review the application of the transfer matrix method to tight-binding models. Following this review, we discuss the physical significance of the various properties of transfer matrices, their eigenvalues, etc. Unfortunately, the transfer matrix method tends to be highly numerically unstable, and in the course of this research we developed an improved, numerically stable, but still exact transfer matrix approach. This new method, the first of its kind, is derived in detail in Chapter 3. Finally, we apply the methods developed to the
study of RTDs of various sorts, briefly examining estimated J-V characteristics but studying in great detail transmission curves, since we can extract more information from them.

Chapter 4 summarizes the contributions of this research and discusses possible directions for future research. The appendix addresses several detailed points omitted from the text of Chapter 3.
CHAPTER 2

BACKGROUND MATERIAL

In this chapter we summarize the common knowledge of the field of resonant tunneling diodes (RTDs). This includes the principles of RTD operation, elementary RTD models, and the growth and fabrication of the devices themselves. We devote extra attention to elementary RTD models since the following chapter presents an advanced RTD simulation using an empirical tight-binding approach.

2.1 RESONANT TUNNELING DIODES: PRINCIPLES OF OPERATION

The RTD was first proposed by Tsu and Esaki [1] and initially realized by Chang, Esaki, and Tsu [2]. The device is a layered structure, consisting in its most basic form of five regions: emitter, emitter barrier, well, collector barrier, and collector. In an electron tunneling device, the emitter, collector, and well are usually made of the same material, the emitter and collector being quite thick (usually over 0.1 μm) and heavily (often degenerately) doped n-type, the well being extremely thin (usually a few nanometers) and most often undoped. The barriers are made of a different material, one whose conduction-band minimum lies higher in energy than that of the emitter, collector, and well, and, like the well, are nanometer-sized and usually undoped. (As electron tunneling devices are the most common, we shall restrict our attention to them.) Figure 2.1 presents a simplified conduction-band diagram of an RTD under flatband (unbiased) conditions. As is readily apparent from fig. 2.1, the idealized RTD is a model realization of the double barrier tunneling problem from elementary quantum mechanics.

RTD operation, illustrated in figures 2.1-2.4, is straightforward. As we know from quantum mechanics, there will be quasi-bound states in the well due to the presence of the barriers. (That is, although the energy spectrum is a continuum, there are only very narrow
regions of energy in which the electron wavefunction in the well will be appreciable and in which the transmission probability through the structure will be significant. Typically these are so narrow that we may consider them to be states.) Under low bias (fig. 2.1 and point 1 of fig. 2.4) the emitter fermi level lies below the first state in the well and consequently there is little tunneling. At resonance (fig. 2.2 and point 2 of fig. 2.4), the lowest state in the well lies between the emitter conduction-band minimum and fermi level, so that many electrons will tunnel, leading to a peak in the current. Finally, beyond resonance (fig. 2.3 and point 3 of fig. 2.4), the first state of the well now lies below the emitter conduction-band minimum while the second state still lies above the emitter fermi level. All electrons therefore have only a small transmission probability, and consequently the current decreases. (Although not shown in fig. 2.4, there will be another peak in the current when the second quasi-bound state lies between the emitter conduction-band minimum and fermi level.) It is appropriate to state some RTD terminology here. The region between points 2 and 3 of fig. 2.4 is referred to as the negative differential resistance (NDR) region of the RTD I-V characteristic. Point 2, where the current reaches a maximum, is called the peak and point 3, where it is a minimum, is called the valley. The peak-to-valley ratio (PVR) is defined as the current at point 2 of fig. 2.4 divided by the current at point 3. In order to more fully understand RTD operation, we shall next consider the simplest possible physically meaningful treatment of the device: the envelope-function model.
Figure 2.1: Schematic conduction-band diagram of a double barrier RTD under zero bias. The conduction-band minimum is indicated by the thick solid line, the quasi-bound states in the well are indicated by the thin solid lines, and the fermi levels in the degenerately doped emitter and collector are indicated by dotted-dashed lines.

Figure 2.2: Schematic conduction-band diagram of a double barrier RTD under resonant bias. Note that the first quasi-bound well state lies in between the emitter fermi level and conduction-band minimum, resulting in a large tunneling current.
**Figure 2.3:** Schematic conduction-band diagram of a double barrier RTD biased past the first resonance. Neither quasi-bound state lies between the emitter fermi level and conduction-band minimum so the tunneling current will be small.

**Figure 2.4:** Schematic I-V characteristic for a double barrier RTD. Points 1-3 of the graph approximately correspond to figures 2.1-2.3, respectively.
2.2 ELEMENTARY MODELING OF RTDs

2.2.1 The Envelope-Function Approximation

As stated above, we wish to simulate the RTD using the simplest possible quantum mechanical model. The schematic conduction-band diagrams presented in figs. 2.1-2.3 suggest that we treat the electrons in the same manner as one does in tunneling problems from elementary quantum mechanics. This is exactly what we shall do. Even though the electrons are not free, the envelope-function (or effective mass) approximation allows us (albeit with many restrictions) to treat them as though they were. That such an approximation is possible should come as no surprise. After all, the translational symmetry of the crystal implies that there will be conserved quantities, \( k \), and that the allowed energies will be functions of these quantities, \( E_n(k) \). It is only natural to expect the energies \( E_n(k) \) to be smooth functions of \( k \) with (in general) local maxima and minima. For values of \( k \) in the vicinity of these local extrema, then, the electrons will behave as if they are free particles with either positive or negative effective mass, according to whether the extremum is a minimum or maximum, respectively.

A brief review of the envelope-function approximation is in order. The Bloch states, \( \psi_{n,k}(r) \), satisfying

\[
H_0 \psi_{n,k}(r) = E_n(k) \psi_{n,k}(r)
\]

(2.1)

where \( H_0(r+R) = H_0(r) \), with \( R \) a direct lattice vector, are also eigenstates of the operator \( E_n(-i\nabla) \) [3]:

\[
E_n(-i\nabla) \psi_{n,k}(r) = E_n(k) \psi_{n,k}(r).
\]

(2.2)

Neglecting band-to-band transitions, the wavepacket
\[ \Psi(r, t) = \sum_{n, k} c_{n, k}(t) \psi_{n, k}(r) \]  

(2.3)

satisfies

\[ \left[ E_n(-i \nabla) - e \phi(r) \right] \Psi(r, t) = -\frac{\hbar}{i} \frac{\partial \Psi}{\partial t} \]  

(2.4)

where \( \phi(r) \) is an externally applied potential[4].

As shown by Madelung[5] and Harrison[6], the usual effective mass equation results when \( E_n(k) \) is expanded about an extremum, say \( k = 0 \), while retaining only terms up to second order and the wavepacket (2.3) is restricted to Bloch states near the extremum. That is, we approximate \( \psi_{n, k}(r) = \psi_{n, 0}(r) e^{i k \cdot r} \) to obtain

\[ \left[ -\frac{\hbar^2}{2 m^*} \nabla^2 - e \phi(r) \right] \chi(r, t) = -\frac{\hbar}{i} \frac{\partial \chi}{\partial t} \]  

(2.5)

where

\[ \chi(r, t) = \sum_{n, k} c_{n, k}(t) e^{i k \cdot r} \]  

(2.6)

Several restrictions on the use of equation (2.5) are readily apparent. It is only justified when \( \chi \) extends over a region at least several lattice constants in size and when the applied potential \( \phi(r) \) is likewise slowly varying[5,6]. Additionally, at a heterointerface between materials whose (lowest) conduction-band minima occur at different points in \( k \)-space, it is not immediately obvious in which minimum in each material to construct the wavepackets. The envelope-function model cannot tell us where to carry out this wavepacket construction, for in deriving it we have discarded nearly all (save for the effective mass and conduction-band offsets) material-specific information. Nevertheless, the envelope-function theory is regularly used to model heterojunctions, so we shall
proceed with the most common heterointerface model, that of BenDaniel and Duke[7].

2.2.2 The BenDaniel-Duke Model

In order to illustrate the BenDaniel-Duke model[7], consider a heterointerface at \( z = z_0 \) with material I on the left \( (z < z_0) \) and material II on the right \( (z > z_0) \). Let \( \chi^{I,II}(r) \), \( m^*_{I,II} \) denote the envelope-function and the effective mass in material I, II, respectively. Then, one has merely to solve the time-independent envelope-function equation

\[
\left[ -\frac{\hbar^2}{2m^*_{I,II}} \nabla^2 - eV(r) \right] \chi^{I,II}(r) = E \chi^{I,II}(r)
\]

(2.7)

with the flux-conserving boundary conditions

\[
\chi^I(r||, z_0) = \chi^{II}(r||, z_0), \quad \frac{1}{m^*_I} \left[ \frac{\partial \chi^I}{\partial z} (r||, z_0) \right] = \frac{1}{m^*_II} \left[ \frac{\partial \chi^{II}}{\partial z} (r||, z_0) \right]
\]

(2.8)

where \( r|| = (x, y) \). In (2.7), note that \( V(r) \) includes any conduction-band discontinuity between materials I and II as well as any externally applied bias. For an unbiased heterojunction, it is customary to set \( V(r) = 0 \) in the (bulk) material with the lowest conduction band.

Most applications of (2.7), (2.8) will be one-dimensional. That is, \( V(r) = V(z) \) only, so that any electric fields (either externally applied or resulting from space-charge in the device) have only \( z \)-components and all heterointerfaces can be considered perfect in the \( x-y \) plane[8]. In these cases, it is obvious that (2.7) separates, and that the envelope function may be written

\[
\chi^{I,II}(r) = e^{i k||r||} \varphi^{I,II}(z)
\]

(2.9)
where $k_{||} = (k_x, k_y)$ is a constant of the motion. Then, the $z$-only envelope-function equations read:

$$
\left[ -\frac{\hbar^2}{2 m^*_{||}} \frac{d^2}{dz^2} - eV(z) \right] \varphi_i(z) = \left( E - \frac{\hbar^2 k_{||}^2}{2 m^*_{||}} \right) \varphi_i(z) = E^I \varphi_i(z)
$$

(2.10)

$$
\left[ -\frac{\hbar^2}{2 m^*_{||}} \frac{d^2}{dz^2} - eV(z) \right] \varphi_{II}(z) = \left( E - \frac{\hbar^2 k_{||}^2}{2 m^*_{||}} \right) \varphi_{II}(z) = E^{II} \varphi_{II}(z)
$$

(2.11)

Note that the eigenvalues $E^I$ and $E^{II}$ in (2.10) and (2.11) are in general different, since the effective masses $m^*_{||}$ are not necessarily identical.

Often, the potential $V(z)$ includes a contribution from space charges and/or an applied bias. In these cases, it is most convenient to discretize $V(z)$ into many small “slices” (each usually one monolayer = 2.83Å in GaAlAs) and solve (2.10) or (2.11) as appropriate in each “slice”, applying boundary conditions (2.8) to connect the envelope-functions in one “slice” to those in adjacent “slices.” The resulting equations are easily rearranged into a 2x2 transfer matrix[9]:

$$
\begin{bmatrix}
A^{(n)} \\
B^{(n)}
\end{bmatrix} =
\begin{bmatrix}
\frac{\alpha}{2} \exp[ z_n (k^{(n+1)} - k^{(n)})] & \frac{\beta}{2} \exp[- z_n (k^{(n+1)} + k^{(n)})] \\
\frac{\beta}{2} \exp[ z_n (k^{(n+1)} + k^{(n)})] & \frac{\alpha}{2} \exp[- z_n (k^{(n+1)} - k^{(n)})]
\end{bmatrix}
\begin{bmatrix}
A^{(n+1)} \\
B^{(n+1)}
\end{bmatrix}
$$

(2.12)

where $z_n$ is the boundary between the n-th and (n+1)-st “slices” and in the j-th “slice” we have:

$$
\varphi^{(j)}(z) = A^{(j)} \exp[k^{(j)} z] + B^{(j)} \exp[- k^{(j)} z]
$$

(2.13)

$$
\alpha = 1 + \frac{m^*_{j} k^{(j+1)}}{m^*_{j+1} k^{(j)}}, \quad \beta = 1 - \frac{m^*_{j} k^{(j+1)}}{m^*_{j+1} k^{(j)}}
$$

(2.14)
\[ k^{(j)} = \sqrt{\frac{2 m^* j}{\hbar^2} [V^{(j)} - E^{(j)}]} \]  

(2.15)

In (2.15), \( V^{(j)} = V[(z_j + z_{j-1})/2] \) and \( E^{(j)} = E - \hbar^2 k^2_{\parallel}/2m^*_j \). [Note that \((z_j + z_{j-1})/2\) is the center of the \( j \)-th “slice.”]

By repeated multiplication of the transfer matrices, we can “propagate” the envelope-function from right to left. For example, consider the case in which layer (or “slice”) 0 represents the bulk emitter of an RTD and layer \( N \) represents the bulk collector. (By “bulk” we mean a region in which band-bending and band discontinuities due to space charges or interfacial disruptions have ceased.) We choose an energy \( E \) so that \( k^{(0)} \) and \( k^{(N)} \) are both purely imaginary (i.e., there are propagating states in both the emitter and collector at this energy). Then, for transmission from layer 0 to layer \( N \) we have \( B^{(N)} = 0 \) and the transmission and reflection coefficients (which are, respectively, ratios of the transmitted and reflected to incident probability current density) are given by:

\[
T = \left| \frac{A^{(N)}}{A^{(0)}} \right|^2 \frac{m^* k^{(N)}}{m^*_N k^{(0)}} \quad R = \left| \frac{B^{(0)}}{A^{(0)}} \right|^2
\]  

(2.16)

where of course \( T + R = 1 \). Note that flux conservation is guaranteed even though the \( z \)-equation eigenvalues \( E^{(j)} \) may differ.

The tunneling current density is easily found in terms of the transmission coefficient:

\[
J_z = \frac{2e}{(2\pi)^3} \frac{1}{\hbar} \int \int \int [f_d(E) - f_v(E)] T(E, k_{||}, V) dE dk_y dk_x
\]  

(2.17)
where \( f_e(E) \) and \( f_c(E) \), the emitter and collector Fermi-Dirac functions, respectively, are given by:

\[
f_e(E) = \frac{1}{1 + \exp[(E - \mu_e)/k_B T]}
\]

\[
f_c(E) = \frac{1}{1 + \exp[(E - \mu_c)/k_B T]}
\]

(2.18) (2.19)

where \( \mu_c = (\mu_e - eV) \) and \(-eV\) is the applied bias. In cases in which the effective mass disparity between the two materials is not large, and/or the temperature is sufficiently low that only states of very small \( k_{||} \) are occupied, we may ignore the \( k_{||} \)-dependence of \( T \) and approximate the current density by the familiar single-integral[1]:

\[
J_z = \frac{e m^*_e k_B T}{2\pi^2 \hbar^3} \int_0^\infty T(E, V) \ln \left| \frac{1 + \exp[(\mu_e - E)/k_B T]}{1 + \exp[(\mu_e - E - eV)/k_B T]} \right| dE
\]

(2.20)

where \( m^*_e \) is the effective mass of the electrons in the bulk emitter.

Finally, we note that the accuracy of the orthodox BenDaniel-Duke model may be improved by incorporating either real- or complex-band nonparabolicity in the expression for \( k \), eq. (2.15). In particular, the dispersion relation for the imaginary band connecting the light-hole valence band and the lowest conduction band at the \( \Gamma \)-point is[10]:

\[
E = E_0 \pm \frac{\hbar^2 \gamma}{m^*} \sqrt{k^2 + \gamma^2}
\]

(2.21)

\[
\gamma = \frac{E_{(\Gamma)}}{\hbar^2 (1/m_e + 1/m_{lh})}
\]

(2.22)
\[ E_0 = E_{c, \text{ min}}^{(\Gamma)} - \frac{E_g^{(\Gamma)}}{1 + m_e/m_{lh}}, \]

where \( m_e \) and \( m_{lh} \) are the electron and light-hole effective masses, respectively; \( E_{c, \text{ min}}^{(\Gamma)} \) and \( E_g^{(\Gamma)} \) are respectively the energy of the conduction band minimum and the energy gap at the \( \Gamma \)-point; and \( \gamma \) is the magnitude of \( k \) at the imaginary band branch point (i.e., where \( \partial k / \partial E = 0 \)) and \( E_0 \) is the energy at that point. In (2.21), for \( E > E_0 \) (branch point-to-conduction band), the positive sign is to be taken and \( m^* = m_e \), whereas for for \( E < E_0 \) (branch point-to-valence band), the negative sign is to be taken and \( m^* = m_{lh} \). Rearranging (2.21)-(2.23), for \( E > E_0 \), we find:

\[ k = \gamma A \sqrt{\frac{m_e^2}{\hbar^4} \left[ \left( E - E_{c, \text{ min}}^{(\Gamma)} \right)^2 + \left( E_{c, \text{ min}}^{(\Gamma)} - E_0 \right)^2 \right] - 1} \]  

(2.24)

We may then use (2.24) in place of (2.15) in the BenDaniel-Duke model, noting that \( (E - E_{c, \text{ min}}^{(\Gamma)}) = (E(\bar{u}) - V(\bar{u})) \).

2.2.3 Space-Charge Regions

A complete solution of the tunneling problem would include self-consistently solving the Schrödinger and Poisson equations. Computational limitations prevent us from doing this, so in order to include space-charge effects, we make the following assumption: the first barrier is impenetrable. Then if \( z_b \) is the location of the first barrier, we may assign an emitter fermi level \( \mu_e \) valid for \( z < z_b \) and a collector fermi level \( \mu_c = (\mu_e - eV) \), where \(-eV\) is the applied bias, valid for \( z > z_b \). Defining the vacuum level \( E_{\text{vac}} \) by:

\[ E_{\text{vac}}(z) = E_c(z) - E_f(z) + \text{constant} \]  

(2.25)
where \( E_f(z) = \{ \mu_e, z < z_b; \mu_c, z > z_b \} \). In this work we only study electron tunneling, so the only charges are the electrons and the ionized donors (we assume full donor ionization). Then, the Poisson equation reads:

\[
- \frac{d}{dz} \left[ \varepsilon \frac{d\phi}{dz} \right] = e (N_D - n)
\]  

(2.26)

where \( \phi = -E_{\text{vac}}/e \), \( \varepsilon \) is the dielectric constant, \( n \) is the electron density, and \( N_D \) is the donor density. We discretize (2.26) over a mesh of points \( z_i, i = 0, 1, \ldots, N \), with constant "slice"-width \( h_s = (z_{i+1} - z_i) \):

\[
\frac{1}{h_s^2} \left[ \varepsilon (z_{i+1}/2) E_{\text{vac}}(z_{i+1}) - [\varepsilon (z_{i+1}/2) + \varepsilon (z_{i-1}/2)] E_{\text{vac}}(z_i) + \varepsilon (z_{i-1}/2) E_{\text{vac}}(z_{i-1}) \right] =
\]

\[
= e^2 \left[ N_D(z_i) - n(z_i) \right], \quad i = 1, 2, \ldots, (N-1)
\]  

(2.27)

with boundary conditions \( E_{\text{vac}}(z_0) = E_{\text{vac},0}, E_{\text{vac}}(z_N) = E_{\text{vac},N} \). (Note that if the emitter and collector are made of the same material and doped identically \( E_{\text{vac},N} = E_{\text{vac},0} = eV \).) In (2.27), note that the electron density, \( n \), is a function of the vacuum level. We usually employ the Joyce-Dixon approximation[11] to relate \( n \) and \( E_{\text{vac}} \), but for low densities, one can use Boltzmann statistics instead. Note, too, that the dielectric constant mesh is offset from the vacuum level and density meshes by \( h_s/2 \). The vacuum level, \( E_{\text{vac}} \), found by solving (2.27) is used as an input to the Schrödinger equation for both the envelope-function and tight-binding (Ch. 3) models. In all the work described in this dissertation, the Poisson equation software employed is based on programs written by Paul van der Waat. An example conduction-band profile calculated from the Poisson equation is shown in fig. 2.5.
Figure 2.5: Conduction-band diagram for a double barrier RTD. The conduction-band minimum is indicated by the solid line while the fermi level is indicated by the dotted-dashed line. Each AlAs barrier is 6 monolayers wide, the GaAs well is 18 monolayers wide, and the GaAs emitter and collector each have 35 monolayer undoped spacers. The emitter and collector are both doped $N_D = 10^{18} \text{ cm}^{-3}$. We take $T = 77K$, and the applied bias is 0.5 V.

2.2.4 Series Resistance

Up to this point we have discussed only idealized RTDs. One very important nonideal effect (which is also easy to model) is that of series resistance. A large part of this resistance is due to the metal-semiconductor contacts to the emitter and collector regions. Likewise, the bulk emitter and collector, although heavily doped, will have finite conductivities, although they will make a smaller contribution to the total resistance than will the contacts. Regardless of its origin, the series resistance is modeled by placing a resistor in series with an ideal RTD. Because the ideal I-V characteristic displays negative
differential resistance, it is apparent that any (nonzero) series resistor will introduce hysteresis. In addition, including series resistance in the simulation will often result in higher calculated valley currents, for the potential drop across the resistor will push the first peak in the ideal RTD characteristic towards the second.

Explicitly, we model the effects of a series resistor by mapping each point of the $J-V$ curve ($V_{\text{RTD}}, J_{\text{RTD}}$) (calculated from either the tight-binding or envelope-function models) to the point ($V_{\text{TOT}}, I_{\text{TOT}}$), where $I_{\text{TOT}} = J_{\text{RTD}}A$ and $V_{\text{TOT}} = V_{\text{RTD}} + J_{\text{RTD}}A R_S$. In these expressions, $A$ is the cross-sectional area of the RTD, $R_S$ is the series resistance, and $V_{\text{TOT}}$ is the total voltage across the actual device (ideal RTD in series with the resistor, $R_S$). From these definitions, it is clear that if the ideal-RTD $J-V$ characteristic, $(V_{\text{RTD}}, J_{\text{RTD}})$, displays NDR the current through the device, $I_{\text{TOT}}$, will not be a single-valued function of the voltage across it, $V_{\text{TOT}}$: points of high current density will receive larger voltage shifts than those of lower current density. Consequently, the $J-V$ characteristic can display hysteresis. This will usually lead to some instability in the device, preventing us from measuring either the true peak or valley current densities.

In figure 2.6 and 2.7 we examine the effect of series resistance on a single barrier tunneling structure. The device consists of a single AlAs barrier 10 monolayers wide, sandwiched between two undoped GaAs spacer layers (each 35 monolayers) and a GaAs emitter and collector each doped to $N_D = 10^{18}$ cm$^{-3}$. The temperature is 77 K. Figure 2.6 presents the $J-V$ characteristics of the device as calculated with the tight-binding model (Ch. 3) without series resistance. (In Chapter 3, sec. 6 we discuss transmission coefficient graphs of single barrier structures, show that they exhibit resonances which make NDR possible and explain the origins of the resonances.) Figure 2.7 presents the I-V characteristics for the same device, assuming a cross-sectional area of 100 $\mu$m$^2$ and a series resistance of 5 $\Omega$. (We choose these values because they are fairly representative of what we have encountered in our laboratory.) The solid line is the I-V characteristic for bias starting at zero and increasing while the dotted line is that for bias starting at $V > 0$ and decreasing; hysteresis is clearly evident. As discussed above, note that the first peak of the ideal device's $J-V$ curve has been pushed towards the second by the series resistance,
**Figure 2.6:**  J-V curve for a 10 monolayer single barrier structure (see text).

**Figure 2.7:**  I-V curve for the device of fig. 2.6 including series resistance (see text).
resulting in a lower peak-to-valley ratio (about 1.36 with series resistance and about 3.19 without). Thus, when the series resistance of the device can be well-established, it is important to include it when comparing theoretical calculations to measured I-V characteristics.

2.3 MBE GROWTH OF GaAs/AlAs RTDs[12]

Molecular beam epitaxy (MBE) is a method of crystal growth with which one can produce crystals having layers of atomic dimensions (i.e., only a few Å thick). The apparatus itself (fig. 2.8) consists of an ultrahigh vacuum (UHV) chamber having a base pressure of around $10^{-10}$ torr, effusion cells which heat elemental sources, and a substrate rotator and heater. There are shutters in front of the source cells to allow the sources to be turned "on" and "off". At such low pressures, the molecules boiling off the surface of a heated source (the "melt") have a mean free path of a few meters, which is far larger than the chamber dimension, so that they truly behave as a beam rather than as a diffusing gas. (Thus, almost as soon as a source shutter is closed, that source may be considered to be "off".) In order to keep molecules which either miss or bounce off the substrate from continuously flying about the chamber, there is a liquid nitrogen cooled shroud (not shown in fig. 2.8) inside the chamber wall; molecules which collide with this shroud stick to it.

In the lattice-matched GaAs/AlAs materials system, crystals are grown with an excess arsenic overpressure (usually 16-22 times the group-III pressure, depending on whether As$_2$ or As$_4$ is the group-V source) in the chamber because at typical growth temperatures (around 600 °C), excess arsenic evaporates from the crystal. (Indeed, heating a GaAs wafer to 400 °C without an arsenic background will cause arsenic to leave the first few monolayers nearest the wafer surface.) Since arsenic behaves in this rather benevolent manner, it is the group-III elements which determine the growth rate. Fortunately, at typical growth temperatures, group-III elements tend to have near-unity sticking coefficients to arsenic, so that to grow GaAs, for example, one need only open the gallium
and arsenic shutters. Dopants, such as Si (n-type) or Be (p-type), may be incorporated into the sample by opening the appropriate shutters as the crystal is growing. In order to enhance epitaxial layer uniformity, the substrate is rotated during growth.

Finally, to achieve atomic layer control over growth, the growth rate and its relation to the beam pressure must be determined, since only the beam pressure is monitored during most growth. The growth rate is found by reflection high energy electron diffraction (RHEED)\cite{13}. Approximately 10 keV electrons are incident on the substrate at a grazing angle. Whenever the surface from which the beam diffracts is ordered, there is a clear pattern having regions of constructive and destructive interference; this is referred to as a “maximum”. As the surface becomes more disordered, the interference pattern becomes progressively washed out, resulting in a “minimum”. Thus, one cycle (maximum-minimum-maximum) indicates the growth of one monolayer of material. In order to relate the growth rate to beam pressure, both are plotted as functions of $1/T$, where $T$ is the absolute temperature. The growth rate is approximately proportional to $\exp[-E/kT]$, where $E$ is, roughly, an activation energy for molecules to escape from the melt. The beam pressure, on the other hand, is approximately proportional to $\sqrt{T} \exp[-E/kT]$, since it necessarily includes a velocity factor. (Recall that a free classical particle has translational kinetic energy -- proportional to $v^2$ -- of $3/2 \ kT$, hence the $\sqrt{T}$ factor.) Since growth is carried out only within a fairly narrow temperature range, however, the $\sqrt{T}$ factor can be treated as a constant, so that the beam pressure is assumed proportional to the growth rate. Thus, MBE achieves the fine control necessary for growing nanostructures such as RTDs.
2.4 FABRICATION OF RTDs[14]

Although not discussed in this dissertation, RTDs were fabricated as a part of this research. These devices, which were grown by Paul van der Wagt on the Harris group System 1 MBE machine, were designed to study tunneling from AlAs through GaAs barriers. (Recall that the GaAs X-valleys lie above those of AlAs.) Unfortunately, these devices did not work. Nevertheless, we shall briefly discuss the processing method employed. The epitaxial layers are grown on an n+ GaAs substrate and on the surface is an InAs/InGaAs contact layer, consisting of 200 Å of n++ InAs above a 10-period n++ superlattice of GaAs and either In$_{0.5}$Ga$_{0.5}$As or In$_{0.7}$Ga$_{0.3}$As, each period being 30 Å. The width of the GaAs in each period varies linearly from 0 Å just beneath the InAs to 27 Å furthest from it[15]. The processing proceeds as follows. First, a backside Au/Ge/Ni/Au (400 Å/100 Å/125 Å/1000 Å) ohmic contact (the thinner gold layer contacting the wafer) is evaporated onto the wafer and alloyed (fig. 2.9a). Then, photoresist is deposited on the
top of the wafer (InAs/InGaAs contact layer side) and developed. A non-alloyed Ti/Au
(200 Å/4800 Å, Ti contacting the InAs) ohmic contact is evaporated onto the wafer (fig.
2.9b) and the photoresist and excess metal are lifted off in acetone. In order to isolate
devices, the wafer is implanted with two successive $1 \times 10^{12}$ atoms/cm$^2$ doses of B+$,$ the
first at 100 keV, the second at 40 keV, with the 4800 Å thick Au contact serving as implant
mask. (This will isolate devices up to about 0.25-0.35 μm beneath the surface.) Because
the InAs/InGaAs contact layer remains conductive after the implants, it is etched off (fig.
2.9c). Other steps, such as depositing metal bonding pads, can be added to the process,
but they were not found necessary for this work. Table 2.1 lists the complete process.

**TABLE 2.1: IMPLANT-ISOLATED RTD PROCESS**

A. **Backside Ohmic Contact**
   1. Remove surface oxide by dipping in 1:1 HCl:H$_2$O for 10 sec-1 min.
      DI rinse and N$_2$ blow dry.
   2. Deposit backside metal: Au/Ge/Ni/Au (400 Å/100 Å/125 Å/1000 Å)
   3. Alloy in RTA at 430 °C for 20 sec, in Ar.

B. **TopsiDe Ohmic Contact: Photolithography**
   1. Singe wafer for 5-10 min. at 120 °C.
   2. Spin on AZ-1370-SF photoresist, 3500 RPM, 30 sec.
   3. Prebake at 85 °C for 20 min.
   4. Soak in chlorobenzene for 8 min.
   5. N$_2$ blow dry.
   6. Continue bake at 85 °C for 5 min.
   7. Expose in OAI aligner, intensity switch at position 2, 11 sec.
   8. Develop in 1:1 AZ concentrate:H$_2$O for 15 sec after clearing (~60 sec total).
   9. DI rinse 90-120 sec followed by N$_2$ blow dry.
   10. Check in microscope; return to step 8 if under-developed.

C. **TopsiDe Ohmic Contact: Deposition, Liftoff, and Isolation**
   1. Remove surface oxide (step A.1) above.
   2. Deposit topside metal: Ti/Au (200 Å/4800 Å).
   3. Liftoff in acetone for 1 hr or more, or use shaker. *NEVER* let the surface dry!
   4. Dip in methanol a few minutes.
   5. DI rinse and N$_2$ blow dry.
   6. Implant with B+$,$ dose $10^{12}$ atoms/cm$^2$, twice. 100 keV first, 40 keV second.
   7. Remove InGaAs layer by etching in 2:1:50 H$_3$PO$_4$:H$_2$O$_2$:H$_2$O at 0 °C.
   Etch rate is approximately 200 Å/min, but use Alpha-Step to calibrate.
Figure 2.9a: GaAs wafer after deposition and alloying of backside ohmic contact.

Figure 2.9b: GaAs wafer after developing photoresist and depositing topside non-alloyed ohmic contact.

Figure 2.9c: GaAs wafer after implantation and removal of InGaAs contact layer. The devices may be measured at this point.
NOTES FOR CHAPTER 2


[8] This is of course not the case in magnetotunneling problems, which we shall not consider here.

[9] Alex Harwit and James S. Harris, Jr., J. Appl. Phys. 60, 3211 (1986). We have modified the results of Harwit and Harris for the case of nonzero $k_{||}$, in which the $z$-equation eigenvalues, $E^{(i)}$, in different “slices” can be different.


[12] The material in this section is based on conversations with Paul van der Wagt and Susan Lord, January 20, 1992.


[14] The implant isolation process described in this section was developed by Kevin Lear and Byung Gook Park while they were members of the Harris group. The process is partially described in Kevin L. Lear, Dissertation, Stanford University (1990), pp. 36-8.

CHAPTER 3
TIGHT-BINDING MODELS

The envelope-function theory discussed in the preceding chapter, although attractive in its simplicity, is completely unsuitable for modeling modern quantum devices. It cannot, for example, correctly describe the state of an electron in the vicinity of a heterointerface. This failure is a result of two restrictions imposed during the derivation of the effective-mass Schrödinger-like equation. First, the said derivation supposes that the potential varies slowly. This is certainly not the case near a heterojunction, where the potential changes abruptly. Second, and more importantly, the presence of a heterointerface disrupts the translational symmetry of the crystal in the direction perpendicular to the interface plane, so that the electron’s state is, in general, a superposition of the allowed bulk and evanescent states on either side of the interface. Hence, an envelope-function description, which only allows one pair of states on either side of the interface, is no longer appropriate.

We must therefore develop another theory for treating quantum devices. Since we must deal with interfaces, an approach based upon localized orbitals is most appropriate. Models of this type are referred to as tight-binding models [1]. These models will also allow us to properly incorporate the symmetries of the crystal from which the device is made, and will in addition automatically take into account effects such as nonparabolicity of the bands, and multiple valleys and bands. The price we pay is increased complexity and some loss of the easy intuitive picture presented by the envelope-function theory.

3.1 UNDERLYING APPROXIMATIONS

In spite of the fact that we are going to relax many of the constraints imposed by the
envelope-function theory, our method makes important approximations as well. Two of the most significant are the “frozen-lattice” and the single-electron approximations. The former neglects the motion of the ions (hence, no phonons), while the latter assumes that the N-electron problem may be reduced to N identical one-electron problems, each with a potential term which includes the interactions with the other electrons in an average way. (The argument in favor of this approximation is usually put forth in terms of a Hartree-Fock-like procedure, even though straightforward Hartree-Fock calculations do not correctly predict the electronic structure of most solids.) The Hamiltonian for our system is therefore given by:

\[ H = \frac{p^2}{2m} + V(r) + H_{SO} \]  \hspace{1cm} (3.1)

with the potential such that in a perfect crystal \( V(r) = V(r + R) \), where \( R \) is a direct lattice vector of the crystal, and \( H_{SO} \) is the spin-orbit Hamiltonian. Note that the mass appearing in (3.1) is the free-electron mass and not the effective mass.

3.2 OPERATORS, ORBITALS, AND SYMMETRIES

In this section, we state and develop several important results which will be needed in succeeding sections of this chapter.

3.2.1 Rotations and Translations of the Coordinate System

Let \( R_n(\mu) \) denote the rotation of the coordinate system about the vector \( n \) by an angle \( \mu \). Let \( r = (x,y,z)^T \) denote the coordinates of a point in the original system and let \( r' = (x',y',z')^T \) denote the coordinates of the same point in the rotated system:
\[ r' = R_n(\mu) \mathbf{r} . \quad (3.2) \]

Written in matrix form, the rotations about the x-, y-, and z-axes are:

\[
R_x(\alpha) = \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos(\alpha) & \sin(\alpha) \\
0 & -\sin(\alpha) & \cos(\alpha)
\end{bmatrix}, \quad R_y(\beta) = \begin{bmatrix}
\cos(\beta) & 0 & -\sin(\beta) \\
0 & 1 & 0 \\
\sin(\beta) & 0 & \cos(\beta)
\end{bmatrix}
\]

\[
R_z(\gamma) = \begin{bmatrix}
\cos(\gamma) & \sin(\gamma) & 0 \\
-\sin(\gamma) & \cos(\gamma) & 0 \\
0 & 0 & 1
\end{bmatrix}
\quad (3.3)
\]

Let \( I_n \) denote inversion about \( n \) (reflection in the plane perpendicular to \( n \)). The matrices for inversions about x, y, and z are:

\[
I_x = \begin{bmatrix}
-1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}, \quad I_y = \begin{bmatrix}
1 & 0 & 0 \\
0 & -1 & 0 \\
0 & 0 & 1
\end{bmatrix}, \quad I_z = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1
\end{bmatrix}
\quad (3.4)
\]

Finally, let \( T_d \) denote translation of the coordinate system by \( -d \). The new coordinates, \( r' \), in terms of the old are:

\[ r' = T_d \mathbf{r} = \mathbf{r} + d . \quad (3.5) \]

\[ \]

3.2.2 \quad \textit{Behavior of the State Vector Under Rotations and Translations} \]

Let \( R \) and \( T \) represent rotations and translations, respectively, of the coordinate system, and let \( R \) and \( T \) be their corresponding quantum-mechanical operators. Let \( \psi(\mathbf{r}) \) be the wavefunction in the original coordinates and let \( \psi'(\mathbf{r}') \) be the wavefunction in the
transformed coordinates. Consider first the case of rotations so that \( r' = R r \). We then have the following results [2]:

\[
\psi'(r') = \psi(R^{-1} r') \tag{3.6}
\]

\[
|\psi'> = R|\psi>
\]

\[
R|r> = |R r>
\tag{3.7}
\]

\[
R^\dagger|r> = |R^{-1} r>
\tag{3.8}
\]

\[
R^\dagger = R^{-1}
\tag{3.9}
\]

Similar results of course hold for translations.

Now consider the case of both translations and rotations. Once again let \( r' = R r \). Let the wavefunction in the original coordinates be \( \phi \) and let \( \phi' \) be the wavefunction in the rotated coordinates. It is natural to demand that \( \phi \) and \( \phi' \) have the same value at corresponding points in space:

\[
\phi'(r' - R d) = \phi(r - d), \tag{3.11}
\]

\[
T(R d) \phi'(r') = T(d) \phi(r)
\tag{3.12}
\]

\[
<r'|T(R d)|\phi'> = <r|T(d)|\phi'>.
\tag{3.13}
\]

Using (3.7) and (3.8),

\[
<r|R^\dagger T(R d)R|\phi'> = <r|T(d)|\phi'>,
\tag{3.14}
\]

and with the aid of (3.10), we finally obtain:

\[
T(R d) = R T(d) R^\dagger.
\tag{3.15}
\]
3.2.3  
Symmetries of the Zinblende Crystal

Most of the commonly used semiconductors, including gallium arsenide, silicon, and germanium, have the zinblende crystal structure, which consists of two interpenetrating face-centered cubic (FCC) Bravais lattices, one offset from the other by \(a/4(1,1,1)\), where \(a\) is the length of a side of the FCC conventional unit cell (cube). Zinblende is thus an FCC lattice with a two-atom basis. We shall take the convention that the anions are located on the Bravais lattice sites. (In the case of an elemental crystal, such as Si, both the anion and cation sites are occupied by the same type of atom.) The relative positions of the four nearest-neighbors (which are all cations) to an anion are:

\[
d_{c(1)} = \frac{a}{4}(1, 1, 1); \quad d_{c(2)} = \frac{a}{4}(1, -1, -1); \quad d_{c(3)} = \frac{a}{4}(-1, 1, -1); \quad d_{c(4)} = \frac{a}{4}(-1, -1, 1).
\]

Note that under the cyclic permutations \(x \rightarrow y \rightarrow z \rightarrow x\), \(d_{c(1)} \rightarrow d_{c(1)}\), and \(d_{c(2)} \rightarrow d_{c(3)} \rightarrow d_{c(4)} \rightarrow d_{c(2)}\).

Inspection of (3.16) reveals that the zinblende crystal has lower symmetry than the FCC lattice. For example, the FCC lattice is invariant under a four-fold rotation about \(x\), \(y\), or \(z\), while zinblende crystal clearly is not. We will primarily be interested in only two of the symmetry operations of the zinblende crystal:

1. Rotation of \(\pm \pi\) about the \(x\)-, \(y\)-, or \(z\)-axis.
2. Rotation of \(\pm \pi/2\) about the \(x\)-, \(y\)-, or \(z\)-axis, followed by a reflection in the plane perpendicular to the rotation axis. (Or, a reflection, followed by a \(\pm \pi/2\) rotation.)

3.2.4  
Lowdin Orbitals

As mentioned in the introduction to this chapter, the model we shall develop will be
based upon localized orbitals. In particular, we will employ localized atomic-like orbitals having \( s^-\), \( p_x^-\), \( p_y^-\), or \( p_z^-\)-like symmetry centered on each anion or cation. Now there will be, in general, a non-zero overlap between an orbital centered on one ion (anion or cation) and one centered on a different ion. Since it is desirable to work with an orthonormal basis, we need to orthogonalize the atomic-like orbitals. In 1950, P. O. Lowdin showed that such a procedure was possible and introduced the method for orthogonalizing the atomic-like orbitals [3]. Furthermore, Slater and Koster have shown that the orbitals so-orthogonalized (Lowdin orbitals) transform under operations of the group of the crystal just as the nonorthogonal functions from which they are constructed [4]. We shall not find it necessary to explicitly refer to Lowdin’s orthogonalization procedure, for, following Slater and Koster, we are going to treat the matrix elements as fitting parameters, as will be explained later. Nevertheless, it is important that we understand that it is, indeed, possible to construct an orthogonal basis having such desirable symmetry properties.

Some useful properties of \( s^-\) and \( p^-\) orbitals (hence, \( s^-\) and \( p^-\)-like Lowdin functions) are listed below:

\[
R_z(\pm \pi)|n\rangle = -|n\rangle, \; n = x, \; y. \tag{3.17}
\]

\[
R_z(\pm \pi)|n'\rangle = |n'\rangle, \; n' = s, \; z. \tag{3.18}
\]

\[
R_y(\pm \pi/2)I_y|x\rangle = \pm|z\rangle \tag{3.19}
\]

\[
R_z(\pm \pi/2)I_x|z\rangle = -|z\rangle \tag{3.20}
\]

\[
R_y(\pm \pi/2)I_j|s\rangle = |s\rangle, \; j = x, \; y, \; z. \tag{3.21}
\]

In (3.19) - (3.21) note that \( I_j \) is the quantum-mechanical operator corresponding to the spatial inversion \( I_j \).

Before proceeding, we should discuss in some detail the specific model we will use. Many tight-binding models use the \( sp^3 \) basis. That is, there are one \( s^-\)-like and three \( p^-\)-like \( (p_x, \; p_y, p_z) \) orbitals centered on each ion in the crystal. This choice of basis is motivated by the fact that the valence electrons of the free atoms of which the crystal is
composed lie in s- and p-orbitals. Since the free-atom p-orbitals are only partially filled, this model will, of course, generate excited states (e.g., conduction bands in a semiconductor). Unfortunately, the \( \text{sp}^3 \) basis does not allow us sufficient freedom to accurately fit the conduction bands. Vogl, Hjalmarsen, and Dow [5] have largely remedied this deficiency with their \( \text{sp}^3\text{s*} \) basis, which has five orbitals per atom. Four of these orbitals are the usual ones from the \( \text{sp}^3 \) model. The fifth is an excited s-like orbital, named s*. It is the extra freedom afforded by the s* orbitals which allows for a superior fit of the conduction bands. Finally, let us remark that the model of Vogl, Hjalmarsen, and Dow [5] includes only same-site and nearest-neighbor interactions. Rather amazingly, this assumption of highly localized orbitals and the resulting small set of matrix elements can quite accurately describe the bands in a wide range of zincblende- and diamond-lattice semiconductors.

### 3.2.5 Matrix Elements and the Nearest-Neighbor Model

In order to diagonalize the Hamiltonian (3.1), we will need to compute its matrix elements with the Lowdin orbitals centered on the various ions. The nearest-neighbor model retains only those matrix elements of \( H \) between orbitals centered on the same, or nearest-neighbor sites. Furthermore, matrix elements between s- and p-like orbitals, or between different p-like orbitals, centered on the same site will vanish [6]. In addition, because both the orbitals and \( H \) should strongly resemble their free-atom counterparts near an ion, we discard matrix elements between s and s* orbitals centered on the same site. Finally, we ignore the spin-orbit interaction in this model.

Before proceeding with the calculation of several examples (we will merely state the results for most matrix elements), let us outline the procedure to be followed. As we will discuss later, it will not be necessary to explicitly calculate any matrix elements. Rather, it will be sufficient to use the symmetries of zincblende to express all of the possible matrix
elements in terms of some minimal set of values, henceforth referred to as “tight-binding parameters.” (This was first suggested by Slater and Koster [4].) Note that because $H$ is real, and because the Lowdin orbitals can be chosen real [7], the tight-binding parameters can be chosen real. This is the choice we shall make.

For the nearest-neighbor only, $sp^3s^*$ model, it turns out that all of the possible matrix elements may be expressed in terms of the following tight-binding parameters (notation follows Ref. [4].)

\[
\begin{align*}
E_{sa,sa}(0,0,0) &= \langle sa; 0|H|sa; 0 \rangle \tag{3.22} \\
E_{s^*a,s^*a}(0,0,0) &= \langle s^*a; 0|H|s^*a; 0 \rangle \tag{3.23} \\
E_{pa,pa}(0,0,0) &= \langle xa; 0|H|xa; 0 \rangle \tag{3.24} \\
E_{sc,sc}(0,0,0) &= \langle sc; d_{c(j)}|H|sc; d_{c(j)} \rangle, j = 1,2,3,4 \tag{3.25} \\
E_{s^*c,s^*c}(0,0,0) &= \langle s^*c; d_{c(j)}|H|s^*c; d_{c(j)} \rangle, j = 1,2,3,4 \tag{3.26} \\
E_{pc,pc}(0,0,0) &= \langle xc; d_{c(j)}|H|xc; d_{c(j)} \rangle, j = 1,2,3,4 \tag{3.27} \\
E_{sa,sc}(1/2,1/2,1/2) &= \langle sa; 0|H|sc; d_{c(1)} \rangle \tag{3.28} \\
E_{sa,s^*c}(1/2,1/2,1/2) &= \langle sa; 0|H|s^*c; d_{c(1)} \rangle \tag{3.29} \\
E_{s^*a,sc}(1/2,1/2,1/2) &= \langle s^*a; 0|H|sc; d_{c(1)} \rangle \tag{3.30} \\
E_{s^*a,s^*c}(1/2,1/2,1/2) &= \langle s^*a; 0|H|s^*c; d_{c(1)} \rangle \tag{3.31} \\
E_{sa,pc}(1/2,1/2,1/2) &= \langle sa; 0|H|xc; d_{c(1)} \rangle \tag{3.32} \\
E_{s^*a,pc}(1/2,1/2,1/2) &= \langle s^*a; 0|H|xc; d_{c(1)} \rangle \tag{3.33} \\
E_{pa,sc}(1/2,1/2,1/2) &= -\langle xa; 0|H|sc; d_{c(1)} \rangle \tag{3.34} \\
E_{pa,s^*c}(1/2,1/2,1/2) &= -\langle xa; 0|H|s^*c; d_{c(1)} \rangle \tag{3.35} \\
E_{x,x}(1/2,1/2,1/2) &= \langle xa; 0|H|xc; d_{c(1)} \rangle \tag{3.36} \\
E_{x,y}(1/2,1/2,1/2) &= \langle xa; 0|H|yc; d_{c(1)} \rangle \tag{3.37}
\end{align*}
\]

In the above equations, $|xc; d_{c(1)}\rangle$, for example, denotes a $p_x$-like Lowdin orbital centered at position $d_{c(1)}$ [see eq. (3.16)]. In (3-34) and (3-35) the negative signs follow because the orientation of the $p_x$-orbital relative to the s-orbital is opposite to that in (3-32) and (3-33).
As an example, let us see which matrix elements we can relate to the tight-binding parameter $E_{sa,pc}(1/2,1/2,1/2)$. From sec. 3.2.3, we know that:

$$ H = R_z(\pi)H R_z^\dagger(\pi) \quad (3.38) $$

Using (3.38) in (3-32) gives:

$$ <sa; 0| R_z(\pi)H R_z^\dagger(\pi)|xc; d_{c(1)} > = E_{sa,pc}(1/2,1/2,1/2). \quad (3.39) $$

Now $|xc; d_{c(1)} > = T(d_{c(1)}) |xc; 0 >$, $R_z^\dagger(\pi) = R_z^{-1}(\pi) = R_z(-\pi)$, and $I = R_z^\dagger(-\pi)R_z(-\pi)$, so that:

$$ <sa; 0| R_z(\pi)H R_z(-\pi)T(d_{c(1)})R_z^\dagger(-\pi)R_z(-\pi)|xc; 0 > = E_{sa,pc}(1/2,1/2,1/2). \quad (3.40) $$

Employing (3.15), (3.16), (3.17), and (3.21) we have:

$$ - <sa; 0| H[T[R_z(-\pi) (a/4,a/4,a/4)]|xc; 0 > = E_{sa,pc}(1/2,1/2,1/2). \quad (3.41) $$

From (3.3) and (3.16), we see that $R_z(-\pi) (a/4,a/4,a/4) = (-a/4,-a/4,a/4) = d_{c(4)}$. Thus,

$$ <sa; 0|H|xc; d_{c(4)} > = -E_{sa,pc}(1/2,1/2,1/2) \quad (3.42) $$

Note that the occurrence of the negative sign could have been deduced equally well by examining the orientation of the $p_x$-orbital relative to the $s$-orbital. We have chosen to employ the more elaborate mathematical apparatus for two reasons. First, it is more elegant. Second, and more important, this apparatus is very useful in deducing matrix elements with vanishing "two-center integrals" (the terminology is that of Slater and Koster [4]) in a second-near neighbor model. (One can use intuitive pictures in this case as well, but significantly more care is required.)

The two remaining matrix elements (those involving $|xc; d_{c(2)} >$ and $|xc; d_{c(3)} >$) may be obtained by repeating the above calculation with the symmetry relation.
\[ H = R_x(\pi)H R_x^\dagger(\pi) \]  
(3.43)

substituted into (3.32) and (3.42), respectively. Matrix elements of the form 
\[ <sa; 0|H|yc; d_{c(j)}> \]  
and 
\[ <sa; 0|H|zc; d_{c(j)}> \]  
are obtained from those found above by the cyclic permutations \( x \rightarrow y \rightarrow z \rightarrow x \). (See sec. 3.2.3 for the results of these permutations on the vectors \( d_{c(j)} \)).

The other symmetry discussed in sec. 3.2.3, namely, a rotation by \( \pm \pi/2 \) about one of the axes followed by an inversion, is useful for finding other matrix elements. For example, substituting the symmetry relation

\[ H = I_y R_y(-\pi/2)H R_y(\pi/2)I_y \]  
(3.44)

(note that \( I_j = I_j^\dagger = I_j^{-1}, j = x,y,z \)) into (3.37) and proceeding in the manner outlined above [equation (3.19) will be useful here] yields:

\[ <za; 0|H|yc; d_{c(4)}> = -E_{x,y}(1/2,1/2,1/2) \]  
(3.45)

As before, cyclic permutations give additional matrix elements. The complete list of same-site and nearest-neighbor matrix elements for the \( sp^3s^* \) model is given below in Table 3.1.
### Table 3.1: Same-Site and Nearest-Neighbor Matrix Elements

#### Same-Site Matrix Elements

<table>
<thead>
<tr>
<th>Expression</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;xa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;ya;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;za;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sc;dc</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*c;dc</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;xc;dc</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;yc;dc</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;zc;dc</td>
<td>H</td>
</tr>
</tbody>
</table>

#### Matrix Elements Between s-like Orbitals

<table>
<thead>
<tr>
<th>Expression</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
</tbody>
</table>

#### Matrix Elements Between p-like Orbitals

<table>
<thead>
<tr>
<th>Expression</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;sa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;s*a;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;xa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;xa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;xa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;xa;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;ya;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;ya;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;ya;0</td>
<td>H</td>
</tr>
<tr>
<td>(&lt;ya;0</td>
<td>H</td>
</tr>
</tbody>
</table>
### TABLE 3.1: CONCLUSION

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle z_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle z_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle z_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle z_0</td>
<td>H</td>
</tr>
</tbody>
</table>

Matrix Elements Between p-like Orbitals

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle x_0</td>
<td>H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Matrix Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>$\langle y_0</td>
<td>H</td>
</tr>
</tbody>
</table>
3.3 PERFECT CRYSTALS: BULK TIGHT-BINDING HAMILTONIANS

We are now in a position to construct the bulk tight-binding Hamiltonian matrix. Since the crystal is perfect, we have:

\[ [H, T(R)] = 0 \]  \hspace{1cm} (3.46)

where \( R \) is a direct lattice vector, so that the solution of (3.1) must be a simultaneous eigenstate of \( H \) and \( T(R) \), with a conserved quantity \( k \); that is, the basis states must satisfy Bloch's Theorem. We apply the Born-von Karman (cyclic) boundary conditions to \( H \). Basis functions satisfying the above requirements are the Bloch sums:

\[ |na; k\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \exp[i k \cdot R_j] |na; R_j\rangle \]  \hspace{1cm} (3.47)

\[ |nc; k\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \exp[i k \cdot (R_j + v)] |nc; R_j + v\rangle \]  \hspace{1cm} (3.48)

where \( n = \{s, x, y, z, s^*\} \), \( N \) is the number of Bravais lattice sites in the crystal, and \( v = d \sigma(1) = (a/4)(1,1,1) \). (Note that we may take the convention that \( R_1 = 0 \).) We employ the linear variational method to find the (approximate) eigenstates of \( H \) in the basis (3.47), (3.48). The general formula for the matrix elements of \( H \), valid for any (zincblende crystal) tight-binding model using Lowdin orbitals is derived below. Note that \( (n, n') \) denote orbital symmetry types, \( (\mu, \mu') \) denote ion type (anion or cation), and matrix elements between states of differing \( k \) in the first Brillouin zone vanish [8].

\[ \langle n'|\mu'; k|H|n\mu; k\rangle = \]

\[ = \frac{1}{N} \sum_{j=1}^{N} \sum_{j'=1}^{N} \exp[i k \cdot (R_j - R_{j'} + v(\delta_{\mu,c} - \delta_{\mu',c}))] \langle n'|\mu'; R_{j'} + \delta_{\mu',c} v|H|n\mu; R_j + \delta_{\mu,c} v\rangle \]  \hspace{1cm} (3.49)
Since we employ the Born-von Karman boundary conditions, define a new vector \( \mathbf{R}_{j''} = \mathbf{R}_j - \mathbf{R}_{j'} \) and replace the sum over \( j \) by one over \( j'' \):

\[
\langle n'\mu';k|H|n\mu;k \rangle = \frac{1}{N} \sum_{j'=1}^{N} \left\{ \sum_{j''=1}^{N} \exp\left[i \mathbf{k} \cdot \left( \mathbf{R}_{j''} + v(\delta_{\mu,c} - \delta_{\mu',c}) \right) \right] \langle n'\mu';\mathbf{R}_{j''} + \delta_{\mu',c}v|H|n\mu;\mathbf{R}_{j''} + \delta_{\mu',c}v \rangle \right\}
\]

(3.50)

The term in curly brackets in (3.50) is independent of \( \mathbf{R}_{j''} \). (This may be shown explicitly by noting that since \( \mathbf{R}_{j''} \) is a direct lattice vector, \( H = T(\mathbf{R}_{j'})HT^\dagger(\mathbf{R}_{j'}) \). Now, \( T^\dagger(\mathbf{R}_{j'}) = T^{-1}(\mathbf{R}_{j'}) = T(-\mathbf{R}_{j'}) \), and \( T(-\mathbf{R}_{j'})|n\mu;\mathbf{R}_{j''} + \mathbf{R}_{j'} + \delta_{\mu,c}v > = |n\mu;\mathbf{R}_{j''} + \delta_{\mu,c}v > \), for example.) It may therefore be taken out of the sum over \( j'' \), which now yields \( N \). Thus, renaming \( j'' \) to be \( j \), we have:

\[
\langle n'\mu';k|H|n\mu;k \rangle = \sum_{j=1}^{N} \exp\left[i \mathbf{k} \cdot \left( \mathbf{R}_j + v(\delta_{\mu,c} - \delta_{\mu',c}) \right) \right] \langle n'\mu';\delta_{\mu',c}v|H|n\mu;\mathbf{R}_j + \delta_{\mu,c}v \rangle
\]

(3.51)

where the sum runs over the neighbors of the orbital \( |n'\mu';\delta_{\mu',c}v > \). Note that the real-space vector in the argument of the exponential in (3.51) is merely the relative vector between the two ions on which the orbitals are centered.

Using equation (3.51) with the summation truncated at nearest-neighbors, and the matrix elements from Table 3.1, we may find the bulk tight-binding Hamiltonian for zincblende crystals in the sp\(^3\)s* model. This matrix is given in Appendix 1. This is a corrected version of that given by Vogl, Hjalmanson, and Dow [5], and it includes matrix elements which they set to zero. All that remains is to specify the tight-binding parameters. Following Slater and Koster [4], we treat them as fitting parameters which will be chosen to reproduce the bulk bandstructure. Our parameters for GaAs and AlAs, which have been chosen to reproduce the lowest conduction band of each material as accurately as possible,
are given in Table 3.2; they are based on those of Refs. [5] and [9]. (The lowest two conduction bands of GaAs and AlAs generated by our parameters are shown in Fig. 3.1.) The energies, effective masses, etc. generated by them are given in Table 3.3. We should note that as the tight-binding parameters are chosen to correctly reproduce the bulk bandstructure, the eigenvalues of the bulk tight-binding Hamiltonian are not predictive. However, we shall see below that the same matrix elements which appear in the perfect-crystal Hamiltonian also show up in the Hamiltonian for interface systems, and in this context, they predict some rather unobvious results.

Before proceeding, we should comment on one important defect of the nearest-neighbor sp$^3$s* Hamiltonian: it cannot correctly reproduce the X-valley transverse effective mass. At the X-points, the determinant $|H - iE|$ is independent of $k_{\parallel}$, the transverse wavevector; that is, the bands from X to W are flat! (For example, the determinant is independent of $k_z$ if $k = 2\pi/a x + k_z z$ [10].) Because the minima in the [001] directions occur somewhat in from the X-points (see table 3.3), the X-valley transverse mass will not actually be infinite. For both GaAs and AlAs, however, the values predicted by the model are unrealistically large. The model predicts $m^{*}_{X,t} = 4.9 \, m_0$ for AlAs and $m^{*}_{X,t} = 5.6 \, m_0$ for GaAs ($m_0$ is the free-electron rest mass), whereas the proper values are about 0.19 $m_0$ and 0.23 $m_0$, respectively. This will have implications for the transfer matrix, to be introduced in the next section; however, as will be explained later, this defect will not affect the utility of the nearest-neighbor model for many important applications.
TABLE 3.2:  **TIGHT-BINDING PARAMETERS**

All values are in eV. Matrix elements not listed are taken to be zero. Our parameters assume that the $\Delta E_c(\Gamma) = 0.6 \Delta E_g(\Gamma)$, and we take the zero of energy to be the top of the GaAs valence bands.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GaAs</th>
<th>AlAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{sa,sa}(0,0,0)$</td>
<td>-8.39000</td>
<td>-8.26631</td>
</tr>
<tr>
<td>$E_{pa,pa}(0,0,0)$</td>
<td>1.07475</td>
<td>0.34429</td>
</tr>
<tr>
<td>$E_{sc,sc}(0,0,0)$</td>
<td>-2.65405</td>
<td>-1.62982</td>
</tr>
<tr>
<td>$E_{pc,pc}(0,0,0)$</td>
<td>3.55475</td>
<td>2.94769</td>
</tr>
<tr>
<td>$E_{s<em>_{a,s}</em>_{a}}(0,0,0)$</td>
<td>8.57475</td>
<td>6.84424</td>
</tr>
<tr>
<td>$E_{s<em>_{c,s}</em>_{c}}(0,0,0)$</td>
<td>6.70475</td>
<td>6.08769</td>
</tr>
<tr>
<td>$4 E_{sa,sc}(1/2,1/2,1/2)$</td>
<td>-6.45130</td>
<td>-6.66420</td>
</tr>
<tr>
<td>$4 E_{xx}(1/2,1/2,1/2)$</td>
<td>1.95460</td>
<td>1.87800</td>
</tr>
<tr>
<td>$4 E_{x,y}(1/2,1/2,1/2)$</td>
<td>4.77000</td>
<td>3.86000</td>
</tr>
<tr>
<td>$4 E_{sa,pc}(1/2,1/2,1/2)$</td>
<td>4.68000</td>
<td>5.60000</td>
</tr>
<tr>
<td>$4 E_{pa,sc}(1/2,1/2,1/2)$</td>
<td>7.70000</td>
<td>7.60000</td>
</tr>
<tr>
<td>$4 E_{s*_{a,pc}}(1/2,1/2,1/2)$</td>
<td>4.85000</td>
<td>4.22000</td>
</tr>
<tr>
<td>$4 E_{pa,s*_{c}}(1/2,1/2,1/2)$</td>
<td>6.90000</td>
<td>8.30000</td>
</tr>
</tbody>
</table>

TABLE 3.3:  **ENERGIES, MINIMA, AND EFFECTIVE MASSES**

Energies are in eV. Wavevectors are in Å\(^{-1}\). Effective masses are expressed in terms of the free-electron rest mass. $E_{X'} = E_{c,MIN}(k_{MIN \text{ in } [001]}) - E_{c,MIN}(\Gamma)$. $E_{XX} = E_{c'}(X) - E_{c,MIN}(k_{MIN \text{ in } [001]})$, where $c'$ denotes the second-lowest conduction band.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>GaAs</th>
<th>AlAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_g$</td>
<td>1.538</td>
<td>3.136</td>
</tr>
<tr>
<td>$E_{c,MIN}(\Gamma)$</td>
<td>1.538</td>
<td>2.497</td>
</tr>
<tr>
<td>$E_{X'}$</td>
<td>0.484</td>
<td>-0.882</td>
</tr>
<tr>
<td>$E_{XX}$</td>
<td>0.268</td>
<td>0.542</td>
</tr>
<tr>
<td>$k_{MIN \text{ in } [001]}$</td>
<td>0.9639</td>
<td>0.8512</td>
</tr>
<tr>
<td>$m^*_{\Gamma}$</td>
<td>0.0689</td>
<td>0.1582</td>
</tr>
<tr>
<td>$m^*_{X,d}$</td>
<td>1.457</td>
<td>1.046</td>
</tr>
</tbody>
</table>
**Figure 3.1a:** The two lowest conduction bands of GaAs from $\Gamma$ to $X$, as reproduced by our tight-binding parameters. The zero of energy is taken to be the top of the GaAs valence bands.

**Figure 3.1b:** The two lowest conduction bands of AlAs from $\Gamma$ to $X$, as reproduced by our tight-binding parameters. The zero of energy is taken to be the top of the GaAs valence bands, and our parameters assume that the $\Delta E_c(\Gamma) = 0.6 \Delta E_g(\Gamma)$. 
3.4 INTERFACE SYSTEMS AND THE TRANSFER MATRIX

In addition to providing convenient descriptions of bulk crystals, tight-binding models are also ideally suited for treating interface systems. Because most quantum devices are grown on [001]-oriented substrates, we shall restrict our attention to [001]-oriented zincblende crystals with interfaces perpendicular to $z$ (i.e., interface planes parallel to the $x$-$y$ plane.) The approach usually taken to treat these systems is the transfer matrix method [11]. Our derivation, given below, will be somewhat different from that of Lee and Joannopoulos [11], and furthermore, it will be restricted to the nearest-neighbor model.

3.4.1 Tight-Binding Hamiltonian for the Interface System and the Single-Monolayer Transfer Matrix

Because the crystal now has one or more interfaces along $z$, $k_z$ is no longer a conserved quantity [12], although $\mathbf{k}_\parallel = k_x \mathbf{x} + k_y \mathbf{y}$ is still conserved. Our basis states must therefore reflect this reduced symmetry. They still satisfy Bloch’s Theorem in the $x$-$y$ plane. We retain the Born-von Karman boundary conditions in the plane and leave the boundary conditions in $z$ unspecified for the moment. For our purposes, it is convenient to introduce the concept of a layer, defined to be an atomic plane of anions and the atomic plane of cations lying directly above it (see fig. 3.2 for the case of GaAs). The Bloch sums in the layer, $L$, are:

$$|\text{n}a; L; \mathbf{k}_\parallel \rangle = \frac{1}{\sqrt{N_\parallel}} \sum_{j=1}^{N_\parallel} \exp\left[i \mathbf{k}_\parallel \cdot \mathbf{R}_j(L)\right] |\text{n}a; L; \mathbf{R}_j(L)\rangle$$  \hspace{1cm} (3.52)

$$|\text{n}c; L; \mathbf{k}_\parallel \rangle = \frac{1}{\sqrt{N_\parallel}} \sum_{j=1}^{N_\parallel} \exp\left[i \mathbf{k}_\parallel \cdot (\mathbf{R}_j(L) + \mathbf{v}_\parallel)\right] |\text{n}c; L; \mathbf{R}_j(L) + \mathbf{v}_\parallel\rangle$$  \hspace{1cm} (3.53)

where $N_\parallel$ is the number of sites in the $x$-$y$ plane, $n$ is, as before, one of \{s, x, y, z, s^\ast\}, $\mathbf{v}_\parallel$
\[ R_j(L) = \begin{cases} R_j(0), & \text{L even} \\ R_j(0) + R_{\text{OFFSET}}, & \text{L odd} \end{cases} \]

(3.54)

where we have, for example, \( R_{\text{OFFSET}} = (a/2)x \). \( R_j \parallel \) is a function of the layer index, \( L \), because in zincblende, the anions (cations) in the layer \( L \) do not lie directly atop the anions (cations) in the layer \( L - 1 \). Note that in zincblende, the spacing between layers is \( a/2 \). Thus, the location of an anion, for example, may be specified either by the combination of the layer index, \( L \), and the planar vector, \( R_j(L) \), or by the corresponding direct lattice vector, \( R_j = R_j(L) + La/2 z \). The former description is more useful in treating interface systems, while the latter is of greater utility with bulk systems, as we have seen above.

As in sec. 3.3, it will be convenient to have an expression for the matrix elements of \( H \) in the basis (3.52), (3.53). Just as we did previously, let \( (n, n') \) denote orbital symmetry types and let \( (\mu, \mu') \) denote ion types. Also, note that the only nonvanishing matrix elements will be between states of the same \( k_{\parallel} \).

\[
\langle n' \mu'; L': k_{\parallel} | H | n\mu; L; k_{\parallel} \rangle = \frac{1}{N_{\parallel}} \sum_{j_1 = 1}^{N_{\parallel}} \sum_{j = 1}^{N_{\parallel}} \exp \left[ i k_{\parallel} \left( R_j(L) - R_{j'}(L') + v_{\parallel}[\delta_{\mu,e} - \delta_{\mu',e}] \right) \right] \times \\
\times \langle n' \mu'; L'; R_{j'}(L') + \delta_{\mu,e} v_{\parallel} | H | n\mu; L; R_j(L) + \delta_{\mu,e} v_{\parallel} \rangle
\]

(3.55)

From (3.54), we may write,

\[
R_{i \parallel}(M) = R_{i \parallel}(0) + \Delta(M) R_{\text{OFFSET}}
\]

(3.56)

where \( \Delta(M) = \{0, M \text{ even}; 1, M \text{ odd} \} \). Substituting (3.56) into (3.55) for both \( R_j(L) \) and \( R_{j'}(L') \), defining \( R_j''(0) = R_j(0) - R_{j'}(0) \), and replacing the sum over \( j \) by a sum over \( j'' \) (we may do this because we employ the cyclic boundary conditions in the plane), we
have:

\[
\langle n'\mu';L';k||H||n\mu;L;k|| \rangle = \\
= \frac{1}{N_{\parallel}} \sum_{j = 1}^{N_{l}} \sum_{j' = 1}^{N_{l}} \exp[i\,k||\cdot (R_{j'})_{\parallel}(0) + v||\cdot (\delta_{\mu,c} - \delta_{\mu',c}) + R_{\parallel}^{\text{OFFSET}}(\Delta(L) - \Delta(L'))] \times \\
\times \langle n'\mu';L';R_{j'}||\parallel(0) + \delta_{\mu',c}v|| + \Delta(L')R_{\parallel}^{\text{OFFSET}}||H||n\mu;L;R_{j'}||\parallel(0) + R_{j'}\parallel(0) + \delta_{\mu,c}v|| + \Delta(L)R_{\parallel}^{\text{OFFSET}} \rangle
\]

(3.57)

Since \( H \) still possesses translational symmetry in the plane, \( H = T(R_{j'}\parallel(0))H \, T^{\dagger}(R_{j'}\parallel(0)) \), and it follows that the matrix element in (3.57) is independent of \( j' \). Thus, everything to the right of the summation over \( j' \) may be pulled out of this sum, which then yields \( N_{\parallel} \), and we are left with a summation over the neighbors of the orbital \( |n'\mu';L';\delta_{\mu,c}v|| + \Delta(L')R_{\parallel}^{\text{OFFSET}}\rangle \):

\[
\langle n'\mu';L';k||H||n\mu;L;k|| \rangle = \sum_{j = 1}^{N_{l}} \exp[i\,k||\cdot (R_{j})_{\parallel}(0) + v||\cdot (\delta_{\mu,c} - \delta_{\mu',c}) + R_{\parallel}^{\text{OFFSET}}(\Delta(L) - \Delta(L'))] \times \\
\times \langle n'\mu';L';\delta_{\mu',c}v|| + \Delta(L')R_{\parallel}^{\text{OFFSET}}||H||n\mu;L;R_{j}\parallel(0) + \delta_{\mu,c}v|| + \Delta(L)R_{\parallel}^{\text{OFFSET}} \rangle
\]

(3.58)

Note that the real-space vector in parentheses in (3.58) is just the projection onto the \( x-y \) plane of the relative vector between the two orbitals. In the nearest-neighbor model, we truncate the sum at closest neighbors to the orbital \( |n'\mu';L';\delta_{\mu,c}v|| + \Delta(L')R_{\parallel}^{\text{OFFSET}}\rangle \). As in sec. 3.3, the matrix elements on the right-hand side of (3.58) are listed in Table 3.1.

We are now prepared to find the tight-binding Hamiltonian for the interface system. The electron wavefunction over the entire crystal is given by:

\[
|\psi_{k||} \rangle = \sum_{L} \sum_{n} \left( C_{L}^{pa} |na;L;k|| + C_{L}^{pc} |nc;L;k|| \right)
\]

(3.59)
where \( n \) once again indexes orbital symmetry types: \( \{s, x, y, z, s^*\} \). Define the transpose of a vector of coefficients by:

\[
[C^\mu L]^T = \begin{bmatrix} C_{L}^{s\mu} & C_{L}^{x\mu} & C_{L}^{y\mu} & C_{L}^{z\mu} & C_{L}^{s*\mu} \end{bmatrix}
\]

(3.60)

where \( \mu \) is either 'a' or 'c', and define \( C_L \) by:

\[
C_L \equiv \begin{bmatrix} C_L^a \\ C_L^c \end{bmatrix}
\]

(3.61)

Employing the linear variational method to find the (approximate) eigenstates of \( H \) in the basis (3.52), (3.53), we obtain the infinite matrix equation for the nearest-neighbor model:

\[
\begin{bmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots \\
H^{(a,c)}_{(L-1,L-1)} & H^{(a,c)}_{(L-1,L-1)} - 1E & H^{(a,c)}_{(L-1,L)} & 0 & \vdots \\
0 & H^{(a,c)}_{(L-1,L)} & H^{(a,c)}_{(L,L)} - 1E & H^{(a,c)}_{(L,L)} & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\end{bmatrix}
\begin{bmatrix}
C_{L-1}^a \\
C_{L-1}^c \\
C_L^a \\
C_L^c \\
\vdots \\
\end{bmatrix}
= \begin{bmatrix}
\vdots \\
0 \\
0 \\
0 \\
\vdots \\
\end{bmatrix}
\]

(3.62)

Note that we have not yet applied boundary conditions along \( z \) (i.e., in \( L \)) to (3.62); we shall take up that issue later. In (3.62), the elements of the (5x5, in the \( sp^3 s^* \) model) matrix \( H_{(L',L)}^{(\mu',\mu)} \), which is a function of \( k \parallel \), are given by:

\[
[H^{(\mu',\mu)}_{(L',L)}]_{(\alpha',\alpha)} = \langle n'\mu';L';k\parallel|H|n\mu;L;k\parallel\rangle
\]

(3.63)

The matrices \( H_{(L',L)}^{(\mu',\mu)} \) in the nearest-neighbor \( sp^3 s^* \) model are given in Table 3.4.

The pairwise three-term recurrence relations in (3.62) may be rearranged in a more useful form:
\[
\begin{bmatrix}
[H^{(a,a)}_{(L,L)}] & [H^{(a,c)}_{(L,L)}] \\
[H^{(c,a)}_{(L-1,L)}] & 0
\end{bmatrix}
\begin{bmatrix}
C_L^a \\
C_L^c
\end{bmatrix} = 
\begin{bmatrix}
0 & [-H^{(c,a)}_{(L-1,L)}]^+ \\
[-H^{(a,c)}_{(L-1,L)}]^+ & [-H^{(c,c)}_{(L-1,L)} - 1E]^-
\end{bmatrix}
\begin{bmatrix}
C_{L-1}^a \\
C_{L-1}^c
\end{bmatrix}.
\]

(3.64)

When the matrix on the left-hand side of (3.64) can be inverted (we shall briefly defer the discussion of the significance of cases in which its inverse does not exist), the (single-monolayer) transfer matrix (also a function of \(k_{||}\)) [11] is given by:

\[
[T_{L,L-1}]_{(1,1)} = -[H^{(c,a)}_{(L-1,L)}]^{-1} [H^{(a,c)}_{(L-1,L-1)}]^+ \\
[T_{L,L-1}]_{(1,2)} = -[H^{(c,a)}_{(L-1,L)}]^+ [H^{(c,c)}_{(L-1,L-1)} - 1E] \\
[T_{L,L-1}]_{(2,1)} = [H^{(a,a)}_{(L,L)}]^{-1} [H^{(a,a)}_{(L-1,L)} - 1E] [H^{(c,a)}_{(L-1,L-1)}]^{-1} [H^{(a,c)}_{(L-1,L-1)}]^+ \\
[T_{L,L-1}]_{(2,2)} = -[H^{(a,c)}_{(L,L)}]^{-1} [H^{(c,a)}_{(L-1,L)}]^+ + [H^{(a,a)}_{(L,L)}]^{-1} [H^{(a,a)}_{(L-1,L)} - 1E] [H^{(c,a)}_{(L-1,L-1)}]^{-1} [H^{(c,c)}_{(L-1,L-1)} - 1E]
\]

(3.65a-d)

where \([T_{L,L-1}]_{(i,j)}\) denotes the \((i,j)\) \((5x5)\) block of the \((10x10)\) matrix \(T_{L,L-1}\). The matrix \(T_{L,L-1}\) is called the transfer matrix because of its action on the vector of coefficients, \(C_L\):

\[
C_L = T_{L,L-1} C_{L-1}
\]

(3.66)

Thus, the knowledge of the wavefunction in the layer \(L-1\) allows us to use the transfer matrix to obtain the wavefunction in the layer \(L\). Note that because we employ a nearest-neighbor model (3.66) is simpler than the more general relations given by Lee and Joannopoulos [11].
Let us now consider the cases in which $T_{LL'}$ fails to exist; that is, the matrices in (3.64) cannot be inverted. Examination of the matrices in Table 3.4 reveals that this will be the case in those regions of the Brillouin zone in which the bands are perfectly flat, such as from X to W. In a perfect crystal with bands which are independent of $k_z$ for a given $k_\parallel$, there are therefore an infinite number of Bloch states at the given $(E, k_\parallel)$. Now, in the basis (3.52), (3.53), there can be at most 10 (in the sp$^3$s* model) linearly independent states confined to a layer L. Since the Bloch states are merely linear combinations of the states (3.47), (3.48), it is obvious that when the bands are flat, there will be an infinite number of states in each layer, L, for a given $(E, k_\parallel)$. Hence, knowledge of the electron wavefunction in the layer $L-1$ is not sufficient to determine it in the layer L, so the transfer matrix will not exist. Physically, we see that perfectly flat bands imply an infinite effective mass, so it should be no surprise that the transfer matrix fails to exist.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3.2}
\caption{GaAs oriented in the [001] direction, illustrating our definition of a layer.}
\end{figure}
TABLE 3.4: BLOCKS OF THE INTERFACE TIGHT-BINDING HAMILTONIAN

\[
H_{(L,L)}^{(a,c)} = \begin{bmatrix}
  v(s,s) f_0 & v(sa,pc) f_2 & v(sa,pc) f_2 & v(sa,pc) f_0 & 0 \\
  -v(pa,sc) f_2 & v(x,x) f_0 & v(x,y) f_0 & v(x,y) f_2 & -v(pa,s^*c) f_2 \\
  -v(pa,sc) f_2 & v(x,y) f_0 & v(x,x) f_0 & v(x,y) f_2 & -v(pa,s^*c) f_2 \\
  -v(pa,sc) f_0 & v(x,y) f_2 & v(x,y) f_2 & v(x,x) f_0 & -v(pa,s^*c) f_0 \\
  0 & v(s^*a,pc) f_2 & v(s^*a,pc) f_2 & v(s^*a,pc) f_0 & 0
\end{bmatrix}
\]

\[
H_{(L-1,L)}^{(c,a)} = \begin{bmatrix}
  v(s,s) f_1 & v(pa,sc) f_3 & -v(pa,sc) f_3 & v(pa,sc) f_1 & 0 \\
  -v(sa,pc) f_3 & v(x,x) f_1 & -v(x,y) f_1 & v(x,y) f_3 & -v(s^*a,pc) f_3 \\
  v(sa,pc) f_3 & v(x,x) f_1 & -v(x,y) f_1 & v(x,y) f_3 & v(s^*a,pc) f_3 \\
  -v(sa,pc) f_1 & v(x,y) f_3 & -v(x,y) f_3 & v(x,x) f_1 & -v(s^*a,pc) f_1 \\
  0 & v(pa,s^*c) f_3 & -v(pa,s^*c) f_3 & v(pa,s^*c) f_1 & 0
\end{bmatrix}
\]

\[
H_{(L,L)}^{(a,a)} = \begin{bmatrix}
  E(sa) & 0 & 0 & 0 & 0 \\
  0 & E(pa) & 0 & 0 & 0 \\
  0 & 0 & E(pa) & 0 & 0 \\
  0 & 0 & 0 & E(pa) & 0 \\
  0 & 0 & 0 & 0 & E(s^*a)
\end{bmatrix}
\]

\[
H_{(L-1,L-1)}^{(c,c)} = \begin{bmatrix}
  E(sc) & 0 & 0 & 0 & 0 \\
  0 & E(pc) & 0 & 0 & 0 \\
  0 & 0 & E(pc) & 0 & 0 \\
  0 & 0 & 0 & E(pc) & 0 \\
  0 & 0 & 0 & 0 & E(s^*c)
\end{bmatrix}
\]

**NOTATION:**

\(E(n\mu) = E_{n\mu,n\mu}(0,0,0)\)

\(v(n'\mu',n\mu) = 2 E_{n'\mu',n\mu}(1/2,1/2,1/2)\) [we omit the (\(\mu',\mu\)) for \(v(s,s), v(x,x),\) and \(v(x,y)\)].

\(f_0(k||) = \cos[(k_x + k_y)a/4]\)

\(f_1(k||) = \cos[(k_x - k_y)a/4]\)

\(f_2(k||) = i \sin[(k_x + k_y)a/4]\)

\(f_3(k||) = i \sin[(k_x - k_y)a/4]\)

The parameters refer to matrix elements taken between orbitals centered in the layers indicated. Thus, the parameters appearing in \(H_{(L-1,L)}^{(c,a)}\) apply to matrix elements between cation orbitals centered in layer \((L-1)\) and anion orbitals centered in layer \(L\).
3.4.2 Eigenvalues, Eigenvectors, and Properties of the Single-Monolayer Transfer Matrix in the Nearest-Neighbor Model

Because the treatment given in sec. 3.4.1 is quite general, we may apply it to bulk crystals as well. For the time being, let us assume that we are working with an infinite perfect crystal. (We have not applied boundary conditions in z.) In this case, all the single-monolayer transfer matrices are identical:

$$ T_{L,L-1} = T_{L+1,L} = T, \quad \forall \ L $$

(3.67)

Suppose that a vector $C_L$ is an eigenstate of $T$:

$$ T \cdot C_L = \lambda \cdot C_L. $$

(3.68)

But $T \cdot C_L = C_{L+1}$ also, so $C_{L+1} = \lambda \cdot C_L$. Then, (3.59) becomes (we now index the state by the transfer matrix eigenvalue as well):

$$ |\psi_{k||\lambda}\rangle = \sum_L \lambda^L \left\{ \sum_n \left[ C_0^{na} |na;L; k||\rangle + C_0^{nc} |nc;L; k||\rangle \right] \right\} $$

(3.69)

If $|\lambda| > 1$, then this state grows as $L \to \infty$ (i.e., as $z \to \infty$). Similarly, if $|\lambda| < 1$, it grows as $L \to -\infty$ (i.e., as $z \to -\infty$). (Note that these states satisfy Bloch's theorem with a complex wavevector, $k_\perp$.) In general, the electron wavefunction can be a superposition of the states (3.69):

$$ |\psi_{k||}\rangle = \sum_\lambda a_\lambda |\psi_{k||,\lambda}\rangle $$

(3.70)

Now, application of any reasonable boundary conditions in z (such as the cyclic boundary conditions) to the perfect crystal will require that $a_\lambda = 0$ for all $|\lambda| \neq 1$, so that the wavefunction does not blow up. Thus, uniform infinite crystals have eigenstates which grow and decay, but they are not allowed states of a perfect finite crystal.
Since these growing/decaying uniform crystal eigenstates evidently play no role in the description of the bulk states, we might wonder when, if ever, they are important. The answer is, of course, in the presence of defects, such as interfaces. Consider the case of a single interface occurring at layer 0. For \( L < 0 \), \( T_{L,L-1} = T_- \) and for \( L > 0 \), \( T_{L,L-1} = T_+ \). (We shall leave the interface transfer matrix, \( T_{0,1} \), unspecified since its precise value is unimportant for our present purposes; we will take up this issue in detail in sec. 3.5.) In particular, note that on the left, the sum over \( L \) in (3.59) or (3.69) runs from \(-\infty\) to \(-1\), while on the right it runs from \(+1\) to \(+\infty\). Therefore those \( T_- \) eigenstates which decay as \( L \to -\infty \) are allowed on the left, and those \( T_+ \) eigenstates which decay as \( L \to +\infty \) are allowed on the right. Therefore the expansion (3.70) for \( L < 0 \) will include those \( T_- \) eigenstates with \(|\lambda_-| \geq 1\), while for \( L > 0 \) it will include those \( T_+ \) eigenstates with \(|\lambda_+| \leq 1\).

where \( \lambda_{\pm} \) is an eigenvalue of \( T_\pm \).

Having discussed the growing/decaying (evanescent) eigenstates of \( T \) in some detail, let us examine the relationship of the \( T \) eigenstates with \(|\lambda| = 1\) to the Bloch states, since once we apply perfect crystal boundary conditions, these \( T \) eigenstates are the only allowed expansion states. Assume that there are \( N_z \) layers in the perfect crystal, and scale the right-hand side of (3.69) accordingly:

\[
|\psi_{k||,\lambda}\rangle = \frac{1}{\sqrt{N_z}} \sum_L \lambda^L \left\{ \sum_n \left[ C_0^{na} |na;L;k||\rangle + C_0^{nc} |nc;L;k||\rangle \right] \right\}
\]

(3.71)

Substitute (3.52) and (3.53) into (3.71), and note that since \(|\lambda| = 1\), we have \( \lambda = e^{i\beta}, \beta \in \text{Re} \) (we now index the state by \( \beta \) rather than by \( \lambda \)):

\[
|\psi_{k||,\beta}\rangle = \frac{1}{\sqrt{N_z N||}} \sum_d \sum_{L=1}^{N_0} \exp[i \left( k|| \cdot R_{j||}(L) + \beta L \right)] \times \left\{ C_0^{na} |na;L;R_{j||}(L)| + C_0^{nc} |nc;L;R_{j||}(L)+v||\rangle \right\}
\]

(3.72)
Now \( N_z N_\parallel = N \), the total number of Bravais lattice sites in the crystal. As discussed in sec. 3.4.1, anions are located at positions \( \mathbf{R}_j = \mathbf{R}_j \parallel (L) + L(a/2)z \) while cations are located at positions \( \mathbf{R}_j + \mathbf{v} \). We may therefore convert the sums over \( L \) and \( j \) into a single sum over all Bravais lattice sites in the crystal, and index the Lowdin orbitals accordingly. Also, define a real quantity \( k_z = 2\beta/a \), so that the argument of the exponential becomes \( \mathbf{k} \cdot \mathbf{R}_j \), where \( \mathbf{k} = (k_\parallel, k_z) \), and index the state by \( \mathbf{k} \):

\[
\psi_\mathbf{k} = \frac{1}{\sqrt{N}} \sum_n \left\{ C_0^a \sum_{j=1}^N \exp[i \mathbf{k} \cdot \mathbf{R}_j] |na; \mathbf{R}_j\rangle + C_0^a \exp[-i k_z \mathbf{a}/4] \sum_{j=1}^N \exp[i \mathbf{k} \cdot (\mathbf{R}_j + \mathbf{v})] |nc; \mathbf{R}_j + \mathbf{v}\rangle \right\}
\]

(3.73)

where we have used \( k_z a + \mathbf{v} = k_z a/4 \). Recall that the Bloch state of wavevector \( \mathbf{k} = (k_\parallel, k_z) \) is merely a superposition of the Bloch sums (3.47), (3.48):

\[
\psi_\mathbf{k}^{BLOCH} = \frac{1}{\sqrt{N}} \sum_n \left\{ B_0^a \sum_{j=1}^N \exp[i \mathbf{k} \cdot \mathbf{R}_j] |na; \mathbf{R}_j\rangle + B_0^a \exp[i \mathbf{k} \cdot (\mathbf{R}_j + \mathbf{v})] |nc; \mathbf{R}_j + \mathbf{v}\rangle \right\}
\]

(3.74)

where the \( B_0^{n\mu} \) are the expansion coefficients which make (3.74) an eigenstate of the Hamiltonian. Consider the case in which the states (3.73) and (3.74) have the same energy. (They are both eigenstates of \( H \), and while they are in different bases, these basis functions are both linear combinations of the same Lowdin orbitals.) Then a comparison of (3.73) and (3.74) reveals two important facts. First, the transfer matrix eigenstate with eigenvalue \( \exp[i k_z a/2] \) [for given \((E, k_\parallel)\)], \( k_z \) Real, just corresponds to a Bloch state of wavevector \( \mathbf{k} = (k_\parallel, k_z) \). Second, the expansion coefficients of the transfer matrix eigenstates are related to the expansion coefficients of the corresponding Bloch state via:
\[ C_0^{na} = B^{na}; \quad C_0^{nc} = \exp\left[\frac{i}{4} k_z \hat{a}\right] B^{nc}. \] (3.75)

From the foregoing discussion, it is apparent that the knowledge of the transfer matrix, \( T(E, k_\parallel) \) (when it exists), is sufficient to describe the bulk crystal. It should therefore have certain special properties. For example, we might expect that for each growing eigenstate of \( T(E, k_\parallel) \) there must be a decaying one. "Time-reversal" symmetry and the symmetries of zincblende lead us to expect that there should be one reverse-propagating state (i.e., a state with \( \partial E / \partial k_z < 0 \)) for each forward-propagating state (i.e., a state with \( \partial E / \partial k_z > 0 \)) [13]. We will now explicitly display these properties.

We show that there is one decaying state for each growing state by proving that transfer matrix eigenvalues come in pairs \( (\lambda, 1/\lambda^*) \). (Note that this says nothing new about eigenvalues with unity magnitude.) Lee and Joannopoulos [11] demonstrate this for transfer matrices, in general. For the nearest-neighbor \( s^3s^* \) model, the explicit relation is:

\[ T_{L,L-1}^{-1} = QT_{L,L-1}^\dagger Q^{-1} \] (3.76)

where

\[ Q = \begin{bmatrix}
Q & \left[\left(H_{L-1,L-1}^{(a,c)}\right)^\dagger\right]^{-1} \\
\left[H_{L,L}^{(a,c)}\right]^{-1} & Q
\end{bmatrix} \] (3.77)

The conclusion that \( \lambda \in \sigma(T) \Rightarrow 1/\lambda^* \in \sigma(T) \), where \( \sigma(T) \) denotes the spectrum of the matrix \( T \), follows from the fact that, if \( \sigma(T) = \{\lambda\} \), \( \Rightarrow \sigma(T^{-1}) = \{\lambda^{-1}\} \) and \( \sigma(T^\dagger) = \{\lambda^*\} \), and (3.76) implies that \( \sigma(T^{-1}) = \sigma(T^\dagger) \).

We can also show that for every forward-propagating state, there is a reverse-propagating state. We do this by proving the more powerful relation \( T \sim T^* \). (Note that this has implications for both the growing/decaying eigenstates of \( T \) as well as the
propagating eigenstates.) This means that the spectra of $T$ and $T^*$, \{\lambda\} and \{\lambda^*\}, respectively, coincide, so the eigenvalues of $T$ must come in pairs ($\lambda, \lambda^*$). For the case of eigenstates with eigenvalues of unity magnitude, the conclusion about the direction of propagation is obvious. In (3.64), call the matrix on the left $M_1$ and call the matrix on the right $M_2$, so that $T = M_1^{-1} M_2$. Van der Waag, Boykin, and Harris [14] give the proof for the nearest-neighbor $sp^3s^*$ model; we briefly sketch their results. From the matrices in Table 3.4, it may be seen that:

$$SM_1S^{-1} = R_1$$  \hspace{1cm} (3.78)
$$SM_2S^{-1} = R_2$$  \hspace{1cm} (3.79)

where $R_1$ and $R_2$ are real (10x10) matrices and

$$S = \begin{bmatrix} S' & 0 \\ 0 & S' \end{bmatrix}$$  \hspace{1cm} (3.80)
$$S' = \text{diag}[ e^{i\pi/4}, e^{-i\pi/4}, e^{i\pi/4}, e^{-i\pi/4}, e^{i\pi/4}, e^{-i\pi/4} ]$$  \hspace{1cm} (3.81)

where \text{diag}[d_{11}, d_{22}, \ldots, d_{nn}] denotes a diagonal matrix with indicated (diagonal) elements. Applying (3.80) as a similarity transformation to $T$ yields:

$$STS^{-1} = [SM_1^{-1}S^{-1}] [SM_2S^{-1}] = R_1^{-1}R_2 = R$$  \hspace{1cm} (3.82)

with $R$ once again a real matrix. Taking the complex conjugate of (3.82) shows that $T^* \sim R$, and, since $T \sim R$ as well, $T \sim T^*$, which completes the proof.

3.4.3 The Usual Transfer Matrix Approach: An Example Calculation of a Tunneling Problem
Having derived the single-monolayer transfer matrix and several of its important properties, we are now in a position to demonstrate how it is used to calculate the wavefunction for an interface system. As in the previous subsection, we shall leave the details of all transfer matrices for non-bulk layers (i.e., interface layers and those layers across which there is a potential drop) to sec. 3.5, since they are not important for our immediate purposes. Instead, we shall concentrate on boundary conditions and the difficulties which arise when one attempts a straightforward transfer matrix calculation. We will demonstrate the straightforward transfer matrix method with a simple tunneling calculation.

The situation we consider is illustrated in Fig. 3.3. Layers with index \( L \leq 0 \) are referred to as "bulk emitter" layers; for these layers, \( T_{L,L-1} = T_{e} \). Layers with index \( L \geq (N+1) \) are called "bulk collector" layers; for these layers, \( T_{L,L-1} = T_{c} \). In the region \( 1 \leq L \leq N \), the layers vary, due to changes in material composition and/or vacuum level, and the various transfer matrices are, in general, different. The straightforward solution to the problem proceeds directly from (3.66):

\[
C_{N+1} = \prod_{L=1}^{N+1} T_{L,L-1} C_0
\]  

(3.83)

In (3.83), the multiplication is performed on the left, the rightmost factor being \( T_{1,0} \). In order to find the wavefunction, we must apply boundary conditions to (3.83); this is most easily done by expanding \( C_0 \) and \( C_{N+1} \) in terms of the eigenstates of \( T_{e} \) and \( T_{c} \), respectively. For this tunneling problem, we restrict our treatment to cases in which there is a single incident state (i.e., a state with \( \partial E/\partial k_x > 0 \)). This will be an eigenstate of \( T_{e} \) with eigenvalue of unity magnitude; call it \( \nu_{e,1} \). From the discussion in sec. 3.4.2, we know that under these circumstances, there will be only one reverse-propagating eigenstate of \( T_{e} \). Call this state \( \nu_{e,6} \). The remaining eight eigenstates of \( T_{e} \) [15] appear as pairs \((\lambda, 1/\lambda^*)\), so that for every eigenstate with \(|\lambda| > 1\), there is a one with \(|\lambda| < 1\). [The
eigenstates can also come in pairs \((\lambda, \lambda^*)\), for \(\lambda\) complex, but this information is not useful in determining the allowed non-propagating states.] The \(\mathcal{T}_e\) eigenstates with \(|\lambda| > 1\) will decay as the layer index approaches minus infinity, so they comprise the remaining allowed expansion states in the emitter region. Call these states \(v_{e,j}; j = 7, 8, 9, 10\). Therefore \(C_0\) is given by the expansion:

\[
C_0 = v_{e,1} + \sum_{j=6}^{10} b_j v_{e,j}
\]

(3.84)

where we assume unity incident current and the \(b_j\) are expansion coefficients, which remain to be determined. On the collector side, we make a similar expansion. From sec. 3.4.2, we know that five of the ten eigenstates of \(\mathcal{T}_c\) will either decay as the layer index approaches positive infinity, or have \(\partial E/\partial k_z > 0\). (It is not important to differentiate the two cases at this time.) Call these states \(v_{c,j}; j = 1, 2, 3, 4, 5\):

\[
C_{N+1} = \sum_{j=1}^{5} b_j v_{c,j}
\]

(3.85)

Substituting (3.84) and (3.85) into (3.83) and rearranging, we obtain the following linear system, the solution of which is the \(b_j\):

\[
\begin{bmatrix}
v_{c,1} & v_{c,2} & \cdots & v_{c,5} & -\mathcal{T} v_{c,6} & -\mathcal{T} v_{c,7} & \cdots & -\mathcal{T} v_{c,10}
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_{10}
\end{bmatrix}
= \mathcal{T} v_{c,1}
\]

(3.86)

where the vertical bar ‘|’ denotes a column of the \((10\times10)\) matrix and

\[
b = \begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_{10}
\end{bmatrix}; \quad \mathcal{T} = \prod_{L=1}^{N+1} \mathcal{T}_{L,L-1}
\]

(3.87)

The solution of (3.86), \(b\), is used to find the transmission and reflection coefficients, and
with the aid of (3.84), (3.85), and (3.66), it may also be used to find the wavefunction in any layer.

Unfortunately, transfer matrix calculations for realistic structures are not as simple as solving equation (3.86), which in practice turns out to be frighteningly numerically unstable. The problem with these types of transfer matrix calculations is the repeated multiplication of transfer matrices, as in (3.83) or (3.87) [16]. Recall from sec. 3.4.2 that each transfer matrix will, in general, have some eigenvalues with magnitude greater than unity. (Physically, these are interface states, and it is quite possible to have both evanescent and Bloch states for a given $E, k_{\parallel}$.) If we imagine the wavefunction in each layer, $L$, decomposed in terms of the eigenstates of $T_{L+1,L}$, it becomes apparent that the eigenstates having eigenvalues with the largest magnitudes will, with repeated multiplication by transfer matrices, come to dominate the expansion. (This reasoning also holds even when $T_{L+1,L} \neq T_{L,L-1}$, provided that they do not differ radically. In these cases, we expect there to be a large overlap between eigenstates of $T_{L+1,L}$ and $T_{L,L-1}$ having eigenvalues with "large" magnitudes.) Thus, transfer matrix calculations tend to be numerically unstable because they provide a very complete description of the solid which includes the (numerically) problematic evanescent states. In fact, when using double precision in the $sp^3s^*$ nearest-neighbor model under the best circumstances (i.e., most favorable $E, k_{\parallel}$), $T$ can be a product of only slightly more than 30 single-monolayer transfer matrices before (3.86) becomes unusable. The effects of space-charge regions could thus not be included in early work on resonant-tunneling diodes (RTDs) [17]. Schulman and Chang did develop a method for dealing with the exponential blowup problem in connection with their Reduced Hamiltonian Method [16], but it was only useful with a simple bulk expansion – no transfer matrices were involved [18]. (Also, the Reduced Hamiltonian Method cannot be used with RTDs or other aperiodic structures having unbound states.)
Layer Index | 0 | 1 | 2 | N−1 | N | N+1

... | | | | | | ...

Non-bulk region

**Figure 3.3:** Layers for the tunneling problem discussed in sec. 3.4.3. The bulk emitter region begins with layer 0, and the bulk collector region begins with layer N+1. The electron is incident from the bulk emitter region.

### 3.4.4 Removal of the Numerical Instability from and Boundary Conditions for the Interface Tight-Binding Hamiltonian

In spite of the difficulties posed by the transfer matrix method, it continued to be used to study semiconductor heterostructures, albeit thin ones! Because of the numerical problems just discussed, the method could only treat unbiased, undoped superlattices and quantum wells (where Schulman and Chang’s method could be used) or very short general structures such as RTDs and biased and/or doped quantum wells and superlattices. Since the accumulation region in front of the emitter barrier of an RTD can be about 0.1 eV deep even under moderate bias (thus leading to a lower effective barrier height), and since this effect cannot be included in a straightforward transfer matrix calculation (due to the length over which one must transfer), current density calculations for RTDs were particularly poor—worse than those calculated with envelope-function models including space-charge...
regions. This situation changed for the better, however, in 1991, when two groups independently came upon different solutions to the numerical instability. Ting, Yu, and McGill [19], following a suggestion by W. R. Frensley, used the method of Lent and Kirkner [20] to express the tunneling problem as a sparse linear system. Boykin, van der Wagt, and Harris (hereafter BWH) [21], on the other hand, modified the transfer matrix method to make it numerically stable. We will confine ourselves to the second of the two methods, discussing first the original form of BWH [21], then a modification to this due to Schulman and Ting [22].

The BWH method, and the original BWH form of the resulting linear system, are derived as below, the method itself being rooted in this principle: *in order to stabilize the transfer matrix method, it is absolutely essential to limit the number of single-monolayer transfer matrix factors in (3.87)* [21]. Suppose that $N_S$ layers may be transferred across without a loss of precision [i.e., equation (3.86) does not blow up]. In this case, rewrite the product of transfer matrices, $\mathcal{T}$, as a product of transfer matrices $\mathcal{T}_j$ each of which is a product of up to $N_S$ matrices $\mathcal{T}_{L_iL_{i-1}}$:

$$\mathcal{T} = \prod_{j=1}^{N_B} \mathcal{T}_j$$  \hspace{1cm} (3.88)

$$\mathcal{T}_j = \prod_{i=1}^{m_j} \mathcal{T}_{L_i(j),L_i(j-1)}, \quad m_j \leq N_S$$  \hspace{1cm} (3.89)

$$\sum_{j=1}^{N_B} m_j = N$$  \hspace{1cm} (3.90)

$$L_i(j) = i + \sum_{k=1}^{j-1} m_k$$  \hspace{1cm} (3.91)

In (3.90), $N$ is the total number of layers and in (3.91), the sum is taken to be zero for $j = 1$. Substituting (3.84), (3.85), and (3.88) into (3.83) and rearranging gives:
\begin{equation}
T_1 \left[ v_{e,1} + \sum_{j=6}^{10} b_j v_{e,j} \right] = \prod_{k=N_B}^{2} T_k^{-1} \left[ \sum_{j=1}^{5} b_j v_{c,j} \right]
\end{equation}

where the multiplication is again on the left, the rightmost matrix being $T^{-1}_{NB}$. Now, expand the right-hand side of (3.92) in terms of the standard basis \{ e_i \}, where the column vector $e_i$ has a 1 in the $i$-th row and zeros elsewhere:

\begin{equation}
T_2^{-1} T_3^{-1} \cdots T_{NB}^{-1} \left[ \sum_{j=1}^{5} b_j v_{c,j} \right] = \sum_{j=1}^{10} c_j^{(2)} e_j
\end{equation}

where the $c_j^{(i)}$ are new expansion coefficients. Previously, there were 10 equations to solve, but now there are 20:

\begin{equation}
T_1 \left[ v_{e,1} + \sum_{j=6}^{10} b_j v_{e,j} \right] = \sum_{j=1}^{10} c_j^{(2)} e_j
\end{equation}

\begin{equation}
T_2 \left[ \sum_{j=1}^{10} c_j^{(2)} e_j \right] = T_3^{-1} T_4^{-1} \cdots T_{NB}^{-1} \left[ \sum_{j=1}^{5} b_j v_{c,j} \right]
\end{equation}

Repeating the above procedure until $T_{NB}$ appears on the left-hand side results in a sparse linear system of $10N_B$ equations, which, arranged in matrix form read:

\begin{equation}
\begin{bmatrix}
\Delta_{1,1} & I & 0 & \cdots & 0 \\
0 & -T_2 & I & \cdots & 0 \\
\vdots & 0 & -T_3 & \ddots & 0 \\
0 & \vdots & \ddots & \ddots & I \\
\Delta_{NB,1} & 0 & \cdots & 0 & -T_{NB}
\end{bmatrix}
\begin{bmatrix}
b \\
c^{(2)} \\
c^{(3)} \\
\vdots \\
c^{(NB)}
\end{bmatrix}
= 
\begin{bmatrix}
T_1 v_{e,1} \\
0 \\
\vdots \\
0
\end{bmatrix}
\end{equation}

where $b$ is defined in (3.87) and the $c_j^{(i)}$ and the $\Delta$'s are given by:
\begin{equation}
    c^{(i)} = \begin{bmatrix}
        c_1^{(i)} \\
        c_2^{(i)} \\
        \vdots \\
        c_{10}^{(i)}
    \end{bmatrix}
\end{equation}

\begin{equation}
    \Delta_{1,1} = \begin{bmatrix}
        0 & 0 & 0 & 0 & -I_1 v_{e,6} & -I_1 v_{e,7} & \cdots & -I_1 v_{e,10}
    \end{bmatrix},
\end{equation}

\begin{equation}
    \Delta_{NB,1} = \begin{bmatrix}
        v_{e,1} & v_{e,2} & \cdots & v_{e,5} & 0 & 0 & 0 & 0
    \end{bmatrix}.
\end{equation}

Note that (3.96) must not be solved by block Gaussian-elimination methods, for they will reintroduce the instabilities. We have found the sparse Gaussian-elimination routine NSPIV [23] to be quite useful for solving (3.96).

Schulman and Ting [22] note that the derivation of the BWH method can be stated more simply as:

\begin{equation}
    T_1 \left[ v_{e,1} + \sum_{j=6}^{10} b_j v_{e,j} \right] = c^{(2)}
\end{equation}

\begin{equation}
    T_j c^{(i)} = c^{(i+1)}, \quad 2 \leq j \leq (NB-1)
\end{equation}

\begin{equation}
    T_{NB} c^{(NB)} = \sum_{j=1}^{5} b_j v_{e,j},
\end{equation}

while the linear system (3.96) may alternatively be written as a banded system:

\begin{equation}
    \begin{bmatrix}
        \Delta_1 & L & 0 & \cdots & 0 \\
        0 & -T_2 & \ddots & \vdots & \vdots \\
        \vdots & \ddots & \ddots & L & 0 \\
        0 & \cdots & 0 & -T_{NB} & \Delta_{NB}
    \end{bmatrix}
    \begin{bmatrix}
        b^{(2)} \\
        c^{(2)} \\
        \vdots \\
        c^{(NB)} \\
        b^{(1)}
    \end{bmatrix}
    = \begin{bmatrix}
        T_1 v_{e,1} \\
        0 \\
        \vdots \\
        0 \\
        0
    \end{bmatrix}.
\end{equation}

Where:
\[ b^{(1)} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_6 \\ b_7 \\ \vdots \\ b_{10} \end{bmatrix}; \quad b^{(2)} = \begin{bmatrix} \end{bmatrix} \]

\[ \Delta_1 = \begin{bmatrix} -L_1 v_{e,6} & -L_1 v_{e,7} & \cdots & -L_1 v_{e,10} \end{bmatrix} \]  

\[ \Delta_{NB} = \begin{bmatrix} v_{c,1} & v_{c,2} & \cdots & v_{c,5} \end{bmatrix} \]  

(3.104)  

(3.105)  

(3.106)  

In the above equations, note that boldface symbols (b) represent 5-element column vectors, bold-italic symbols (v) represent 10-element column vectors, underlined-italic symbols (L) represent 10x10 matrices, and bold-underlined symbols (Δ) represent 10x5 matrices. 

Finally, we return to the question of boundary conditions for the Hamiltonian matrix (3.62). Suppose that for \( L \leq 0 \) and \( L \geq N+1 \) the system is bulk-like. Following van der Wagt, Boykin, and Harris [14], we apply boundary conditions (3.84) and (3.85) to (3.62) to obtain the finite matrix equation:

\[
\begin{bmatrix}
[H_{(0,0)}^{(c,c)} - 1E] & H_{(0,1)}^{(c,a)} & 0 & \cdots \\
[H_{(0,1)}^{(c,a)}]^{+} & [H_{(1,1)}^{(a,a)} - 1E] & H_{(1,1)}^{(a,c)} & \cdots \\
0 & \ddots & \ddots & \ddots \\
\vdots & 0 & [H_{(NB,NB+1)}^{(c,a)}]^{+} & [H_{(NB+1,NB+1)}^{(a,a)} - 1E]
\end{bmatrix}
\times
\]
\[
\begin{bmatrix}
 v^{(2)}_{0,1} + \sum_{i=6}^{10} b_i v^{(2)}_{0,i} \\
 c^{(1)}_1 \\
 c^{(2)}_1 \\
 \vdots \\
 c^{(1)}_{NB} \\
 c^{(2)}_{NB} \\
 \sum_{i=1}^{5} b_i v^{(1)}_{NB+1,i}
\end{bmatrix}
\times
\begin{bmatrix}
 -H^{(a,c)}_{(0,0)} \\
 v^{(1)}_{0,1} + \sum_{i=6}^{10} b_i v^{(1)}_{0,i} \\
 0 \\
 \vdots \\
 -\sum_{i=1}^{5} b_i H^{(a,c)}_{(NB+1,NB+1)} v^{(2)}_{NB+1,i}
\end{bmatrix}
\]

(3.107)

where:

\[
v_{L,i} = \begin{bmatrix} v^{(1)}_{L,i} \\ v^{(2)}_{L,i} \end{bmatrix}
\]

(3.108)

and L is the layer index. Equation (3.107) bears a striking resemblance to eq. (3.103) when all of the transfer matrices in (3.103) are single-monolayer transfer matrices. This is not surprising, since the two equations are merely different ways of solving the same problem. Indeed, by rearranging the first and last block rows of (3.107), and by performing block operations on the system (e.g., multiply the first block row by \([H^{(c,a)}_{(0,1)}]^{-1}\)), equation (3.107) may be transformed into (3.103) or (3.96).

We have therefore seen the development of the tight-binding model for interface systems, from the determination of the matrix elements up to problems encountered in its actual use in practice. Other than minor matters, such as the derivation of transmission and reflection coefficients, we are ready to perform actual calculations (and numerically well-behaved ones at that!). In the next section, we shall examine in detail predictions of the model and their physical significance.
3.5 RESONANT TUNNELING DIODES

In this section, we first fill in some details left out of the discussion of the tunneling problem in sec. 3.4. Then, we examine predictions of the tight-binding model for GaAs/AlAs RTDs and discuss their physical significance.

3.5.1 Specification of the Non-Bulk Transfer Matrices

In our treatment of the tunneling problem, there are two basic types of non-bulk transfer matrices: those which transfer across an interface and those which transfer across a region over which there is a potential drop. Consider first the case of interfaces. In fig. 3.1, suppose that the cation plane in the layer $L$ is aluminum instead of gallium. An interface between GaAs and AlAs thus occurs between layers $L-1$ and $L$: we assume that there is no potential drop across any layer. On either side of the interface, we take:

$$T_{M,M-1} = T_{\text{GaAs}}, \quad M \leq (L-1)$$

$$T_{M,M-1} = T_{\text{AlAs}}, \quad M \geq (L+1)$$

(3.109)  (3.110)

where $T_{\text{GaAs}}$ and $T_{\text{AlAs}}$ are the bulk transfer matrices for GaAs and AlAs, respectively. For the transfer matrix $T_{L,L-1}$ we take:

$$H_{L,L}^{(a,a)} = H_{\text{AlGaAs}}^{(a,a)}; \quad H_{L,L}^{(a,c)} = H_{\text{AlAs}}^{(a,c)}; \quad H_{L-1,L}^{(c,a)} = H_{\text{GaAs}}^{(c,a)}$$

$$H_{L-1,L-1}^{(a,c)} = H_{\text{GaAs}}^{(a,c)}; \quad H_{L-1,L-1}^{(c,c)} = H_{\text{GaAs}}^{(c,c)}$$

(3.111)

where the subscript indicates the parameters to be used. The parameters for the 'AlGaAs'-subscripted matrix are taken to be the average of the GaAs and AlAs parameters. Other choices are, of course, possible – one could, for example, use averaged parameters in the
matrices $H_{(L-1,L)}^{(c,a)}$ and $H_{(L,L)}^{(a,c)}$. (Our initial work [21] in fact used this choice. We have since adopted the present scheme because it is simpler and more in the spirit of the nearest-neighbor only approximation.) Note that because the interface parameters are assumed, no choice is without problems.

Potential drops across a layer are treated in a similar manner. The applied potential is a function of $z$ only, since the crystal is assumed to be perfect in the $x$-$y$ plane. Taylor-expanding the applied potential $\Delta V(z)$ at the midpoint between each pair of layers $L$ and $L-1$, $z_L = (L-1/2)a/2$, we have:

$$\Delta V(z) = \sum_{n=0}^{\infty} \Delta V_{n}^{(L)} (z - z_L)^n$$

(3.112)

We retain only the lowest-order term in each Taylor expansion (3.112), so that the only constituent matrices of $T_{L,L-1}$ to be modified are:

$$H_{(L,L)}^{(a,a)} = [H_{(L,L)}]_0 - \Delta V_0^{(L)} \mathbb{1} ; \quad H_{(L-1,L-1)}^{(c,c)} = [H_{(L-1,L-1)}]_0 - \Delta V_0^{(L)} \mathbb{1}$$

(3.113)

where the subscript '0' on a matrix indicates that the parameters to be used are the appropriate bulk (or averaged, in the case of an interface as discussed above) parameters.

### 3.5.2 The Transmission Coefficient and the Current Density

The transmission and reflection coefficients for an electron of energy $E$ and parallel wavevector $k_\parallel$ are derived from the probability current density, $\mathcal{J}$. As usual, the system under consideration is perfect in the $x$-$y$ plane, in which we apply the cyclic boundary conditions, and is non-bulklike along $z$ in the region $0 < z < Na/2$; the crystal extends from $-\infty$ to $+\infty$ in $z$ (see fig. 3.3). The probability current density,
\[ J(r,t) = \frac{\hbar}{2m} \left[ \Psi^* \nabla \Psi(r,t) - \Psi \nabla \Psi^*(r,t) \right] . \]  

(3.114)

obeys the continuity equation:

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot J(r,t) = 0 \]

(3.115)

We are only concerned with the steady-state behavior of the system, so the time derivative in (3.115) vanishes. Consider a volume, \( V \), defined by: \( 0 \leq x \leq N_x a_x \), \( 0 \leq y \leq N_y a_y \), \( z_- \leq z \leq z_+ \), where \( N_x(y) \) is the number of cells in the \( x \) (\( y \)) direction and \( a_x(y) \) is their spacing (the cells need not be primitive unit cells), and \( z_- \) and \( z_+ \) are points in the bulklike regions \( z \leq 0 \) and \( z \geq (N+1)a/2 \), respectively. Integrating (3.115) over this volume and employing Gauss' Law, we have:

\[ \int_V \nabla \cdot J \, dr = \int_{\partial V} J \cdot \hat{n} \, dS = 0 \]

(3.116)

The surface integral in (3.116) is:

\[ \int_{0,0}^{z_+} \int_{N_x a_x}^{N_y a_y} \left[ J_x(x,y,z_+) - J_x(x,y,z_-) \right] \, dy \, dx + \int_{z_-}^{z_+} \int_{0}^{N_y a_y} \left[ J_y(x,N_y a_y,z) - J_y(x,0,z) \right] \, dx \, dz + \]

\[ + \int_{z_-}^{z_+} \int_{0}^{N_x a_x} \left[ J_z(N_x a_x,y,z) - J_z(0,y,z) \right] \, dy \, dz = 0 \]

(3.117)

Because of the application of cyclic boundary conditions in the \( x-y \) plane, we see that \( J_y(x,y,z) = J_y(x,y+N_y a_y,z) \) and \( J_x(x,y,z) = J_x(x+N_x a_x,y,z) \), so that the last two integrals in
(3.117) vanish. Thus, the quantity

\[
J_z = \int_{0}^{N_x a_x} \int_{0}^{N_y a_y} J_d(x,y,z) \, dy \, dx, \quad \forall \ z
\]

\[
(3.118)
\]
is independent of z. The probability current density in the z-direction is therefore given by:

\[
J_z = \frac{1}{m L_z} \langle \psi | p_z | \psi \rangle,
\]

\[
(3.119)
\]

where \( L_z \) is a normalization length. The transmission and reflection coefficients are simply:

\[
T = \frac{J_{z, \text{TRANS}}}{J_{z, \text{INC}}}
\]

\[
(3.120)
\]

\[
R = \frac{J_{z, \text{REFL}}}{J_{z, \text{INC}}}
\]

\[
(3.121)
\]

where \( J_{z, \text{INC}} \) is the incident current, \( J_{z, \text{TRANS}} \) is the transmitted current, and \( J_{z, \text{REFL}} \) is the reflected current.

We must now calculate the expectation value in (3.119). Since \( z_- \) and \( z_+ \) are arbitrary points in the bulklike regions, let us place them far enough inside their respective bulk regions that in each case the wavefunction is (essentially) a superposition of Bloch states only – that is, all of the evanescent states have decayed far from the interfaces. We may then calculate the matrix element as if the crystal were perfect in each region. For example, in the transmitted region, we have:

\[
|\Psi_{E,k\parallel}^{\text{TRANS}}\rangle = \sum_i b_i^{\text{TRANS}} |E; k_i\rangle
\]

\[
(3.122)
\]

where the state \( |E; k_i\rangle \) is a Bloch state of energy E and wavevector \( k_i = (k_{\parallel}, k_{z,i}) \). We must
now find the matrix elements $\langle E; k'|p_z|E; k\rangle$. Note that $p_z$ is translationally invariant: $p_z = T(d)p_z T^\dagger(d)$, for all $d$. Since for Bloch states $T(-R)|E; k\rangle = \exp[i k R] |E; k\rangle$, where $R$ is a direct lattice vector, we have:

$$
\langle E; k'|p_z|E; k\rangle = \langle E; k'|p_z|E; k\rangle_{\text{CELL}} \sum_{j=1}^{N} \exp[i (k - k') \cdot R_j]
$$

(3.123)

In (3.123), we have used the translational properties of the Bloch states to express the integral over the crystal as an integral over a unit cell and a sum over cells. The integral over a cell is independent of the sum over cells and for both $k$ and $k'$ in the first Brillouin zone, then, the sum is just $N \delta_{k,k'}$, where $N$ is the number of cells [24]:

$$
\langle E; k'|p_z|E; k\rangle = \langle E; k|p_z|E; k\rangle \delta_{k,k'}
$$

(3.124)

The expectation value in (3.124) is, not surprisingly, just [25]:

$$
\langle E; k|p_z|E; k\rangle = \frac{m}{\hbar} \frac{\partial E}{\partial k_z}
$$

(3.125)

Therefore, (3.119) becomes:

$$
J_z^{\text{TRANS}} = \frac{1}{\hbar L_z} \sum_i |b_i^{\text{TRANS}}|^2 \frac{\partial E}{\partial k_{z,i}}
$$

(3.126)

Note that the index, $i$, in (3.126) is composite, involving both wavevectors and bands. A similar result holds for the reflected current. Assuming the incident state to be a single Bloch state of $z$-wavevector $k_z^{\text{INC}}$ and unity incident current,

$$
T = \sum_i |b_i^{\text{TRANS}}|^2 \frac{|\partial E/\partial k_{z,i}^{\text{TRANS}}|}{|\partial E/\partial k_{z,i}^{\text{INC}}|}
$$

(3.127)
\[ R = \sum_i |b_i^{\text{REFL}}|^2 \left| \frac{\partial E/\partial k_{z,i}^{\text{REFL}}}{\partial E/\partial k_z^{\text{INC}}} \right| \]  

(3.128)

For the case considered in sec. 3.4.3, where only one incident state is kinematically possible, (3.128) reduces to:

\[ R = |b_{e,d}|^2. \]  

(3.129)

We note here that Osbourn and Smith [26] have proven reciprocity theorems for (3.127) and (3.128).

We now employ the transmission coefficient (3.127) to calculate the tunneling current for the device:

\[ j_z = \frac{2e}{(2\pi)^2 \hbar} \int_{BZ} \left( \int [f_e(E) - f_c[E]] T(E,k_{\|},V) dE \right) dk_y dk_x \]  

(3.130)

where \( f_e(E) \) and \( f_c(E) \), the Fermi-Dirac functions for the emitter and collector bulk-like regions, respectively, are:

\[ f_e(E) = \frac{1}{1 + \exp[(E - \mu_c)/k_B T]} \]  

(3.131)

\[ f_c(E) = \frac{1}{1 + \exp[(E - \mu_e)/k_B T]}, \]  

(3.132)

and \(-eV\) is the applied bias and \( \mu_e \) and \( \mu_c = (-eV + \mu_e) \) are the respective emitter and collector chemical potentials. In (3.130), 'BZ' indicates that the integral is to be taken over the Brillouin zone. Note that in (3.130) we are summing transmission probabilities instead
of transmission \textit{amplitudes}, which amounts to postulating that the phases of the (independent) electrons leaving the emitter Fermi sea are completely unrelated. By discarding all phase information, we have therefore introduced the necessary irreversibility into the description of the device.

As the reader is doubtless aware by now, none of the calculations we have performed have been particularly computationally convenient. Equation (3.130) is, alas, no exception: being a triple integral, it is a numerically unfriendly beast, exceedingly time-consuming to compute. It is therefore time for an uncontrolled approximation. In (3.130), we make the replacement:

$$T(E, k_{||}, V) \rightarrow T(E - E_{||}, 0, V)$$

(3.133)

where $E_{||} = \hbar^2 k_{||}^2 / 2m^*_{\Gamma, GaAs}$. In making this replacement, we make two approximations. The first, and more defensible, is that the electrons have small parallel wavevector, so that they may be considered "free" particles in the $x$-$y$ plane. The second is that the transmission coefficient depends mostly on $k_{z, \text{INC}}$ and only weakly on $k_{||}$. Equation (3.133) allows us to turn the triple integral into the more manageable and familiar single integral [27]:

$$J_z = \frac{e m^*_\Gamma, GaAs k_B T}{2 \pi^2 \hbar^3} \int_0^\infty T(E, V) \ln \left\{ \frac{1 + \exp[(\mu_e - E)k_B T]}{1 + \exp(\mu_e - E - eV)k_B T} \right\} dE$$

(3.134)

The approximations just made will be examined in detail (insofar as our computational resources will allow such investigation) in a later section. For the present, we shall move on to a discussion of the predictions of the tight-binding model, beginning first with the current-voltage characteristic, which is what one actually measures in practice.

3.5.3 \textit{Simulation Results: Current-Voltage Characteristics}
Figure 3.4 is the J-V characteristics of a triple-barrier RTD calculated with both the tight-binding (solid line) and envelope-function (broken line) models. The device's emitter, collector, and wells are made of GaAs and its barriers are AlAs. The first barrier is 2 ML (monolayers; 1 ML = 2.83 Å) thick, the first well 11 ML, the second barrier 4 ML, the second well 14 ML and the third well 6 ML. Additionally, the undoped emitter spacer is 35 ML while the undoped collector spacer is 88 ML. The emitter and collector are both doped N-type with $N_D = 10^{18}$ cm$^{-3}$, and the temperature is 77 K. In both cases, the classical Poisson equation (see Ch. 2) was solved to obtain the potential profile for the Schrödinger equation: the two equations were not solved self-consistently. In both calculations, the electrons are incident from the GaAs $\Gamma$-valley and in the envelope-function calculation, a $\Gamma$-point profile is assumed with no nonparabolicity corrections for either real- or complex-bands.

**Figure 3.4:** J-V plot for the triple-barrier RTD discussed in sec 3.5.3 calculated with both the tight-binding (solid line) and envelope-function (broken line) models.
The two graphs of fig. 3.4 are quite different even though both have two peaks. The first peaks resemble each other qualitatively and semi-quantitatively, but the second peaks are completely different. The similarity of the first peak of the tight-binding model to its envelope-function counterpart suggests that it is probably due to a resonance in one of the GaAs wells, and that the wavefunction of most electrons will be mainly composed of $\Gamma$-valley states. Likewise, the great dissimilarity between the two second peaks, along with the fact that the tight-binding model automatically includes the effects of the longitudinal X-valleys[28], suggests that the tight-binding wavefunction in this case will be more X-like than $\Gamma$-like[29]. In the next subsection, we shall demonstrate that this is indeed the case.

For the present, however, we will be content to explore two less dramatic differences: the lower resonant biases in the tight-binding model relative to the envelope-function model and the higher peak and valley currents.

Both of these differences result from the fact that the RTD in the tight-binding model is more transparent than in the envelope-function model. Recall that the envelope-function calculation assumes $\Gamma$-state tunneling throughout and includes no corrections for either conduction-band or imaginary-band nonparabolicity. In contrast, the tight-binding calculation automatically addresses both. This is particularly important in the barriers, where the tight-binding $\Gamma$-states, because they obey the proper imaginary-band dispersion relation, decay more slowly than do their envelope-function counterparts. This tends to make the tight-binding RTD more transparent than the envelope-function RTD, and hence leads to lower peak biases and higher transmission, both on- and off-resonance. On the other hand, the electrons in the well of the envelope-function RTD will have higher group velocities than they will in the well of the tight-binding RTD, once again because the envelope-function calculation's (real) conduction band dispersion relation is parabolic. This tends to have somewhat the opposite effect on both peak position and current, but it should be less important since the well states are propagating unlike the evanescent barrier states.
Let us also briefly note that when series resistance is included, the tight-binding model will predict significantly more realistic I-V characteristics than will the envelope-function model. Not only are the peaks in the tight-binding J-V curve closer together than the corresponding peaks in the envelope-function curve, there are additional peaks as well. Recall from Ch. 2 that series resistance tends to decrease the peak-to-valley ratio by pushing one peak into another since the J-V point \((V_{\text{RTD}}, J_{\text{RTD}})\) is mapped to the I-V point \((V_{\text{RTD}} + J_{\text{RTD}}AR_S, J_{\text{RTD}}A)\), where A is the device cross-sectional area and \(R_S\) is the series resistance. In fig. 3.4, we see that although this will not make much difference in the envelope-function calculation, including even a small series resistance in the tight-binding calculation will skew the first peak into the second because they are so close.

3.5.4 Simulation Results: Transmission Coefficients

While examination of the current-voltage characteristics produced by the tight-binding model yields some insight, an even greater understanding of the situation may be achieved by studying the transmission coefficients predicted by the model, even though they cannot be measured in practice. Our convention will be to graph transmission versus applied bias for an electron of a given energy, \(E\), and parallel wavevector, \(k_{||}\). We choose this convention over the usual practice, graphing transmission versus energy for a given \(k_{||}\) and applied bias for two reasons. First, and most important, in any real experiment, one can only inject electrons with energies less than about 100 meV above the conduction-band minimum in the emitter bulklike region. Thus, graphing \(T\) vs. \(E\) for \((E-E_{c,\text{MIN}}) > 0.1\) eV is utterly meaningless. Second, the \(T\) vs. \(V\) curves often bear a strong resemblance to the J-V curves, thus allowing us to infer the nature of the various peaks in the J-V curve.

Figure 3.5 is a graph of transmission versus applied bias for the triple-barrier RTD discussed in sec. 3.5.3. The incident electron has \(k_{||} = 0\) and \(E = 5.506\) meV above the GaAs conduction-band minimum in the emitter \((k_z^{\text{INC}} = 0.01\ \text{Å}^{-1})\). In table 3.5, we
present the expansion of the wavefunction in certain layers of the RTD in terms of the bulk transfer matrix eigenstates for those layers. (The square magnitudes of the expansion coefficients do not add to unity because the transfer matrix eigenstates are not necessarily orthogonal – recall that the transfer matrix is manifestly non-Hermitian. This also means that the decomposition must be computed as an \( Ax = b \) problem.) Figure 3.6 presents the electron probability density in the structure at two different biases of interest, clearly showing in which regions the electron becomes localized. (Note that because these are conduction-band states the cation density will usually be greater than the anion density.) To aid us in interpreting the wavefunction expansions in table 3.5, we graph the complex bands of GaAs and AlAs in fig. 3.7. These are just the (in general complex) wavevectors, \( k_z \), extracted from the eigenvalues of the single-monolayer transfer matrices for GaAs and AlAs, with \( k_\parallel = 0 \). (Recall that the transfer matrix eigenvalues take the form \( \exp[i \, k_z a/2] \), where \( k_z \) is in general complex.)

\[ \text{Figure 3.5: Transmission coefficient versus applied bias for the triple-barrier RTD discussed in sec. 3.5.3 and 3.5.4 calculated with the tight-binding model. The incident electron has } k_\parallel = 0 \text{ and } (E - E_{\text{MIN}}) = 5.506 \text{ meV.} \]
Figure 3.6: Electron probability density for the triple-barrier structure of sec. 3.5.3 at biases of 0.550 V (top) and 0.941 V (bottom). Solid line: Γ-point conduction band profile. Dotted line: cation density. Dotted-dashed line: anion density. At 0.550 V, the electron is clearly localized in the second well of the RTD, while at 0.941 V it is mostly in the last barrier.
Comparison of the transmission coefficient graph (fig. 3.5) and the tunneling current graph (fig. 3.4) reveals that the two main tunneling current peaks at 0.52 V and 0.92 V (with current densities of 1.12x10^5 A/cm^2 and 3.20x10^5 A/cm^2, respectively) coincide nicely with the transmission peaks at 0.550 V and 0.941 V, respectively. Examining the wavefunction decompositions at these two biases (table 3.5) and the electron probability densities (fig. 3.6) will help determine the nature of these features. The decomposition at 0.550 V shows that this resonance is associated with the second well of the RTD, and that in this well, the wavefunction is mostly composed of Γ-states. We see, also, that the wavefunction is mostly Γ-like in the first well and second barrier. Only in the last barrier is it mainly composed of X-states. (That is, longitudinal-X valley states; since k_∥ = 0, the transverse-X valleys play no role.) The transmission resonance at 0.941 V, on the other hand, is more associated with the last barrier, although the electron has a significant probability to be in the wells. Furthermore, the wavefunction in the second barrier, second well, and last barrier is mostly composed of X-valley states. Since the first barrier is only 2 ML thick, it will not be very effective at confining the electrons, so that under some circumstances the second barrier, second well, and third barrier effectively comprise the RTD. These conclusions hold for electrons with larger energies as well; fig. 3.8 is the transmission graph for the same RTD, but for an electron incident with k_∥ = 0 and E = 21.702 meV above the GaAs conduction-band minimum in the emitter (k_z^{INC} = 0.02 Å⁻¹). Wavefunction decompositions for this electron (not listed) are similar. Finally, note that the lower, broader transmission resonance coincides with the larger of the two tunneling current peaks.
Figure 3.7: The complex bandstructures of GaAs (top) and AlAs (bottom) for $k_{\parallel} = 0$, as generated by the respective bulk transfer matrices using our parameters (Table 3.2). Solid lines are bands for purely-real or purely-imaginary $k_{zz}$, while broken lines (of the same type) represent the real and imaginary parts of complex $k_z$. 
We may gain further insight by examining other features of the J-V curve, such as the bump at 0.37 V and the plateau from about 1.0 V to 1.15 V. The former feature corresponds well to the two transmission peaks at 0.437 V and 0.457 V. Both of these peaks are associated primarily with the last barrier, and secondarily with the last well. In both cases, X-valley states are the major contributors to the wavefunction in the second barrier, second well, and last barrier, although in some instances Γ-valley states make significant contributions. The anti-resonances, such as the one at 0.443 V are also of interest. Ko and Inkson [31] note that this is due to destructive interference between Γ and longitudinal-X states in the AlAs and may be understood in terms of Fano resonances, which occur when a discrete state is coupled to a continuum. (Recall that the X-point profile has barriers in the GaAs and wells in the AlAs, so that there are quasi-bound states in the AlAs in addition to the more familiar quasi-bound states in the GaAs.) Fano [32] demonstrates that the contributions from the discrete state and the continuum states to the total transition amplitude (out of some initial state) will constructively interfere on one side of the resonance and destructively interfere on the other, thus producing the peculiar anti-resonances seen in the transmission graphs. Finally, we can learn more about the plateau by examining the wavefunction expansion at 1.127 V in table 3.5. There, we see that the electron is mainly in the first well and second barrier, and that it tends to be of a mixed composition in the first well and barrier, more X-like in the last two barriers, and more Γ-like in the second well.

We should note that our results discussed above do not entirely agree with those of Rousseau, et al. [17], who find that the electron typically takes an all-Γ or a Γ₁-X₁RTD-C₂ path (where 'E' and 'C' denote emitter and collector bulk regions, respectively) tunneling path through the RTD. While we do find largely-Γ and largely-X wavefunctions (at, say, biases of 0.550 V and 0.941 V, respectively, for the RTD of fig. 3.6), we find that the electron is mainly X-like in the last barrier, even when it is mostly Γ-like elsewhere. Furthermore, we find instances such as the "plateau" region discussed above where the wavefunction is
generally mixed in some regions, $\Gamma$-like in others, and $X$-like in yet others. Also, we find that $X$-like and mixed transmission resonances can lead to peaks in the $J$-$V$ curve. (It is, however, doubtful that NDR due to these effects can be easily observed in practice, since we have not taken into account such nonidealities as series resistance.) In considering the differences between our work and that of Rousseau, et al. [17], it must be remembered that our transmission coefficients are for biased RTDs including space-charge regions, while theirs are for unbiased structures with no space-charge regions. Our parameterization also places the AlAs $X$-valley above the GaAs $\Gamma$-valley while theirs places it below. Since the application of a potential drop will disrupt the $z$-symmetry as well, a wavefunction which is $\Gamma$-like on one end of a biased region can have $X$-components on the other end of the region. Nevertheless, most of the $X$-like transmission resonances are significantly lower than those due to $\Gamma$-like behavior, which partially supports the conclusions of Rousseau, et al.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure38.png}
\caption{
Transmission coefficient versus applied bias for the triple-barrier RTD discussed in sec. 3.5.3 and 3.5.4 calculated with the tight-binding model. The incident electron has $k_{\parallel} = 0$ and $(E - E_{c, \text{MIN}}) = 21.702 \text{ meV}.$
}\end{figure}
TABLE 3.5:  WAVEFUNCTION DECOMPOSITIONS FOR FIG. 3.5

In this table, the ‘Coefficient’ column lists the square magnitude of the expansion coefficient and the ‘State’ column lists the wavevector as \( \text{Re}\{kz\}, \text{Im}\{kz\} \). Only states having square magnitudes of expansion coefficients greater than 0.05% of the largest squared expansion coefficient are listed.

\( BIAS: 0.437 \text{ V} \)

<table>
<thead>
<tr>
<th>Last layer of Barrier 1</th>
<th>Last layer of Well 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density:</strong> 0.00572</td>
<td><strong>Density:</strong> 0.00294</td>
</tr>
<tr>
<td><strong>Coefficient</strong></td>
<td><strong>Coefficient</strong></td>
</tr>
<tr>
<td>0.0000995 (0.8459, -0.1118)</td>
<td>0.000100 (0.9147, -0.3379)</td>
</tr>
<tr>
<td>0.001052 (-0.8459, -0.1118)</td>
<td>0.000106 (-0.9147, -0.3379)</td>
</tr>
<tr>
<td>0.000872 (0.8459, 0.1118)</td>
<td>0.002527 (0.0343, 0.0000)</td>
</tr>
<tr>
<td>0.000907 (-0.8459, 0.1118)</td>
<td>0.001318 (-0.0343, 0.0000)</td>
</tr>
<tr>
<td>0.001147 (0.0000, -0.1354)</td>
<td></td>
</tr>
<tr>
<td>0.003468 (0.0000, 0.1354)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last layer of Barrier 2</th>
<th>Last layer of Well 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density:</strong> 0.00086</td>
<td><strong>Density:</strong> 0.01510</td>
</tr>
<tr>
<td><strong>Coefficient</strong></td>
<td><strong>Coefficient</strong></td>
</tr>
<tr>
<td>0.000218 (0.8508, -0.0204)</td>
<td>0.003468 (0.9178, -0.3187)</td>
</tr>
<tr>
<td>0.000085 (0.8508, 0.0204)</td>
<td>0.004273 (-0.9178, -0.3187)</td>
</tr>
<tr>
<td>0.001579 (-0.8508, -0.0204)</td>
<td>0.002152 (0.0480, 0.0000)</td>
</tr>
<tr>
<td>0.001066 (-0.8508, 0.0204)</td>
<td>0.001203 (-0.0480, 0.0000)</td>
</tr>
<tr>
<td>0.000784 (0.0000, -0.1341)</td>
<td></td>
</tr>
<tr>
<td>0.000101 (0.0000, 0.1341)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last layer of Barrier 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density:</strong> 0.20418</td>
<td></td>
</tr>
<tr>
<td><strong>Coefficient</strong></td>
<td></td>
</tr>
<tr>
<td>0.029767 (0.7282, 0.0000)</td>
<td></td>
</tr>
<tr>
<td>0.035560 (-0.7282, 0.0000)</td>
<td></td>
</tr>
<tr>
<td>0.047544 (0.9933, 0.0000)</td>
<td></td>
</tr>
<tr>
<td>0.052865 (-0.9933, 0.0000)</td>
<td></td>
</tr>
<tr>
<td>0.000553 (0.0000, -0.1322)</td>
<td></td>
</tr>
<tr>
<td>0.000051 (0.0000, 0.1322)</td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>State</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>0.001345</td>
<td>(0.8462,-0.1087)</td>
</tr>
<tr>
<td>0.001400</td>
<td>(-0.8462,-0.1087)</td>
</tr>
<tr>
<td>0.001179</td>
<td>(0.8462,0.1087)</td>
</tr>
<tr>
<td>0.001214</td>
<td>(-0.8462,0.1087)</td>
</tr>
<tr>
<td>0.001455</td>
<td>(0.0000,-0.1353)</td>
</tr>
<tr>
<td>0.004356</td>
<td>(0.0000,0.1353)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000183</td>
<td>(0.8225,0.0000)</td>
<td>0.009246</td>
<td>(0.9182,-0.3164)</td>
</tr>
<tr>
<td>0.000946</td>
<td>(-0.8225,0.0000)</td>
<td>0.009815</td>
<td>(-0.9182,-0.3164)</td>
</tr>
<tr>
<td>0.000009</td>
<td>(0.8804,0.0000)</td>
<td>0.001448</td>
<td>(0.0494,0.0000)</td>
</tr>
<tr>
<td>0.000343</td>
<td>(-0.8804,0.0000)</td>
<td>0.000582</td>
<td>(-0.0494,0.0000)</td>
</tr>
<tr>
<td>0.000510</td>
<td>(0.0000,-0.1340)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.000124</td>
<td>(0.0000,0.1340)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.016274</td>
<td>(0.7213,0.0000)</td>
</tr>
<tr>
<td>0.015476</td>
<td>(-0.7213,0.0000)</td>
</tr>
<tr>
<td>0.021612</td>
<td>(1.0034,0.0000)</td>
</tr>
<tr>
<td>0.017435</td>
<td>(-1.0034,0.0000)</td>
</tr>
<tr>
<td>0.000324</td>
<td>(0.0000,-0.1319)</td>
</tr>
<tr>
<td>0.000036</td>
<td>(0.0000,0.1319)</td>
</tr>
</tbody>
</table>
TABLE 3.5: **CONTINUED**

**BIAS: 0.550 V**

<table>
<thead>
<tr>
<th>Last layer of Barrier 1</th>
<th>Last layer of Well 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density:</strong> 0.60144</td>
<td><strong>Density:</strong> 5.01271</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.203092</td>
<td>(0.8474, -0.0928)</td>
<td>0.298480</td>
<td>(0.9159, -0.3302)</td>
</tr>
<tr>
<td>0.217788</td>
<td>(-0.8474, -0.0928)</td>
<td>0.297643</td>
<td>(-0.9159, -0.3302)</td>
</tr>
<tr>
<td>0.116050</td>
<td>(0.8474, 0.0928)</td>
<td>0.995925</td>
<td>(0.0404, 0.0000)</td>
</tr>
<tr>
<td>0.125894</td>
<td>(-0.8474, 0.0928)</td>
<td>0.806393</td>
<td>(-0.0404, 0.0000)</td>
</tr>
<tr>
<td>0.427735</td>
<td>(0.0000, -0.1350)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.146988</td>
<td>(0.0000, 0.1350)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last layer of Barrier 2</th>
<th>Last layer of Well 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density:</strong> 5.18467</td>
<td><strong>Density:</strong> 18.87624</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.942511</td>
<td>(0.7724, 0.0000)</td>
<td>0.317607</td>
<td>(0.9199, -0.3056)</td>
</tr>
<tr>
<td>0.968380</td>
<td>(-0.7724, 0.0000)</td>
<td>0.188610</td>
<td>(-0.9199, -0.3056)</td>
</tr>
<tr>
<td>0.198535</td>
<td>(0.9362, 0.0000)</td>
<td>10.300329</td>
<td>(0.0554, 0.0000)</td>
</tr>
<tr>
<td>0.190804</td>
<td>(-0.9362, 0.0000)</td>
<td>10.146392</td>
<td>(-0.0554, 0.0000)</td>
</tr>
<tr>
<td>5.446380</td>
<td>(0.0000, -0.1334)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.096746</td>
<td>(0.0000, 0.1334)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Last layer of Barrier 3 | Density: 8.02227 |

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.162088</td>
<td>(0.6940, 0.0000)</td>
</tr>
<tr>
<td>1.901163</td>
<td>(-0.6940, 0.0000)</td>
</tr>
<tr>
<td>1.335023</td>
<td>(1.0529, 0.0000)</td>
</tr>
<tr>
<td>1.526040</td>
<td>(-1.0529, 0.0000)</td>
</tr>
<tr>
<td>0.022679</td>
<td>(0.0000, -0.1307)</td>
</tr>
<tr>
<td>0.179274</td>
<td>(0.0000, 0.1307)</td>
</tr>
</tbody>
</table>
TABLE 3.5:  \textit{CONTINUED}

\textit{BIAS: 0.941 V}

\textit{Last layer of Barrier 1}
Density: 0.22423

\begin{tabular}{ll}
\hline
Coefficient & State \\
0.171670 & (0.7872,0.0000) \\
0.179013 & (0.7872,0.0000) \\
0.088572 & (0.9190,0.0000) \\
0.105323 & (0.9190,0.0000) \\
0.133663 & (0.0000,-0.1336) \\
0.065329 & (0.0000,0.1336) \\
\hline
\end{tabular}

\textit{Last layer of Well 1}
Density: 0.53788

\begin{tabular}{ll}
\hline
Coefficient & State \\
0.034639 & (0.9202,-0.3036) \\
0.052083 & (-0.9202,-0.3036) \\
0.236300 & (0.0565,0.0000) \\
0.163047 & (-0.0565,0.0000) \\
\hline
\end{tabular}

\textit{Last layer of Barrier 2}
Density: 0.17637

\begin{tabular}{ll}
\hline
Coefficient & State \\
0.032330 & (0.6892,0.0000) \\
0.075323 & (-0.6892,0.0000) \\
0.063946 & (1.0652,0.0000) \\
0.077717 & (-1.0652,0.0000) \\
0.013453 & (0.0000,-0.1305) \\
0.019549 & (0.0000,0.1305) \\
\hline
\end{tabular}

\textit{Last layer of Well 2}
Density: 1.71795

\begin{tabular}{ll}
\hline
Coefficient & State \\
0.827725 & (0.9275,-0.2573) \\
0.954216 & (-0.9275,-0.2573) \\
0.065087 & (0.0767,0.0000) \\
0.001207 & (-0.0767,0.0000) \\
\hline
\end{tabular}

\textit{Last layer of Barrier 3}
Density: 3.57945

\begin{tabular}{ll}
\hline
Coefficient & State \\
0.810838 & (0.6168,0.0000) \\
1.035870 & (-0.6168,0.0000) \\
0.250170 & (1.1101,-0.1046) \\
0.023861 & (1.1101,0.1046) \\
0.026828 & (0.0000,-0.1250) \\
0.000898 & (0.0000,0.1250) \\
\hline
\end{tabular}
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.124346</td>
<td>(0.7543, 0.0000)</td>
<td>1.279798</td>
<td>(0.9222, -0.2909)</td>
</tr>
<tr>
<td>0.124340</td>
<td>(-0.7543, 0.0000)</td>
<td>1.291610</td>
<td>(-0.9222, -0.2909)</td>
</tr>
<tr>
<td>0.064687</td>
<td>(0.9583, 0.0000)</td>
<td>1.154091</td>
<td>(0.0628, 0.0000)</td>
</tr>
<tr>
<td>0.068935</td>
<td>(-0.9583, 0.0000)</td>
<td>1.118476</td>
<td>(-0.0628, 0.0000)</td>
</tr>
<tr>
<td>0.580173</td>
<td>(0.0000, -0.1329)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.000280</td>
<td>(0.0000, 0.1329)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.656226</td>
<td>(0.6645, 0.0000)</td>
<td>0.007462</td>
<td>(0.9315, -0.2317)</td>
</tr>
<tr>
<td>1.684960</td>
<td>(-0.6645, 0.0000)</td>
<td>0.030006</td>
<td>(-0.9315, -0.2317)</td>
</tr>
<tr>
<td>1.830179</td>
<td>(1.1101, -0.0612)</td>
<td>0.207969</td>
<td>(0.0854, 0.0000)</td>
</tr>
<tr>
<td>0.851447</td>
<td>(1.1101, 0.0612)</td>
<td>0.176181</td>
<td>(-0.0854, 0.0000)</td>
</tr>
<tr>
<td>0.033986</td>
<td>(0.0000, -0.1290)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.160177</td>
<td>(0.0000, 0.1290)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>State</th>
<th>Coefficient</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.029490</td>
<td>(0.5902, 0.0000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.063504</td>
<td>(-0.5902, 0.0000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.008204</td>
<td>(1.1101, -0.1119)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.000218</td>
<td>(1.1101, 0.1119)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.007133</td>
<td>(0.0000, -0.1219)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.006223</td>
<td>(0.0000, 0.1219)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.6 SINGLE-BARRIER DIODES

The observation of negative differential resistance in the I-V characteristic of a single barrier GaAs/AlAs/GaAs tunneling structure by Beresford, et al. [33] highlights the importance of bandstructure effects in tunnel diodes. Rousseau, et al. [17] calculate current-voltage characteristics for a single-barrier diode and Ko and Inkson [31] compute transmission coefficients, however neither of these two theoretical investigations is satisfactory, for neither includes space-charge regions. (The latter work even treats the devices in the flat-band approximation.) As we shall presently demonstrate, the inclusion of space-charge regions can have a profound effect on the predictions of the model.

The four different GaAs/AlAs/GaAs single-barrier heterostructures we consider have a single undoped AlAs barrier of either 5, 10, 20, or 50 monolayers (1 monolayer = 2.83 Å) surrounded by two undoped GaAs spacers (each 35 monolayers) and a GaAs emitter and collector each doped to \( N_D = 10^{18} \text{ cm}^{-3} \). We take \( T = 77 \text{ K} \) throughout. In Figures 3.9-3.12 we examine the tunneling of a low-energy electron incident on each of the different barriers. The electron has \( k_{\parallel} = 0 \) and \( E - E_C = 5.5058 \text{ meV} \) (\( k_{z,\text{inc}} = 0.01 \text{ Å}^{-1} \)). The solid lines are the transmission coefficients calculated with the tight-binding model while the dashed lines are transmission coefficients calculated with the envelope-function model for an electron with \( k_{\parallel} = 0 \) and \( E - E_C = 5.5058 \text{ meV} \) as well. (The envelope-function model assumes \( \Gamma \)-point tunneling and does not include nonparabolicity corrections). In general, we see that the two models agree more closely for thinner barriers than they do for the thicker ones, the tight-binding model predicting significantly higher transmission than does the envelope-function model. This disagreement is particularly striking for the 50 monolayer device, where at higher biases the tight-binding model predicts transmission coefficients well over ten orders of magnitude greater than those predicted by the envelope-function calculation.
Figure 3.9: Transmission coefficient versus applied bias through a 5 monolayer AlAs barrier for an electron incident from the GaAs Γ-valley with $k_{||} = 0$ and $E - E_C = 5.5058 \text{ meV}$, as calculated with the tight-binding (solid line) and envelope-function (dashed line) models. See text for details of the barrier structure.

Figure 3.10: Transmission coefficient for the electron of Fig. 3.9 through a 10 monolayer AlAs barrier.
Figure 3.11: Transmission coefficient for the electron of Fig. 3.9 through a 20 monolayer AlAs barrier.

Figure 3.12: Transmission coefficient for the electron of Fig. 3.9 through a 50 monolayer AlAs barrier.
Let us now examine several important trends and features evident in these graphs. First, notice that at low biases the tight-binding and envelope-function curves parallel each other, but the transmission as calculated in the tight-binding model is higher than that in the envelope-function model, the discrepancy between the predictions of the two models increasing dramatically for thicker barriers. The fact that the two curves parallel each other at low biases suggests that $\Gamma$-state tunneling is most important here. Decomposing the tight-binding wavefunction in the last layers of both the accumulation region and the barrier reveals that this is indeed the case: the wavefunction is more $\Gamma$-like at the lowest biases (even in the barrier) and still has significant $\Gamma$-state contributions at somewhat higher biases in this regime. Examination of the complex bandstructure of AlAs in our parameterization (Fig. 3.7) will help explain both why $\Gamma$-state tunneling is most important at low biases and why the tight-binding model predicts significantly higher transmission than does the envelope-function model. Referring to Fig. 3.7, we see that an electron incident on an AlAs barrier from the GaAs $\Gamma$-minimum has an energy such that in the barrier (where the electron is in a superposition of $\Gamma$- and $X$-valley states), the imaginary parts of the wavevectors of its primary $X$-valley components are nearly equal in magnitude to the (purely imaginary) wavevectors of its $\Gamma$-valley components. Above this energy, the $X$-valley states decay more slowly than do the $\Gamma$-states, and they quickly become propagating. Thus, the $\Gamma$-states will be important only at low biases, where they decay less rapidly than the $X$-states. Conversely, the $X$-valley states in the barrier are always significant, becoming totally so for thicker barriers and at higher biases. (At a given bias, more potential will be dropped across a thick barrier than a thin one.) The difference in transmission, even at very low biases (for which $\Gamma$-valley transport dominates), is due to the tight-binding model's incorporation of imaginary-band nonparabolicity, so that at a given energy the tight-binding $\Gamma$-states will decay much less rapidly than will their
envelope-function counterparts. For a given barrier width, this results in increased
transmission as compared to the envelope-function model, the increase of course being
greater for thicker barriers.

The resonance/antiresonance pairs present in all these graphs -- the antiresonances
are due to destructive interference between $\Gamma$- and longitudinal $X$-states in the AlAs[31] and
may be understood in terms of Fano resonances[32], which occur when a discrete state is
coupled to a continuum -- deserve attention as well, since with regard to them our
simulations seem on first inspection at variance with those of Ko and Inkson[31]. Careful
consideration, however, shows that these differences are due to the fact that our
calculations include the space-charge regions on either side of the barrier while those of
Ref. 31 were performed under flatband conditions. First, while Ko and Inkson’s[31]
calculations show a sharp transition from $\Gamma$-like to $X$-like tunneling as the incident electron
energy is increased above the AlAs $X$-valley conduction-band edge, we do not observe
such a change here because we consider transmission versus bias for electrons of a given
energy. Due to the bias, it is often the case that the bulk emitter $\Gamma$-point minimum lies
below the $X$-point minimum in the emitter side of the barrier but above the $X$-point
minimum in the collector side of the barrier. Thus, the $X$-point profile in the AlAs begins
as a triangular barrier and ends up as a triangular well. Hence, at all but the lowest biases,
even low-energy electrons will be in AlAs $X$-valley Bloch states in the collector end of the
barrier, and we will not observe sharp transitions from $\Gamma$-like to $X$-like tunneling.
Furthermore, it follows that Fano resonances will occur in thin barrier structures even for
low-energy electrons. Additionally, our simulations indicate that the Fano resonances exist
in very thick barriers, while the results of Ref. 31 predict that they will be absent due to the
fact that the $\Gamma$-state contribution to the electron wavefunction becomes negligible compared
to the $X$-state contribution (in the terminology of Ref. 31, “the $\Gamma$-channel disappears”).
Once again, this difference is due to the fact that Ko and Inkson only consider devices
under flatband conditions. In this case, the $X$-state contribution to the electron
wavefunction consists of states which are either only evanescent or only propagating throughout the entire barrier. Now in a thick barrier, even for high-energy electrons, the AlAs Γ-valley states will always be evanescent and will contribute little toward the wavefunction as compared to the propagating X-valley states. This in effect deprives the system of a continuum, and, consequently, the Fano resonances do not appear[31]. In contrast, our simulations show the Fano resonances persisting in even the 50 monolayer device until about 0.80 V. Note that it is at approximately this bias that the X-valleys in both the emitter side of the AlAs barrier and the GaAs undoped collector spacer become even with the bulk emitter conduction band minimum, so that there are purely propagating states throughout the barrier and propagating X-valley states in the GaAs collector. (The Fano resonances persist beyond this point in thin barriers because the Γ-states will not decay appreciably over the length of the barrier.) At lower biases, then, the emitter side of the AlAs will have only evanescent states, so that the Γ-state contribution to the electron wavefunction will be sufficient in comparison to that of the X-states that Fano resonances will occur. In addition, at higher biases the collector side of the barrier will have evanescent as well as propagating X-valley states, again providing a continuum so that the Fano resonances can reappear. Also, we can see that for thick barriers the Fano resonances will persist at higher biases as the barrier width is increased. This is merely due to electrostatics: for a given bias, the wider the (insulating) barrier, the greater the potential that appears across it, and the smaller the potential that appears across the space-charge regions on either side of it. A higher bias is thus necessary before the X-valley minimum in the emitter side of the AlAs drops below the bulk emitter conduction band minimum, and the Fano resonances will therefore persist at higher biases.

More may be learned about the nature of the tunneling by examining the peaks of the transmission graphs in detail. The 5 monolayer barrier (fig. 3.9) is rather interesting in that both the envelope-function and tight-binding calculations predict a gentle transmission resonance at about 0.4 V bias. Since the envelope-function model includes only Γ-valley effects, we expect that this weak resonance is associated with tunneling over the triangular
well-shaped accumulation region in front of the barrier. Using the tight-binding model, we may verify this conclusion by decomposing the wavefunction (at the resonant bias) in the last layer of both the accumulation region and the barrier in terms of the bulk transfer matrix eigenstates appropriate for each layer. Doing so, we find that the electron density in the last layer of the accumulation region in about 36 times as great as it is in the last layer of the barrier, and that in the former layer the wavefunction is mostly composed of $\Gamma$-valley states ($k_z = \pm 0.0362 \ \text{Å}^{-1}$), while in the latter it is mainly composed of $X$-valley states ($k_z = \pm 0.7749 \ \text{Å}^{-1}$, $k_z = \pm 0.9332 \ \text{Å}^{-1}$) but with significant evanescent $\Gamma$-valley components ($k_z = \pm i 0.1334 \ \text{Å}^{-1}$). Performing similar decompositions for the other two peaks of the tight-binding curve (at about 0.60 and 0.81 V, respectively), we find that in both cases the resonance is associated with the barrier rather than the accumulation region and that in both the last layer of the accumulation region and the last layer of the barrier the wavefunction is mostly composed of $X$-valley states. (These are evanescent in the accumulation region.)

In Figures 3.13-3.15 we examine relatively high-energy electron transmission through the 5, 10, and 20 monolayer barriers, respectively. The electron under consideration has $k_{\parallel} = 0$ and $E - E_C = 63.9777 \ \text{meV}$ ($k_{z, \text{inc}} = 0.035 \ \text{Å}^{-1}$), placing it slightly above the bulk emitter Fermi level. As before, we see that the envelope-function and tight-binding models agree most closely for the 5 monolayer barrier device, and that the disagreement between the two models increases with increasing barrier thickness. We see, too, that the high-energy tight-binding curves are quite similar to those for lower energy, the major difference being that corresponding features in the high-energy graphs appear at lower biases than in the low-energy curves, as one should expect. (We do not present the high-energy results for the 50 monolayer barrier device because they display these same general trends relative to the low-energy case as do those for the thinner barrier structures.) Wavefunction decompositions like those discussed above reveal that all of the resonances for each device are associated with the $X$-valley states in the barrier rather than the accumulation layer.
Figure 3.13: Transmission coefficient versus applied bias through a 5 monolayer AlAs barrier for an electron incident from the GaAs Γ-valley with \( k_\parallel = 0 \) and \( E - E_C = 63.9777 \) meV, as calculated with the tight-binding (solid line) and envelope-function (dashed line) models. See text for details of the barrier structure.

Figure 3.14: Transmission coefficient for the electron of Fig. 3.13 through a 10 monolayer AlAs barrier.
**Figure 3.15:** Transmission coefficient for the electron of Fig. 3.13 through a 20 monolayer AlAs barrier.

**Figure 3.16:** Transmission versus applied bias through a 10 monolayer AlAs barrier calculated with the tight-binding model for $k_\parallel = 0$, $E - E_C = 21.7020$ meV (solid line) and $k_\parallel = (0, 0.01)$ Å$^{-1}$, $E - E_C = 26.9507$ meV (dashed-dotted line), as well as transmission for $E - E_C = 21.7020$ meV as calculated with the envelope-function model (dashed line).
In Fig. 3.16 above, we examine the role of nonzero $k_\parallel$ on tunneling. In fig. 3.16, the solid and dashed-dotted curves are calculated with the tight-binding model, representing electrons with $k_{inc} = (0, 0.02) \text{Å}^{-1}$, $E - E_C = 21.7020$ meV and $k_{inc} = (0, 0.01, 0.02) \text{Å}^{-1}$, $E - E_C = 26.9507$ meV, respectively. The dashed curve is calculated with the envelope-function model, representing an electron with $k_\parallel = 0$ and $E - E_C = 21.7020$ meV[34]. Each of the four peaks in each tight-binding curve is associated with X-valley states in the barrier. There is a downward shift of the $k_\parallel \neq 0$ curve with respect to the $k_\parallel = 0$ case because the AlAs X-valley transverse effective mass is greater than the GaAs $\Gamma$-valley mass. (Recall from Sec. 3.3 that in the nearest-neighbor-only sp$^3$s* tight-binding model, the X-valley transverse mass is in fact infinite, regardless of the values of the matrix elements.) Therefore, in the barrier at a given bias, the electron with $k_\parallel \neq 0$ will be in a superposition of states of greater z-velocity than the electron with $k_\parallel = 0$. (In effective-mass language, when the electron has a larger transverse mass in the barrier than in the emitter and $k_\parallel$ is conserved, more kinetic energy will go toward the z-velocity than in the case $k_\parallel = 0$.) Thus, the resonance should occur at a lower bias. Note that there will still be a downward shift in a model which correctly reproduces the X-valley transverse mass, but the shift will of course be smaller.

We have seen that transmission in single AlAs barriers via X-valley AlAs states is important, even for devices with relatively thin barriers, and that it dominates in devices with thick barriers. These simulations indicate that in a realistic experiment it will not be possible to observe the sharp transition from $\Gamma$-like to X-like tunneling predicted in Ref. 31. Also, they indicate that in a system under bias Fano resonances can persist even in thick barrier systems up to fairly high biases. Both of these latter results differ from what one would expect from Ref. 31 and are directly attributable to the inclusion of space-charge regions in our model. Finally, we have shown that nonzero $k_\parallel$ results in a downward shift of the transmission graph due to the disparity between the GaAs $\Gamma$-valley effective-mass and AlAs X-valley transverse effective masses.
3.7 THE SINGLE INTEGRAL APPROXIMATION

In sec. 3.5.2, we noted that due to computational limitations we were forced to approximate the current-density triple integral by a single integral. We shall now briefly examine the validity of that replacement. Equation (3.133) is the critical step, in which we assume that the transmission coefficient is insensitive to $k_{\parallel}$. However, we have seen in sec. 3.6 that this is clearly not the case. In fig. 3.16 we saw that for single-barrier diodes the transmission-versus-bias graph for an electron with $k_{\text{inc}} = (k_{\parallel}, k_{z,\text{inc}})$, $k_{\parallel} \neq 0$ was shifted downward in bias relative to that for an electron with $k_{\text{inc}} = (0, k_{z,\text{inc}})$, due to the difference between the transverse effective masses in the barrier and bulk regions. Since this effect is purely a function of the materials from which the device is fabricated, it should show up in other structures as well. This is indeed the case, as shown in fig. 3.17, for a double-barrier RTD having 6 monolayer AlAs barriers and an 18 monolayer GaAs well, with 35 monolayer undoped emitter and collector spacers. As in fig. 3.16, the two curves are rather similar in shape, but the curve for the electron with $k_{\parallel} \neq 0$ is clearly shifted downward in bias relative to that for the electron with $k_{\parallel} = 0$.

The presence of the shifts downward in bias discussed above makes it clear that the approximation (3.133) is not especially good. While lack of a sufficiently fast computer prevents us from quantifying the error introduced by the single-integral approximation, we can, with the aid of figs. 3.16 and 3.17, draw some qualitative conclusions. In the first place, the shifts downward in bias imply that the current peaks calculated with the triple integral (3.130) will occur at lower biases than those calculated with the single integral (3.134). More importantly, the current peaks will tend to be lower and broader than when calculated with the single integral approximation. (The occurrence of the broadening is readily seen if one imagines building up the I-V curve by superimposing a series of appropriately weighted transmission curves for various $E$ and $k_{\parallel}$. Note with reference to figs. 3.16 and 3.17 that in both cases all electrons have $E$ well below $\mu_e$ and so have equal weights.) Thus, the correct expression (3.130) should predict more realistic peak current densities and peak-to-valley ratios than the approximate one (3.134).
Figure 3.17: Transmission versus applied bias for a double barrier RTD having 6 monolayer AlAs barriers, an 18 monolayer GaAs well, and 35 monolayer undoped emitter and collector spacers. The emitter and collector are both doped to $N_D = 10^{18}$ cm$^{-3}$, and $T = 77$K. The solid curve is for an electron with $k_{\text{inc}} = (0, 0, 0.02)$ Å$^{-1}$ ($E - E_C = 21.7020$ meV) while the dashed curve is for an electron with $k_{\text{inc}} = (0.01, 0.01, 0.02)$ Å$^{-1}$ ($E - E_C = 32.1379$ meV).

3.8 CRITICISMS OF THE PRESENT TIGHT-BINDING MODEL

While the tight-binding model presented here constitutes a great improvement over the more usual envelope-function approach, it is still far from perfect, for it does not include many scattering mechanisms which are present in actual devices, it only crudely estimates the I-V characteristics because of insufficient computing power, and the underlying tight-binding Hamiltonian is quite simply incomplete. The most glaring scattering mechanism omitted is the electron-phonon interaction. It is absolutely necessary
to include this in a model if one is to correctly simulate the room-temperature I-V characteristics of an RTD. For this reason, we have restricted our studies to devices at 77K. While simulating devices at this temperature reduces the errors resulting from the neglect of the electron-phonon interaction, it does not eliminate them. Spontaneous phonon emission will occur with the same probability at any temperature, and should be important for electrons in the collector space-charge region and perhaps in the well, too.

While inclusion of the electron-phonon interaction would doubtless improve the model, this enhancement presents a very difficult problem both physically and computationally. There are, however, other changes which can be made with far less difficulty. Within the context of the present model, the current should be calculated more accurately (with the triple integral), as we noted in sec. 3.7. Additionally, the underlying tight-binding Hamiltonian could be expanded to allow an accurate fit of the X-valley transverse mass. One way this could be accomplished is by including second-near neighbor interactions in the sp$^3$s* model. An alternative (and better) approach is to retain only same-site and nearest-neighbor matrix elements but to include more orbitals in the basis [35]. Finally, to properly treat the valence bands, the spin-orbit interaction would need to be included as well.

We have not attempted any of the above enhancements since we are currently at the limits of our computer capabilities, at least insofar as RTD current-voltage calculations are concerned. For the purposes of calculating energy levels in superlattices and quantum wells (see the Appendix), the improvements to the underlying tight-binding Hamiltonian (inclusion of the spin-orbit interaction and expansion of the basis) are certainly possible with our present computers and indeed merit serious consideration. Addition of the electron-phonon interaction represents such a significant change in the model that we cannot at the present quantitatively estimate the extra computer power needed. We can only be certain that it will be much greater than that needed for any of the other enhancements.
3.9 CONCLUSIONS

In this chapter we have applied an empirical tight-binding model to the study of electron tunneling in heterostructures. In the process, we have developed the first numerically stable tight-binding transfer matrix method and have demonstrated its relationship to the bulk Hamiltonian. In comparing tight-binding and envelope-function calculations of tunneling currents and transmission coefficients, we have seen that the former model predicts more closely spaced peaks and higher peak-to-valley ratios than does the latter. We have seen that transmission coefficients calculated with the tight-binding model exhibit sharp resonance/antiresonance pairs due to the interaction between Γ- and X-valley states in the AlAs region(s), and have investigated the contributions of states from these valleys to the total transmission. Additionally, taking advantage of the fact that the tight-binding model naturally incorporates \( k_\parallel \), we have briefly investigated tunneling of electrons with differing \( k_\parallel \) (but identical \( k_{z,inc} \)) and have seen that due to differences between the transverse effective masses in GaAs and AlAs the transmission coefficient graphs for \( k_\parallel \neq 0 \) are shifted downward in bias relative to those for \( k_\parallel = 0 \).
NOTES FOR CHAPTER 3


[6] For example, define $V$ by:

$$V = \langle za;0|H|xa;0 \rangle.$$ 

Now, $H = R_z(\pi)H R_z^{\dagger}(\pi)$ so that:

$$V = \langle za;0|R_z(\pi)H R_z^{\dagger}(\pi)|xa;0 \rangle.$$ 

But $R_z^{\dagger}(\pi) = R_z(-\pi)$ and $R_z(-\pi)|xa;0 \rangle = -|xa;0 \rangle$ [eq. (3.17)] and $\langle za;0|R_z(\pi)$

$$=[R_z(-\pi)|za;0 \rangle]^{\dagger} = \langle za;0|$$

[eq. (3.18)]. Thus,

$$V = -\langle za;0|H|xa;0 \rangle = -V,$$

which leads to the conclusion that $V = 0$.

[7] This assertion follows from the fact that the atomic orbitals of which the Lowdin functions are composed can be chosen real.

[8] This last assertion follows from arguments analogous to those used in going from eq. (3.50) to (3.51) and Ashcroft and Mermin, Appendix F.


[10] We used the symbolic computation package, Maple (version 4.2), running on an Amiga 2500 microcomputer to show this explicitly. Maple was developed by the Symbolic Computation Group, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada.

[12] This is of course not the case when we are dealing with a superlattice. Nevertheless, the method outlined here is very useful for treating superlattice problems.

[13] "Time-reversal" symmetry (complex-conjugation of the time-independent Schrödinger equation) for a model with no spin-orbit coupling gives:

$$E_n(k||, k_z) = E_n(-k||, -k_z).$$

Since zincblende is invariant under a rotation of $\pi$ about the $z$-axis,

$$E_n(k||, k_z) = E_n(-k||, k_z).$$


$$E_n(k||, k_z) = E_n(k||, -k_z).$$

Taking partial derivatives with respect to $k_z$ shows that the states $E_n(k||, k_z)$ and $E_n(k||, -k_z)$ have $z$-velocities equal in magnitude but opposite in direction.

[14] Jan P. A. van der Wagt, Timothy B. Boykin, and James S. Harris, Jr., to be published.

[15] We consider only those cases in which the transfer matrix, $T$, is not defective. We can see that it can be defective at any point for which the bulk dispersion relation $E_n(k)$ has $\partial E_n/\partial k_z = 0$, for at these points there is but one eigenvalue with unity magnitude for each maximum/minimum of the $E_n(k)$ curve, and, if the transfer matrix is not to be defective, there should be two eigenvectors, one representing a forward-propagating state, one representing a reverse-propagating state. (Note that this state cannot really be said to "propagate" as its velocity is zero.) $T$ is obviously non-Hermitian, so there is no guarantee that two eigenvectors having the same eigenvalue will be linearly independent. Note that as the defects in $T$ appear when there are propagating states with zero velocity, and since any state with zero velocity makes no contribution to the transmission coefficient, we are not particularly interested in these cases anyway.


[17] See, for example, Kenneth V. Rousseau, K. L. Wang, and J. N. Schulman, Appl. Phys. Lett. 54, 1341 (1989). The authors do not specifically state that they have neglected space-charge regions, but as their calculations transfer across at most about 60 monolayers, it is clear that these effects have not been included.


[24] Ashcroft and Mermin, Appendix F.

[25] Ashcroft and Mermin, Appendix E.


[28] It also, of course, includes the effects of the transverse-X valleys, but since they will not be populated even at room temperature, and we assume \( k_\parallel \) conservation, they do not play a role in the calculation.

[29] This inclusion of multi-valley effects has lead some writers to state that the tight-binding model takes into account "\( \Gamma\)-X" or "intervalley scattering". The use of the term "scattering" is very misleading, for it implies that there is a potential or quasi-particle which effects the scattering, which is decidedly not the case. Here, the heterointerface merely disrupts the translational symmetry in \( z \), so that the wavefunction is in general a superposition of various (possibly complex) \( k_z \)-states.

[30] This problem is covered in any good introductory quantum mechanics text or problem book. See, for example, S. Flugge, Practical Quantum Mechanics, New York: Springer-Verlag (1974), problem 25. Although our conclusions are drawn for the even states, they hold for the odd ones as well.


[34] Although most envelope-function models do not incorporate the effects of nonzero \( k_\parallel \) it is a straightforward matter to include them. See, for example, D. J. BenDaniel and C. B. Duke, Phys. Rev. 152, 683 (1966); and Gerald Bastard, Wave Mechanics Applied to Semiconductor Heterostructures, (Les Editions de Physique, Les Ulis, France, 1988), Ch. III, Sec. II.3.

CHAPTER 4

CONCLUSIONS

4.1: SUMMARY

Motivated by a lack of suitable, physically reasonable, models for nanostructures, we have applied tight-binding methods to resonant tunneling diode (RTD) simulation. This method automatically incorporates the bandstructures of the materials from which the RTD is fabricated, thus avoiding many of the more questionable assumptions and practices of the older, more common, envelope-function approaches. In order to carry out the calculation, we developed the first numerically-stable transfer-matrix method. Although we have only discussed this method in the context of heterostructure device modeling, it is generally applicable to any problem stated in terms of a localized orbital basis. With it, transfers across essentially unlimited numbers of layers are possible. Thanks to this improved method, our tight-binding simulations were the first transfer-matrix calculations to include the long space-charge regions which form on either side of the RTD and play a very significant role in its J-V characteristics. In addition, we have published the first RTD J-V characteristics for devices including these space-charge regions calculated with a tight-binding model. By including space-charge effects in a tight-binding model, our calculations have produced more realistic predicted J-V curves than was previously possible.

Through the use of this improved model, we have gained further insight into the nature of resonant tunneling, especially concerning the role the X-valleys of the AlAs barriers play in the process. Our simulations indicate that these valleys should contribute extra resonances which do not occur in models which include only the Γ-valley of AlAs in both the transmission coefficient and J-V curves. We have shown that when the series resistance of an RTD is taken into account, these additional peaks result in a much more
realistic predicted J-V characteristic, since the series resistance "pushes" one peak toward the next. We have also studied the transmission of electrons through single AlAs barriers and have demonstrated that resonances due to the AlAs X-valley states will persist up to fairly high biases, even for fairly thick barriers. Finally, we have briefly examined the role of nonzero $k_{\parallel}$ on the transmission coefficient and have shown that it can cause a significant shift in the transmission curve when the transverse effective mass in the barrier differs from that in the emitter.

4.2: SUGGESTIONS FOR FUTURE WORK

4.2.1: Improvements Within the Context of the Tight-Binding Model

Even within the present model, there is room for improvement. First and foremost is the correct calculation of J-V characteristics. This will require computing a triple integral instead of a single one, necessitating much faster computers than we currently have. Another possible enhancement is the self-consistent solution of the Schrödinger and Poisson equations. This, too, will require much faster computers, particularly for calculating J-V characteristics.

In order to make the simulation more useful, particularly for modeling superlattices and quantum wells, the spin-orbit interaction should be included. In the present model, which does not include spin-orbit, the valence bands are not treated properly. First, there is no split-off band, and second, the symmetry of the bands is not correct. For example, at the $\Gamma$-point, the valence bands are either purely $p_{x^*}$, $p_{y^*}$, $p_{z^*}$-like, whereas the spin-orbit interaction will mix the various $p$-orbitals. In addition, as noted in Ch. 3, the present model cannot correctly reproduce the X-valley transverse mass. In order to remedy this shortcoming, it will be necessary either to expand the basis or to include second-nearest-neighbor interactions in the $s^3 s^*$ model. Making both of these enhancements will require
approximately 64 times the computing speed necessary for the current model, since each will result in doubling the dimension of the transfer matrix, and matrix operations scale as the cube of the dimension. (Specifically, an sp³s* model with second nearest-neighbor interactions and spin-orbit will have a 40x40 transfer matrix instead of the current model's 10x10 matrix. A spin-orbit, nearest-neighbor only model with more orbitals than used in the sp³s* basis should also have a transfer matrix of approximately the same size.) Note, however, that this situation is not as bad as it at first sounds. For quantum wells and superlattices, one is typically interested in finding the eigen-energies instead of calculating J-V characteristics; the former do not require numerical integration whereas the latter do. Since the cost of finding eigen-energies is of the same order of magnitude as that of finding transmission coefficients, which is about 1/120 the (time) cost of a crude estimate of the J-V characteristic, it should be possible to calculate eigenstates of quantum wells and superlattices with a model incorporating the spin-orbit interaction and an expanded basis and nearest-neighbor coupling (or the sp³s* basis with second nearest-neighbor interactions) in about the same time it now takes to compute J-V curves (around 60 hours on a DECstation 3100).

4.2.2 Other Improvements

Perhaps the most important enhancement outside of the context of the elastic model is the inclusion of the electron-phonon interaction. Even at low temperatures, phonon emission can be important, and should therefore be included. For accurate modeling of room-temperature characteristics, we feel that this will be essential. (Indeed, recent experiments have shown that optical phonon scattering plays a significant role in determining the RTD valley current at temperatures above 200 K[1]). Note, however, that as the present model does not properly reproduce the X-valley transverse effective-mass, it will only be worthwhile to add the electron-phonon interaction to one of the expanded models discussed above. The electron-phonon problem is very difficult, and its inclusion
would represent a much greater departure from the present model than would the other enhancements. We thus have no quantitative estimate of the computing power necessary to incorporate the electron-phonon interaction. However, we expect that it will be much greater than that required for the other improvements. Finally, one could also add scattering due to impurities or interface roughness.
NOTES FOR CHAPTER 4

APPENDIX

A.1: BULK TIGHT-BINDING HAMILTONIAN FOR THE \( sp^3s^* \) MODEL

The corrected version of the bulk tight-binding Hamiltonian of Vogl, Hjalmarson and Dow [1] is given below. Note that we have re-ordered the basis. The 10x10 matrix is most conveniently written in terms of four 5x5 blocks:

\[
H(k) = \begin{bmatrix}
H_{aa}(k) & H_{ac}(k) \\
[H_{ac}(k)]^\dagger & H_{cc}(k)
\end{bmatrix}
\]

(A.1)

where the 5x5 block matrices in (A.1) are:

\[
H_{aa}(k) = \begin{bmatrix}
E_{sa,sa}(0,0,0) & 0 & 0 & 0 & 0 \\
0 & E_{pa,pa}(0,0,0) & 0 & 0 & 0 \\
0 & 0 & E_{pa,pa}(0,0,0) & 0 & 0 \\
0 & 0 & 0 & E_{pa,pa}(0,0,0) & 0 \\
0 & 0 & 0 & 0 & E_{s^*a,s^*a}(0,0,0)
\end{bmatrix}
\]

(A.2)

\[
H_{cc}(k) = \begin{bmatrix}
E_{sc,sc}(0,0,0) & 0 & 0 & 0 & 0 \\
0 & E_{pc,pc}(0,0,0) & 0 & 0 & 0 \\
0 & 0 & E_{pc,pc}(0,0,0) & 0 & 0 \\
0 & 0 & 0 & E_{pc,pc}(0,0,0) & 0 \\
0 & 0 & 0 & 0 & E_{s^*c,s^*c}(0,0,0)
\end{bmatrix}
\]

(A.3)

\[
H_{ac}(k) = \begin{bmatrix}
V_{sa,sc} g_0(k) & V_{sa,pc} g_1(k) & V_{sa,pc} g_2(k) & V_{sa,pc} g_3(k) & V_{sa,s^*c} g_0(k) \\
-V_{pa,sc} g_1(k) & V_{x,x} g_0(k) & V_{x,y} g_3(k) & V_{x,y} g_2(k) & -V_{pa,s^*c} g_1(k) \\
-V_{pa,sc} g_2(k) & V_{x,y} g_3(k) & V_{x,x} g_0(k) & V_{x,y} g_1(k) & -V_{pa,s^*c} g_2(k) \\
-V_{pa,sc} g_3(k) & V_{x,x} g_1(k) & V_{x,y} g_3(k) & V_{x,x} g_0(k) & -V_{pa,s^*c} g_3(k) \\
V_{s^*a,sc} g_0(k) & V_{s^*a,pc} g_1(k) & V_{s^*a,pc} g_2(k) & V_{s^*a,pc} g_3(k) & V_{s^*a,s^*c} g_0(k)
\end{bmatrix}
\]

(A.4)
where in (A.4) $V(n\mu, n'\mu') = 4E_{n\mu,n'\mu'}(1/2, 1/2, 1/2)$. The functions $g_i(k)$ are:

\begin{align*}
  g_0(k) &= \cos(k_{x4}) \cos(k_{y4}) \cos(k_{z4}) - i \sin(k_{x4}) \sin(k_{y4}) \sin(k_{z4}) \\
  g_1(k) &= -\cos(k_{x4}) \sin(k_{y4}) \sin(k_{z4}) + i \sin(k_{x4}) \cos(k_{y4}) \cos(k_{z4}) \\
  g_2(k) &= -\sin(k_{x4}) \cos(k_{y4}) \sin(k_{z4}) + i \cos(k_{x4}) \sin(k_{y4}) \cos(k_{z4}) \\
  g_3(k) &= -\sin(k_{x4}) \sin(k_{y4}) \cos(k_{z4}) + i \cos(k_{x4}) \cos(k_{y4}) \sin(k_{z4})
\end{align*}

(A.5)  

(A.6)  

(A.7)  

(A.8)

where $a$ is the conventional unit cell lattice constant.

### A.2: THE TIGHT-BINDING METHOD APPLIED TO SUPERLATTICES

Let the superlattice consist of $N_c$ cells and let there be $N_L$ principal layers in each cell, numbered $l = 0, 1, 2, \ldots, N_{L-1}$. Due to the periodicity of the superlattice, the wavefunction will be a Bloch state:

\[
\Psi_\mathbf{q}(r) = \frac{1}{\gamma N_c} \sum_{j=0}^{N_c-1} \exp[i \mathbf{q} \mathbf{Z}_j] \left\{ \sum_{l=0}^{N_L-1} \sum_{n, \alpha} B_{l,n,\alpha} \phi_{\mathbf{q}l}^{(n,\alpha)}(r - Z_j - z_i) \right\}
\]

(A.9)

where $Z_j = jd = jN_L a$ is a direct lattice vector parallel to $z$, $d$ is the cell width, $a$ is the principal layer width, the B's are expansion coefficients, and the $\phi$'s are Bloch states in the plane of (orbital, atom) type $(n, \alpha)$, centered in layer $l$ of cell $j$, which obey Bloch's theorem in the plane:
\[ \phi_{q||}^{(n,\alpha)}(r + R_{||}) = \exp[i \cdot q_{||} \cdot R_{||}] \phi_{q||}^{(n,\alpha)}(r) \]  

(A.10)

where \( R_{||} \) is a direct lattice vector in the \( x\)-\( y \) plane. Using (A.10), it is easy to verify that (A.9) is a Bloch state. Let \( R_J = R_{||} + Z_J \) be a direct lattice vector of the superlattice. Then, we have:

\[ \Psi_q(r + R_J) = \frac{\exp[i \cdot q_{||} \cdot R_{||}]}{\sqrt{N_c}} \sum_{j = 0}^{N_c - 1} \exp[i \cdot q_z \cdot Z_j] \left\{ \sum_{l = 0}^{N_e - 1} \sum_{n, \alpha} B_{l,n,\alpha} \phi_{q||}^{(n,\alpha)}(r + Z_j - Z_j - z_l) \right\} \]

(A.11)

where we have already employed (A.10). Since we employ the Born-von Karman boundary conditions, we may convert the sum over cells \( j \) to one over cells \( j' \), where \( Z_{j'} = Z_j - Z_J \). Doing so, we extract a factor \( \exp[i \cdot q_z \cdot Z_j] \) from each term in the sum and so obtain:

\[ \Psi_q(r + R_J) = \exp[i \cdot q \cdot R_J] \Psi_q(r) \]

(A.12)

which is just Bloch's Theorem.

Written in terms of the same basis functions \( \phi \), the transfer matrix eigenstate for the superlattice is:

\[ \Psi_{q||}^{\text{TRANSFER}}(r) = \frac{1}{\sqrt{N_c}} \sum_{j = 0}^{N_c - 1} \left\{ \sum_{l = 0}^{N_e - 1} \sum_{n, \alpha} C_{l,n,\alpha}^{(j)} \phi_{q||}^{(n,\alpha)}(r - Z_j - z_l) \right\} \]

(A.13)

If (A.13) and (A.9) are to be identical, then, we must have:

\[ C_{l,n,\alpha}^{(j)} = \exp[i \cdot q_z \cdot Z_j] B_{l,n,\alpha} \]

(A.14)
Having found the expression for the $C$'s, we now demonstrate the application of the transfer-matrix method to the superlattice problem. We might first imagine diagonalizing the single-principal layer transfer matrix $T_{0,-1}(E, k_{||})$ to obtain a set of eigenvectors $v_1, v_2, \ldots, v_{N'}$, where $N'$ is the total number of orbitals per layer and using this set as the basis in which to expand the state in layer 0. (Because there are no semi-infinite regions of one material or the other in the superlattice, we must use all states in the expansion.) We might equally well note that due to the periodicity of the structure:

$$T_{NL}(N_{L-1}) (E, k_{||}) = T_{0,-1}(E, k_{||}),$$

(A.15)

so that the expansion states in the layer $N_{L}$ will be the same $v$'s.

This is not, however, the best strategy. Recall from linear algebra that since the eigenvectors of a diagonalizable matrix form a basis and since one basis is as good as any other, we are free to choose the one most convenient for our purposes. This turns out to be the $\phi$'s. Using them, the expansion vectors will be the standard basis $\{e_j\}$, where the column vector $e_j$ has a 1 in the $j$-th row and zeros elsewhere, and we must therefore solve:

$$\sum_{j=1}^{N'} c_j^{(1)} e_j = T_{NL}(N_{L-1}) \cdots T_2, i T_1, 0 \sum_{j=1}^{N} c_j^{(1)} e_j.$$

(A.16)

Now, from (A.14), $C_{l,n,\alpha}^{(1)} = \exp[i q z d] C_{l,n,\alpha}^{(j-1)}$, from which it follows that $c_j^{(1)} = \exp[i q z d] c_j^{(1)}$. Substituting this expression into (A.16), renaming the product of transfer matrices $T_{TOT}$, and rearranging, we have:

$$[T_{TOT} - \exp[i q z d] I] \begin{bmatrix} c_1^{(1)} \\ c_2^{(1)} \\ \vdots \\ c_{N'}^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

(A.17)
Note that if the product of transfer matrices in (A.17) is sufficiently large, it will be necessary to employ the improved transfer matrix method from sec. 3.4.4. One implementation of this method with (A.17) results in a system identical to (3.96) with the following exceptions: (i) it is homogeneous, since it comes from the homogeneous system (A.17); (ii) all coefficients represent vectors in the same basis, so that instead of $b$ we have $c^{(l)}$; and (iii) the matrices $\Delta_{1,1}$ and $\Delta_{NB,1}$ are now given by:

$$\Delta_{1,1} = -l_1$$

(A.18)

$$\Delta_{NB,1} = \exp[i qd] L.$$  

(A.19)

In either case, the superlattice eigen-energies are those energies for which the matrix of the entire linear system is singular.
NOTES FOR APPENDIX