University of Bath

PHD

Relation between band-structure and impact ionization in semiconductors.

Mallinson, James Richard

Award date:
1975

Awarding institution:
University of Bath

Link to publication

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 11. May. 2021

RELATION BETWEEN BAND-STRUCTURE AND IMPACT IONIZATION IN SEMICONDUCTORS

submitted by

James Richard Mallinson

for the degree of Ph.D.
of the University of Bath

1975

Copyright

Attention is drawn to the fact that copyright of this thesis rests with its author. This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

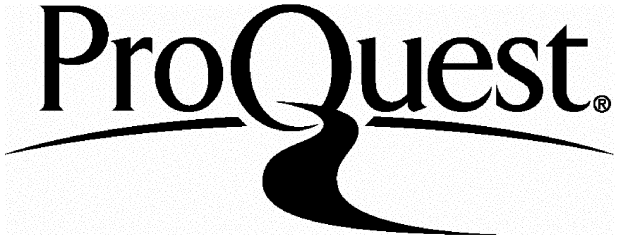
J. R. Mallinson.

ProQuest Number: U413943

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

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



ProQuest U413943

Published by ProQuest LLC(2015). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

An investigation is carried out into the relation between impact ionization threshold positions and the detailed band structure of a number of semiconductors. The band structures of the semiconductors investigated are reproduced, from the data published by previous workers, by the Empirical Pseudopotential Method (EPM). From these detailed band structures, the impact ionization threshold positions are calculated by a method developed in the present work, referred to as the Envelope Method, and are compared with the values calculated by using two different approximate band structure models.

From the EPM, the plane-wave expansions of the wavefunctions of the electron states involved in each impact ionization threshold are then calculated. These wavefunctions are then used to evaluate the sizes of the matrix element (overlap integral) of the coulomb interaction corresponding to each threshold position determined. The relative significance of the threshold positions, particularly the lowest threshold positions, are compared with each other to determine the lowest significant threshold position.

It is shown that it is dangerous to rely on impact ionization threshold values determined by approximate band structure models, and that realistic band structures should be used which are in substantial agreement with experimental data. It is also shown that it is necessary to consider the sizes of the matrix element of the coulomb interaction, since many impact ionization threshold positions have corresponding matrix element sizes which are insignificant.

Acknowledgements

I wish first to express my gratitude to Dr. R.A. Ballinger. I am indebted to him for suggesting the topics investigated, and for many helpful discussions on them, and for his continued encouragement and advice.

I am also indebted to Dr. K.G. Major of Bath University and to Dr. F.W.G. Rose of the Hirst Research Centre, G.E.C., Wembley for their advice and many helpful discussions throughout this investigation. I also wish to thank Mr. C.F. Fisher of Bath University for his helpful suggestions on the geometrical properties of envelopes.

Finally, I wish to acknowledge the financial support provided by the Science Research Council as a CAPS studentship in co-operation with the Hirst Research Centre, G.E.C., Wembley.

CONTENTS

1. INTRODUCTION

Impact ionization by electrons in a semiconductor is the process by which a high energy, or hot, electron in a conduction band can interact with an electron in a valence band to produce an electron-hole pair. If the ionization process does not involve phonons, then both the energy and wavevector of the electron states involved must be conserved. Thus, if a hot electron initially in a state H in the conduction band (see figure 1.1) is to move to a state I in the same conduction band, then the second electron involved must have its initial state somewhere on the broken curve of figure 1.1; that is on the curve obtained by displacing the conduction band by the vector $\vec{HI}$. Since this displaced curve intersects with the valence band at V_1 and V_2, impact ionization is possible in this case. The second electron is promoted from the valence band to the conduction band in the process either from V_1 to C_1 or from V_2 to C_2. Now, instead of one current carrier, the hot electron, there are three current carriers, two conduction electrons and one valence hole. Impact ionization by holes is the similar process involving holes instead of electrons, and producing two valence holes and one conduction electron as current carriers.

Clearly, not all conduction states containing a hot electron can partake in impact ionization (II), since the excess energy needed must be at least equal to the energy gap. An electron which is not hot enough to partake in II may gain sufficient energy from some process to enable it to move into a state in which II is possible. The minimum energy at which an electron can partake in II is the threshold energy, and it is this parameter which is important in semiconductor theories. There is not just one II threshold energy, but many due to multiple conduction and valence bands, and multiple conduction band minima. These other threshold energies, although higher, are no less

important than the lowest threshold energy. Also, for some conduction bands, there may be threshold energies beyond the

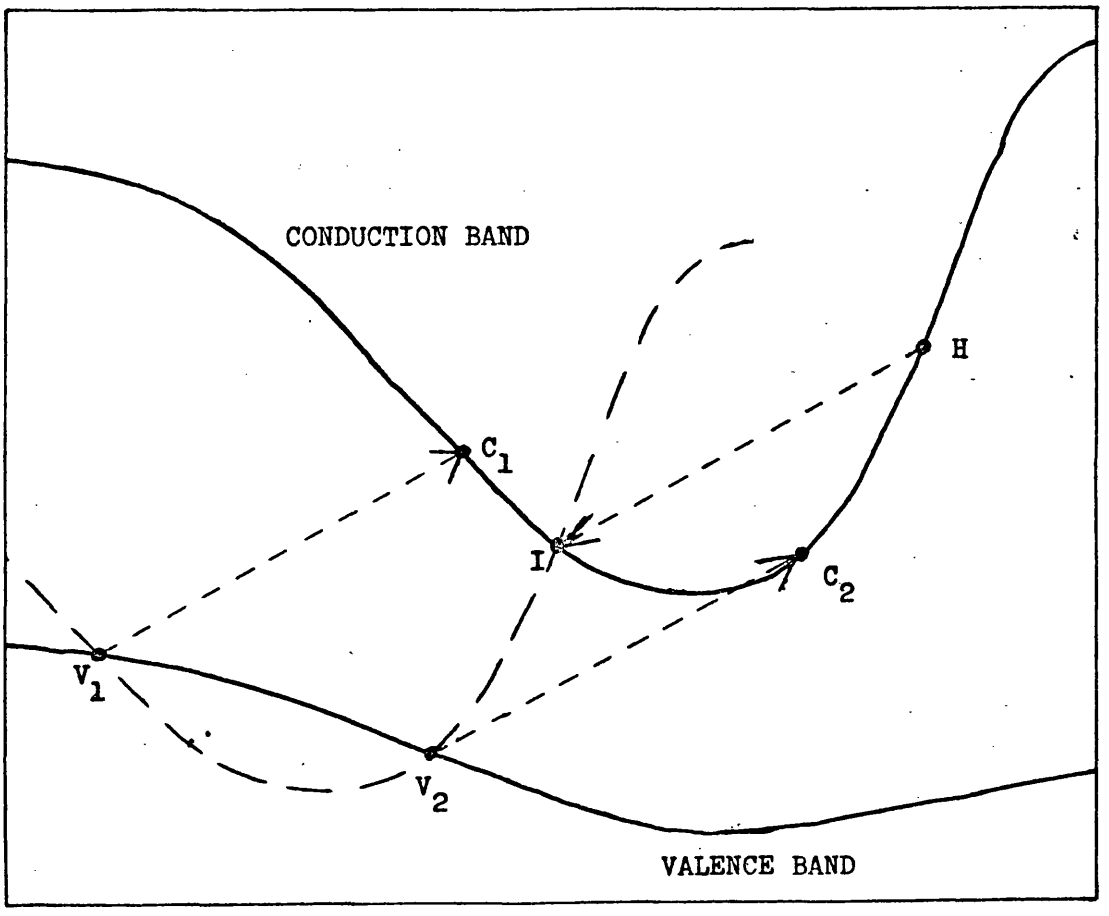


Figure 1.1 A phonon-less impact ionization process showing the conservation of energy and wave-vector of the electron states involved.

first after which II processes are no longer possible, so that finite II windows exist in certain bands.

In the theories of Wolff, Shockley and Baraff [1-3] concerning ionization rates and problems related to p-n junctions, and also in the calculation of ionization rates and avalanche breakdown [4-8], the threshold energy for II was taken as an adjustable parameter. Calculations of this parameter, even by using very simplified band structure models, was not considered. Instead, this parameter was chosen to enable good agreement to be obtained between experimental evidence and

and the related theories.

The first attempt to calculate II threshold energies was made by Tewordt [9] whose treatment was later extended and generalized by Franz [10]. These were not very realistic attempts as they assumed parabolic energy bands in direct gap semiconductors, but were better than chosing the threshold energy arbitrarily. A slightly better attempt was made by Dexter [11], by considering indirect gap semiconductors, which gave rise to further II thresholds due to inter-valley transitions. These methods were used by Antoneik, Beattie and Hodgkinson [12-16] in their investigations of quantum yield, in which II threshold energies are also important, rather than choose II threshold values arbitrarily.

Although various values of II threshold energies were given by the authors of references [9-16], no formulae by which these values were calculated were explicitly quoted. A formula was first quoted by Beattie and Landsberg [17], but was for parabolic energy bands and a direct gap. Hauser [18] and Huldt [19] extended the formula to include indirect band gaps, again for parabolic energy bands. Camphausen and Hearn [20] also tried to extend the formula using a Kane [21] band structure for the conduction bands, but apparently failed to succeed.

The main reasons for using parabolic energy bands instead of realistic band structures was partly due to simplicity of use, but mainly due to the lack of knowledge concerning realistic band structures of semiconductors. It was this lack of knowledge which prevented any calculations of II threshold energies for realistic band structures. As a result of this, a wide range of II threshold energies have, in the past, been used, as can be seen from the review article by Mahadevan et. al [22]. It is this lack of knowledge which prompted this research project. The

main purpose of this work is to investigate the relationship between II threshold energies and the detailed band structures of various semiconductors, considering both the II threshold data and the corresponding probability of the transition.

There are several methods now used to calculate band structures (see for example, G.C. Fletcher 'The Electron Band Theory of Solids' p.67ff. [23]), of which those mainly used are the Augmented Plane-Wave (A.P.W.) method, the Orthogonalized Plane-Wave (O.P.W.) method, their variants which includes the Empirical Pseudopotential (E.P.) method, and the Korringa, Kohn and Rostoker (K.K.R.) variational method. A brief description of these methods is given in the introduction to Chapter 2, outlining the reasons for using the E.P. method in the present work to reproduce the detailed band structures of all the relevant semiconductors. Since this project originated from interest shown by G.E.C., the semiconductors investigated are mainly those in which G.E.C. are interested; namely Silicon, Germanium, 3C Silicon-Carbide and the III-V compounds. The method of computing the band structures by the E.P. method is that as used by Brust [24] and by Cohen and Bergstresser [25], and the band structures of Silicon and Germanium are reproduced from the data of Cohen and Bergstresser. The band structures of Gallium-Phosphide and Gallium-Arsenide are reproduced from the data of Walter and Cohen [26].

Band structure calculations for 3C Silicon-Carbide have recently been performed by Junginger and van Haeringen [27] and by Hemstreet and Fong [28, 29]. The calculations of Hemstreet and Fong include an additional term to the pseudo-Hamiltonian, that of a nonlocal, angular-momentum-dependent, potential term, which at the time of calculation was thought to be significant. The analysis of the E.P. method is given in Chapter 2, including the analysis of the nonlocal potential as described by Hemstreet

and Fong [30].

Before any detailed calculations of 3C Silicon-Carbide were performed, a pilot study was done to determine the effect of the nonlocal term. The results of this pilot study, which are presented at the end of Chapter 2, show that the nonlocal term has a negligible effect, and also revealed a mistake in the work of Hemstreet and Fong. Since the effect of the nonlocal term is negligible, it is not included in the full band structure calculations of 3C Silicon-Carbide.

It is necessary to use realistic band structures in order to obtain accurate calculations of II threshold energies and the associated transition probabilities. The accuracy required cannot be achieved if approximate band structure models, such as those used by Tewordt and Franz, are used. In order to achieve the required accuracy, the method of determing II threshold energies as described by Franz is further generalized to take into consideration the detailed band structures. This method, which has been presented previously [31], is based upon the conservation of energy and wavevector of all four electron states involved in an II process, and is presented in detail in Chapter 3. The method is referred to as the Envelope Method, for which the one dimensional case only is investigated, but the complete generalization to three dimensions is, in principle, straightforward. Also presented in Chapter 3 are formulae for calculating II threshold energies using approximate band structure models. Included is a discussion of the formula described by Camphausen and Hearn.

While the method of calculating II threshold energies was being developed in the present work, Anderson and Crowell [32] were investigating II thresholds, also taking into consideration realistic band structures. It was not known that Anderson and Crowell were investigating II thresholds

until the present work was well advanced. A brief description, together with the advantages and disadvantages, of their method compared with the method presented in this work, is given in the introduction to Chapter 3.

The method of determining II threshold energies presented in this work required analytic expressions of the energy bands being investigated. The band structure calculations do not give analytic expressions of the energy bands, but a set of discrete energy levels. To obtain the analytic expressions of the energy bands, the required expressions have to be fitted through the appropriate set of discrete energy levels. The method of curve fitting, and the form of analytic expressions used, is given in Chapter 4. One dimensional analytic expressions are fitted to the energy levels concerned in each of the three principal symmetry directions considered, the Γ - X, Γ - L and Γ - K - X directions.

Associated with every transition there is a probability that the process will occur. While it is of use to have accurate II threshold values, it is also useful to know the probability associated with each threshold. These probabilities depend basically upon three factors:

(1) The probabilities of the initial states being occupied and the final states being empty.

(2) The transition probability.

(3) The density of states of the energy bands involved in a transition for hot electron energies just above threshold energy.

Calculation of the probability of occupation of the initial states is dependent upon the process by which an electron gains sufficient energy to enable it to partake in an II process. If the excess energy is gained through thermal heating, then the probability of occupation is the Fermi-Dirac probability, which is straightforward to calculate,

being of exponential form which is well known (for details see, for example, P.T. Landsberg 'Solid State Theory : Methods and Applications' p.266ff. [33]). If the excess energy is gained through the effect of an electric field, then the probability of occupation is related to the probability that the electron will travel the required distance without an intervening collision with the lattice. The calculation of this first factor is not considered in the present work.

The second factor, the transition probability, depends on the size of the matrix element of the coulomb interaction between the states involved in a transition. Calculations of the matrix elements are made using the theory developed by Beattie and Landsberg [17], which has since been widely used [34-37]. Their theory for calculating the matrix elements of the coulomb interaction is based on the states of the crystal involved in transitions being described by orthonormal, one-electron functions. Only the electrons which partake in a transition are assumed to have their states changed, while all other electrons are assumed to be unaffected. The matrix element is thus obtained as a multiple sum over reciprocal lattice vectors. This theory also gives reasons why Umklapp processes can be considered negligible.

It is on the basis of this theory that other workers have always considered Umklapp processes to be negligible in the past (see for example, reference [16, 18 and 19]). The present investiation reveals that Umklapp processes are not necessarily negligible, and that some Umklapp processes are far more probable than some Normal processes. The analysis of the matrix elements based on the theory of Beattie and Landsberg is presented in Chapter 5. Also presented is the calculation of the coefficients appearing in the quadruple sum over reciprocal lattice vectors, and a brief discussion on the importance of Umklapp processes.

The third factor is based on the number of states, in each energy

band involved in a transition, which are able to partake in an ionization transition for a hot electron just above a threshold energy. No detailed investigations into this factor have been performed, but Dexter [11], assuming parabolic energy bands, stated that the transition probability increases quadratically with increasing energy just above threshold due to this factor. In the past, it was not known whether this factor proves significant in the total probability. In Chapter 6, a simplified calculation of this factor is presented. Near the energies of all the states involved in a transition, parabolic energy bands are assumed, and a formula to calculate the number of states in which the promoted electron, or hole, may lie, which are able to partake in an impact ionization transition is presented, for a hot electron, or hole, just above threshold. It is shown that this factor, for Silicon, is unimportant when it is compared with the differing sizes of the matrix elements.

Computational details of the calculations of the band structures, threshold data and matrix elements are presented in Chapter 7. Detailed results, with discussions, are presented in the following chapters: Silicon in Chapter 8, Germanium in Chapter 9, 3C Silicon-Carbide in Chapter 10, Gallium-Phosphide in Chapter 11 and Gallium-Arsenide in Chapter 12. The wavevector and energy are given for all states involved, followed by the matrix elements and the error in energy conservation associated with each threshold. The ratio between the II threshold energy and the energy band gap is also given, together with the comparable ratios calculated from approximate band structure models.

In the chapters on Silicon and Germanium, results from a preliminary study [31] are also presented. This study investigates the sensitivity of the various II threshold energies to the precise details of the band structure. The band structures considered for Silicon are those of Cohen

and Bergstresser [25] and of Stukel and Euwema [38], and for Germanium are those of Cohen and Bergstresser and of Stukel [39]. These band structures are reproduced, as accurately as possible, from the figures presented by the relevant authors.

The conclusions of this project are presented in Chapter 13, together with some possible ideas for future work in this field.

2. THE EMPIRICAL PSEUDOPOTENTIAL METHOD

2.1 Introduction

To obtain realistic band structures, the Schrödinger Equation has, by some method, to be solved to a reasonable degree of accuracy. The main methods now used to calculate band structures are the Augmented Plane-Wave (A.P.W.) method, the Orthogonalized Plane-Wave (O.P.W.) method, their variants which includes the Empirical Pseudopotential (E.P.) method, and the Korringa, Kohn and Rostoker (K.K.R.) variational method. While all these methods give reasonable accuracy, there are advantages and disadvantages associated with them which makes some of them unsuitable for calculating the band structures of semiconductors.

The K.K.R. method [40, 41] and the A.P.W. method [42] assume that the potential energy of the crystal is spherically symmetrical about the atomic cores within what is termed as muffin-tin (M.T.) spheres, and constant in the region outside these spheres. While this assumption proves to be sufficiently accurate for metals, it is not particularly accurate for semiconductors. For semiconductors with the zinc-blende type lattice, the assumption of non-overlapping M.T. spheres about both atoms in the unit cell [43] results in a relatively large volume of the crystal in which the potential is assumed constant. This relatively large volume of constant potential results in a loss of accuracy, and is why the K.K.R. and A.P.W. methods are not used for semiconductors. Attempts are being made to make these methods suitable for semiconductors by adopting warped M.T. potentials, but while the convergence of the methods are fairly rapid, the methods also involve perturbation methods, which are computationally difficult to handle.

In the O.P.W. method [44], the wave-functions representing the valence electrons are approximated by a combination of plane-waves, chosen

such that they are orthogonal to the wave-functions representing the core electrons. While this method is simpler to operate than the K.K.R. and A.P.W. methods, it is harder to justify theoretically. For example, it is not at all clear that the orthogonalized plane-waves are the correct wave-functions for the valence electrons.

In the E.P. method, the complete wave-function is that of the O.P.W. method, but written in a slightly different form. By substituting this wave-function into the Schrödinger Equation, it can be reduced to a combination of plane-waves. In doing this an extra term is introduced into the Schrödinger Equation which can be regarded as an extra potential added to the crystal potential. The combination of the two potentials forms a weak, slowly varying potential, referred to as a pseudo-potential, for which only a few terms in the expansion of the wave-function are needed.

Due to the drawbacks of the K.K.R, A.P.W. and O.P.W. methods, the E.P. method is used in the present work to calculate the band structures of all the relevant semiconductors. The E.P. method, which is based upon the O.P.W. method proposed by Herring [44], was introduced by Phillips and Kleinman [45-47] and has since been developed by various workers [48, 49]. Accurate band structure calculations for Silicon, Germanium and the III-V compounds have recently been obtained using this method [24-26, 50-52], and some of these band structures have been reproduced in the present work. The potential used by the above workers is spherically symmetric about the atomic cores, and the analysis of the method of determining the matrix elements of the secular equations is presented in the next section.

In the calculations of the band structure of 3C Silicon-Carbide by Hemstreet and Fong [28, 29], the E.P. method is modified to account for the lack of cancellation of the full crystal potential for p-valence

states in the carbon cores. This modification takes the form of a nonlocal, angular-momentum-dependent, repulsive potential, which is added to the local, spherically symmetric potential, and the analysis of this nonlocal term is presented in §2.3. The effect of this term is thought to have a significant effect on the calculations of 3C Silicon-Carbide [28, 29], since it has a significant effect on the calculations of Diamond [30] and of Potassium [53]. A pilot study in the present work shows that the effect is negligible, and also reveals a mistake in previous calculations of 3C Silicon-Carbide [28, 29]. The results of this pilot study are presented in §2.7, and consequently the nonlocal term is not included in the full band structure calculations.

Once the matrix elements of the secular equations have been determined, the eigenvalues and corresponding eigenvectors then have to be calculated. To obtain the degree of accuracy required, the size of the matrix will be very large. As electronic digital computers are used to evaluate the eigenvalues and corresponding eigenvectors, the size of the matrix is restricted by the capacity of the computer used, and also by the time consumption. Taking these factors into consideration, it is necessary to reduce the size of the matrix while retaining the accuracy obtained by the large matrix. This is done by means of a form of perturbation theory introduced by Löwdin [54] and used frequently in band structure calculations [24-26, 28-30]. The method is presented in §2.4, which includes the method of retrieving the coefficients of the eigenvectors of the original matrix.

Several methods of calculating the eigenvalues and corresponding eigenvectors of the secular equations are available, and the method used in this work is presented in §2.5. While sections 2.2 to 2.4 are completely general, particular forms have to be chosen to reproduce the band structure calculations of previous workers. These forms,

together with some other computational considerations are presented in §2.6.

As calculations of accurate band structures have been carried out previously, and the method used in this work is a reproduction of the efforts of previous workers, no account of the problems of convergence or of accuracy at this stage of the calculations are considered. The accuracy of the calculations are the same as those of the results being reproduced.

2.2 Local Pseudopotential Analysis

The Schrödinger Equation can be written in the form

$$\nabla^2 \psi + \{E - V(\mathbf{r})\} \psi = 0 \qquad \text{(in atomic units)} \qquad 2.2.1$$

where ψ are the wave-functions,

∇^2 the kinetic energy,

$V(\mathbf{r})$ the crystal potential

and E the energy levels.

In the O.P.W. method, the wave-function is expanded in terms of plane-waves, from which is subtracted a number of Block sums; that is

$$\psi(\mathbf{r}) = \sum_m a_m \psi_{\mathbf{k}m}(\mathbf{r}) \qquad 2.2.2$$

where $$\psi_{\mathbf{k}m}(\mathbf{r}) = e^{i\mathbf{k}_m \cdot \mathbf{r}} - \sum_s \alpha_{ms} b_{s\mathbf{k}}(\mathbf{r}) \qquad 2.2.3$$

and $\mathbf{k}_m = \mathbf{k} + 2\pi \mathbf{K}_m$, $\mathbf{K}_m$ a reciprocal lattice vector.

If $\phi_s(\mathbf{r})$ is an atomic wave-function, then the Block sum

$$b_{s\mathbf{k}}(\mathbf{r}) = N^{-\frac{1}{2}} \sum_n e^{i\mathbf{k}.\mathbf{R}_n} \phi_s(\mathbf{r}-\mathbf{R}_n) \qquad 2.2.4$$

where N is the number of unit cells in the crystal and $\mathbf{R}_n$ the positions of the atomic cores.

Combining equations 2.2.3 and 2.2.4, and rewriting them gives

$$\psi(\mathbf{r}) = \Phi(\mathbf{r}) - \sum_s \sum_m a_m \alpha_{ms} b_{s\mathbf{k}}(\mathbf{r}) \qquad 2.2.5$$

where $\Phi(\mathbf{r}) = \sum_m a_m e^{i\mathbf{k}_m \cdot \mathbf{r}}$. 2.2.6

If $\psi(\mathbf{r})$ satisfies the Schrödinger Equation, then

$$\nabla^2\psi(\mathbf{r}) + \left\{ E - V(\mathbf{r}) \right\} \psi(\mathbf{r}) = 0 \quad . \qquad 2.2.7$$

Using the result that

$$\nabla^2\phi_s + \left\{ E_s - V(\mathbf{r}) \right\} \phi_s = 0$$

then, after some algebra, equation 2.2.1 can be written as

$$\nabla^2\Phi + \left\{ E - V(\mathbf{r}) - V_R(\mathbf{r},\mathbf{k}) \right\} \Phi = 0 \qquad 2.2.8$$

where $V_R(\mathbf{r},\mathbf{k})\ \Phi = \sum_s (E-E_s) b_{s\mathbf{k}}(\mathbf{r}) \int \Phi(\mathbf{r}')\ b^*_{s\mathbf{k}}(\mathbf{r}')\ d\mathbf{r}'$. 2.2.9

The quantity $V_R(\mathbf{r})$ can be regarded as an extra, positive potential added to the crystal potential $V(\mathbf{r})$. It thus reduces the potential $V(\mathbf{r})$, and consequently, in the region where $V(\mathbf{r})$ is rapidly varying, the cancellation is almost complete. This leaves a small, slowly varying potential

$$V_p(\mathbf{r},\mathbf{k}) = V(\mathbf{r}) + V_R(\mathbf{r},\mathbf{k}) \qquad 2.2.10$$

which is known as the pseudo-potential.

This potential can be expanded in terms of a sum of local, spherically symmetric terms $V_L(\mathbf{r})$ plus a nonlocal, angular-momentum-dependent term $V_{NL}(\mathbf{r})$; that is

$$V_p(\mathbf{r},\mathbf{k}) = V_L(\mathbf{r}) + V_{NL}(\mathbf{r}) \qquad 2.2.11$$

where $V_L(\mathbf{r}) = \sum_{j=1}^{N} \sum_{i=1}^{L} v_i (\mathbf{r} - \mathbf{R}_j - \mathbf{T}_i)$. 2.2.12

The sum of local terms $r_i(\mathbf{r})$ is over the number of atoms per unit cell, L, and over the number of unit cells in the crystal, N. $\mathbf{R}_j + \mathbf{T}_i$ is the position of the i^{TH} atom (of L) in the j^{TH} unit cell (of N).

Equation 2.2.8 can now be written

$$\nabla^2 \sum_m a_m e^{i\mathbf{k}_m \cdot \mathbf{r}} + \left\{ E - V_L(\mathbf{r}) - V_{NL}(\mathbf{r}) \right\} \sum_m a_m e^{i\mathbf{k}_m \cdot \mathbf{r}} = 0 \ . \qquad 2.2.13$$

The nonlocal potential term, $V_{NL}(\mathbf{r})$, is not considered here, as it is analysed in the next section.

Consider then, the kinetic energy term and the local potential term. Let the crystal have volume Ω, and a unit cell of the crystal have volume Ω_o. Consider the matrix element of the secular equations between the plane-waves with reciprocal lattice vectors $\mathbf{k}_n = \mathbf{k} + \mathbf{K}_n$ and $\mathbf{k}_m = \mathbf{k} + \mathbf{K}_m$, and denote all terms between these reciprocal lattice vectors by the suffices nm.

The matrix element of the secular equations for the kinetic energy term of 2.2.13 is

$$T_{nm} = \langle \mathbf{k} + \mathbf{K}_n \left| \nabla^2 \right| \mathbf{k} + \mathbf{K}_m \rangle$$

which becomes, after performing the differentiation and using the orthogonality property of the plane-waves,

$$T_{nm} = \left| \mathbf{k} + \mathbf{K}_m \right|^2 \delta_{nm} \qquad 2.2.14$$

where δ_{nm} is the Kronecker delta.

Consider now the local potential of equation 2.2.12. This can be expanded in terms of reciprocal lattice vectors, $\mathbf{K}$, by taking the Fourier Transform, that is

$$v_i(\mathbf{K}) = \frac{N.L}{\Omega} \int_\Omega v_i(\mathbf{r}) \, e^{-i\mathbf{K}.\mathbf{r}} \, d\mathbf{r} \ . \qquad 2.2.15$$

Taking the unit cell in which $\mathbf{R}_j = 0$, this can be written

$$v_i(\mathbf{K}) = \frac{L}{\Omega_o} \int_\Omega v_i(\mathbf{r} - \mathbf{T}_i) \, e^{-i\mathbf{K}.(\mathbf{r} - \mathbf{T}_i)} \, d\mathbf{r} \, e^{-i\mathbf{K}.\mathbf{T}_i}$$

and by taking the inverse Fourier Transform,

$$V_L(\mathbf{r}) = \sum_{\mathbf{K}} \left\{ \frac{1}{L} \sum_{i=1}^{L} v_i(\mathbf{K}) \, e^{-i\mathbf{K}.\mathbf{T}_i} \right\} e^{i\mathbf{K}.\mathbf{r}} \ . \qquad 2.2.16$$

Thus the matrix element of the potential of 2.2.16 is

$$V_{L_{nm}} = \langle \mathbf{k} + \mathbf{K}_n | V_L(\mathbf{r}) | \mathbf{k} + \mathbf{K}_m \rangle$$

which, by substituting 2.2.16 for $V_L(\mathbf{r})$, interchanging the order of summation and integration, and by using the orthogonality property of the plane-waves, gives

$$V_{L_{nm}} = \frac{1}{L} \sum_{i=1}^{L} v_i(\mathbf{K}_n - \mathbf{K}_m) \, e^{-i(\mathbf{K}_n - \mathbf{K}_m)\cdot \mathbf{T}_i} \qquad 2.2.17.$$

Combining equations 2.2.14 and 2.2.17 gives the matrix element of the secular equation for the local pseudopotential, namely

$$H_{L_{nm}} = |\mathbf{k}+\mathbf{K}_m|^2 \, \delta_{nm} + \frac{1}{L} \sum_{i=1}^{L} v_i(\mathbf{K}_n - \mathbf{K}_m) \, e^{-i(\mathbf{K}_n - \mathbf{K}_m)\cdot \mathbf{T}_i} \qquad 2.2.18.$$

The form is left completely general here, but is expanded in §2.6 for the particular case used in this work.

2.3 Nonlocal Pseudopotential Analysis

The pseudopotential can be written, as was discribed in the previous section, as a sum of local, spherically symmetric potentials plus a nonlocal, angular-momentum-dependent potential, namely

$$V_p(\mathbf{r},\mathbf{k}) = V_L(\mathbf{r}) + V_{NL}(\mathbf{r}) \quad .$$

Since the nonlocal term is assumed to operate only on p states in the core regions, it is chosen to have the form

$$V_{NL}(\mathbf{r}) = \sum_{\ell} \left\{ \sum_{i=1}^{L} \hat{P}_\ell^* \, U_\ell(|\mathbf{r}-\mathbf{T}_i|) \, \hat{P}_\ell \right\} . \qquad 2.3.1.$$

The projection operator $\hat{P}_\ell$ and its hermitian conjugate $\hat{P}_\ell^*$ operate only on the ℓ^{TH} spherical harmonic component around $\mathbf{T}_i$. $U_\ell(|\mathbf{r}-\mathbf{T}_i|)$ is the corresponding core potential associated with the ℓ^{TH} spherical harmonic, and is assumed to be spherically symmetric about the atom centred at $\mathbf{T}_i$.

The matrix element of the secular equations for this nonlocal potential term is then

$$V_{NL_{nm}} = \langle \mathbf{k} + \mathbf{K}_n | V_{NL}(\mathbf{r}) | \mathbf{k} + \mathbf{K}_m \rangle$$

$$= \frac{1}{\Omega_o} \int_{\Omega_o} e^{-i(\mathbf{k}+\mathbf{K}_n).\mathbf{r}} \left\{ \sum_{\ell} \sum_{i=1}^{L} \hat{P}^*_{\ell} U_{\ell}(|\mathbf{r} - \mathbf{T}_i|) \hat{P}_{\ell} \right\} e^{i(\mathbf{k}+\mathbf{K}_m).\mathbf{r}} \, d\mathbf{r}$$

$$= \frac{1}{\Omega_o} \sum_{\ell} \int_{\Omega_o} \sum_{i=1}^{L} \left\{ \hat{P}_{\ell} e^{i(\mathbf{k}+\mathbf{K}_n).(\mathbf{r}-\mathbf{T}_i)} \right\}^* e^{-i(\mathbf{k}+\mathbf{K}_n).\mathbf{T}_i} U_{\ell}(|\mathbf{r} - \mathbf{T}_i|)$$

$$\times \left\{ \hat{P}_{\ell} \, e^{i(\mathbf{k}+\mathbf{K}_m).(\mathbf{r}-\mathbf{T}_i)} \right\} e^{i(\mathbf{k}+\mathbf{K}_m).\mathbf{T}_i} \, d\mathbf{r}$$

and hence

$$V_{NL_{nm}} = \frac{1}{\Omega_o} \sum_{\ell} \sum_{i=1} e^{-i(\mathbf{K}_n - \mathbf{K}_m).\mathbf{T}_i} \int_{\Omega_o} \left\{ \hat{P}_{\ell} \, e^{i(\mathbf{k}+\mathbf{K}_n).\mathbf{r}_i} \right\}^* U_{\ell}(|\mathbf{r}_i|)$$

$$\times \left\{ \hat{P}_{\ell} \, e^{i(\mathbf{k}+\mathbf{K}_m).\mathbf{r}_i} \right\} d\mathbf{r}_i \qquad 2.3.2.$$

where $\mathbf{r}_i = \mathbf{r} - \mathbf{T}_i$, and the spherically symmetric potentials are non-overlapping.

Now expanding the exponentials as an infinite sum of Bessel Functions, namely

$$e^{i(\mathbf{k}+\mathbf{K}).\mathbf{r}} = 4\pi \sum_{\ell=0}^{\infty} \sum_{j=-\ell}^{\ell} (i)^j \, j_{\ell}(|\mathbf{k}+\mathbf{K}|r) Y^*_{\ell j}(\theta_{\mathbf{K}}, \phi_{\mathbf{K}}) Y_{\ell j}(\theta, \phi) \qquad 2.3.3.$$

where $j_{\ell}(r)$ is the spherical Bessel function of order ℓ and $Y_{\ell j}(\theta, \phi)$ is the spherical surface harmonic,

then $$\hat{P}_{\ell} \, e^{i(\mathbf{k}+\mathbf{K}).\mathbf{r}} = 4\pi \sum_{j=-\ell}^{\ell} (i)^j \, j_{\ell}(|\mathbf{k}+\mathbf{K}|r) Y^*_{\ell j}(\theta_{\mathbf{K}}, \phi_{\mathbf{K}}) Y_{\ell j}(\theta, \phi) \qquad 2.3.4.$$

and so, substituting 2.3.4 into 2.3.2 gives

$$V_{NL_{nm}} = \frac{1}{\Omega_o} \sum_{\ell} \sum_{i=1}^{L} e^{-i(\mathbf{K}_n - \mathbf{K}_m).\mathbf{T}_i} \int_{\Omega_o}$$

$$\times \left\{ 4\pi \sum_{j=-\ell}^{\ell} (-i)^j \, j_{\ell}(|\mathbf{k}+\mathbf{K}_n| r_i) Y_{\ell j}(\theta_{\mathbf{K}_n}, \phi_{\mathbf{K}_n}) Y^*_{\ell j}(\theta, \phi) \right\} U_{\ell}(|\mathbf{r}_i|)$$

$$\times \left\{ 4\pi \sum_{j=-\ell}^{\ell} (i)^j \, j_{\ell}(|\mathbf{k}+\mathbf{K}_m| r_i) Y^*_{\ell j}(\theta_{\mathbf{K}_m}, \phi_{\mathbf{K}_m}) Y_{\ell j}(\theta, \phi) \right\} d\mathbf{r}_i \qquad 2.3.5$$

Separating the integral over the unit cell centred on $\mathbf{T}_i$ into its

radial and angular component, and rearranging the terms slightly gives

$$V_{NL_{nm}} = \frac{(4\pi)^2}{\Omega_o} \sum_{\ell} \sum_{i=1}^{L} e^{-i(K_n-K_m).T_i} \int_0^{R_c} j_\ell(|k+K_n|r_i) j_\ell(|k+K_m|r_i) U_\ell(|r_i|) r_i^2 \, dr_i$$

$$\times \sum_{j=-\ell}^{\ell} \sum_{k=-\ell}^{\ell} Y_{\ell j}(\theta_{K_n},\phi_{K_n}) Y^*_{\ell k}(\theta_{K_m},\phi_{K_m}) \int_0^{2\pi}\int_0^{\pi} Y^*_{\ell j}(\theta,\phi) Y_{\ell k}(\theta,\phi)$$

$$\times \sin\theta d\theta d\phi \qquad 2.3.6.$$

where R_c is the maximum radius around the atoms for which the potential is non-zero.

Two properties of spherical surface harmonics can now be used, namely

$$\int_0^{2\pi}\int_0^{\pi} Y^*_{\ell j}(\theta,\phi) Y_{\ell' j'}(\theta,\phi) \sin\theta d\theta d\phi = \delta_{\ell\ell'}\delta_{jj'} \qquad 2.3.7.$$

and $$\sum_{j=-\ell}^{\ell} Y^*_{\ell j}(\theta_K,\phi_K) Y_{\ell j}(\theta_{K'},\phi_{K'}) = \left(\frac{2\ell+1}{4\pi}\right)^{\frac{1}{2}} Y_{\ell o}(\theta_{KK'})$$

$$= \left(\frac{2\ell+1}{4\pi}\right) P_\ell(\cos\theta_{KK'}) \qquad 2.3.8.$$

where $P_\ell(\cos\theta)$ is the Legendre Polynomial of order ℓ and $\theta_{KK'}$ is the angle between the reciprocal lattice vectors $k+K$ and $k+K'$. Using 2.3.7. in equation 2.3.6 reduces the double sum, over j and k, to a single sum, over j, and then using 2.3.8. eliminates this single sum to give, after slight rearrangement,

$$V_{NL_{nm}} = \frac{4\pi}{\Omega_o} \sum_{\ell} (2\ell+1) P_\ell(\cos\theta_{K_nK_m}) \sum_{i=1}^{L} e^{-i(K_n-K_m).T_i} \times$$

$$\times \int_0^{R_c} j_\ell(|k+K_n|r_i) j_\ell(|k+K_m|r_i) U_\ell(|r_i|) r_i^2 dr_i \qquad 2.3.9$$

Again, the form is left completely general here, but is expanded in §2.6 for the particular case used in this work. Combining equation

2.3.9 with equation 2.2.18 gives the complete matrix element of the secular equations, namely

$$H_{nm} = |k+K_m|^2 \delta_{nm} + \frac{1}{L} \sum_{i=1}^{L} r_i(K_n-K_m)\, e^{-i(K_n-K_m).T_i} +$$

$$+ \frac{4\pi}{\Omega_o} \sum_{\ell} (2\ell+1)\, P_\ell(\cos\theta_{K_n K_m}) \sum_{i=1}^{L} e^{-i(K_n-K_m).T_i} \times$$

$$\times \int_o^{R_c} j_\ell(|k+K_n|r_i) j_\ell(|k+K_m|r_i) U_\ell(|r_i|) r_i^2\, dr_i \qquad 2.3.10$$

2.4 Perturbation Theory

The perturbation theory due to Löwdin [54] is based on the variational principle. It is assumed that the wave-function can be formed by a linear combination of known, orthonormal functions

$$\psi = \sum_{n=1}^{N} C_n \psi_n^{(o)} \quad .$$

The matrix element of the total hermitian operator H between the states represented by the wave-functions $\psi_n^{(o)}$ and $\psi_m^{(o)}$ is

$$H_{nm} = \int \psi_n^{(o)*} H \psi_m^{(o)}\, d\tau \quad .$$

The coefficients C_m can be determined by the variational principle, which gives the system of linear equations

$$\sum_{m=1}^{N} (H_{nm} - E\delta_{nm})\, C_m = 0 \quad . \qquad 2.4.1.$$

The condition for the existence of a non-trivial solution is that the determinant of the matrix $(H_{nm}-E\delta_{nm})$ is zero. To satisfy this condition, the values of E must be the eigenvalues of the matrix H_{nm}, and the coefficients are then the corresponding eigenvectors.

Now, let it be assumed that the linear combination of functions

forming the wave-function can be divided into two distinct classes A and B. Let the main interest lie in the states in class A, and attempt to derive a formula by treating the states in class B as a perturbation. Equation 2.4.1. can be written as

$$(E-H_{nn})C_n = \sum_m^A H'_{nm} C_m + \sum_m^B H'_{nm} C_m \qquad 2.4.2$$

where $H'_{nm} = H_{nm}(1-\delta_{nm})$

and the first sum being over the states in class A and the second sum being over the states of class B.

Using the notation

$$h_{nm} = H_{nm}/(E-H_{nn}) \qquad 2.4.3$$

equation 2.4.2 can be written as

$$C_n = \sum_m^A h'_{nm} C_m + \sum_m^B h'_{nm} C_m \quad . \qquad 2.4.4$$

The states in class B can now be eliminated by a process of iteration. The coefficients C_m occurring in the sum over the states in class B in equation 2.4.4 can be expressed by equation 2.4.4 itself, giving

$$C_n = \sum_m^A h'_{nm} C_m + \sum_m^B h'_{nm} \left\{ \sum_i^A h'_{mi} C_i + \sum_i^B h'_{mi} C_i \right\}$$

$$= \sum_m^A h'_{nm} C_m + \sum_i^A \sum_m^B h'_{nm} h'_{mi} C_i + \sum_m^B \sum_i^B h'_{nm} h'_{mi} C_i \quad . \qquad 2.4.5$$

Repeating this process for equation 2.4.5, and subsequently by repeated use, gives the formal expansion

$$C_n = \sum_m^A h'_{nm} C_m + \sum_m^A \left\{ \sum_i^B h'_{ni} h'_{im} + \sum_{i,j}^B h'_{ni} h'_{ij} h'_{jm} + \ldots \right\} C_m \qquad 2.4.6$$

Using equation 2.4.3, and introducing the notation

$$U_{nm} = H_{nm} + \sum_{i}^{B} \frac{H^1_{ni} H^1_{im}}{E-H_{ii}} + \sum_{i,j}^{B} \frac{H^1_{ni} H^1_{ij} H^1_{jm}}{(E-H_{ii})(E-H_{jj})} + \ldots \qquad 2.4.7$$

equation 2.4.6 becomes

$$C_n = \sum_{m}^{A} \frac{U_{nm}-H_{nm}\delta_{nm}}{E-H_{nn}} C_m \qquad 2.4.8$$

For the two cases of n in A or B, there are the two corresponding basic formulae

$$\sum_{m}^{A} (U_{nm}-E\delta_{nm}) C_m = 0 \qquad \text{for n in A} \qquad 2.4.9$$

$$C_n = \sum_{m}^{A} \frac{U_{nm} C_m}{E-H_{nn}} \qquad \text{for n in B} \qquad 2.4.10$$

The problem of determining the eigenvalues and corresponding eigenvectors of the matrix in equation 2.4.1 is now reduced to determining the eigenvalues and corresponding eigenvectors of the matrix in equation 2.4.9. This now gives fewer eigenvalues and truncated eigenvectors, but the coefficients of the corresponding original eigenvectors can be determined from equation 2.4.10.

2.5 Method of Solution of the Secular Equations

This work is primarily concerned in the investigation of impact ionization threshold data, and not in investigating the calculation of eigenvalues and eigenvectors of matrices. For this reason the method used is chosen purely on grounds of convenience and reliability. There are many methods available for calculating eigenvalues and eigenvectors of matrices, of which most are programmed for use on an electronic digital computer. Some of these methods are for use with general matrices, while a few are for use only with symmetric or hermitian matrices. Since the matrix here is hermitian, one of these latter methods is used, since they are quicker and more accurate than the methods for use with general matrices.

In trying to find a method which has already been programmed for a digital computer, it was discovered that most of these methods are programmed

in computer languages other than the one used here. Since translation is a lengthy procedure, these methods had to be rejected. There are only a few methods available which are programmed in the correct computer language. Of these, some are methods for use with general matrices, and some are methods which do not include the calculation of the eigenvectors, all of which are not suitable.

Of the very few methods remaining, the computer program available for one of the methods is known to be unreliable, which virtually restricted the choice to one method. The method used reduces the symmetric or hermitian matrix into tri-diagonal form by the procedure due to Householder, followed by the QR algorithm for determining the eigenvalues and eigenvectors of a tri-diagonal matrix (see for example, J.H. Wilkinson 'The Algebraic Eigenvalue Problem' pp 290-299, pp 515-521.[55]). All the eigenvalues of a matrix are calculated using this method, and the eigenvectors can be determined with very little extra effort.

However, this method, as programmed for the digital computer, requires that the matrix is symmetric, and not hermitian. This presents no difficulty, as a hermitian matrix can easily be transformed into a symmetric matrix. This is done by writing the hermitian matrix as a sum of its real and imaginary parts, that is

$$H = A + i\,B \qquad 2.5.1$$

where A is real and symmetric, and B is real and skew-symmetric.

If the eigenvalues of H are λ_j with corresponding eigenvectors $\mathbf{W}_j = \mathbf{U}_j + i\mathbf{V}_j$, then

$$H\mathbf{W}_j = \lambda_j \mathbf{W}_j$$

and from 2.5.1

$$(A + i\,B)\,\mathbf{W}_j = \lambda_j\,\mathbf{W}_j \qquad 2.5.2$$

which can be written as

$$A\mathbf{W}_j - B(-i\mathbf{W}_j) = \lambda_j \mathbf{W}_j \qquad 2.5.3$$

Multiplying equation 2.5.2 throughout by (-i) gives

$$(B - i\,A)\mathbf{W}_j = -i\,\lambda_j \mathbf{W}_j$$

which can be written as

$$B\mathbf{W}_j + A(-i\,\mathbf{W}_j) = \lambda_i(-i\,\mathbf{W}_j) \quad . \qquad 2.5.4$$

Thus, combining 2.5.3. and 2.5.4. gives

$$\begin{bmatrix} A & -B \\ B & A \end{bmatrix} \begin{bmatrix} \mathbf{W}_j \\ -i\mathbf{W}_j \end{bmatrix} = \lambda_j \begin{bmatrix} \mathbf{W}_j \\ -i\mathbf{W}_j \end{bmatrix} \qquad 2.5.5$$

This is now a symmetric matrix, twice the size of the hermitian matrix, and thus having twice the number of eigenvalues and eigenvectors, with the eigenvectors containing twice the number of elements. The eigenvalues and eigenvectors occur in pairs of complex conjugates, so instead of one eigenvalue λ_j with corresponding eigenvector $\mathbf{W}_j$, there is now two eigenvalues λ_j and $\bar{\lambda}_j$ with corresponding eigenvectors $\{\mathbf{W}_j, -i\mathbf{W}_j\}$ and $\{\bar{\mathbf{W}}_j, i\bar{\mathbf{W}}_j\}$.

Since the eigenvalues of a symmetric matrix are all real, $\lambda_j = \bar{\lambda}_j$ and the eigenvalues are repeated. The corresponding eigenvectors can be identified by comparing the top half of the vector with the bottom half. If the top half of one vector is (i) times the bottom half, then it corresponds to the first of the repeated eigenvalues. If the eigenvectors determined are real, then since

$$\begin{bmatrix} \mathbf{W}_j \\ -i\mathbf{W}_j \end{bmatrix} = \begin{bmatrix} \mathbf{U}_j + i\mathbf{V}_j \\ \mathbf{V}_j - i\mathbf{U}_j \end{bmatrix}$$

this gives $\{\mathbf{U}_j, \mathbf{V}_j\}$ as the eigenvector, from which the original eigen-

vector is the top half of the vector plus (i) times the bottom half.

2.6 Some Computational Considerations

All the cubic semiconducting materials investigated in the present work are of the diamond or zinc-blende type lattice, which have the Face Centred Cubic structure with two atoms per unit cell. That is, in sections 2.2 and 2.3

$$L = 2 \tag{2.6.1}$$

and $$\Omega_o = \frac{1}{4} a^3 \tag{2.6.2}$$

where a is the length of the unit cube.

If the origin of the co-ordinate system is taken midway between the two atoms, as in reference [24], then

$$\left.\begin{aligned} \boldsymbol{\tau}_1 &= \frac{a}{8}(1,1,1) = \boldsymbol{\tau} \\ \text{and} \quad \boldsymbol{\tau}_2 &= -\boldsymbol{\tau}_1 = -\boldsymbol{\tau} \end{aligned}\right\} \tag{2.6.3}$$

and the potential can be written as a sum of a symmetric plus an antisymmetric potential, namely

$$\left.\begin{aligned} V_s(\mathbf{K}_n-\mathbf{K}_m) &= \tfrac{1}{2}\{v_1(\mathbf{K}_n-\mathbf{K}_m) + v_2(\mathbf{K}_n-\mathbf{K}_m)\} \\ \text{and} \quad V_A(\mathbf{K}_n-\mathbf{K}_m) &= \tfrac{1}{2}\{v_2(\mathbf{K}_n-\mathbf{K}_m) - v_1(\mathbf{K}_n-\mathbf{K}_m)\} \end{aligned}\right\} \tag{2.6.4}$$

In the nonlocal potential term used by Hemstreet and Fong [28,29], the core potential associated with the first spherical harmonic only is allowed to be non-zero, for which the Legendre Polynomial

$$P_1(\cos\theta_{\mathbf{K}_n\mathbf{K}_m}) = \cos\theta_{\mathbf{K}_n\mathbf{K}_m} \tag{2.6.5}$$

and the potential is of the form

$$U(|\mathbf{r}_i|) = \begin{cases} A_i\, r_i\, e^{-\alpha r_i} & \text{for } r_i \leq R_c \\ 0 & \text{for } r_i > R_c \end{cases} \tag{2.6.6}$$

where, for silicon-carbide, $A_1 = A$ and $A_2 = 0$ 2.6.7

Introducing the notation

$$I_{nm} = \int_o^{R_c} j_1(|k+K_n|r) j_1(|k+K_m|r) r^3 \, e^{-\alpha r} dr \qquad 2.6.8$$

and using equations 2.6.1 to 2.6.8 in equation 2.3.10 gives, after rearrangement of the terms,

$$H_{nm} = |k+K_m|^2 \delta_{nm} + \{V_S(K_n-K_m)\cos[(K_n-K_m).t] + iV_A(K_n-K_m)\sin[(K_n-K_m).t]\}$$

$$+ \frac{48\pi}{a^3} AI_{nm} \cos\theta_{K_n K_m} \{\cos[(K_n-K_m).t] - i\sin[(K_n-K_m).t]\} \qquad 2.6.9$$

This is the form of the matrix elements used in the present calculations. The perturbation theory used assumes that the plane-waves, $e^{i(k+K_n).r}$, used in the calculations have been ordered such that for $n > m$, $|k+K_n|^2 \geq |k+K_m|^2$. Writing $K_n = \frac{2\pi}{a} G_n$, the matrix elements representing the states in class A are formed from those reciprocal lattice vectors G_n such that

$$|k + G_n|^2 \leq E_1 \quad .$$

The matrix elements representing the states in class B are formed from those reciprocal lattice vectors G_n such that

$$E_1 < |k+ G_n|^2 \leq E_2 \quad .$$

The effect of all other reciprocal lattice vectors is neglected. The values of E_1 and E_2 are taken as those used by previous workers, which for Silicon, Germanium and the III-V compounds (references [25,26]) the values $E_1 = 7$ and $E_2 = 21$ are taken, and for 3C Silicon-Carbide (references [28,29]) the values $E_1 = 10$ and $E_2 = 27$ are taken.

The perturbation to the matrix elements representing the states in class A takes the form of an infinite series of sums over the matrix elements between the classes A and B. This series is truncated by Brust [24] after the first two terms, to enable the calculations to be feasible. Thus

equation 2.4.7 reduces to

$$U_{nm} = H_{nm} + \sum_{i}^{B} \frac{H'_{ni} H'_{im}}{E-H_{ii}} \qquad 2.6.10$$

which is then used in equations 2.4.9 and 2.4.10 to determine the eigenvalues and the original expansion of the corresponding eigenvectors.

The lowest four eigenvalues represent the valence states, while the higher eigenvalues represent the conduction states. Since the interest is in the valence and lower conduction states, the lowest eight eigenvalues and corresponding eigenvectors only are produced. To enable all the eigenvalues to be calculated at the same time, the eigenvalue dependence of the matrix elements in equation 2.6.10 is removed. If this is not done, the eigenvalues have to be determined individually by an iterative proceedure. The dependence is removed by making the substitutions made by Brust [24], namely;

Off the diagonal:

E is replaced by $\bar{E}$, an average of the lowest eight energy levels at each point in the first Brillouin zone. The value $\bar{E} = 2$ is used, as is used by Brust.

On the diagonal:

E is replaced by $\bar{E} = H_{nn} = |k+K_n|^2$, essentially the kinetic energy of the principal plane-wave in the expansion of the wave-function.

This gives the equations 2.4.9 and 2.4.10 as

$$\sum_{m}^{A}\left\{\left\{H_{nm} + \sum_{i}^{B} \frac{H'_{ni} H'_{im}}{\bar{E}-H_{ii}}\right\} - E\,\delta_{nm}\right\} C_m = 0 \qquad 2.6.11$$

and $$C_n = \sum_{m}^{A}\left\{H_{nm} + \sum_{i}^{B} \frac{H'_{ni} H'_{im}}{\bar{E}-H_{ii}}\right\}\frac{C_m}{E-H_{nn}} \qquad 2.6.12$$

where $$\bar{E} = \begin{cases} 2 & \text{if } n \neq m \\ |k+K_n|^2 & \text{if } n = m \end{cases} . \qquad 2.6.13$$

The first seven of the reciprocal lattice vectors have squared

magnitudes 0, 3, 4, 8, 11, 12 and 16, and only these are allowed to have non-zero potentials. The symmetric structure factors for $|\mathbf{G}|^2 = 4$ and $|\mathbf{G}|^2 = 12$ are zero, and so the corresponding potentials need not be considered. However, to be consistent with the work of Hemstreet, Fong and Cohen [30] in their calculations of diamond, the symmetric structure factor for $|\mathbf{G}|^2 = 12$ is set to unity. The antisymmetric structure factors for $|\mathbf{G}|^2 = 0$, $|\mathbf{G}|^2 = 8$ and $|\mathbf{G}|^2 = 16$ are zero, and again the corresponding potentials need not be considered. The symmetric potential $V_s(|\mathbf{G}|^2 = 0)$ is made zero since it merely adds a constant to all energy levels. Hence there are only five symmetric and four antisymmetric form factors to be considered, namely $V_s(|\mathbf{G}|^2 = 3)$, $V_s(8)$, $V_s(11)$, $V_s(12)$, $V_s(16)$, $V_A(3)$, $V_A(4)$, $V_A(11)$ and $V_A(12)$.

These nine local form factors, together with the three nonlocal parameters, the lattice constant, and the values of E_1 and E_2 are the only parameters required to perform band structure calculations. The nonlocal parameters are A and α, for which the product $A.^{1/\alpha}$ roughly represents the 'strength' of the nonlocal potential, and R_c, the free ion core radius.

All the details presented in this section are included in the computer program written to calculate the band structures of cubic semiconductors with the diamond or zinc-blende type lattice. Further computational details are given in Chapter 7.

2.7 Results of the Pilot Study on the Nonlocal Potential Term

The nonlocal potential term, added to the local potential and kinetic energy terms, as presented in sections 2.3 and 2.6, is

$$V_{NL_{nm}} = \frac{48\pi}{a^3} AI_{nm} \cos\theta_{nm}\{\cos[(\mathbf{K}_n-\mathbf{K}_m).\mathbf{t}] - i \sin[(\mathbf{K}_n-\mathbf{K}_m).\mathbf{t}]\} \qquad 2.7.1$$

where θ_{nm} is the angle between the reciprocal lattice vectors $k + K_n$ and $k + K_m$, and

$$I_{nm} = \int_o^{R_c} j_1(|k+K_n|r) j_1(|k+K_m|r)\ r^3\ e^{-\alpha r}\ dr \qquad 2.7.2$$

where A, α and R_c are the nonlocal parameters.

In this pilot study to determine the effect of the nonlocal potential on the band structure of Silicon-Carbide, the data of Hemstreet and Fong [29] is used.

For 3C SiC, the lattice constant a = 4.35Å, and the local form factors used are; $V_S(|G|^2 = 3) = -0.419$, $V_S(8) = 0.101$, $V_S(11) = 0.118$, $V_A(3) = 0.001$, $V_A(4) = 0.080$ and $V_A(11) = 0.051$, all expressed in Rydbergs, and all other form factors being zero. The nonlocal parameters used are; $R_c = 0.2$Å, approximately equal to the free ion core radius of carbon, A = -0.128 Ryd. and $\alpha = 1.02$Å^{-1}. The number of plane-waves being treated exactly and through perturbation are determined by the values $E_1 = 10$ and $E_2 = 27$, described in section 2.6.

These values are used in the computer program to calculate the energy eigenvalues at selected symmetry points within the first Brillouin zone. For this pilot study, the energy eigenvalues are calculated at the symmetry points Γ, X and L, and the key energy gaps are then determined. These energy gaps are in disagreement with the energy gaps calculated by Hemstreet and Fong, both with and without the effect of the nonlocal potential. The energy gaps calculated here are tabulated, together with those calculated by Hemstreet and Fong, in Table 2.1.

Since there is disagreement in the values where there should be agreement, the energy eigenvalues at Γ, X and L are calculated again, but this time without the nonlocal potential (that is, with A = 0 in equation 2.7.2). The energy gaps determined from the calculations, both

with and without the nonlocal potential, are in very close agreement. The maximum difference between the energy eigenvalues at the three symmetry points considered are; at Γ, 0.000045eV: at X, 0.000045eV: and at L, 0.000044eV. This apparent insignificance of the effect of the nonlocal potential is consequently investigated.

Table 2.1

Key energy gaps of 3C SiC at Γ, X and L points (expressed in eV's)

	$\Gamma_{15}-\Gamma_1$	$\Gamma_{15}-\Gamma_{15}$	L_3-L_1	L_3-L_3	X_5-X_1	X_5-X_3	$\Gamma_{15}-L_1$	$\Gamma_{15}-X_1$
Present calculations	5.95	6.50	6.03	9.19	6.37	9.41	4.39	2.36
Hemstreet and Fong with V_{NL}	5.90	6.47	5.97	9.08	6.13	9.21	4.39	2.33
Hemstreet and Fong Without V_{NL}	5.92	6.49	6.02	9.18	6.37	9.40	4.38	2.35

This is done by investigating the size of the matrix elements, both with and without the nonlocal potential term included. Since the effect of the nonlocal term, as calculated in reference [29], is reported to be greatest at the point X, the energy eigenvalues at that point are investigated. This is done both manually and by using the digital computer. With the values of E_1 and E_2 given above, at the point X, 40 plane-waves are treated exactly, and a further 110 plane-waves are treated through perturbation theory.

All the matrix elements are obtained from the digital computer by slightly modifying the computer program. The size of the integrals occurring in the nonlocal potential terms are also obtained. Since the matrix is of size 40 x 40, all the matrix elements cannot be obtained manually.

Instead, only two matrix elements are fully investigated, one diagonal and one off-diagonal element. The two matrix elements are also fully investigated by using the digital computer. The diagonal matrix element corresponds to the plane-wave having reciprocal lattice vector $(k+K_{17})$, where $k = \frac{2\pi}{a}(1,0,0)$ and $K_{17} = \frac{2\pi}{a}(1, -1, -1)$. The off-diagonal matrix element corresponds to the plane-waves having reciprocal lattice vectors $(k+K_{17})$ and $(k+K_{11})$, where $K_{11} = \frac{2\pi}{a}(-2,2,0)$.

From the computer results, it is seen that all the values of the integral of equation 2.7.2 are small, and are in the range 2.05×10^{-6} Ryd to 1.83×10^{-5} Ryd. When the values of the integral are multiplied by the appropriate constant and structure factors, a few of the nonlocal potential terms become zero. The non-zero terms, of which there are many, have values in the range 4.4×10^{-7} Ryd to 4.3×10^{-6} Ryd. These values are computed to an accuracy of 10 decimal places. The values of the two matrix elements calculated manually are calculated to an accuracy of 8 decimal places, and to within this accuracy are in agreement with the computer calculations. The values of the integral, the nonlocal potential term and the local potential term for both the matrix elements investigated are tabulated in Table 2.2

Table 2.2

Values of Matrix Elements (expressed in Rydbergs)

Plane-Wave numbers, n,m	I_{nm}	$V_{NL_{nm}}$	$H_{L_{nm}}$
n = 17 m = 17	1.156×10^{-5}	-2.711×10^{-5}	3.4217
n = 17 m = 11	1.062×10^{-5}	$(1+i)\ 1.291 \times 10^{-6}$	0.0

The size of the nonlocal potential terms are much too small to have any significant effect on the energy eigenvalues as calculated without the nonlocal potential. Correspondence with Professor Hemstreet

was then entered into, explaining the results of this pilot study, and asking for his reasons for the discrepancies between the two sets of results. The ensuing correspondence has revealed that an error in his calculation of the integral of equation 2.7.2 has now been discovered.

The method by which Professor Hemstreet calculated this integral was to calculate the two integrals

$$I_{nm}^{(1)} = \int_0^\infty j_1(|k+K_n|r) j_1(|k+K_m|r) r^3 \, e^{-\alpha r} dr$$

and $$I_{nm}^{(2)} = \int_{R_c}^\infty j_1(|k+K_n|r) j_1(|k+K_m|r) r^3 \, e^{-\alpha r} dr$$

and the required integral was the determined via $I_{nm} = I_{nm}^{(1)} - I_{nm}^{(2)}$. The method of evaluating the above two integrals was by Gauss-Laguerre quadrature. The accuracy of this method had previously been checked for several representative matrix elements, but on further checks it was discovered that the integrals $I_{nm}^{(1)}$ and $I_{nm}^{(2)}$ were very sensitive to $|k+K_n|$ and $|k+K_m|$, the arguments of the Spherical Bessel Functions. For larger values of $|k+K_n|$ and $|k+K_m|$, the method of quadrature was less accurate, thereby introducing significant errors into the value of the integral I_{nm}.

The values of some of the nonlocal potential terms were thought to be as large as $\sim$0.001 Ryd., but were, in effect, the result of the inaccuracies of the method used for computing the integrals. Hence, Professor Hemstreet concludes that the nonlocal term probably does not have any effect on the band structure of 3C SiC, and accepts the results of this pilot study as probably being correct. Consequently, the nonlocal potential term will not be included in the full band structure calculations of 3C SiC.

3. THE ENVELOPE METHOD FOR DETERMINING IMPACT IONIZATION THRESHOLD ENERGIES

3.1 Introduction

Until recently, very little work has been done to determine accurate impact ionization (II) threshold energies, E_T, for realistic band structures. Previous calculations either assumed model band structures, or used parabolic band approximations to the energy band structures in the regions of the energy band extrema. These approximations are usually made by using suitable effective masses at the conduction band minimum and the valence band maximum.

A graphical method of calculating values of E_T has been provided by Tewordt [9], which has been extended and generalized by Franz [10]. This method, while considering the detailed band structure for the valence bands, assumes a parabolic conduction band based on the conduction band minimum. It was reasonable to make such approximations at the time, since not much was known about the detailed band structures of semiconductors. These approximations are no longer reasonable, as the knowledge of the detailed band structures of a number of semiconductors has become more extensive in the past few years.

Recent calculations of II threshold values have taken into consideration the detail of realistic band structures. The method developed in the present work is a generalization of the method developed by Franz, for which the Franz parabolic construction is a special, simplified case. This method is based on the conservation of energy and wavevector of all four states involved in an II process, and is presented in detail in sections 3.2 and 3.3 for electron II processes. The method applies equally as well for hole II processes.

When the present theory was in an advanced state of development, it was learnt that other workers had also been considering the same

problem. Anderson and Crowell [32] have developed a step-by-step graphical procedure based on two separate criteria. Firstly, they use the fact that the group velocities of three of the four states involved in an II process must be identical at threshold. Secondly, they invoke the conservation of energy and wavevector of all four states involved in the process. Using this procedure, they have obtained the first reliable estimates of E_T values for realistic band structures of a number of semiconductors.

The accuracy associated with their results ($\pm$ 0.2eV for each value of E_T) however, is not very great. The reason for this is probably due to their using a simple graphical technique, as opposed to applying the procedure to a digital computer, which could give much greater accuracy. The method presented here, while also being basically a graphical procedure is programmed for use on a digital computer, and so the results obtained are more accurate than those of Anderson and Crowell.

The method developed and presented here, which has been presented previously [31], is an alternative to the method developed by Anderson and Crowell. It has some advantages over the method developed by Anderson and Crowell, but it also has some disadvantages. Since both methods are basically graphical, the associated errors are of the same magnitude. Programming the method for use on a digital computer gives greater accuracy, and will apply to both methods. However, an advantage of the method developed here is that it is apparently simpler to program for use on a digital computer.

In numerical calculations, analytic representations of the energy bands are required for both methods. While good accuracy can be achieved in curve fitting, it is well known that large errors may occur in the

derivatives of such curves. The method developed by Anderson and Crowell depends upon the derivatives of the energy bands, and consequently may be subject to large errors in their threshold values. In fact, threshold values which do not really exist may be found due to errors in the derivatives of the energy bands. The method developed here, while making use of the derivatives of the energy bands, does not depend upon them to the same extent as does the method developed by Anderson and Crowell, and is therefore not subject to the associated errors to the same extent. This is another slight advantage of the method developed here over the method developed by Anderson and Crowell.

The method developed here restricts the final states involved in the II process to lie in the same energy band, although the method can be further generalized to lift this restriction, but the method developed by Anderson and Crowell does not have this restriction. Also, Anderson and Crowell allow for the inclusion of the emission or absorbtion of phonons in their method, which is not considered in the method developed here. These are two disadvantages of the method developed here over the method developed by Anderson and Crowell. However, in the results obtained by Anderson and Crowell, they have assumed the final states involved in the II processes to lie in the same energy band, and have not considered the emission or absorbtion of phonons in the process.

While reliable estimates of the value E_T for realistic band structures are calculated, values of E_T corresponding to approximate band structure models are also calculated. The approximations made, and the approximate formulae, are presented in section 3.4, which includes the parabolic band approximation applied to indirect gap semiconductors. Some computational considerations of the programming of the method developed here for use on a digital computer are presented in section 3.5.

Camphausen and Hearn [20] have also developed a method of obtaining II threshold values. Their method is based upon a Kane [21] band structure, but the formulae they quote do not appear to be correct. These are investigated further in section 3.6, and a simplified correct version of the formulae is presented.

3.2 The Basis of the Envelope Method

Impact ionization processes not involving phonons must conserve the energy and wavevector of the initial and final electron states. Thus, if a hot electron initially in a state, represented by the point H, on a conduction band (see figure 3.1(a)) is to move to a state, represented by the point I, in a second conduction band, not necessarily the same conduction band, then the second electron involved, to move to a state in the second conduction band, must have its initial state somewhere on the dotted curve of figure 3.1(a). That is, on the curve in E-$\mathbf{k}$ space obtained by displacing the whole of the second conduction band by the vector $\vec{HI}$. Since the displaced conduction band in figure 3.1(a) intersects the valence band at the states V_1 and V_2, impact ionization is possible in this case. The second electron is promoted from the valence band to the second conduction band in the ionization process either from V_1 to C_1 or from V_2 to C_2.

For a fixed position of H, suppose that the final state of the hot electron at I is allowed to vary within the second conduction band. The displaced conduction bands corresponding to the different positions of I then generate the shaded region of E-$\mathbf{k}$ space shown in figure 3.1(a). All states in the valence band lying within this region (that is between V_3 and V_4) can therefore partake in impact ionization processes with a hot electron initially at H.

The shaded region in figure 3.1(b) corresponds to a different,

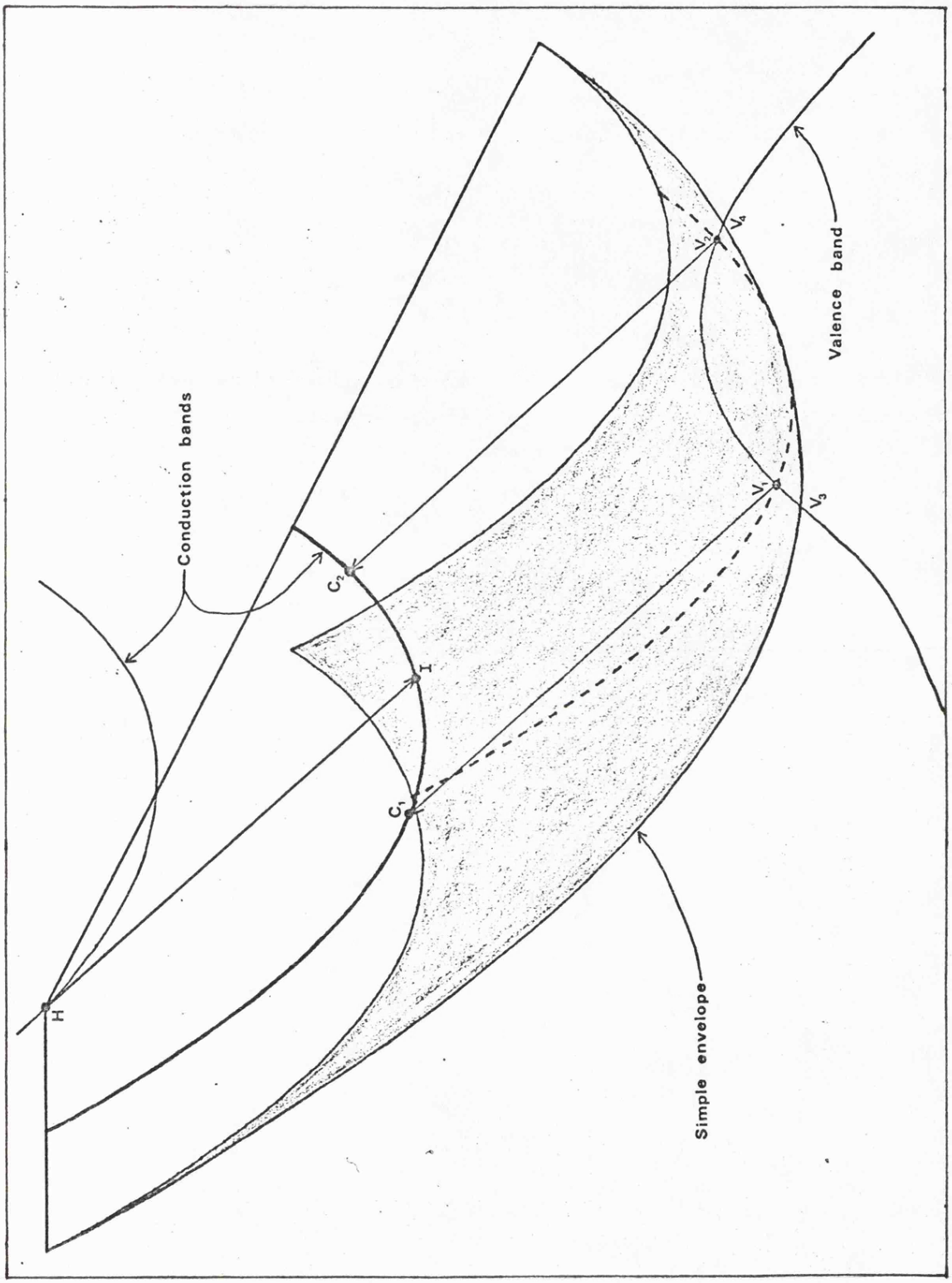


Figure 3.1(a) The simple envelope generated by displaced conduction bands, showing the position of the hot electron, H, above threshold. The dotted curve shows the conduction band displaced by $\overrightarrow{HI}$.

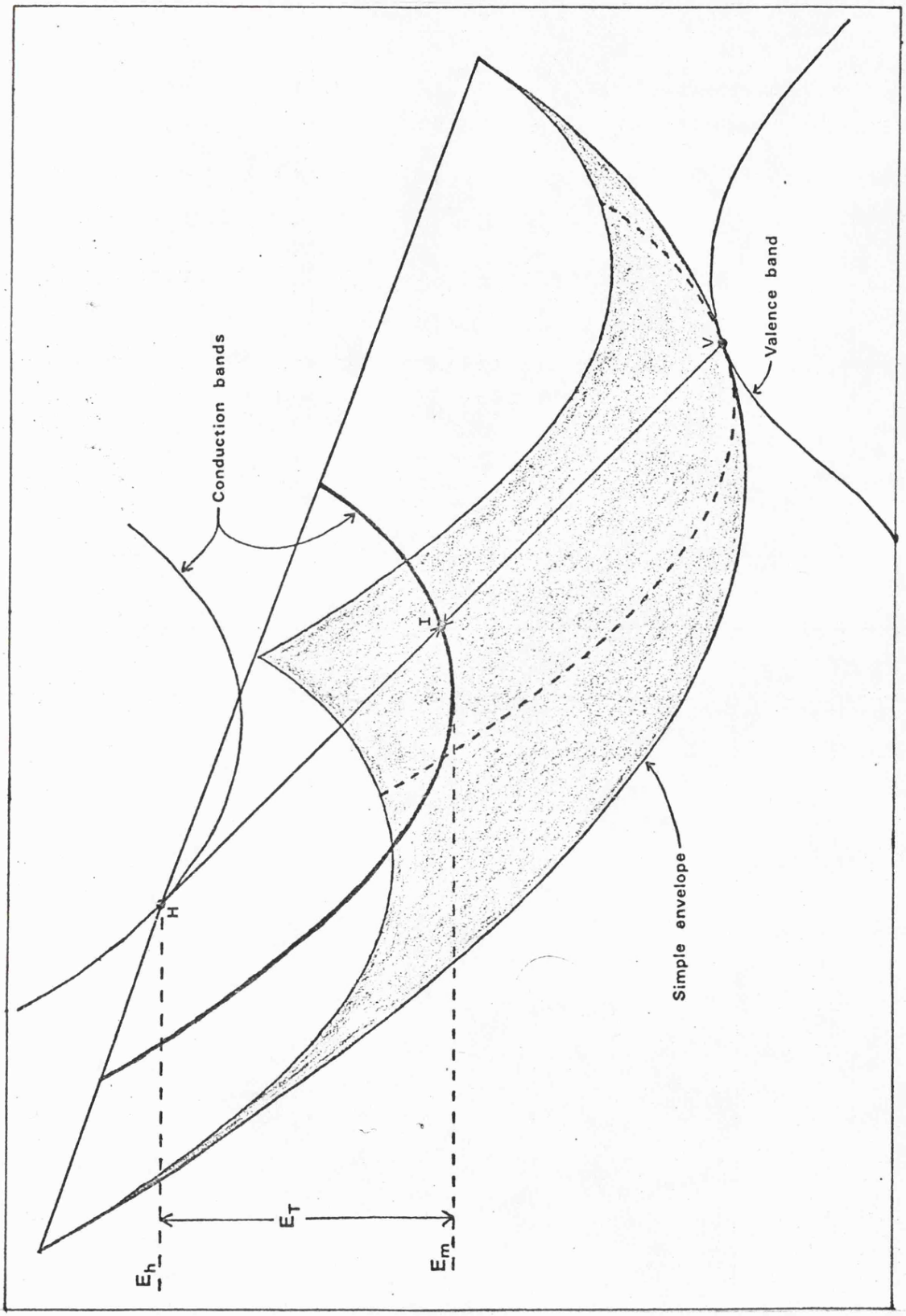


Figure 3.1(b) The simple envelope generated by displaced conduction bands, showing the position of the hot electron, H, at threshold.

but important, position of H. As the position of H is varied, the region covered by the displaced conduction bands moves relative to the first conduction band, but its shape remains unchanged. Figures 3.1(a) and 3.1(b) illustrate this fact. They also show that when H moves to lower energies, the minimum of the shaded region moves to higher energies, and fewer states in the valence band are available for impact ionization. Ultimately a position is reached when the lower boundary of the shaded region just touches the valence band at one point. This situation provides a threshold of the type being sought, since a further reduction in the energy of the initial position of the hot electron will not allow impact ionization to take place. The threshold situation, determining the value of E_T, is illustrated in figure 3.1(b). The lower boundary of the shaded region is tangential to the valence band at V. It is also noted that the final states after ionization are both at I, and that the gradient of the second conduction band at I is equal to that of the valence band at V. This is the first criteria used by Anderson and Crowell.

The problem of finding thresholds is thus reduced to that of determining the shape and position of the lower edge of the shaded region for any initial position, H, of the hot electron. The independence of the shape of this curve of the position of H greatly simplifies this problem. The curve is just the envelope of a family of displaced conduction bands, and may be calculated for any known band structure. The analysis presented here is for a one-dimensional band structure; the generalization to a higher number of $\mathbf{k}$-space dimensions is, in principle, straightforward.

Let the energy-wavevector relation for the first conduction band in which the hot electron is initially be given by

$$E = \phi_1(k) \qquad 3.2.1$$

Let the energy-wavevector relation for the second conduction band in which both electrons are finally be given by

$$E = \phi_2(k) \qquad 3.2.2$$

The displaced conduction band of figure 3.1(a) has the equation

$$E = \phi_2(k-\{k_i-k_h\}) + \phi_2(k_i) - \phi_1(k_h) \qquad 3.2.3$$

where k_i and k_h are the k coordinates of the points I and H respectively. The required envelope is obtained by finding, for any fixed k and k_h, the minimum value taken by E in 3.2.3 as k_i is varied. This will be when $\partial E/\partial k_i = 0$, that is when

$$\phi_2'(k_i) = \phi_2'(k-\{k_i-k_h\}) \qquad 3.2.4$$

Solving 3.2.4 for k_i and substituting into 3.2.3 gives the equation of the required envelope.

One obvious solution of 3.2.4 is always given when the arguments are equal, that is when

$$k_i = \frac{1}{2}(k + k_h) \qquad 3.2.5$$

and the equation of the envelope is then given by

$$E = 2\phi_2(\tfrac{1}{2}\{k + k_h\}) - \phi_1(k_h) \quad . \qquad 3.2.6$$

This formula should be regarded as an E-k relationship in which the coordinates of H appear as variable parameters. It shows that the envelope is merely a magnified and translated version of the second conduction band, to which it corresponds. This property can be seen in figure 3.1. It can also be seen from the figure that the states on this envelope may be found by joining the state H to states in the second conduction band, and then doubling the displacement. That is, by applying the displacement $\vec{HI}$ to each state I of the second

conduction band. Equation 3.2.6 may be used to verify this fact.

It should be noted that the state H can lie in any conduction band. It has been taken to lie in a different conduction band to that in which both the final states of the electrons lie after ionization. The state H could, for example,lie in the same conduction band, that is on $E = \phi_2(k)$. The shape of the envelope of $E = \phi_2(k)$ is not affected by the position of H, providing that both the final states after ionization lie on $E = \phi_2(k)$, that is in the same conduction band. Thresholds involving interband transitions of this type are thus within the framework of the envelope method. The extension of the analysis to include more complicated interband transitions is not considered in the present work.

Let the initial state of the hot electron be taken to lie in the same conduction band as the final states of the electrons, that is $\phi_1(k) \equiv \phi_2(k)$. If a parabolic relation with a suitable effective mass, m_c, is appropriate for $\phi_2(k)$, namely

$$E = \frac{\hbar^2}{2m_c}(k-k_m)^2 + E_m \qquad 3.2.7$$

with E_m the minimum energy and k_m the corresponding wavevector, then 3.2.6 gives the envelope in this case as

$$E = \frac{\hbar^2}{4m_c}(k-\{2k_m-k_h\})^2 + 2E_m-E_h \qquad 2.3.8$$

where E_h is the initial energy of the hot electron. This is just the "half-slope" parabola given by the Franz construction [10].

For more general band structures, there may exist solutions of 3.2.4 other than the obvious one given in 3.2.5. This is not the case if $\phi_2'(k)$ is a monatonic function of k. Thus, for a simple energy band without inflexions, 3.2.5 gives the only solution, as in the case illustrated in figure 3.1. This type of envelope is referred to

as a Simple Envelope (SE), to distinguish it from other possible solutions of 3.2.4 which are investigated in the next section. These solutions in no way affect the basis of the method, which is to vary the position of H until the shaded region becomes tangential to the valence band. They merely make the lower boundary of the shaded region more complicated.

3.3 The Envelope Method for Realistic Energy Bands

Careful examination of the minimization procedure covered in equations 3.2.3 to 3.2.6, shows that the curvature of the second conduction band is an important factor in the determination of thresholds by the envelope method. To determine whether the solution given by 3.2.6 is a minimum solution, it is necessary to consider the curvature of the simple envelope. That is by considering

$$\frac{\partial^2 E}{\partial k^2} = \frac{1}{2}\,\phi_2''\,\left(\frac{1}{2}\{k+k_h\}\right) \quad . \qquad 3.3.1$$

A minimum solution can only be given when this is positive. Thus it is seen that the envelope given in 3.2.6 can only give a lower boundary to the region of displaced conduction bands when $\phi_2''(k_i)$ is positive.

Corresponding to any position of I in the conduction band given by $E = \phi_2(k)$, there is always a simple envelope solution at a displacement $2\vec{HI}$ from H. This solution however, is of little consequence when the curvature at I is negative. The lower boundary curve must be provided by another solution, or solutions, of 3.2.4 in this case. Any realistic conduction band will have points of inflexion, unlike the simple case considered in figure 3.1. The analysis of the previous section must therefore be extended, to consider the envelopes provided by these other solutions.

Suppose a solution of 3.2.4 exists at a value of k_i given by

$$k_i = \frac{1}{2}(k+k_h) + K \qquad 3.3.2$$

where K is non-zero. Inserting this solution into 3.2.4 gives

$$\phi_2^1\left(\frac{1}{2}\{k+k_h\} + K\right) = \phi_2^1\left(\frac{1}{2}\{k+k_h\} - K\right) \quad . \qquad 3.3.3$$

It is therefore seen that another solution of 3.2.4 also exists for

$$k_i = \frac{1}{2}(k+k_h) - K \quad . \qquad 3.3.4$$

Using either 3.3.2 or 3.3.4 in 3.2.3, it is seen that for <u>both</u> these solutions, the envelope is given by

$$E = \phi_2\left(\frac{1}{2}\{k+k_h\} + K\right) + \phi_2\left(\frac{1}{2}\{k+k_h\} - K\right) - \phi_1(k_h) \quad . \qquad 3.3.5$$

This type of envelope is referred to as a Double Envelope (DE), since each point on it is generated twice, for the two different values of k_i. Let these values of k_i given in 3.3.2 and 3.3.4 be denoted by k_{i1} and k_{i2} respectively. For each such pair of states, I_1 and I_2, on the second conduction band with equal gradients, there is a point on the double envelope given by 3.3.5. Relative to the position of H, the coordinates of this double envelope point, V say, are

$$\left.\begin{aligned} E_V &= \phi_2(k_{i1}) + \phi_2(k_{i2}) \\ k_V &= k_{i1} + k_{i2} \end{aligned}\right\} \qquad 3.3.6$$

It is seen that V is a vertex of the parallelogram HI_1I_2V. This gives a simple geometrical method for locating double envelopes, which is illustrated in figure 3.2.

This figure shows the shaded region of displaced conduction bands together with all envelope solutions. The <u>shape</u> of the shaded region is again independent of the position of H. The simple envelope

is again seen as a magnified version of the conduction band. As expected, it does not provide the lower boundary of the shaded region at all its points. Double envelopes are not generated continuously, but are seen to terminate, by merging

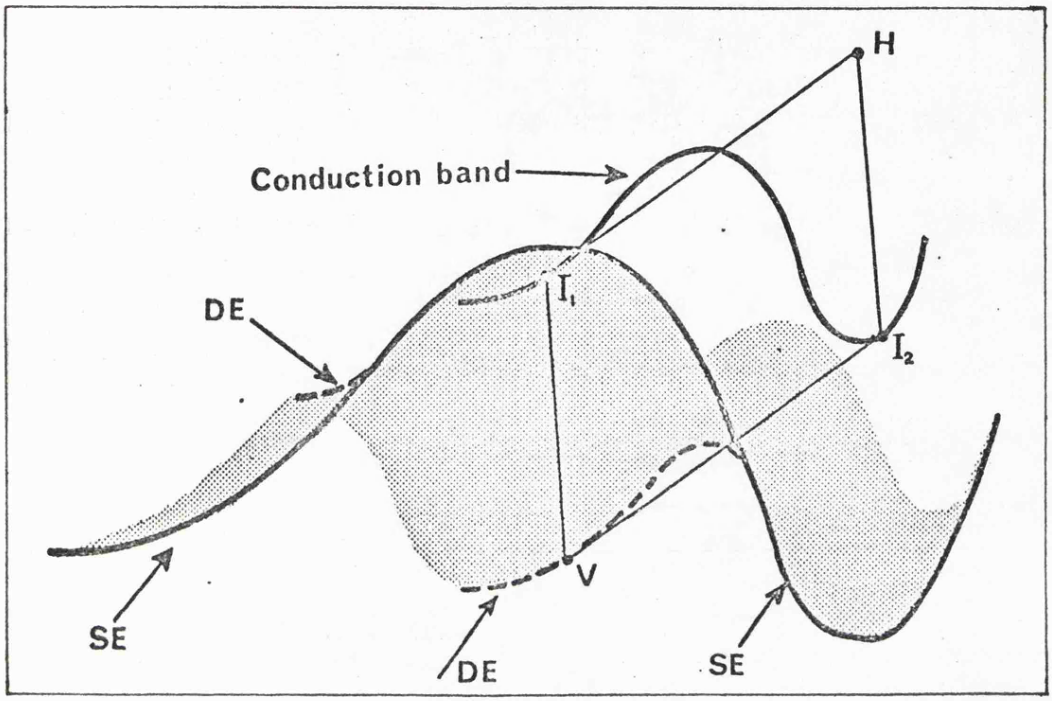


Figure 3.2 The simple envelope (SE) and double envelopes (DE) generated by a conduction band with inflexions.

into the simple envelope, at points of inflexion of the latter. This is to be expected from the previous remarks concerning the curvature.

The parallelogram HI_1I_2V shows that if V is a state in the valence band, then there are two possible interpretations of the ionization process. The hot electron at H can move to the state I_1 or I_2, while the valence electron at V is promoted to the state I_2 or I_1 respectively. The gradient of the double envelope at V is the same as that of the conduction band at I_1 and I_2. Thus for thresholds arising from double envelopes, the group velocities of three out of the four states involved are again identical. The example illustrated

in figure 3.2 corresponds to an intervalley transition.

It is seen in figure 3.2 that there is a section of the lower boundary of the region of displaced conduction bands which does not correspond to either a simple envelope solution or a double envelope solution. This is due to the finite length of the conduction band being considered, and does not occur when the length of the conduction band is assumed to be infinite. For the envelope method to include Umklapp-type processes, it is necessary to consider energy bands in an extended zone scheme. Band structures in the directions of symmetry in **k**-space are periodic in such a scheme. In the principal symmetry directions, the bands are also symmetric about the centre of the first Brillouin zone.

Let the conduction bands represented by $E = \phi_2(k)$ be periodic with period p and symmetric about $k = 0$.

That is,

$$\phi_2(k+p) = \phi_2(k) \qquad 3.3.7$$

$$\phi_2(-k) = \phi_2(k) \qquad 3.3.8$$

Using 3.3.7 and 3.3.8, it is seen that the envelope solutions given by 3.2.6 and 3.3.5 are periodic with period 2p, and symmetric about $k = 0$ relative to the k coordinate of H. However, by replacing k by k+p in 3.2.4 by using 3.3.7, it is seen that the solutions 3.2.5, 3.3.2 and 3.3.4 become

for S.E.'s $\quad k_i = \frac{1}{2}(k+k_h+p) \qquad 3.3.9$

for D.E.'s $\quad k_i = \frac{1}{2}(k+k_h+p) \pm K \qquad 3.3.10$

Substituting these values of k_i in 3.2.3, the envelope solutions 3.2.6 and 3.3.5 become

$$E = 2\phi_2\left(\tfrac{1}{2}\{k+k_h+p\}\right) - \phi_1(k_h) \qquad 3.3.11$$

for simple envelopes, and

$$E = \phi_2(\tfrac{1}{2}\{k+k_h+p\}+K) + \phi_2(\tfrac{1}{2}\{k+k_h+p\}-K)-\phi_1(k_h) \qquad 3.3.12$$

for double envlopes. Thus, it is seen that there are twice as many envelope solutions as before, and when 3.2.6 is combined with 3.3.11, and 3.3.5 is combined with 3.3.12, the complete envelope pattern is given by

$$E = 2\phi_2(\tfrac{1}{2}\{k+k_h+np\}) - \phi_1(k_h) \qquad n = 0,1 \qquad 3.3.13$$

for simple envelopes, and

$$E = \phi_2(\tfrac{1}{2}\{k+k_h+np\}+K) + \phi_2(\tfrac{1}{2}\{k+k_h+np\}-K) - \phi_1(k_h) \quad n = 0,1 \quad 3.3.14$$

for double envelopes. By use of 3.3.7 and 3.3.8 it is seen that the complete envelope pattern given by 3.3.13 and 3.3.14 has the same periodicity and symmetry properties as the conduction band $E = \phi_2(k)$. This is illustrated in figure 3.3

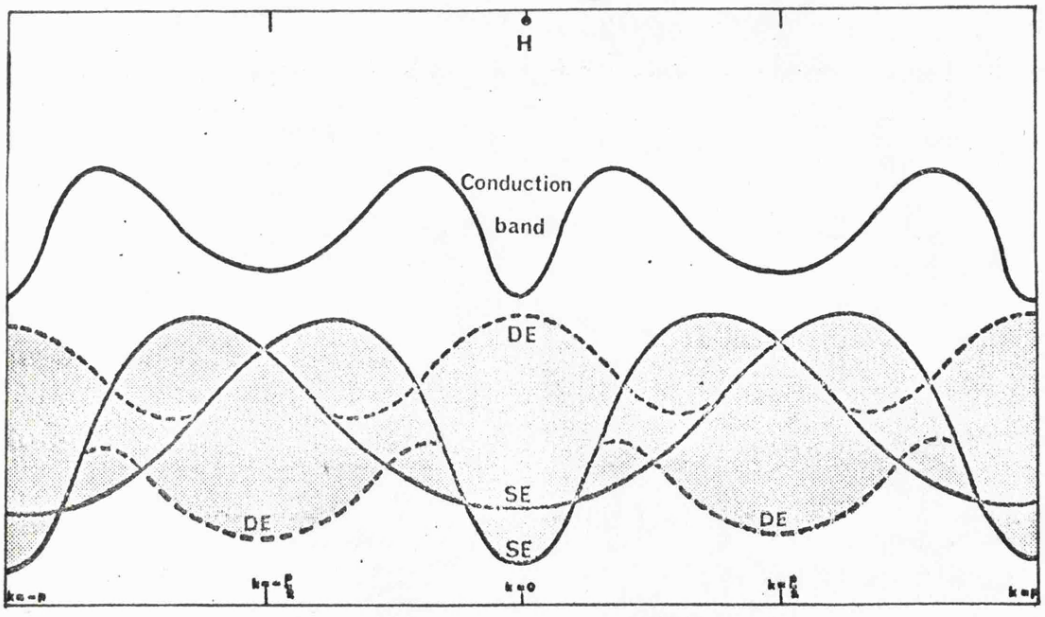


Figure 3.3 The envelope pattern generated by a conduction band having period p, and symmetric about k = 0. The simple envelope (SE) solutions are given by 3.3.13 and the double envelope (DE) solutions by 3.3.14.

3.4 Approximate Band Structure Models

While reliable estimates of impact ionization threshold energies, E_T, for realistic band structures are calculated by the method described in the two previous sections, corresponding values of E_T using approximate band structure models are also calculated. This is done in order to obtain a comparison between the different values of E_T calculated, and so to determine the reliability of earlier estimates of the threshold energy. Two approximate band structure models are therefore considered.

The first approximate band structure model considered is that for which the conduction band is assumed to be parabolic, of the form given in 3.2.7, and the valence bands are treated exactly. The envelope method then simplifies, and reduces to the method developed by Tewordt and extended by Franz. This approximation is referred to as the Franz construction. The method of determining the values of E_T by this method is the same as that for the envelope method before simplification, and is outlined in the next section.

The second approximate band structure model considered is that of the effective mass approximation. Let the conduction band on which the hot electron lies initially be approximated by the parabola

$$E_1 = \frac{\hbar^2}{2m_{c1}} (k-k_n)^2 + E_n \qquad 3.4.1$$

and the conduction band on which both electrons lie finally be approximated by the parabola

$$E_2 = \frac{\hbar^2}{2m_{c2}} (k-k_m)^2 + E_m \qquad 3.4.2$$

Here, (k_n, E_n) and (k_m, E_m) are the coordinates of the first and

second conduction band minima respectively, with corresponding effective masses m_{c1} and m_{c2}. Further, let the valence band be approximated by the parabola

$$E_v = \frac{-\hbar^2}{2m_v} k^2 \qquad 3.4.3$$

where m_v is taken to be positive.

Using 3.4.1 and 3.4.2 in the simple envelope equation, given by 3.2.6, for the energy bands $\phi_1(k)$ and $\phi_2(k)$ respectively, gives the simple envelope in this case as

$$E_{nv}(k) = \frac{\hbar^2}{4m_{c2}} (k+k_h-2k_m)^2 + 2E_m-E_h \qquad 3.4.4$$

where (k_h,E_h) are the coordinates of the initial position of the hot electron. The condition for a threshold situation is when the envelope just touches the valence band. That is when

$$E_{nv}(k) = E_v(k) \qquad 3.4.5$$

and

$$\frac{\partial}{\partial k} E_{nv}(k) = \frac{\partial}{\partial k} E_v(k) \qquad 3.4.6$$

Substituting 3.4.3 and 3.4.4 into 3.4.5 and 3.4.6 gives

$$\frac{\hbar^2}{4m_{c2}} (k+k_h-2k_m)^2 + 2E_m-E_h = \frac{-\hbar^2}{2m_v} k^2 \qquad 3.4.7$$

and

$$\frac{\hbar^2}{2m_{c2}} (k+k_h-2k_m) = \frac{-\hbar^2}{m_v} k \qquad 3.4.8$$

respectively, where E_h is given by

$$E_h = \frac{\hbar^2}{2m_{c1}} (k_h-k_n)^2 + E_n \qquad 3.4.9$$

Rearranging 3.4.8 to obtain an explicit equation for k, and then substituting into 3.4.7 gives, after some simplification,

$$\hbar^2(2k_m-k_h)^2 = 2(m_v+2m_{c2})(E_h-2E_m) \quad . \qquad 3.4.10$$

Now rearranging 3.4.9 to obtain k_h in terms of E_h, and then substituting into 3.4.10 gives

$$\hbar^2\left\{(2k_m-k_n) \mp \sqrt{\frac{2m_{c1}}{\hbar^2}(E_h-E_n)}\right\}^2 = 2(m_v+2m_{c2})(E_h-2E_m)$$

which, upon expanding the left hand side, and rearranging the terms, gives

$$\pm\sqrt{2m_{c1}\hbar^2}\,(2k_m-k_n)(E_h-E_n)^{\frac{1}{2}} = \frac{1}{2}\hbar^2(2k_m-k_n)^2 + m_{c1}(E_h-E_n) - (m_v+2m_{c2})(E_h-2E_m) \qquad 3.4.11$$

Introducing, for ease of writing and clarity, the notation

$$\left.\begin{aligned} X &= (2k_m - k_n) \\ \text{and} \qquad M &= (m_v + 2m_{c2}) \end{aligned}\right\} \qquad 3.4.12$$

equation 3.4.11 becomes, upon squaring,

$$2m_{c1}\hbar^2(E_h-E_m)X = \frac{1}{4}\hbar^4 X^2 + \hbar^2\{m_{c1}(E_h-E_n)-M(E_h-2E_m)\}X + \{m_{c1}(E_h-E_n)- M(E_h-2E_m)\}^2 \qquad 3.4.13$$

By expanding the square on the right hand side, and rearranging the terms, 3.4.13 can be written in the form of a quadratic equation in E_h which has the form

$$a E_h^2 - 2bE_h + c = 0 \qquad 3.4.14$$

where

$$\left.\begin{aligned} a &= (M-m_{c1})^2 \\ b &= (M-m_{c1})(2ME_m - m_{c1}E_n) - \tfrac{1}{2}\hbar^2(M+m_{c1})X \\ c &= (2ME_m - m_{c1}E_n)^2 + \hbar^2(2ME_m + m_{c1}E_n)\,X + \tfrac{1}{4}\hbar^4 X^2 \end{aligned}\right\} \quad 3.4.15$$

Equation 3.4.14 then has the two solutions given by

$$E_h = \left[\frac{b}{a}\right] \pm \sqrt{\left[\frac{b}{a}\right]^2 - \left[\frac{c}{a}\right]} \quad 3.4.16$$

Substituting the values of a, b and c given by 3.4.15 into 3.4.16, the two values of the hot electron energy, E_T, corresponding to a threshold situation are then given, after a little algebra and by substituting back for the notation of 3.4.12, by the equation

$$E_T = E_h - E_m = E_m + \frac{m_{c1}(2E_m - E_n)}{(m_v - m_{c1} + 2m_{c2})} + \frac{\hbar^2(2k_m - k_n)^2}{2(m_v - m_{c1} + 2m_{c2})^2}\left[(m_v + m_{c1} + 2m_{c2}) \pm \right.$$
$$\left. \pm\, 2\sqrt{m_{c1}(m_v + 2m_{c2})}\left\{1 + \frac{2(m_v - m_{c1} + 2m_{c2})(2E_m - E_n)}{\hbar^2(2k_m - k_n)^2}\right\}^{\frac{1}{2}}\right] \quad 3.4.17$$

Hence, when the conduction and valence bands are approximated by parabolae, the values of E_T can be easily calculated from the expression given by 3.4.17. It should be noted that the expression gives two values of E_T for a given set of parabolic bands. The smaller value corresponds to a threshold value of the type being sought, that is the onset of impact ionization processes, while the larger value corresponds to an anti-threshold value of the type described by Anderson and Crowell. For hot electron energies greater than this larger value, impact ionization processes are no longer possible; thus it is seen that for parabolic energy bands, a "pair-production window" exists as described by Camphausen and Hearn.

The expression given by 3.4.17 allows for two conduction bands to be considered, but if only one conduction band is allowed for, as is more often the case, equation 3.4.17 can be simplified slightly. By writing $m_{c1} = m_{c2} = m_c$, $E_m = E_n$ and $k_m = k_n$, substituting into 3.4.17, and introducing the familiar notation $\mu = {}^{m_c}/_{m_v}$, 3.4.17 can then be written

$$E_T = \frac{(1+2\mu)}{(1+\mu)} E_m + \frac{\hbar^2 k_m^2}{2m_v(1+\mu)^2} \left[(1+3\mu) \pm 2\mu^{\frac{1}{2}} (1+2\mu)^{\frac{1}{2}} \left\{ 1+\frac{2m_v(1+\mu)E_m}{\hbar^2 k_m^2} \right\}^{\frac{1}{2}} \right] \qquad 3.4.18$$

Since most approximate band structure models assume just one conduction band based on the minimum, the expression given by 3.4.18 is used in determing the values of E_T.

For a direct band gap, $k_m = 0$ and $E_m = E_G$, and then 3.4.18 reduces to the well known form

$$\frac{E_T}{E_G} = \frac{1+2\mu}{1+\mu} \qquad 3.4.19$$

It should be noted that for indirect band gaps, E_m is only equal to E_G in the appropriate $\mathbf{k}$ direction.

3.5 Some Computational Considerations

In the analysis of the envelope method presented in sections 3.2 and 3.3, it is required that the final states of both electrons involved in the ionization process lie in the same conduction band. In the computer program written to calculate the impact ionization threshold values, the final states of both electrons are taken to lie in the lowest-lying conduction band. This assumption is also made in the calculations of Anderson and Crowell. The initial state of the hot electron is, however, allowed to lie in any conduction band, and all possible bands are investigated to determine threshold values in addition to the absolute minimum

threshold value.

The computer program is also used to determine impact ionization threshold values by hot holes, which is done by simply reversing the roles of the conduction and valence bands throughout. The final states of both holes are then taken to lie in the highest-lying valence band. The program is also capable of determing some anti-threshold values, and hence some "pair-production windows". The anti-threshold values correspond to the situations where a valence band just touches the uppermost envelope of the shaded region. The calculation of these anti-threshold values, and hence pair-production windows, is not considered in detail here, although a few anti-threshold values are determined.

To be able to apply the envelope method, analytic expressions of all the relevant energy bands are required. These are determined by the method described in Chapter 4, and are of the form of even Fourier series. The computer program has been written to accept these Fourier series, and also to accept parabolic expressions. This is done to allow calculations to be performed of threshold values corresponding to the Franz construction and to the effective mass approximation. However, it is quicker, and easier, to calculate the threshold values corresponding to the effective mass approximation by using a programmable, desk-top calculator, for which a simple program has been written to evaluate equation 3.4.18.

The computer program operates in a chosen symmetry direction in $\underline{k}$-space, and makes full use of the periodicity and all symmetry properties associated with that particular direction. The three principal symmetry directions Γ-Δ-X, Γ-Λ -L and Γ-Σ-K-S-X only are investigated. The threshold values are then determined by using an iterative procedure based on the variation of k_h. The energy bands in which the initiating and promoted

particles lie are provided as data, together with a sensible first iterate for k_h, as described in Chapter 4. The program then automatically investigates the intersection of the envelope with the appropriate energy band until a threshold situation is obtained. At each stage of the iteration the program only calculates those parts of the envelope which are appropriate.

The coordinates of any point on the envelope are determined through equations 3.3.13 and 3.3.14. Points on the simple envelope with n = 0 are obtained by using the 'double displacement' property discussed in relation to 3.2.6. Points on the double envelope with n = 0 are obtained from the parallelogram vertex given by 3.3.6, using positions of equal gradient on the appropriate energy band which are separated by less than the distance p. The envelopes for n = 1 are calculated by a displacement of those for n = 0 through a distance p. Further computational details are given in Chapter 7.

3.6 A Discussion of the Short Note by Camphausen and Hearn

In the short note by Camphausen and Hearn [20], two equations are derived by which impact ionization threshold values may be determined. The derivation is based on a Kane [21] band structure for the conduction band with the inclusion of arbitrary spin-orbit coupling energy. The equations thus derived are

$$\left.\begin{aligned} & k(E_t)\{\mu k'(F_t)\}^2 - k(F_t)\{1+2\mu k'(F_t)\}^2 = 0 \\ & E_t = 1+2F_t + k(F_t)\,\mu^{-1}\{k'(F_t)\}^{-2} \end{aligned}\right\} \qquad 3.6.1$$

where all energies are measured in terms of the band gap, the function k(E) gives the modulus of the wavevector of an electron of energy E in the conduction band, the prime denoting differentiation, and μ is the ratio of effective masses m_c/m_v. In terms of absolute energies,

$$E_t = \frac{E_h - E_G}{E_G} \quad \text{and} \quad F_t = \frac{E_f - E_G}{E_G}$$

where E_h and E_f are the initial and final energies of the hot electron respectively.

It is claimed that for a parabolic conduction band, equations 3.6.1 give the familiar result

$$E_t = \frac{1+2\mu}{1+\mu}$$

For a parabolic conduction band, the wavevector - energy relation is given by

$$k(E) = \sqrt{\frac{2m_c E_G}{\hbar^2}}\, E^{\frac{1}{2}} \qquad 3.6.2$$

where E is measured in terms of the band gap.

Differentiating gives

$$k'(E) = \frac{1}{2}\sqrt{\frac{2m_c E_G}{\hbar^2}}\, E^{-\frac{1}{2}} \qquad 3.6.3$$

Substituting 3.6.2 and 3.6.3 into the equations 3.6.1 and introducing the notation

$$A = \sqrt{2m_c E_G/\hbar^2}$$

gives

$$\left.\begin{aligned} &\frac{A^3\mu^2 E_t^{\frac{1}{2}}}{4F_t} - AF_t^{\frac{1}{2}}\left\{1+\mu AF_t^{-\frac{1}{2}}\right\}^2 = 0 \\ &E_t = 1+2F_t + \frac{4F^{3/2}}{\mu A} \end{aligned}\right\} \qquad 3.6.4.$$

From work by Beattie and Landsberg [17] it has been shown that for parabolic energy bands of the type being used here, in the same notation,

$$E_t = \frac{1+2\mu}{1+\mu} \quad \text{and} \quad F_t = \frac{\mu^2}{(1+\mu)(1+2\mu)}$$

Substituting these values into equations 3.6.4 gives, after some simplification, the equations

$$\left.\begin{aligned} &A^2(1+\mu)^2(1+2\mu)^3 - 4\mu\left\{1+A(1+\mu)^{\frac{1}{2}}(1+2\mu)^{\frac{1}{2}}\right\}^2 = 0 \\ &A = \frac{4\mu}{(1+\mu)^{\frac{1}{2}}(1+2\mu)^{\frac{1}{2}}} \end{aligned}\right\} \qquad 3.6.5$$

By substituting the second of these equations into the first equation to eliminate A, and after a little algebra to simplify the equation, an equation in μ is obtained, namely

$$4\mu(1+2\mu)^2\ (1+\mu) - (1+4\mu)^2 = 0$$

This is a quartic equation in μ, and can be written more explicitly as

$$16\mu^4 + 32\mu^3 + 4\mu^2 - 4\mu - 1 = 0 \qquad 3.6.6$$

Thus, for equations 3.6.1 to be consistent with each other for parabolic energy bands, equation 3.6.6 must be satisfied. This is true for only <u>one</u> positive real value of μ ($\mu = 0.37$), and not for <u>any</u> positive real value of μ as is required. It follows that since 3.6.6 is not satisfied for <u>all</u> values of μ, except just one value, then equations 3.6.1 must be inconsistent. If they are inconsistent, and hence incorrect, for the simplified case of parabolic energy bands, they must also be incorrect for the Kane band structure considered.

To try to determine in what way the equations 3.6.1 are incorrect, equivalent equations are derived using a slightly simpler band structure. Instead of using a Kane band structure, the final states of the electrons are assumed to lie in a parabolic energy band given by 3.6.2. The initial state of the hot electron, however, is not restricted to lie in any particular form of energy band. In the same notation as that of Camphausen

and Hearn, the equations governing the conservation of energy and wavevector are

$$\left.\begin{aligned} k(E_t) - 2k(F_t) - k(E_v) &= 0 \\ E_t &= 1 + 2F_t + E_v \end{aligned}\right\} \qquad 3.6.7$$

To eliminate the dependence of the valence band from these equations, the fact that the group velocity of the promoted electron remains unaltered is used. This is equivalent to

$$k'(E_v) = k'(F_t)$$

which, from 3.6.3 gives

$$E_v = \mu^{-1} F_t$$

and from 3.6.2 gives

$$k(E_v) = 2k'(E_v)E_v = 2\mu^{-1} k'(F_t)F_t$$

Also, from 3.6.2 and 3.6.3, it is seen that

$$F_t = \frac{1}{2}k(F_t)\{k'(F_t)\}^{-1}$$

and hence that

$$\left.\begin{aligned} E_v &= \frac{1}{2} k(F_t)\{\mu k'(F_t)\}^{-1} \\ k(E_v) &= \mu^{-1}k(F_t) \end{aligned}\right\} \qquad 3.6.8$$

Thus, substituting equations 3.6.8 into equations 3.6.7 and rearranging the terms slightly, give the equations

$$\left.\begin{aligned} \mu k(E_t) - k(F_t)\{1+2\mu\} &= 0 \\ E_t &= 1+2F_t+k(F_t)\mu^{-1}\{2k'(F_t)\}^{-1} \end{aligned}\right\} \qquad 3.6.9$$

Equations 3.6.9 are the equivalent equations to those given by Camphausen and Hearn, but clearly are not in agreement with them. Also, if parabolic energy bands are used throughout, it can be verified that equations 3.6.9 are consistent and reduce to the well known result of

$$E_t = \frac{1+2\mu}{1+\mu} \quad .$$

4. CURVE FITTING TECHNIQUES AS USED FOR THE ENERGY BANDS

4.1 Introduction

The Empirical Pseudopotential Method, as described in Chapter 2, is used to obtain the realistic band structures of the semiconductors being investigated. A band structure is given as a set of discrete energy levels, the eigenvalues of the secular equations, at as many discrete points within the first Brillouin zone as is required. Thus, for each energy band to be investigated there corresponds energy levels at a set of discrete points in **k**-space. However, the method of determining the impact ionization threshold values, as presented in the previous chapter, requires analytic expressions for the energy bands being investigated. Two different types of expressions can be obtained, using two different methods of approach.

The first method which can be used is that by which simple interpolation schemes are employed to approximate to the energy of a particular energy band. This method requires several different schemes to be fitted to the available data points in order to obtain the energy values to sufficient accuracy. More than one scheme is necessary because schemes which are valid at intermediate points near the middle of the range of data points are different from schemes which are valid at intermediate points near the end of the range of data points. Thus, given a set of data points, this method requires a test to be made to determine the position of the point, at which the energy value is to be calculated, in relation to the set of data points. Once this position is determined, and the appropriate interpolation scheme is selected, tests have to be made to determine which data points are to be used in the scheme. After all these tests have been performed, the energy value on the required energy band can then be calculated. It is seen that this method is a lengthy and complicated procedure, and since it will be necessary to

calculate the energy, on a given band, at a large number of points, it is considered to be impracticable.

The second method which can be used is that by which one analytic expression is fitted to each energy band, and which is valid over the whole range of data points. This is done by first selecting a suitable analytic expression, and then employing a suitable curve-fitting routine by which a good approximation to the data points is obtained. Once a sufficiently accurate analytic expression is determined, the calculation of the energy, on any energy band, at any point within the range of data points becomes a simple task of evaluating the appropriate analytic expression. The method of determining the analytic expressions used in curve-fitting is more complicated than that of the interpolation schemes, but interpolation schemes are more complicated to use in order to calculate the required energy of a given energy band than is the single analytic expression. Since the analytic expression chosen will be used to calculate the energy on a given energy band a large number of times, the method of curve-fitting is chosen to approximate to the energy bands.

In a curve-fitting problem, there are several different methods by which an analytic expression can be fitted to a set of data points, and there are many different analytic expressions which can be used. The method selected by which the analytic expressions are fitted to a set of data points, is that of multiple regression, and the analysis of this method is described in the next section. The particular form of analytic expression to use is subject to an investigation between two possible alternatives; a polynomial consisting of even powers only, and a Fourier series consisting of cosine terms only. The results of this investigation are presented in section 4.3, from which the even Fourier series is chosen as the analytic expression to be used in the curve-fitting routine.

A curve-fitting routine can only ensure that the values of the analytic expressions used are in good agreement with the data points used, but

cannot ensure that they are in good agreement elsewhere. The classic example of this arises when the number of coefficients in the analytic expression is the same as the number of data points. A curve-fitting routine is then able to fit the analytic expression to the data points exactly, but it is most likely to be rapidly oscillating between the data points, so giving a meaningless expression. This situation does not arise in the multiple regression routine, as there has to be fewer coefficients in the analytic expression than there are data points. However, to ensure no large deviations from the expected shape of the energy bands occur, the resulting analytic expressions of the energy bands are drawn by using a graph plotter attached to a digital computer.

A computer program is written to obtain the graphs of the analytic expressions fitted to the energy bands, and also to draw three additional graphs, which are used in conjunction with the method of determining the impact ionization threshold values. The first of these additional graphs draws the energy bands in an extended zone scheme together with the reflection of the conduction band about the conduction band minimum, and of the valence band about the valence band maximum. The other two additional graphs draw the envelopes of the conduction and valence bands respectively, using the appropriate equations for the envelopes as presented in the previous chapter. These additional graphs provide an approximate graphical technique of determining impact ionization threshold values, and is described in section 4.4. Some computational considerations required in using the multiple regression routine are presented in section 4.5, together with some considerations of the computer program written to draw the graphs of the energy bands.

4.2 Multiple Regression Routine

Multiple regression has several uses in statistics, of which one that is widely used is as a method of fitting analytic expressions to

sets of data points. The method employed expresses a dependent variable, Y, as a linear combination of a number of linearly independent variables, x_i (i=1 to n), which are thought to influence the behaviour of Y. This gives an approximation to Y given by the equation

$$\hat{Y} = b_o + \sum_{i=1}^{n} b_i x_i \qquad 4.2.1$$

where $\hat{Y}$ is the estimate of the dependent variable, and the b_i are the regression coefficients. These coefficients are estimated by minimizing the expression

$$\sum_{j=1}^{m} (Y_j - \hat{Y}_j)^2 \qquad 4.2.2$$

where the sum extends over all the observations, or data points, of the dependent variable. The estimate of the regression coefficients are just the least squares estimates.

When this method is applied to curve fitting, the analytic expression of the curve being fitted is required to be in the form of equation 4.2.1. This is achieved if the analytic expression is naturally of that form, or if it is in a form which can be reduced to that of 4.2.1 by transformations. Thus, if the analytic expression of the curve is a polynomial given by

$$y = b_o + \sum_{i=1}^{n} b_i z^i \qquad 4.2.3$$

then the transformations $x_i = z^i$ (i=1 to n) are performed to give the regression equation identical to 4.2.1. Similarly, if the analytic expression of the curve is an even Fourier series given by

$$y = b_o + \sum_{i=1}^{n} b_i \cos(i\pi z) \qquad 4.2.4$$

then the transformations $x_i = \cos(i\pi z)$ are performed to give the required form of equation 4.2.1.

Before the multiple regression calculations are performed, various preliminary statistical calculations have to be performed. Let it be assumed that there are n independent variables, or terms allowed in the analytic expression of the curve, and let there be p dependent variables, or energy bands, which are being considered. For each variable let there be m observations, that is distinct points within the first Brillouin zone at which energy values are calculated on each energy band being considered. Then, using the notation that x_{ij} is the j^{TH} observation (of m) of the i^{TH} variable (of n + p), the preliminary statistical calculations can be performed.

The first statistical quantity calculated is the sample mean, $\bar{x}_i$, of each variable x_i, and is given by the equation

$$\bar{x}_i = \frac{1}{m} \sum_{j=1}^{m} x_{ij} \qquad (i = 1 \text{ to } n + p) \qquad 4.2.5$$

The sample variance, S_i^2, of each variable measuring the dispersion of the observations about the mean is then calculated by

$$S_i^2 = \frac{1}{m-1} \sum_{j=1}^{m} (x_{ij} - \bar{x}_j)^2 \qquad (i = 1 \text{ to } n + p) \qquad 4.2.6$$

from which the sample standard deviation, S_i, of each variable is given. The next quantity calculated is the matrix of the sums of squares and cross products about the mean. This matrix is given by $A = [a_{ij}]$ where

$$a_{ij} = \sum_{k=1}^{m} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \qquad (i,j = 1 \text{ to } n + p) \qquad 4.2.7$$

where the diagonal elements, a_{ii}, of the matrix contain the sums of squares of the observations corresponding to the variable x_i, and the off-diagonal elements, a_{ij}, contain the cross products between the variables x_i and x_j, summed over the observations.

From this cross products matrix, the covariance matrix is calculated,

for which an element, c_{ij}, gives a measure of the coherence between the variables x_i and x_j. When i = j the covariance becomes the variance of the variable x_i, and the matrix is given by $c = [c_{ij}]$ where

$$c_{ij} = a_{ij}/(m-1) \qquad (i,j = 1 \text{ to } n + p) . \qquad 4.2.8$$

The covariance between two variables can be related to the standard deviations of those two variables by the inequality

$$- S_i S_j \leqslant c_{ij} \leqslant S_i S_j \qquad 4.2.9$$

By dividing c_{ij} by $S_i S_j$ the covariance is transformed into the unit quantity known as the correlation coefficient of the variables x_i and x_j. This coefficient must have a value in the range [-1, +1], and is insensitive to the scales of measurement of the variables. The correlation coefficients can be interpreted as being a measure of the strength of the linear relationship between the two variables, and is given by $R = [r_{ij}]$ where

$$r_{ij} = \frac{c_{ij}}{S_i S_j} = \frac{a_{ij}}{\sqrt{a_{ii} a_{jj}}} \quad (i,j = 1 \text{ to } n + p) \qquad 4.2.10$$

The last preliminary statistical calculation to be performed is a simple regression of every variable, x_i, onto every other variable, x_j, for which a corresponding simple regression coefficient matrix is given by $G = [g_{ij}]$ where

$$g_{ij} = a_{ij}/a_{jj} \qquad (i,j = 1 \text{ to } n + p) \qquad 4.2.11$$

This gives the matrix of simple regression coefficients of rows on columns, for which there is a similar matrix of simple regression coefficients of columns on rows which is given by $H = [h_{ij}]$ where

$$h_{ij} = a_{ij}/a_{ii} \qquad (i,j = 1 \text{ to } n + p) \qquad 4.2.12$$

The matrix H is just the transpose of the matrix G, since it is seen from 4.2.7 that the matrix A is symmetric. These calculations form the basis of the multiple regression method, from which the multiple regression coefficients are calculated in a manner similar to that described by Efroymson [56].

This method considers one dependent variable at a time, and the calculation of the regression coefficients is performed by a stepwise procedure in which one independent variable at a time is either rejected from or accepted to the regression equation. At any step in this procedure, to determine whether an independent variable is rejected from or accepted to the regression equation, a vector is calculated from the equation

$$V_i = d_{i\ell}d_{\ell i}/d_{ii} \quad (i = 1 \text{ to } n) \qquad 4.2.13$$

where ℓ is the subscript of the dependent variable. The matrix $D = [d_{ij}]$ is the correlation matrix whose rows and columns correspond to the independent variables included in the regression equation by all previous steps of the procedure. The vector, V, contains the change in the cross products between all the independent variables, x_i, and the dependent variable, x_ℓ, being considered. A value V_i is negative if the variable x_i is in the regression equation and positive if the variable is not in the regression equation.

The procedure first seeks to reject a variable, x_i, from the regression equation, which is done if the corresponding value of V_i is negative and sufficiently small in modulus. Only if a variable is not rejected does the procedure seek to accept a variable x_i to the regression equation, which is done if the corresponding value of V_i is positive and sufficiently large. That is, a variable rejected from the regression equation causes the least increase in the variance, while a variable accepted to the

regression equation causes the greatest decrease in the variance.

The measure by which the values of V_i for the rejection or acceptance of a variable in the regression are considered to be sufficiently small or large is known as the F-distribution. This distribution, as used in multiple regression, is a measure of the ratio of the variances of two different regression equations, or a measure of the change in variance between two different regression equations. The F-distribution is a basic statistical quantity (see for example, G.B. Wetherill, 'Elementary Statistical Methods', pp 150-152.[57]). Two different values of the F-distribution are used, one as a level for rejecting a variable from the regression equation which has an insignificant effect on the regression, and the other as a level for accepting a variable to the regression equation which has a significant effect on the regression. If these two values are denoted by F_r and F_a respectively, then variables are rejected from or accepted to the regression equation when $|V_i| < F_r$ or $V_i > F_a$ respectively, where the V_i are given by 4.2.13.

When no independent variables are rejected from or accepted to the regression equation by the above procedure, the current regression equation gives the best approximation to the observations for the given values of the F-distribution. The stepwise procedure is terminated at this stage, and the regression coefficients are then calculated from the equation

$$b_i = d_{i\ell} S_\ell / S_i \qquad (i = 1 \text{ to } k) \qquad 4.2.14$$

where k is the number of independent variables included in the regression equation, excluding the regression constant. The regression constant is then calculated from the equation

$$b_o = \bar{x}_\ell - \sum_{i=1}^{k} b_i \bar{x}_i \qquad 4.2.15$$

and hence the estimated value of the dependent variable, x_ℓ, is given, for any values of the independent variables, by substituting the values of b_i, as given by equations 4.2.14 and 4.2.15, into equation 4.2.1.

Once the regression coefficients are determined, other statistical quantities can be calculated, most of which are not required for the purposes of curve-fitting. The quantities of importance in curve-fitting include the residual mean square error given by

$$RMS = \frac{m-1}{m-k-1} \; S_\ell^2 \; d_{\ell\ell} \qquad 4.2.16$$

the regression estimates at each observation given by

$$\hat{x}_{\ell j} = b_o + \sum_{i=1}^{k} b_i \, x_{ij} \qquad (j = 1 \text{ to } m) \qquad 4.2.17$$

and the residual error at each observation given by

$$\varepsilon_j = x_{\ell j} - \hat{x}_{\ell j} \qquad (j = 1 \text{ to } m) \qquad 4.2.18$$

where ℓ is the subscript of the dependent variable. The residual error gives a guide to the accuracy of the regression equation, and is kept as small as possible.

When the regression equation corresponding to one dependent variable has been determined, together with all the associated calculations, the multiple regression procedure then considers the next dependent variable. Since all the preliminary calculations are performed before the start of the multiple regression procedure, they need not be repeated, and the procedure is repeated starting from equation 4.2.13. When all dependent variables have been considered and their corresponding regression equations determined, the procedure is terminated.

4.3 Investigation of the type of Analytic Expressions Considered

There are two possible analytic expressions which are considered, for use in the multiple regression routine, to be fitted to the energy

bands; a polynomial consisting of even powers only, and a Fourier series consisting of cosine terms only. Thus the polynomial is given by

$$y = a_o + \sum_{n=1}^{m} a_n x^{2n} \qquad (0 \leqslant x \leqslant 1) \qquad 4.3.1$$

and the Fourier series is given by

$$y = a_o + \sum_{n=1}^{m} a_n \cos\left[\frac{n\pi x}{\ell}\right] \qquad (0 \leqslant x \leqslant 1) \qquad 4.3.2$$

where ℓ is half the period of the curve being considered, which for realistic energy bands can take the values $\ell = 1$ or $\ell = 2$. In the curve-fitting routine, it is required to approximate to the true curve as accurately as possible, and also to approximate to the first and second derivatives of the true curve accurately. This is done in order to obtain accurate threshold values as calculated by the Envelope Method presented in the previous chapter, since the Envelope Method makes use of the derivatives of the energy bands.

The accuracy of the two analytic expressions considered, given by 4.3.1 and 4.3.2, is investigated before deciding which one to use. This is done by comparing the values, first and second derivatives of the approximating equations with those of the curve given by

$$y = x \sin(n\pi x) \qquad (0 \leqslant x \leqslant 1) \qquad 4.3.3$$

The first case considered is with $n = 1$ in 4.3.3, which corresponds to simple forms of energy bands having at most two extrema values and one point of inflexion. In order to fit the analytic expressions given by 4.3.1 and by 4.3.2 (with $\ell=2$) to the function given by 4.3.3, eleven equi-spaced data points are taken for use in the multiple regression routine described in the previous section.

The graphs of the resulting regression equations are shown in figure 4.1(a), and both the graphs of the analytic expressions are

seen to be in excellent agreement with the true curve. The graphs of the first and second derivatives of the regression equations and the true function are shown in figures 4.1(b) and 4.1(c) respectively, as good agreement is also required in these quantities. Indeed, it is seen that the agreement between both analytic expressions and the true function is very good. Thus, for energy bands having this sort of shape, either of the two analytic expressions considered will approximate accurately the true energy band, its first and second derivatives.

The next case to consider is with n = 2 in equation 4.3.3, which corresponds to a more realistic

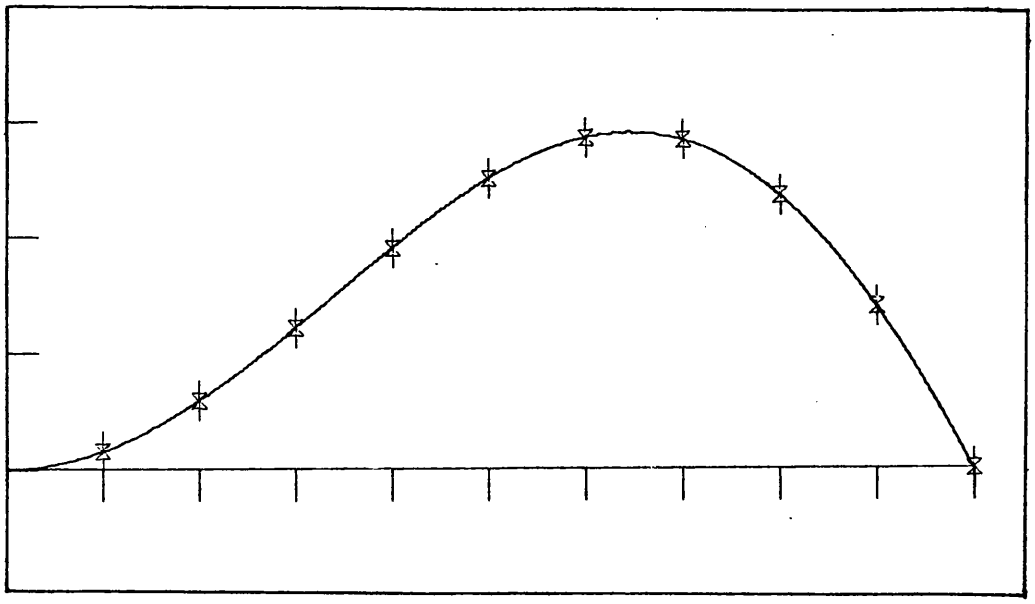


Figure 4.1(a) Plot of curves fitted to the function $f(x) = x \sin(\pi x)$, with 11 equi-spaced points between 0 and 1.

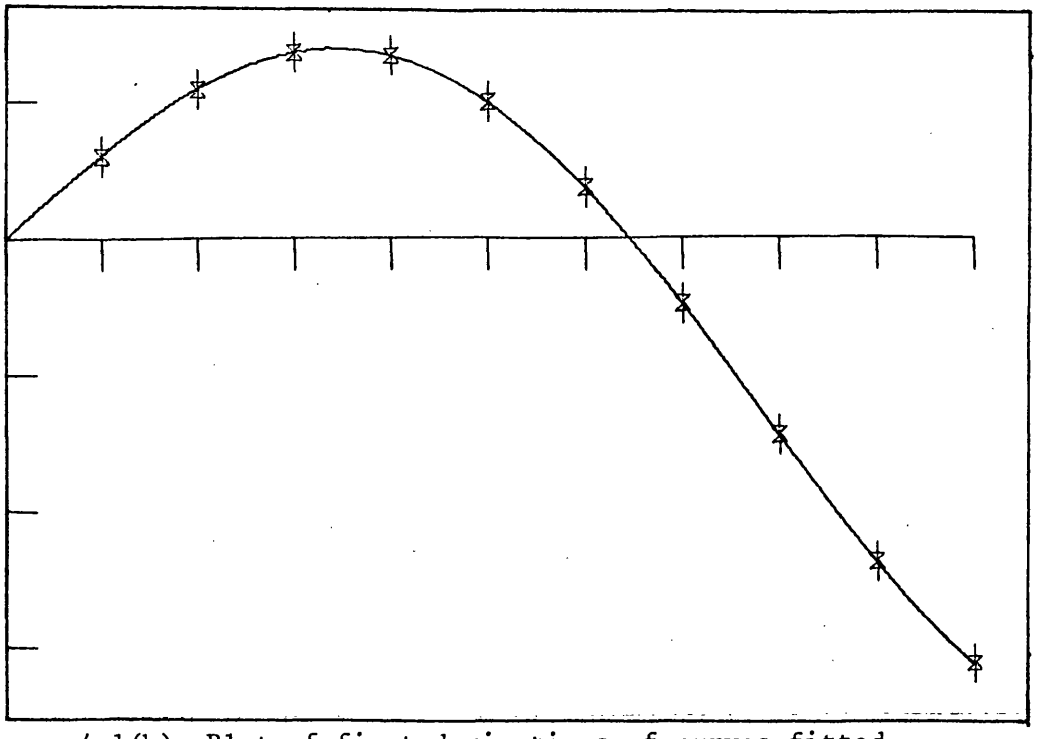


Figure 4.1(b) Plot of first derivatives of curves fitted to the function $f(x) = x \sin(\pi x)$, with 11 equi-spaced points between 0 and 1.

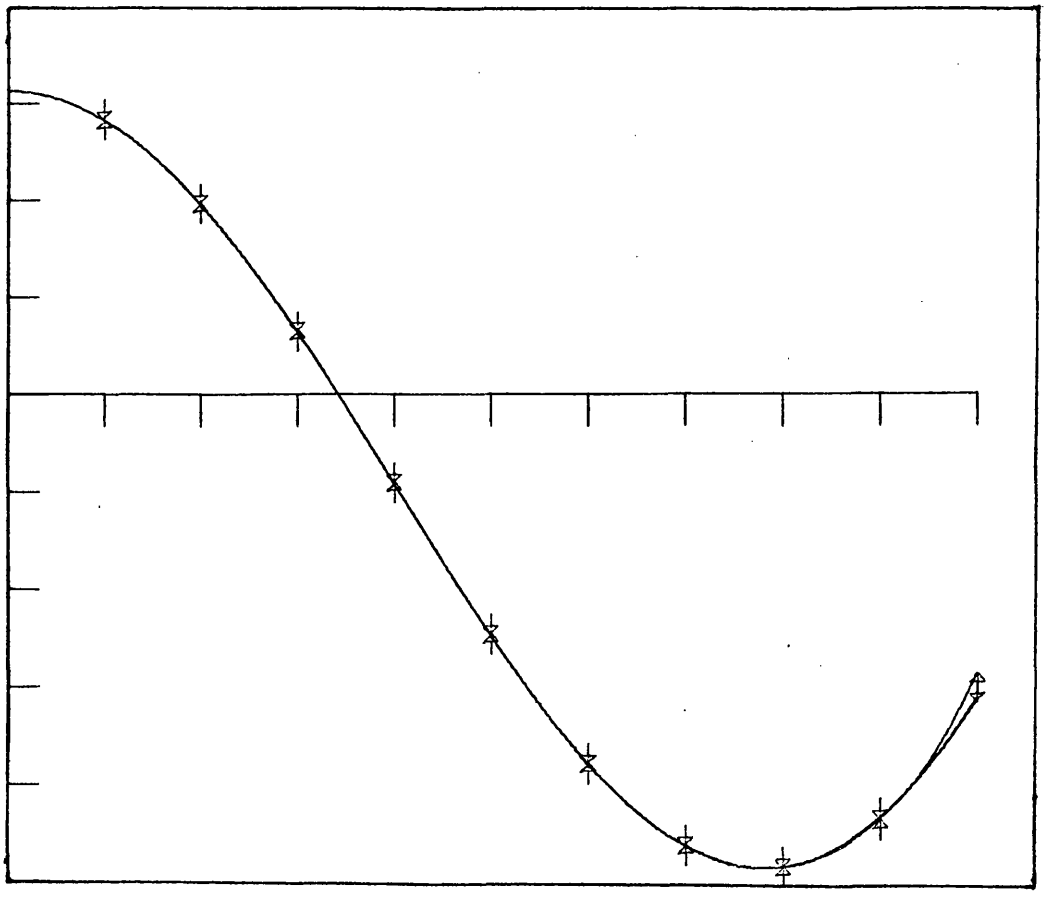


Figure 4.1(c) Plot of second derivatives of curves fitted to the function $f(x) = x \sin(\pi x)$, with 11 equi-spaced points between 0 and 1.

shape for an energy band, having three extrema values and at least two points of inflexion. Again, eleven equi-spaced data points are used in order to fit the analytic expressions given by 4.3.1 and 4.3.2 (again with $\ell = 2$) to the function given by 4.3.3. The graphs of the resulting regression equations are drawn, together with that of the true function, and also the graphs of the first and second derivatives, corresponding to all three analytic expressions, are drawn. These are shown in figures 4.2(a), (b) and (c) respectively, and it is seen that the results for the first two graphs are similar to those of the first case; that is excellent agreement is obtained between the values of the functions, and very good agreement between the corresponding first derivatives.

However, for the second derivatives, there is excellent agreement between the Fourier series and the true function, but the agreement with the polynomial is not particularly good. While it is in good agreement over the smaller values in the

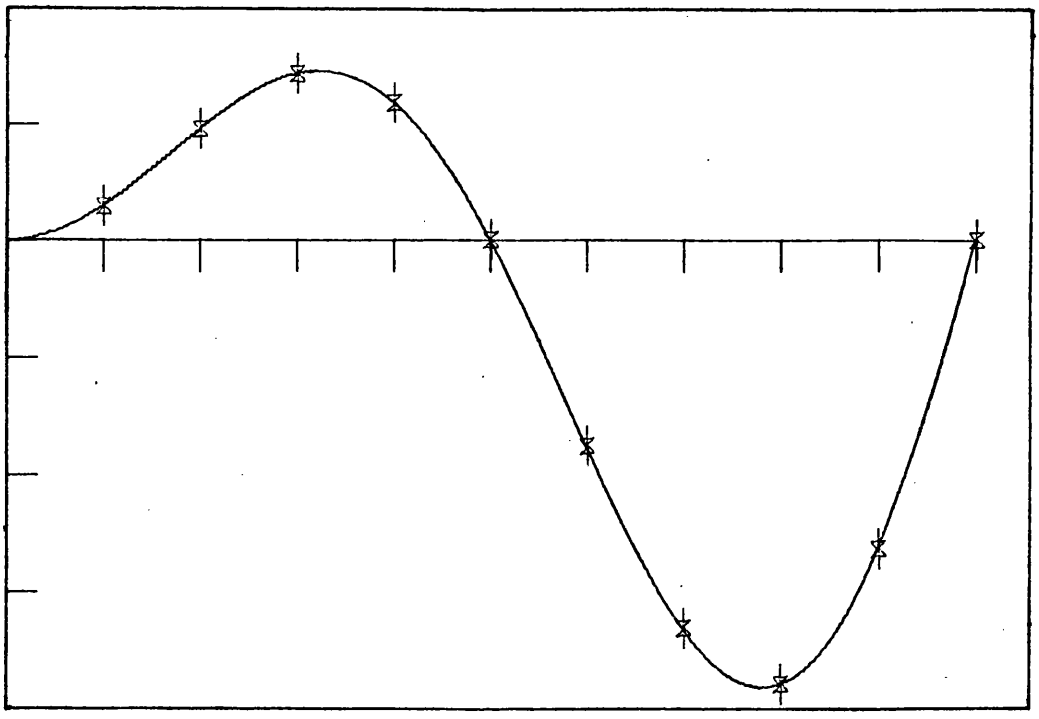


Figure 4.2(a) Plot of curves fitted to the function $f(x) = x \sin(2\pi x)$, with 11 equi-spaced points between 0 and 1.

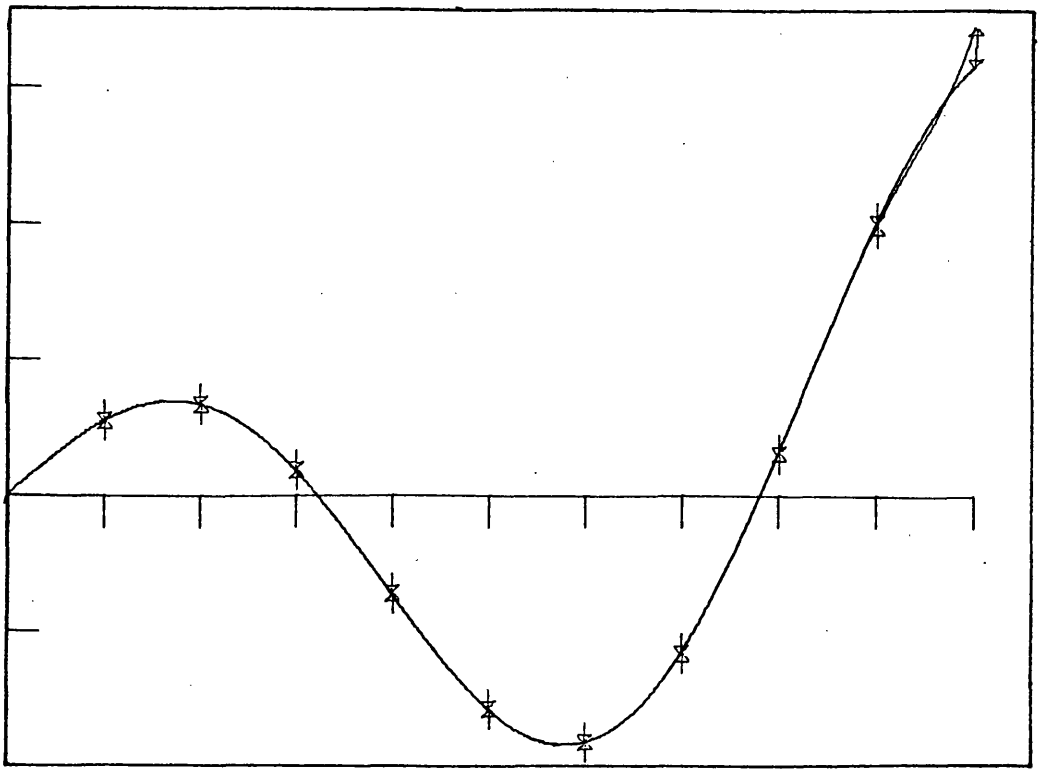


Figure 4.2(b) Plot of first derivatives of curves fitted to the function $f(x) = x \sin(2\pi x)$, with 11 equi-spaced points between 0 and 1.

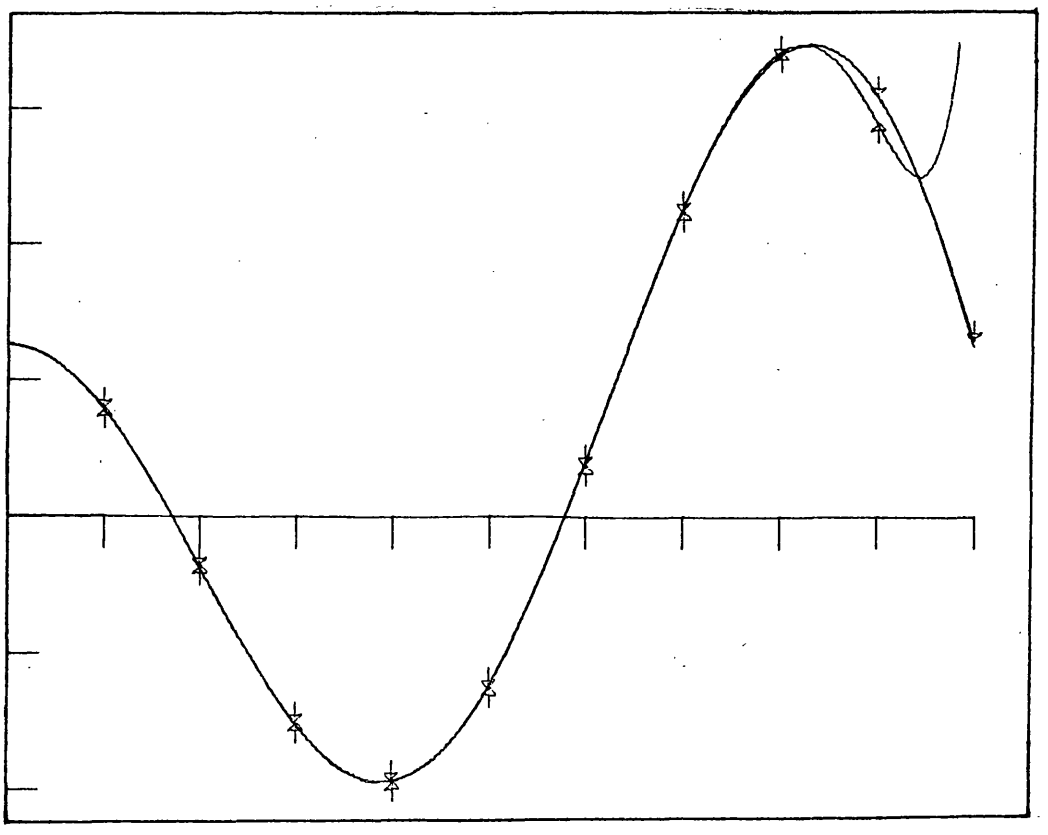


Figure 4.2(c) Plot of second derivatives of curves fitted to the function $f(x) = x \sin(2\pi x)$, with 11 equi-spaced points between 0 and 1.

range of values of the graph, it is a bad approximation at the higher values in the range. This seems to suggest that the polynomial expression is not very good to use as the approximation to the energy bands, but the Fourier series is an excellent approximation.

With only eleven data points being used, the multiple regression routine allows at most nine terms to be taken in the analytic expression; the constant, a_o, and terms up to and including m = 8 in equations 4.3.1 and 4.3.2. The accuracy of the polynomial expression may increase if more data points are taken for use in the multiple regression routine, and hence the number of terms allowed to be taken in the analytic expressions. By taking 21 data points for use in the multiple regression routine, the number of terms allowed in the analytic expressions is at most 19. However, further restrictions in the multiple regression routine allow at most 16 terms to be taken in the Fourier series, and so the same number of terms is also taken in the polynomial expansion.

The third case considered is thus that with n = 2 in equation 4.3.3, as in the previous case, but with 21 equi-spaced data points used in the multiple regression routine. The procedure followed in the first two cases is thus repeated for this case, and the resulting graphs of function values, first and second derivatives are drawn for all three analytic expressions, and are shown in figures 4.3(a), (b) and (c) respectively. Once more, the curves corresponding to the Fourier series are in excellent agreement with those corresponding to the true function.

The curves corresponding to the polynomial

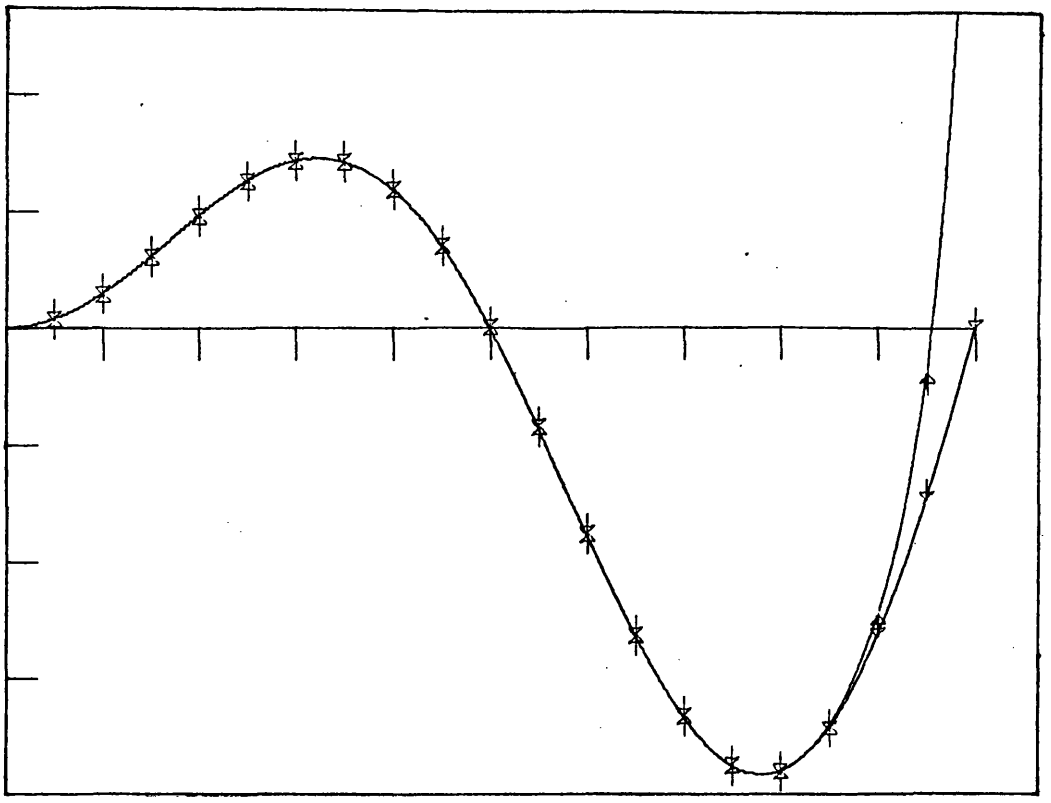


Figure 4.3(a) Plot of curves fitted to the function $f(x) = x \sin(2\pi x)$, with 21 equi-spaced points between 0 and 1.

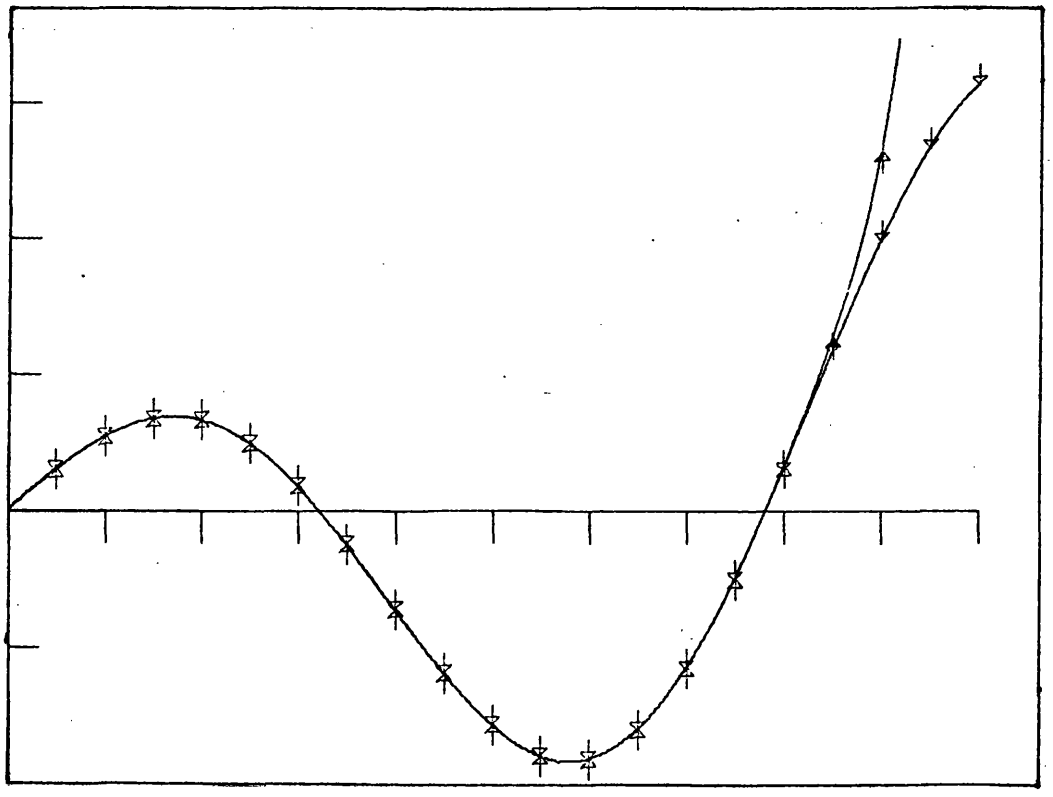


Figure 4.3(b) Plot of first derivatives of curves fitted to the function $f(x) = x \sin(2\pi x)$, with 21 equi-spaced points between 0 and 1.

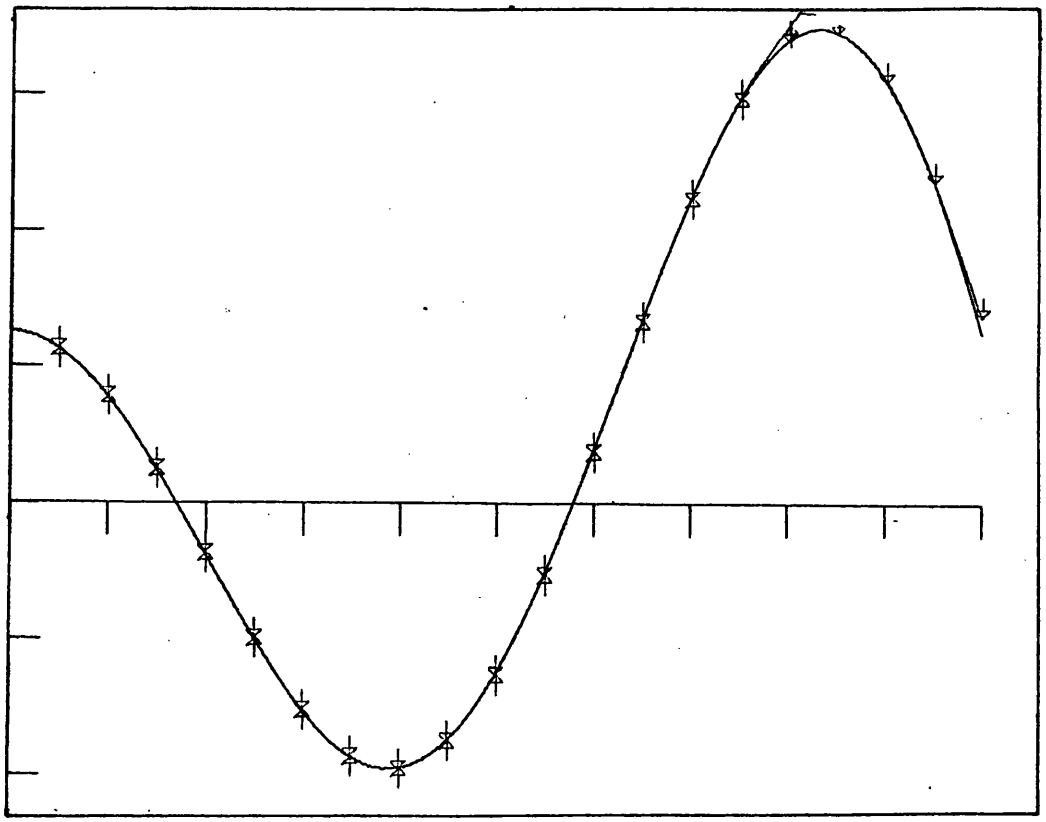


Figure 4.3(c) Plot of second derivatives of curves fitted to the function $f(x) = x \sin(2\pi x)$, with 21 equi-spaced points between 0 and 1.

expression however, are in worse agreement with those corresponding to the true function than in the previous case when only 11 data points were used. The approximation is still good over the smaller values in the range of values of the graphs, but is not at all good at the higher values in the range. Thus, by taking more data points, and hence more terms in the polynomial expansion than were taken previously, the resulting approximation to the true function is worse, not better, than before .

Hence, from all three cases in which the Fourier series of 4.3.2 (with $\ell = 2$) is used to approximate the function given by 4.3.3, excellent agreement is achieved in the values of the function, first and second derivatives. The polynomial expression only gives excellent

agreement in the values of the function, first and second derivatives for the first case, and for each successive case the polynomial approximation becomes worse. Thus, from these results, it is seen that the type of analytic expression to use in approximating the energy bands is a Fourier series, and the form given by equation 4.3.2 will be used in the multiple regression routine.

4.4 Graphical Method of Approximating the Impact Ionization Threshold Values

When analytic expressions are used to approximate the energy bands using a set of discrete data points, errors are certain to occur. While these errors can be minimized at the data point, the errors at intermediate points cannot be guaranteed to be as small. There is a possibility that, while good agreement is achieved at the data points, large variations may occur between the data points. To ensure this sort of situation does not remain unnoticed, a graph of the energy bands, together with all the data points, is drawn using a graph plotter attached to a digital computer. If an analytic expression is seen to draw an energy band which is not in agreement with the energy band produced by the calculations of the author whose work is being reproduced, then the analytic expression is rejected and a new one calculated.

Once the analytic expressions approximating the energy bands are obtained to sufficient accuracy, the calculation of the impact ionization threshold positions can be performed. The computer program written to calculate the impact ionization threshold values, as described briefly in the previous chapter, requires as input data, an approximation to the initial $\mathbf{k}$ coordinates of the hot and promoted electrons, or holes. These positions are provided by a graphical technique, for which a computer program is written to draw the required

graphs. This program also includes the procedure for drawing the energy bands and data points which show whether the analytic expressions being used are sufficiently accurate.

The graphical technique requires three graphs to be drawn, two of which are the envelope patterns of the conduction and valence bands. The third graph is of the energy bands in an extended zone scheme, extending over two complete zones, and additionally, a second set of curves which do not occur in practice. These additional curves correspond to the reflection of the conduction bands about the conduction band minimum, and to the reflection of the valence bands about the valence band maximum. These curves are the equivalent constructions, in one dimension, to the reflected paraboloid constructed by Franz [10], and are constructed in a similar manner.

If a hot electron is initially at a point H in a conduction band with coordinates (k_h, E_h), then the lowest possible position that the envelope, of the lowest lying conduction band, can take is given by displacing the point on the envelope corresponding to the conduction band minimum by the vector $\vec{HM}$, where M is the position of the conduction band minimum, from the conduction band minimum. If the coordinates of the conduction band minimum are (k_m, E_m), then for every point H in a conduction band, there exists a point A, displaced from M by the vector $\vec{HM}$, with coordinates $(2k_m-k_h, 2E_m-E_h)$ on which the minimum of the envelope is centred. The curve traced by all such points is the reflection of the corresponding conduction band about the conduction band minimum. The same procedure applies to a hot hole initially in a valence band, for which the curves are the reflections of the corresponding valence bands about the valence band maximum.

The method of determining approximate positions of impact ionization thresholds, as described by Franz, can now be employed. This method

involves moving the minimum of the envelope around the reflected conduction bands until a point on the envelope just touches a point on a valence band. When this situation arises, the initial position of the hot electron corresponds to a threshold position. The **k** coordinates of the envelope minimum and the point at which the envelope and valence band just touch are then read off the graph, from which the initial **k** coordinates of the hot and promoted electrons are determined. Once this is done, the same procedure is repeated for another threshold situation, and is repeated until all possible threshold situations have been determined. When all such threshold positions have been determined, the values are used as input data to the computer program which then calculates the positions more accurately.

4.5 Some Computational Considerations

A computer program to perform the multiple regression calculations described in section 4.2 is provided as a standard program by I.C.L. [58] for use on their Systems 4 computers, and is therefore used in the present work. The program is written in a manner in which the operations required to be performed are specified by submitting the appropriate data input. In addition to the basic multiple regression calculations, other connected calculations are included, if specified, and the amount of results produced is varied depending upon the data input. The two main parts of the data input are the observations of the dependent variables and the independent variables, and the required regression equations.

The independent variable is the coordinate of the point in the first Brillouin zone at which the energy values are calculated, and is scaled to be the proportional distance from the centre of the zone to the edge of the zone. That is, all values of the independent variable lie between 0 and 1, and are chosen to be equi-spaced along the chosen

symmetry direction in **k**-space. The dependent variables are the energy values at these points in **k**-space corresponding to the energy bands being investigated. The regression equation is supplied as data by means of a series of transformations of the independent variable. The number of terms allowed in the regression equation is limited by the number of observations, but is limited even more by the number of transformations the computer program allows.

For the Fourier series given by 4.3.2, two transformations for each term are required to obtain the required equation. The first transformation multiplies the value of the independent variable, x, by the appropriate constant to obtain the new variable

$$u_i = \frac{i\pi x}{\ell} \qquad (i = 1 \text{ to } n) \qquad 4.5.1$$

where ℓ takes the value of either 1 or 2. The cosine of this transformed variable is then taken as the second transformation to obtain another new variable

$$Z_i = \cos(u_i) \qquad (i = 1 \text{ to } n) \qquad 4.5.2.$$

from which the regression equation is given by

$$y = b_o + \sum_{i=1}^{n} b_i Z_i \quad .$$

The number of terms allowed in the regression equation, due to the restriction on the number of transformations allowed, is 16, that is the regression constant plus 15 cosine terms.

The two different values of ℓ which are used in 4.5.1, provide the choice of three different series which can be used to approximate a given energy band. If the energy band has a minimum at the zone boundary, then the series with $\ell = 1$ is chosen to ensure that the minimum is reproduced, and all values if i between 1 and 15 are included. When the gradient of the energy band at the zone edge is

not zero, then one of the series with $\ell = 2$ is chosen. The two possible series correspond to the choice of the values of i which are taken, one series having all values of i between 1 and 15, and the other having only odd values of i between 1 and 29.

Once the regression equation has been calculated by the computer program, the regression coefficients and the residual errors at each observation are printed out. By studying the residual errors, a guide to the expected accuracy is obtained, and a regression equation is accepted as being sufficiently accurate at this stage of the calculations if all the residual errors are less than about 0.01eV. Further details of the multiple regression procedure are given in Chapter 12 of reference [58], which includes some of the options available for the computer program. All other options, including the permissible transformations, are given in Chapter 2 of reference [58], and in Chapter 1 is described how to run the program.

A computer program is written to draw the energy bands and the other curves related to the calculation of the approximate positions of impact ionization thresholds. This program requires as input data, all the observations used in the multiple regression program, together with the regression coefficients of the Fourier series corresponding to all the energy bands being considered. Options are available to specify which graphs are to be drawn by the program. Further computational details are given in Chapter 7.

5. THE COULOMB INTERACTION MATRIX ELEMENTS

5.1 Introduction

Associated with every collision process in semiconductors, there is a probability of that collision occurring. One of the factors occurring in the probability is the matrix element governed by the states of the particles involved in the collision. For impact ionization collisions, both by electrons and by holes, the matrix element is that of the coulomb interaction between the initial and final states involved. For the impact ionization thresholds determined by the method presented in Chapter 3, the sizes of the corresponding matrix elements are calculated by the method presented here.

The method used to perform these calculations comes from the theory by Beattie and Landsberg [17] which has been widely used [34-37]. In the analysis of the matrix element, which is based on a one-electron approximation to the wave-function, a quadruple sum over reciprocal lattice vectors is obtained. By the reasons given by Beattie and Landsberg, this quadruple sum can be reduced to the product of two double sums, and Umklapp processes can be neglected. The analysis of this method is presented in the next section, together with a discussion of the conclusions reached by Beattie and Landsberg concerning the importance of Umklapp processes. Some computational considerations are presented in section 5.3, which include the simplification made by Beattie and Landsberg.

In Chapter 2, the analysis of the method of calculating the energy eigenvalues and corresponding eigenvectors, of any cubic semiconductor, at any point within the first Brillouin zone, was presented. This analysis included the recovery of the original expansion of the eigenvectors from just the basis expansion, which was done through the

perturbation theory developed by Löwdin [54]. The differences in using the original full expansion of the eigenvectors and the basis expansion is investigated in a pilot study. The results of this pilot study are presented in section 5.4, giving the comparisons between the two expansions used. The computational time required for each expansion is also investigated, and it is shown that the original expansion of the eigenvectors can be neglected. On the results of this pilot study, the matrix elements associated with all impact ionization thresholds will be calculated by using the basis expansion of the eigenvectors only.

5.2 Analysis of the Coulomb Interaction Matrix Elements

The method used in the present work to determine the size of the matrix element corresponding to an impact ionization threshold transition, is that developed by Beattie and Landsberg [17], and later reiterated by Landsberg [35]. The theory is based on the states of the crystal involved in the transition being described by orthonormal one-electron functions. Only the electrons which partake in the transition are assumed to have their states altered, while all other electron states are assumed to be unaffected. The perturbation operator, U, can then have a non-zero matrix element only for the term involving the coulomb interation. That is,

$$U(\mathbf{r}_1,\mathbf{r}_2) = \frac{e^2 \exp\{-\lambda|\mathbf{r}_1-\mathbf{r}_2|\}}{\varepsilon|\mathbf{r}_1-\mathbf{r}_2|} \qquad 5.2.1.$$

where λ^{-1} is the screening radius, ε is the dielectric constant (equal to the square of the refractive index), and $\mathbf{r}_1$ and $\mathbf{r}_2$ are the initial positions (in real space) of the electrons involved in the collision.

The matrix element, after summation over spin variables, can

then be written as

$$U_{if} = \iint \{ \psi_1^*(\mathbf{r}_1)\psi_2^*(\mathbf{r}_2)\Delta_{21} - \psi_2^*(\mathbf{r}_1)\psi_1^*(\mathbf{r}_2)\Delta_{12} \} \times$$

$$\times U(\mathbf{r}_1,\mathbf{r}_2)\psi_3(\mathbf{r}_1)\psi_4(\mathbf{r}_2)\, d_1\, d_2$$

$$\equiv M_1\,\Delta_{21} - M_2\,\Delta_{12} \qquad 5.2.2.$$

where $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ are the states of the hot and promoted electrons respectively before the collision, $\psi_3(\mathbf{r})$ and $\psi_4(\mathbf{r})$ are the states of the hot and promoted electrons respectively after the collision. Here, Δ_{21} and Δ_{12} are the spin variables, the possible assignments for which are given in Table 5.1, using the convention of Landsberg [35], that the initial state of the hot electron always refers to "spin up".

Table 5.1

Spin assignments

Initial States 1	Initial States 2	Final States 3	Final States 4	Δ_{21}	Δ_{12}	U_{if}	
↑	↑	↑	↑	1	1	M_1-M_2	Like spins
↑	↓	↑	↓	1	0	M_1	Unlike spins
↑	↓	↓	↑	0	1	M_2	Unlike spins
↑	other assignments			0	0	0	

If the electron states are represented by a sum of plane-waves, namely

$$\psi_n(\mathbf{r}) = V^{-\frac{1}{2}} \sum_m a_{n,m}\, e^{i(\mathbf{k}+\mathbf{K}_m).\mathbf{r}} \qquad n = 1,2,3,4. \qquad 5.2.3$$

where $\mathbf{K}_m$ is a reciprocal lattice vector and $\mathbf{k}$ is a reciprocal lattice vector within the first Brillouin zone, then the first integral of the matrix element given by 5.2.2. can be written

$$M_1 = \frac{e^2}{\varepsilon V^2}\iint\{\sum_i a^*_{1,i}\ e^{-i(k_1+K_i).r_1}\sum_j a^*_{2,j}\ e^{-i(k_2+K_j).r_2}\}\times$$

$$\times\frac{e^{-\lambda|r_1-r_2|}}{|r_1-r_2|}\{\sum_\ell a_{3,\ell}\ e^{i(k_3+K_\ell).r_1}\sum_m a_{4,m}\ e^{i(k_4+K_m).r_2}\}d_1 d_2 \qquad 5.2.4$$

Performing a fourier analysis on the coulomb potential, that is

$$\frac{e^{-\lambda r}}{r} = \frac{4\pi}{V}\sum_k \frac{e^{ik.r}}{k^2+\lambda^2}$$

substituting into 5.2.4 and using the orthogonality property of plane-waves, gives

$$M_1 = \frac{4\pi e^2}{\varepsilon V}\sum_{i,j,\ell,m}\frac{a^*_{1,i}a^*_{2,j}a_{3,\ell}a_{4,m}}{|k_1-k_3+K_i-K_\ell|^2+\lambda^2}\cdot\delta(k_1+k_2-k_3-k_4+K_i+K_j-K_\ell-K_m) \qquad 5.2.5$$

Thus the matrix element is determined from a quadruple sum over reciprocal lattice vectors and involving the coefficients of the plane-wave representations of the electron states involved in the collision. The terms in this sum can be rearranged in the fashion used by Beattie and Landsberg, and written as the sum of three summations, namely

$$M_1 = M_{1a} + M_{1b} + M_{1c} \qquad 5.2.6$$

where

$$\left.\begin{array}{ll} \text{for } M_{1a} & K_i = K_\ell,\ K_j = K_m \\ \text{for } M_{1b} & K_i + K_j = K_\ell + K_m,\ K_i \neq K_\ell \\ \text{for } M_{1c} & K_i + K_j \neq K_\ell + K_m,\ K_i \neq K_\ell \end{array}\right\} \qquad 5.2.7$$

Also, introducing the notation

$$\left.\begin{aligned} F_{13} &= \sum_i a^*_{1,i}\, a_{3,i} \\ F_{24} &= \sum_i a^*_{2,i} a_{4,i} \end{aligned}\right\} \qquad 5.2.8$$

it is seen that the first sum of 5.2.6, M_{1a}, can be written

$$M_{1a} = \frac{4\pi e^2}{\varepsilon V} \frac{F_{13}\, F_{24}}{|\mathbf{k}_1-\mathbf{k}_3|^2+\lambda^2}\, \delta(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3-\mathbf{k}_4) \quad . \qquad 5.2.9$$

A completely similar argument can be applied to the second integral, M_2, of the matrix element, giving equations similar to 5.2.5, 5.2.8 and 5.2.9, namely

$$M_2 = \frac{4\pi e^2}{\varepsilon V} \sum_{i,j,\ell,m} \frac{a^*_{1,i}\, a^*_{2,j}\, a_{3,\ell}\, a_{4,m}}{|\mathbf{k}_1-\mathbf{k}_4+\mathbf{K}_i-\mathbf{K}_m|^2 + \lambda^2}\, \delta(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3-\mathbf{k}_4+\mathbf{K}_i+\mathbf{K}_j-\mathbf{K}_\ell-\mathbf{K}_m) \qquad 5.2.10$$

$$\left.\begin{aligned} F_{14} &= \sum_i a^*_{1,i}\, a_{4,i} \\ F_{23} &= \sum_i a^*_{2,i}\, a_{3,i} \end{aligned}\right\} \qquad 5.2.11$$

$$M_{2a} = \frac{4\pi e^2}{\varepsilon V} \frac{F_{14} F_{23}}{|\mathbf{k}_1-\mathbf{k}_4|^2+\lambda^2} \cdot \delta(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3-\mathbf{k}_4) \qquad 5.2.12$$

where the terms in the sum M_2 are rearranged in exactly the same manner as those of M_1.

An approximation to the quadruple sum of 5.2.5 is thus obtained, provided it can be shown that the sums M_{1b} and M_{1c} are negligible. Beattie and Landsberg [17] give reasons why these sums can be neglected, and Landsberg [35] gives the same reasons. The sum M_{1b} can be neglected because the denominator of 5.2.5 is then of the form $|\mathbf{k}_1-\mathbf{k}_3+\mathbf{L}|^2$, where $\mathbf{L}$ is a non-zero reciprocal lattice vector, and is thus smaller than M_{1a} by several powers of 10. The sum M_{1c}, which corresponds to Umklapp

processes, can be neglected because of the same reason, and also because, in direct gap semiconductors, the initial and final states involved are pushed far from the band edge, and so are weighted with very small Fermi-Dirac probabilities.

The assumptions concerning the sums M_{1b} and M_{1c} are made without concerning the numerator of 5.2.5, namely the product of the plane-wave coefficients $a_{i,j}(i = 1,2,3,4)$. If the coefficients corresponding to the terms occurring in the sum M_{1a} are smaller than those corresponding to the terms occurring in the sums M_{1b} and M_{1c}, then the latter two sums may not be negligible. In this case, the sums M_{1b} and M_{1c} may be of comparable sizes with the sum M_{1a}, or may even be larger. Also, there are many more terms in the sums M_{1b} and M_{1c} than there are in the sum M_{1a}, and even if the coefficients corresponding to the three sums are all of comparable size, then again the sums M_{1b} and M_{1c} may be of comparable sizes with the sum M_{1a}. It is therefore erroneous to neglect these two sums purely on the grounds of a large denominator in all the terms.

The assumption that very small Fermi-Dirac probabilities occur in the sum M_{1c} may be true for direct gap semiconductors with a parabolic conduction band, but is not true for indirect gap semiconductors, or for some direct gap semiconductors where the detailed band structure is taken into consideration. In indirect gap semiconductors, the lowest impact ionization threshold usually corresponds to an Umklapp process, as can be seen from the results of Anderson and Crowell [32], and as will be seen from the results presented later in this work. Also, Umklapp processes are often interlaced with normal processes when the threshold energies are ordered in ascending order of magnitude, even in direct gap semiconductors. It is therefore erroneous to neglect the

sum M_{1c}, and also Umklapp processes in general, on the grounds of being weighted with very small Fermi-Dirac probabilities. The most important threshold energies are the lowest ones, and possibly those, if any, which are close to the lowest threshold energy, irrespective of whether they are Umklapp or normal processes.

For these reasons, the sizes of the matrix elements as calculated from 5.2.5. and 5.2.9 are investigated in a pilot study, the results of which are presented in section 5.4. Consequently, the size of the matrix element is calculated, for all impact ionization thresholds determined, by both the equations, 5.2.5 and 5.2.9. A continual comparison between the sizes of the matrix element, as calculated by these two equation, is thus obtained.

5.3 Some Computational Considerations

In the analysis of the pseudopotential method presented in Chapter 2, the one-electron approximation to the wave-function was made. That is, the wave-function, $\psi_n(\mathbf{r})$, to an electron state with an energy E_n is represented by a sum of plane-waves,

$$\psi_n(\mathbf{r}) = V^{-1/2} \sum_m a_{n,m} e^{i(\mathbf{k}+\mathbf{K}_m).\mathbf{r}}$$

The pseudopotential method requires a system of secular equations to be solved, for which the energies, E_n, are given by the eigenvalues of the resulting secular determinant. The corresponding eigenvectors are then the coefficients of the plane-waves used in the approximation to the wave-function. The same one-electron approximation to the wave-function is also made in the analysis of the matrix element of the coulomb interaction presented in the previous section. The coefficients, $a_{n,m}$, appearing in the summations given by 5.2.5, 5.2.8 and 5.2.9 are thus the eigenvectors, corresponding to the appropriate eigenvalues,

as calculated by the pseudopotential method.

However, the analysis of Chapter 2 gives two possible expansions of the wave-functions which are used in calculating the size of the matrix element. The first is just the expansion of eigenvectors obtained from the basis matrix; which are the coefficients of the plane-waves treated exactly in the pseudopotential method. The second is the original expansion of eigenvectors obtained by applying the perturbation theory developed by Löwdin as presented in section 2.4; which are the coefficients of the plane-waves treated both exactly and through the perturbation theory.

With many more plane-waves involved in the original expansion than in the basis expansion of the wave-function, the corresponding calculation of the matrix element of the coulomb interaction is a much more lengthy process. The improved accuracy obtained by using the original expansion is therefore investigated, for a few impact ionization thresholds, in a pilot study. The results of this pilot study are presented in the next section, which show that the increased accuracy is not significant, and the resulting saving of computer time is vast.

In the computer program written to calculate the size of the matrix element of the coulomb interaction, only the first integral of 5.2.2. is considered, and the calculation is performed in the two ways as given by 5.2.5 and 5.2.9. These values are printed out, for comparison, together with the values of F_{13} and F_{24} as given by equation 5.2.8. The electron states involved in some transitions will be degenerate, so providing more than one matrix element for that transition. When one, or more, of the electron states involved is degenerate, all possible combinations of individual states are investigated. This gives a range of sizes of the matrix element of the coulomb interaction, corresponding to the

different states in which the electrons lie. Further computational details are given in Chapter 7.

5.4 Results of the Pilot Study

To obtain the size of the matrix element of the coulomb interaction corresponding to an impact ionization threshold, the plane-wave coefficients forming the approximation to the wave-functions of the electron states involved are first calculated by the method described in Chapter 2. The original expansion of the eigenvectors is obtained, from which the size of the matrix element is calculated by the two methods described previously. Also, the coefficients corresponding to the basis expansion of the eigenvectors are extracted from the original expansion, and the size of the matrix element is calculated by the same two methods as before. Since the original expansion of eigenvectors is automatically normalized, the basis expansion, therefore, is not, but is normalized manually. The sizes of the matrix elements calculated are presented in Table 5.2, indicating the type of threshold investigated.

It is seen that the sizes of the matrix element corresponding to the threshold in Silicon, as calculated by 5.2.5, are in agreement to the same order of magnitude, while the sizes of the matrix elements corresponding to the thresholds in Germanium, as calculated by 5.2.5, are in excellent agreement. The threshold in Silicon is an Umklapp process, for which the calculation of the size of the matrix element by 5.2.9 gives a value of zero. This clearly is not in agreement with the finite size of the matrix element given by 5.2.5 which is of the same order of magnitude as the sizes of the matrix elements for the thresholds in Germanium. Also, this matrix element corresponds to the lowest impact ionization threshold in Silicon, and is therefore

Table 5.2

The Matrix Elements of the Coulomb Interaction corresponding to Thresholds Investigated

Threshold Investigated	Ratio E_T/E_G	Calculations involving the basis expansion				Calculations involving the original expansion			
		F_{13}	F_{24}	M_1 by 5.2.9	M_1 by 5.2.5	F_{13}	F_{24}	M_1 by 5.2.9	M_1 by 5.2.5
Si. Γ - X Direction Lowest threshold. Umklapp process involving 3 states	1.10	3.04×10^{-1}	1.19×10^{-11}	0	1.42×10^{-10}	3.12×10^{-1}	1.18×10^{-11}	0	1.89×10^{-10}
Ge. Γ - X Direction Lowest Normal process. Involving 4 states	1.28	9.00×10^{-10}	2.75×10^{-12}	4.86×10^{-21}	3.53×10^{-11}	8.77×10^{-10}	2.77×10^{-12}	4.76×10^{-21}	3.26×10^{-11}
Ge. Γ - X Direction Normal process about energy gap at Γ involving 3 states	1.99	9.05×10^{-1}	2.51×10^{-11}	7.49×10^{-10}	7.59×10^{-10}	9.03×10^{-1}	2.50×10^{-11}	7.47×10^{-10}	7.56×10^{-10}

weighted with a Fermi-Dirac probability which is larger than for any other threshold.

In the calculations of Germanium, the first threshold investigated is the lowest threshold corresponding to a normal process. It is seen that the size of the matrix element given by 5.2.9 is not in agreement with that given by 5.2.5, both in the original expansion and in the basis expansion calculations. However, in the calculations of the second threshold in Germainium, the sizes of the matrix element given by 5.2.9 are in excellent agreement with those given by 5.2.5. For all the results, it is seen that the agreement of the original expansion calculations with the basis expansion calculations is excellent.

The calculations of this pilot study are performed on an I.C.L. System 4-50 and 4-70 computer. For the matrix element coresponding to the threshold in Silicon, the time taken for the basis expansion calculations is between 2 and 3 minutes on a 4-50, while the time taken for the original expansion calculations is about 14 hours on a 4-50 plus about 7 hours on a 4-70, which is the equivalent to nearly 70 hours on a 4-50. The computer program written to perform these calculations, however, is inefficient, and is slightly modified to improve the efficiency as much as possible. Even with the improved efficiency obtained, the times taken for the two basis expansion calculations and the two original expansion calculations of Germanium are about 1 minute, 3 minutes, 37 hours and 45 hours respectively on a 4-50. These differences in computer time are not unexpected, as the number of terms summed by 5.2.5 is of the order of $(20)^4$ for the basis expansion calculations, and $(110)^4$ for the original expansion calculations.

Clearly, with the accuracy obtained by using the basis expansion, together with the vast saving in computer time, the basis expansion of the eigenvectors only need be considered. Therefore, on the basis of this pilot study, the sizes of the matrix element, for all impact ionization thresholds determined, will be calculated using the basis expansion of the eigenvectors. Also, it is seen that the assumptions made by Beattie and Landsberg in deriving the equation 5.2.9 are not necessarily valid. It is seen that the contribution to the size of the matrix element, M, of the two sums M_{1b} and M_{1c} is not negligible for the first two thresholds investigated, but is in fact dominant. Consequently, the sizes of the matrix element, for all thresholds determined, will be calculated by using the equation 5.2.5, as the values given by 5.2.9 are unreliable. However, calculations by equation 5.2.9 are also performed to give a continual comparison of the sizes of the matrix elements as calculated by the two equations.

6. THE INCREASE IN THE TRANSITION PROBABILITY JUST ABOVE THRESHOLD

6.1 Introduction

Another factor on which the probability of an impact ionization threshold depends, is the number of states which are able to partake in an ionization transition for a hot electron, or hole, just above the threshold energy. This factor determines the rate of increase of the transition probability as a hot electron increases its energy from that at threshold, to an energy just above the threshold energy. No detailed investigations into this factor have been performed, but Dexter [11] has stated that, if parabolic energy bands are assumed, then the transition probability increases quadratically with increasing energy just above threshold.

In the past, it was not known whether this factor proves significant in the total probability of an impact ionization threshold transition. For example, if two different threshold situations have almost equal energies and almost equal matrix elements, then the threshold for which the rate of increase of the transition probability is the greater is more probable to occur than the other threshold. Since there is this uncertainty in the importance of this factor on the total probability, it is therefore investigated in the present work.

A simplified calculation of the rate of increase of the number of valence states able to partake in impact ionization transitions for a hot electron just above threshold is performed, the analysis of which is presented in the next section. These calculations are based upon the assumption that the energy bands are approximated by parabolae in the regions centred on the states involved in the impact ionization threshold transition. The parabolae are obtained from Taylor series expansions of the energy bands involved in the transition. A formula is thus obtained

by which the number of valence states able to partake in impact ionization for a hot electron just above threshold can be calculated.

This formula is programmed for use on the digital computer, and used to obtain graphs of the excess energy above threshold versus the number of valence states able to partake in impact ionization. One graph is thus obtained for each impact ionization threshold, and is compared with the graphs corresponding to all the other thresholds. Some computational considerations are presented in section 6.3, and the results of this investigation for silicon are presented in section 6.4. From these results, it is seen that the rate of increase in the number of valence states able to partake in impact ionization does not vary greatly between thresholds. The small variation is thus considered to be insignificant when compared with the variation in the sizes of the matrix elements of the coulomb interaction. Since this factor in the total probability of a threshold transition is insignificant in silicon, it will not be calculated for the thresholds of the other semiconductors being investigated.

6.2 Analysis of the Number of States able to partake in Impact Ionization just above Threshold

Analytic expressions in the form of even Fourier series are used to represent the energy bands in the calculations of the impact ionization thresholds by the envelope method, as presented in Chapter 3. Once an impact ionization threshold position has been determined by this method, Taylor series expansions about each of the four states involved in the transition can be easily obtained by evaluating the coordinates of the states, and the respective first and second derivatives. Thus, for a hot electron initially in a state with wavevector-energy coordinates (k_h, E_h), the first and second derivatives, E_h' and E_h'' respectively, can be easily calculated from the appropriate Fourier series used to represent

the energy band. The Taylor series expansion about the hot electron state, truncated after the quadratic term, is then given by

$$E_{ih}(k) = \frac{1}{2} E_h''(k-k_h)^2 + E_h'(k-k_h) + E_h \qquad 6.2.1$$

Similar Taylor series expansions are obtained for the final state of the hot electron, and the initial and final states of the promoted electron. Thus, if the coordinates, first and second derivatives of the final state of the hot electron are (k_1,E_1,E_1',E_1'') respectively, and those of the initial and final states of the promoted electron are (k_v,E_v,E_v',E_v'') and (k_2,E_2,E_2',E_2'') respectively, then the corresponding Taylor series are given by

$$E_{fh}(k) = \frac{1}{2} E_1''(k-k_1)^2 + E_1'(k-k_1) + E_1 \qquad 6.2.2$$

$$E_{iv}(k) = \frac{1}{2} E_v''(k-k_v)^2 + E_v'(k-k_v) + E_v \qquad 6.2.3$$

$$E_{fv}(k) = \frac{1}{2} E_2''(k-k_2)^2 + E_2'(k-k_2) + E_2 \qquad 6.2.4$$

Now, in any transition which does not involve interactions with phonons or photons, the energy and wavevector of the states involved must be conserved, that is

$$E_h + E_v = E_1 + E_2 \qquad 6.2.5$$

$$k_h + k_v = k_1 + k_2 \qquad 6.2.6$$

Also, as was shown in the envelope method, the group velocities of three of the four states involved in the transition must be identical at threshold. Since the group velocity of an electron state is just the first derivative of the energy-wavevector curve, this gives

$$E_v' = E_1' = E_2' \qquad 6.2.7$$

With the Taylor series approximations about the states involved in an impact ionization threshold transition given above, the equation for the double envelope, as presented in Chapter 3, becomes

$$E_{nv}(k) = E_{fh}(\tfrac{1}{2}[k+k_h] + K) + E_{fv}(\tfrac{1}{2}[k+k_h] - K) - E_{ih}(k_h) \qquad 6.2.8$$

where K is given by

$$E'_{fh}(\tfrac{1}{2}[k+k_h] + K) = E'_{fv}(\tfrac{1}{2}[k+k_h] - K) \qquad 6.2.9$$

The double envelope given by 6.2.8 touches the valence band at the point (k_v, E_v); that is

$$\left.\begin{aligned} E_{iv}(k_v) - E_{nv}(k_v) &= 0 \\ E'_{iv}(k_v) - E'_{nv}(k_v) &= 0 \end{aligned}\right\} \qquad 6.2.10$$

By increasing the energy of the hot electron by a small amount, δE, the position of the lower boundary of the double envelope is lowered by the same amount. This causes the double envelope to intersect the valence band in two points, k_{v1} and k_{v2}, which gives

$$E_{iv}(k_{v1}) - E_{nv}(k_{v1}) = E_{iv}(k_{v2}) - E_{nv}(k_{v2}) = 0 \qquad 6.2.11$$

The states in the proportion of the valence band lying above the envelope, between the points k_{v1} and k_{v2} are now able to partake in impact ionization transitions. The number of states, N, involved is directly proportional to the amount of wavevector space of the valence band lying above the envelope; that is

$$N \propto |k_{v1} - k_{v2}| \qquad 6.2.12$$

The value of $|k_{v1} - k_{v2}|$, and hence of the number of valence states

able to partake in impact ionization, can be obtained in terms of the known parameters, (k, E, E', E''), of the electron states involved at threshold, and the increase in energy, δE, above threshold of the hot electron.

When the initial hot electron energy is increased by the small amount δE, its initial position moves to the new state with coordinates $(k_h+\delta k, E_h+\delta E)$ in the conduction band given by 6.2.1. The double envelope, given by equation 6.2.8, moves its position by the same amount, and its equation then becomes

$$E_{nv}(k) = E_{fh}\left(\tfrac{1}{2}\left[k+k_h+\delta k\right] + K\right) + E_{fv}\left(\tfrac{1}{2}\left[k+k_h+\delta k\right] - K\right) - E_{ih}(k_h+\delta k) \qquad 6.2.13$$

where the relation between δk and δE is determined from equation 6.2.1 and is given by

$$\delta E = \frac{1}{2} E_h'' \, \delta k^2 + E_h' \, \delta k \qquad 6.2.14$$

By using equation 6.2.9 with $(k_h+\delta k)$ in place of k_h, K is eliminated from equation 6.2.13. Thus, differentiating equations 6.2.2 and 6.2.4, and substituting into the modified equation 6.2.9, gives

$$E_1''\left(\tfrac{1}{2}\left[k+k_h+\delta k\right] + K-k_1\right) + E_1' = E_2''\left(\tfrac{1}{2}\left[k+k_h+\delta k\right] - K-k_1\right) + E_2'$$

But since $E_1' = E_2'$, this equation reduces to the explicit expression for K given by

$$K = \left[k_1E_1'' - k_2E_2'' + \tfrac{1}{2}(k+k_h+\delta k)(E_2'' - E_1'')\right] / (E_1''+E_2'') \qquad 6.2.15$$

Substituting this value of K into equation 6.2.13 gives

$$E_{nv}(k) = E_{fh}\left(\frac{\left[k+k_h+\delta k\right]E_2'' + k_1E_1'' - k_2E_2''}{(E_1'' + E_2'')}\right) + E_{fv}\left(\frac{\left[k+k_h+\delta k\right]E_1''-k_1E_1''+k_2E_2''}{(E_1'' + E_2'')}\right) - E_{ih}(k_h+\delta k)$$

Using equations 6.2.1, 6.2.2 and 6.2.4 in the above equation, and invoking the conservation of energy and wavevector and the equal group velocities as given by equations 6.2.5, 6.2.6 and 6.2.7, gives, after simplification

$$E_{nv}(k) = \frac{E_1'' E_2'' [k+\delta k-k_v]^2}{2(E_1''+E_2'')} + E_1'[k+\delta k-k_v] + E_v - \frac{1}{2} E_h'' \delta k^2 - E_h' \delta k \qquad 6.2.16$$

Now, to determine the number of valence states lying above this envelope, the values of k_{v1} and k_{v2} used in equation 6.2.12 are determined from equation 6.2.11. Thus, employing equations 6.2.3 and 6.2.16 in equation 6.2.11, an equation is obtained which, after simplification, can be written in the form of a quadratic equation in $(k-k_v)$, which has the form

$$a(k-k_v)^2 + 2b(k-k_v) + c = 0 \qquad 6.2.17$$

where $$a = \frac{E_1'' E_2''}{E_1'' + E_2''} - E_v$$

$$b = \frac{E_1'' E_2'' \delta k}{E_1'' + E_2''}$$

and $$c = \left[\frac{E_1'' E_2''}{E_1'' + E_2''} - E_h''\right] \delta k^2 + 2(E_1'-E_h') \delta k$$

Equation 6.2.17 has the two solutions given by

$$(k-k_v) = -\left[\frac{b}{a}\right] \pm \sqrt{\left[\frac{b}{a}\right]^2 - \left[\frac{c}{a}\right]}$$

where a, b and c are given by equations 6.2.18.

These two solutions correspond to the values k_{v1} and k_{v2}, therefore the difference between the two solutions $|k_{v1}-k_{v2}|$ is given by

$$|k_{v1}-k_{v2}| = 2\left[\left[\frac{b}{a}\right]^2 - \left[\frac{c}{a}\right]\right]^{\frac{1}{2}} \qquad 6.2.19$$

Thus, by substituting the values of a, b and c, given by equations 6.2.18, into the equation 6.2.19, after a little algebra and simplification, and by using equation 6.2.14, gives the equation

$$|k_{v1}-k_{v2}|^2 = \frac{8(E''_1+E''_2)}{[E''_1E''_2-E''_v(E''_1+E''_2)]}\left[\frac{[E''_1E''_2(E''_h+E''_v)-E''_hE''_v(E''_1+E''_2)]}{E''_h[E''_1E''_2-E''_v(E''_1+E''_2)]}\delta E - \right.$$

$$\left. -\left[\frac{E''_1E''_2E''_vE'_h}{E''_h[E''_1E''_2-E''_v(E''_1+E''_2)]} + E'_1\right]\delta k\right] \qquad 6.2.20$$

But, by equation 6.2.14, δk can be eliminated from equation 6.2.20, since

$$\delta k = \frac{-E'_h \pm [(E'_h)^2 + 2E''_h\,\delta E]^{\frac{1}{2}}}{E''_h} \qquad 6.2.21$$

Taking the root which gives the smaller increase in k for a corresponding increase in E, that is taking the positive square root sign if E'_h is positive, and the negative square root sign if E'_h is negative, and substituting into equation 6.2.20, gives

$$|k_{v1}-k_{v2}|^2 = \frac{8(E''_1+E''_2)}{[E''_1E''_2-E''_v(E''_1+E''_2)]}\left\{\frac{[E''_1E''_2(E''_h+E''_v)-E''_hE''_v(E''_1+E''_2)]}{E''_h[E''_1E''_2-E''_v(E''_1+E''_2)]}\delta E + \right.$$

$$+\left[\frac{E''_1E''_2E''_vE'_h}{E''_h[E''_1E''_2-E''_v(E''_1+E''_2)]} + E'_1\right]\left[\frac{E'_h}{E''_h} - \right.$$

$$\left.\left. - \frac{\text{sign }(E'_h)}{E''_h}\,[(E'_h)^2+2E''_h\,\delta E]^{\frac{1}{2}}\right]\right\} \qquad 6.2.22$$

This equation gives the value of $|k_{v1}-k_{v2}|$ for any value of δE above the threshold energy. That is, the proportion of

wavevector space of the valence band able to partake in impact ionization transitions corresponding to any increase in energy of the hot electron above threshold. In the derivation of the equation, it was assumed that the impact ionization threshold was given by a double envelope. However, if the threshold is given by a simple envelope, equation 6.2.22 can be simplified slightly. For a simple envelope, the final electron states coincide, and thus $E_{fh}(k) \equiv E_{fv}(k)$ which gives, in equation 6.2.22, $E_1'' = E_2''$. Substituting this into equation 6.2.22 and simplifying gives

$$|k_{v1}-k_{v2}|^2 = \frac{16}{E_h''(E_1''-2E_v'')^2}\left\{\left[E_1''(E_h''+E_v'')-2E_h''E_v''\right]\delta E + \left[\frac{E_1''E_v''E_h'}{E_h''} + E_1'(E_1''-2E_v'')\right]\left[E_h'-\text{sign}(E_h')\left[(E_h')^2+2E_h''\delta E\right]^{\frac{1}{2}}\right]\right\} \qquad 6.2.23$$

6.3 Some Computational Considerations

A computer program is written to draw the graphs of the excess energy above threshold versus the proportion of wavevector space of the valence band able to partake in impact ionization, as determined by equation 6.2.22. Thus, one graph is drawn for each impact ionization threshold initiated by a hot electron. The computer program is also used to draw the equivalent graphs for each impact ionization threshold initiated by a hot hole, which is done by simply reversing the roles of the valence and conduction bands throughout. The graphs drawn are then of the excess energy above threshold versus the proportion of wavevector space of the conduction band able to partake in impact ionization.

In order to obtain an easy comparison between the different rates of increase in the proportion of wavevector space able to partake in

impact ionization just above threshold, more than one graph is drawn in the same figure. The graphs corresponding to all the thresholds determined, initiated both by electrons and by holes, in one symmetry direction in reciprocal lattice space, are drawn in the same figure. Thus, since three symmetry directions are being investigated, the Γ-Δ-X, Γ-Λ-L, and Γ-Σ-K-S-X directions, three figures are produced, each containing as many graphs as there are thresholds in that symmetry direction.

The formula derived in the previous section, by which these graphs are drawn, is approximate, and only valid for small values of δE. The errors involved will be small for sufficiently small values of δE, and will increase as δE increases. While the formula derived is intended as an approximation to the rate of increase in the total probability above threshold, it is desirable to keep the errors to a minimum. Thus, the range of values of δE over which the graphs are drawn is kept reasonably small. However, errors are incurred in determining the impact ionization threshold data, and if the values of δE are restricted too much, then the errors in the threshold data may dominate. The range of values of δE over which the graphs are drawn, must therefore be sufficiently large to overcome the errors associated with the threshold data. Hence, to obtain a set of reliable graphs which show the behaviour of the rate of increase in the total probability above threshold, the range of values of δE is chosen to be $0 \leqslant \delta E \leqslant 0.25$eV. Further computational details are given in Chapter 7.

6.4 Results for Silicon

The impact ionization threshold values are determined by the envelope method, as presented in Chapter 3, and the sizes of the matrix elements of the coulomb interaction between the electron states involved

by the method as described in Chapter 5. The energy and wavevector coordinates, the first and second derivatives are calculated, for each threshold determined, for use in the equation 6.2.22 to enable the graphs of the rate of increase in the total probability to be drawn. The threshold values for silicon are thus determined and the corresponding graphs then drawn. The threshold values and matrix elements are presented in detail in Chapter 8, but a summary is presented in Table 6.1 below, giving the ratio between the threshold energy above the energy band gap

Table 6.1

Rates of Increase in the Total Probability

Symmetry direction	Graph number	Initiating particle	Ratio E_T/E_G	$\lvert k_{v1}-k_{v2}\rvert$ for $\delta E = 0.25$	Largest value of Matrix Element
Γ-Δ-X	3	Electron	1.05	0.233	1.3×10^{-12}
	1	Electron	1.06	0.175	9.5×10^{-2}
	4	Electron	1.47	0.275	6.1×10^{-10}
	2	Electron	1.49	0.203	1.6×10^{-11}
	8	Hole	1.82	0.387	2.2×10^{-2}
	7	Hole	2.18	0.363	7.4×10^{-2}
	6	Hole	2.51	0.340	1.6×10^{-11}
	5	Hole	3.26	0.316	7.4×10^{-2}
Γ-Λ-L	5	Hole	2.90	0.500	5.1×10^{-3}
	4	Electron	3.55	0.337	2.5×10^{-2}
	1	Electron	3.59	0.142	1.5×10^{-10}
	6	Electron	4.03	0.716	7.6×10^{-3}
	2	Electron	4.07	0.226	4.5×10^{-12}
	3	Hole	4.52	0.320	1.6×10^{-2}
Γ-Σ-K-S-X	11	Hole	1.64	0.427	7.2×10^{-14}
	1	Hole	1.73	0.114	2.0×10^{-13}
	10	Electron	2.08	0.281	2.1×10^{-15}
	8	Hole	2.28	0.246	2.6×10^{-13}
	7	Electron	2.46	0.205	5.7×10^{-12}
	5	Electron	2.52	0.137	2.3×10^{-3}
	9	Hole	2.60	0.270	1.7×10^{-2}
	3	Hole	2.62	0.121	8.8×10^{-2}
	2	Hole	3.01	0.115	4.5×10^{-12}
	12	Electron	3.15	0.436	1.5×10^{-15}
	6	Electron	3.96	0.188	9.9×10^{-2}
	4	Electron	4.09	0.122	3.4×10^{-11}

and the energy band gap.

For the Γ-Δ-X symmetry direction, eight thresholds are determined, four being initiated by a hot electron, and four by a hot hole. The graphs corresponding to each threshold are then drawn (see figure 6.1) and labelled in ascending order of rates of increase in the total probability. The proportion of wavevector space able to contain the promoted particle in an ionization process is calculated for a hot electron, or hole, initially at an energy of 0.25eV above threshold, and these values are presented in Table 6.1. It is seen that, for the Γ-Δ-X direction, the difference between the smallest and largest increases in rates is a factor of nearly $2\frac{1}{4}$.

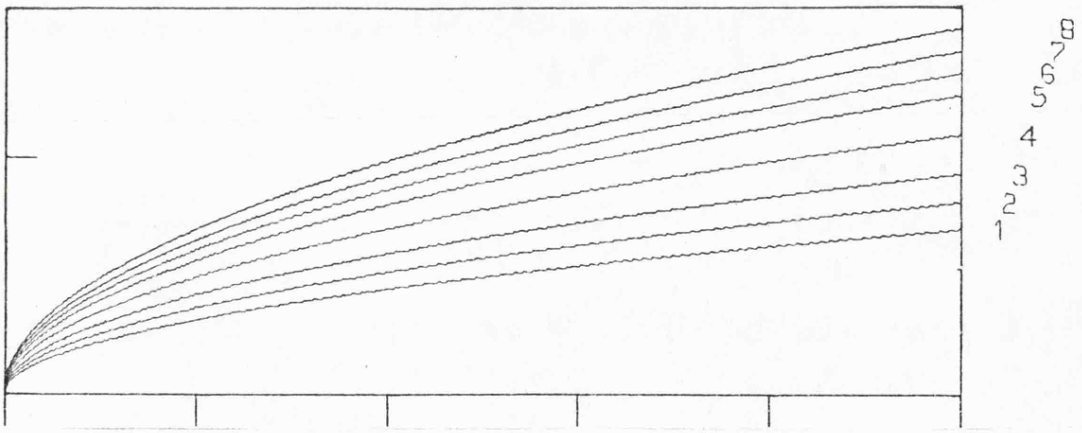


Figure 6.1 Plot of proportion of valence states able to partake in impact ionization versus excess energy above threshold, for thresholds along the Γ-Δ-X axis.

However, the difference between the increases in rates corresponding to thresholds having approximately equal energies is a factor of less than $1\frac{1}{2}$. These small variations, taken on their own, indicate that the rate of increase in the total probability above threshold is not significant.

The same procedure of determining the threshold values and the graphs of the rates of increase in the total probability is applied

to the other two symmetry directions being investigated, the Γ-Λ-L and Γ-Σ-K-S-X directions. For the Γ-Λ-L direction, it is seen from Table 6.1 and figure 6.2 that the difference between the smallest and largest increases in rates is a factor of just over 5, while the difference corresponding to thresholds having approximately equal energies is a factor of less than 3¼. For the Γ-Σ-K-S-X direction, it is seen from Table 6.1 and figure 6.3.that the corresponding differences are a factor of nearly 4 and nearly 2¼ respectively. These variations, while still small, are larger than those in the Γ-Δ-X direction, and it is harder to justify their insignificance in the total probability.

To do this, the sizes of the matrix element of the coulomb interaction between the electron states involved in a threshold transition must be considered. The matrix elements of interest are those which correspond to thresholds with almost equal

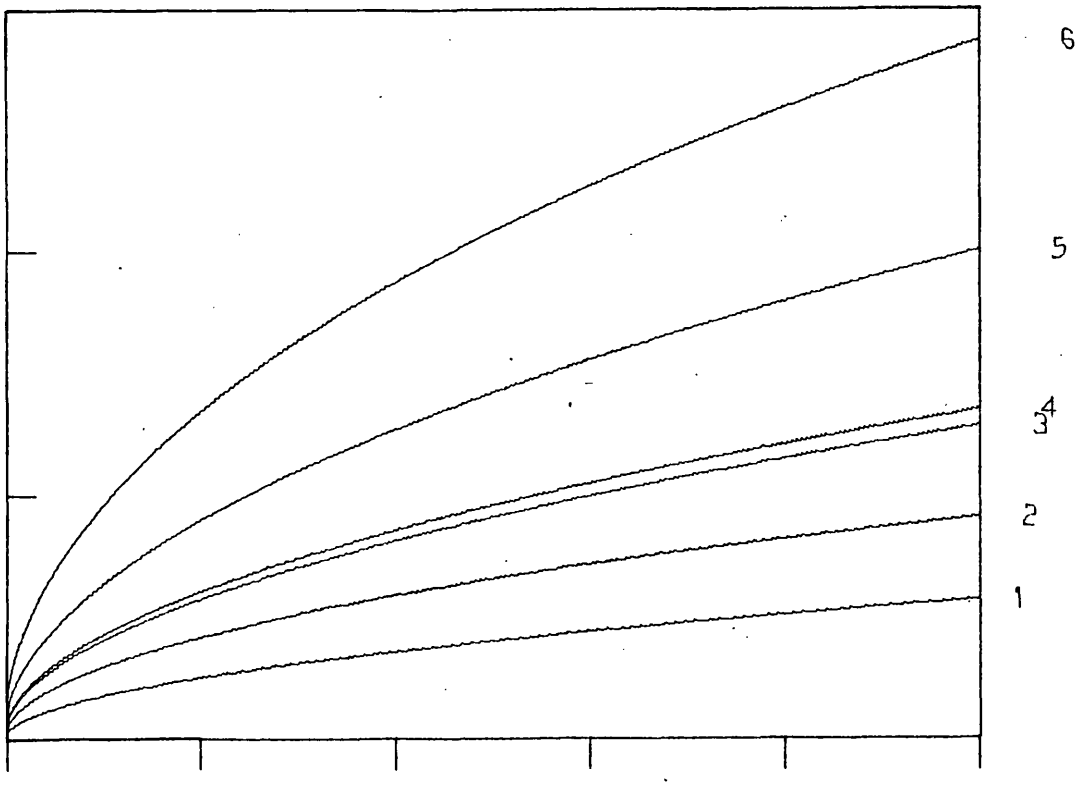


Figure 6.2 Plot of proportion of valence states able to partake in impact ionization versus excess energy above threshold, for thresholds along the Γ-Λ-L axis.

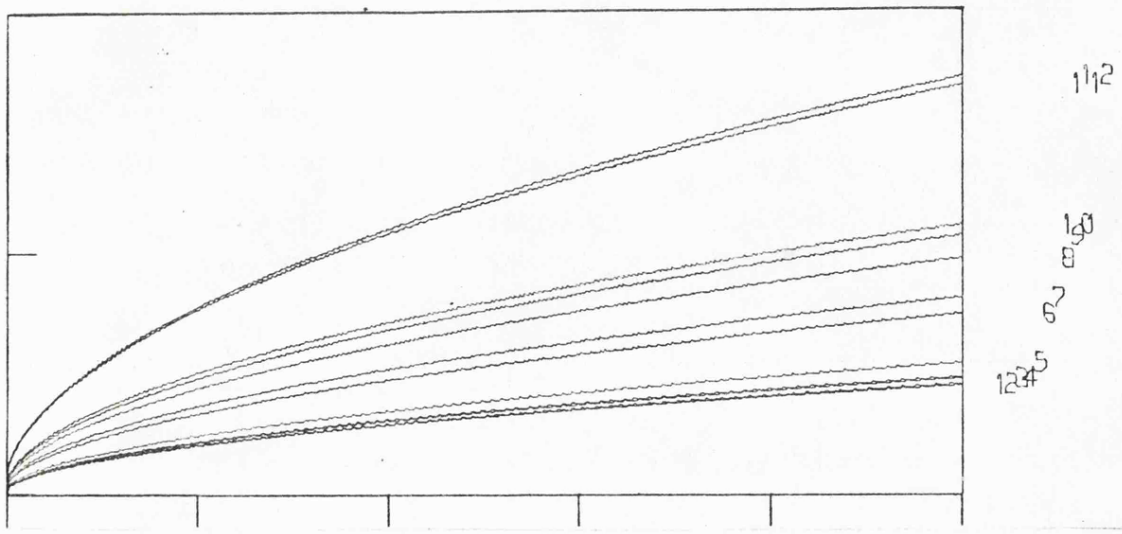


Figure 6.3 Plot of proportion of valence states able to partake in impact ionization versus excess energy above threshold, for thresholds along the Γ-Σ-K-S-X axis.

energies. For the thresholds in question in the Γ-Δ-X direction, the smaller difference between the sizes of matrix elements is a factor of nearly 40, which makes the difference between the increases in rates of the total probability even less significant. In the Γ-Λ-L direction, the difference in the sizes of the matrix elements is a factor of approximately 10^8, which makes the differences in the rates of increase in the total probability totally irrelevant. The thresholds in the Γ-Σ-K-S-X direction which have almost equal energies, have matrix elements differing by a factor greater than 5, which is larger than the difference between the corresponding rates of increase in the total probabilities. Thus, while the rates of increase are not insignificant in this case, they do not contribute in a significant manner to the total probability, since the sizes of the matrix elements have a greater effect.

It is concluded that, for silicon, the rate of increase in the total probability just above threshold is not a significant factor in

the probability that one impact ionization threshold will occur in preference to another threshold. On the basis of these results for silicon, the number of states able to partake in impact ionization for a hot electron, or hole, just above threshold will not be calculated for any of the other semiconductors being investigated.

7. SOME DETAILS OF THE COMPUTER PROGRAMS USED IN THE PRESENT WORK

7.1 Introduction

In Chapters 2 to 6 the different sections of the process of calculating the impact ionization threshold data was presented. For each semiconductor being investigated, this procedure involves calculating the energy band structure by the Empirical Pseudopotential (E.P.) Method, fitting analytic expressions to the energy bands, determining the impact ionization threshold values by the Envelope Method, and calculating the corresponding sizes of matrix elements of the coulomb interaction and the rates of increase in the total probabilities for hot electrons, or holes, just above threshold. To perform these calculations, several computer programs had to be written, and some of the details of these programs are presented in this chapter.

The procedure commences with the reproduction of the required band structure by the E.P. method as described in Chapter 2, the computer program being described briefly in the next section. The energy bands thus obtained have then to be approximated by analytic expressions, for which a multiple regression routine is used, as described in Chapter 4 and in reference [58]. The graphs of these analytic expressions of the energy bands are then drawn, together with additional graphs used to obtain approximate impact ionization threshold positions. The computer program for this routine, also described in Chapter 4, is described briefly in section 7.3. The threshold positions are then determined more accurately by the Envelope Method, described in Chapter 3, and a brief description of the computer program is given in section 7.4.

The E.P. method is then used again to calculate the coefficients of the plane-waves occuring in the wave-function expansions of the electron states involved in the ionization threshold. From these wave-functions,

the corresponding matrix elements of the coulomb interaction between the electron states involved in the ionization threshold process are then calculated by the method described in Chapter 5, using the computer program described briefly in section 7.5. Finally, for the thresholds in silicon only, the rates of increase in the total probabilities for hot electrons, and holes, just above threshold are calculated and the graphs drawn by the method described in Chapter 6, the computer program being described briefly in section 7.6.

In each of the following sections, a brief description of the methods involved in the calculations performed by the computer programs are given. Some details of the input and output operations are also given, together with the computer facilities required for each program. All the programs are written in FORTRAN IV for use on an I.C.L. Systems 4 computer. While standard Fortran is not used, the programs should not need many alterations to enable them to be used on other types of computer. Full details of the operations of the computer programs may be obtained upon request.

7.2 The Empirical Pseudopotential Method

The computer program written to perform the band structure calculations can be considered as being made up from three basic sections. The first section calculates the matrix elements of the secular equations from the given input data, the second section then calculates all the eigenvalues and, if required, the corresponding eigenvectors. If the eigenvectors are calculated, then the third section, if required, recovers the original expansion of the eigenvectors by use of the perturbation theory of Löwdin [50]. These three sections are written as subroutines which are accessed in turn by the main program. The main program also performs all the input and output operations, initializes all the reciprocal lattice vectors which are allowed to be used in the program,

and selects those which are to be used in the band structure calculations.

All the reciprocal lattice vectors which are used by the program are stored and initialized in a block data subroutine, which contains all the vectors $\mathbf{G}$ such that $|\mathbf{G}|^2 \leq 40$. The main program, after performing all the input operations and some of the corresponding output operations, then selects those reciprocal lattice vectors which are to be used in the band structure calculations. It then proceeds further to order them in ascending order of squared magnitude, and to separate those which are to be treated exactly in the basis matrix, and those which are to be treated through perturbation. The first subroutine is then accessed by the main program, which calculates the matrix elements of the secular equations by using the relevant equations given in section 2.6. Amongst these calculations is the integral occurring in the nonlocal potential term, which is evaluated numerically by Chebyshev integration (see for example, F.B. Hildebrand 'Introduction to Numerical Analysis' Second Edition p414ff [59]).

The second subroutine is then accessed, and transforms the symmetric matrix into tri-diagonal form by Householder's method. This subroutine then accesses the subroutine which performs the Q.R. algorithm to evaluate all the eigenvalues of the matrix, together with the corresponding eigenvectors, if required. The eigenvalues are then ordered in ascending order of magnitude by another subroutine. The main program then prints out the eigenvalues before accessing the third routine, if required, which calculates the original expansion of the eigenvectors using the perturbation theory of Löwdin. These eigenvectors are then printed out by the main program, if they are required to be calculated.

All the data for input to the program is submitted through the

card reader, most of which is printed out on the line printer immediately. The data is submitted in the following order:-

(1) The name of the material being considered.

(2) The options which specify some of the operations to be performed.

(3) The lattice constant.

(4) The local symmetric form factors.

(5) The local antisymmetric form factors, if they are not all zero.

(6) The nonlocal potential parameters.

(7) The energy of the valence band maximum if the first point at which the energy levels are to be calculated in not at Γ.

(8) The cut-off points for the inclusion of plane-waves is the basis matrix and perturbation treatment.

(9) A dummy input card describing the data points submitted.

(10) The data set reference number which is used to output the results to a magnetic medium for later use.

(11) The number of results, if any, to be added to.

(12) The set of data points within the first Brillouin zone at which the energy levels are to be calculated.

The name of the material, lattice constant, local form factors, non-local parameters if not all zero, and the plane-wave cut-off points are printed out immediately. After the eigenvalues have been calculated, the number of plane-waves included in the basis matrix and through perturbation are printed out together with the position in the first Brillouin zone at which the energy levels are calculated, and the eight lowest energy levels. If the eigenvectors are calculated, then all the coefficients determined are printed out for the vectors corresponding to the eight lowest eigenvalues, together with the plane-waves they correspond to.

The minimum computer requirements of the program are; one card reader, two line printer files, which is logically two line printers but physically only one, and a main core store size of 154K bytes (1K = 1024). If the eigenvectors are to be output to a magnetic medium, then the program requires in addition, either disc storage space or a magnetic tape. If the eigenvectors are to be output to magnetic tape and added to a set of existing results, then two magnetic tapes are required.

7.3 The Graph Plotting of the Energy Bands

Once the energy levels at a set of discrete points within the first Brillouin zone have been determined and the analytic expressions have been fitted to the corresponding energy bands, the graphs of these analytic expressions are drawn. The computer program written to draw these graphs also draws some additional graphs and performs some extra calculations. The extra graphs are used to determine approximate impact ionization threshold positions by the envelope method, while the extra calculations determine the equations of the parabolic approximations to the energy bands about the energy band extrema. These equations can then be used in the Franz construction or parabolic band approximation in order to obtain approximate values of the impact ionization thresholds.

The computer program can be considered as being made up from four basic sections, one calculational section and three graph plotting sections. The first section calculates the equations of the parabolic band approximations to each energy band, and is always performed. The three graph plotting sections are all optional, the first of which plots the energy bands and data points, the second plots the energy bands and reflected energy bands in an extended zone scheme, and the third plots the envelopes of the conduction and valence bands. The options on the graph plotting sections allow them to be included in any combination.

The main program, after performing all the input operations, and the initial output operations, accesses the subroutine which calculates and prints out the equations of the parabolic band approximations to each energy band extremum. The graph of the energy bands and data points is then plotted by the use of three subroutines called one after the other. The first opens the graph plotting file and plots the title, the second plots all the data points within the bounds of the graph, and the third draws the energy bands. The energy bands are then drawn again on a second graph, but this time in an extended zone scheme covering two complete zones. On the same graph are also drawn the valence bands reflected about the valence band maximum, and the conduction bands reflected about the conduction band minimum.

Two more graphs are then drawn, the first is of the envelope of the conduction bands and the second is of the envelope of the valence band. These two graphs are drawn on the same scale as the graph of the energy bands and reflected energy bands, which are used to obtain approximate impact ionization threshold positions. The values on the envelopes are calculated according to the appropriate equations presented in section 3.3, and the k coordinates from which the envelope energy is calculated are also determined by the appropriate equations. For the double envelopes, this latter calculation involves an iterative process, for which Newton's method is used to obtain the two positions of equal gradient.

All the data for input to the program is submitted through the card reader, and in the following order:-

(1) The name of the material being considered.

(2) The symmetry direction being considered.

(3) The number of energy bands being drawn and the number of data points on each energy band.

(4) Approximate k coordinates of the energy band extreme.

(5) The k coordinates followed by all the energy levels for each data point.

(6) The Fourier series of all the energy bands, giving the number of terms, the order and coefficients of each term.

(7) The options for which graphs are to be drawn.

(8) The number of valence bands, and the conduction band on which the conduction band minimum lies.

The name of the material, the symmetry direction being considered, the number of energy bands and the number of data points on each energy band are printed out immediately. The only other output is that of the equations of the parabolae, used to approximate the energy bands, for each energy band.

The computer requirements of the program are; one card reader, one line printer, one graph plotter, the facility to read and write directly from and to the main core store, and a main core store size of 40K bytes.

7.4 The Envelope Method for determining Impact Ionization Threshold Energies

The impact ionization thresholds of the semiconductors being investigated are determined approximately by the use of the graphs produced by the graph plotting program. The computer program written to calculate impact ionization thresholds accurately uses these approximate positions as the initial step in an iterative method. At each step of this iterative method, the program investigates the intersection of the envelope with the valence band, finding a position at which the gradients are equal. This involves another iterative process for which, at each step, the positions and values on the conduction band, from which the envelope is constructed, have to be determined. For a double envelope, this involves

yet another iterative process which determines the two distinct points of equal gradient on the conduction band, from which the double envelope is constructed.

When a threshold position is determined sufficiently accurately, the position is printed out, and extra threshold data is calculated and printed out. Another approximate threshold position is then considered, and the process is repeated for each threshold position to be determined accurately. To perform the required calculations, several subroutines have been written, in which most of the calculations are performed. The main program performs all the input operations, most of the output operations, a few calculations, and accesses the required subroutines.

The main program, after performing nearly all the input operations, determines whether the ionization thresholds to be calculated accurately are initiated by electrons or by holes. If the thresholds are initiated by holes, then the roles of the valence and conduction bands are reversed, and the program then proceeds as though the initiating particles are electrons. The initial positions of the hot and promoted electrons, or holes, involved in the threshold are then input to the program, and the iteration proceedure started. The main program performs any necessary iterations of the position of the hot electron to ensure that its energy above the conduction band minimum is greater than the energy gap. Once the electron has sufficient energy, the subroutine which investigates the intersection of the envelope with the valence band is accessed.

This subroutine performs the iteration of the position of the valence electron, while keeping the position of the hot electron fixed. The iteration is to determine the position at which the envelope and valence band have equal gradients, which in turn determines whether an impact ionization process is possible. Newton's method together with Aitken's acceleration method are used as the method of iteration, for which, at each step, the positions and values on the conduction band,

from which the envelope is constructed, have to be calculated. The values from which the simple envelope is constructed are straightforward to calculate, being the points midway between the positions of the hot and valence electrons. The values from which the double envelope is constructed are determined from the positions on the conduction band where the gradients are equal. These positions are calculated by an iterative method, for which Newton's method is used, and the appropriate equations given in section 3.3 are used to evaluate the values on the envelopes.

When the required position of the valence electron has been calculated, the energy difference between the envelope and valence band is calculated. If this energy difference is sufficiently small, then the threshold position is considered to be sufficiently accurate, and the iteration process terminated. Otherwise, the main program accesses the subroutine which performs the next iteration of the position of the hot electron. This iteration is just a constant step length, until the threshold position has been passed, upon which the step length is halved at each iteration. When a new hot electron position has been calculated, the main program loops back to access the subroutine which investigates the intersection of the envelope with the valence band, and the process is repeated.

Once the position of the hot electron has been determined sufficiently accurately, all that remains is to calculate the rest of the threshold data. The initial positions of the hot and promoted electrons are already known, the corresponding energies are easily calculated, and the final positions and energies of the two electrons are then calculated. If the threshold was determined by a simple envelope, then the required values are calculated in the main program, otherwise they are calculated in a subroutine accessed by the main program. When these values have been calculated, they are printed out, the necessary adjustments being made

if the threshold was initiated by a hole. The main program then accesses the subroutine which calculates and prints out the values of the first and second derivatives of the four states involved in the ionization threshold process.

The main program finally determines whether the ionization process corresponds to a Normal process or an Umklapp process before looping back to consider the next threshold. Throughout the iterative process of the position of the hot electron, checks are made to ensure too much time is not being used, or the hot electron position is not moving too far from the centre of the first Brillouin zone. If either of these situations arise, the iterative proceedure is terminated and an appropriate error message printed out.

All the data for input to the program is submitted through the card reader, and in the following order:-

(1) A title card describing the thresholds to be determined, for which the first four columns determine whether the initiating particle is an electron or a hole.

(2) The number of valence bands and the number of conduction bands.;

(3) The group theory label and Fourier series of all the energy bands, giving the number of terms, the order and coefficients of each term.

(4) The energy band numbers and k coordinates of the conduction band minimum and valence band maximum.

(5) The overall conduction band minimum and valence band maximum of the material, and the proportional length of k-space within the first Brillouin zone.

(6) The error tolerances in position and energy.

(7) The threshold data, giving the type of envelope to use, the energy band numbers and initial positions of the hot and promoted electrons, and the initial hot electron position step length.

As many threshold position data cards as required may be submitted at the same time. The title card describing the thresholds to be determined is printed out at the beginning of each threshold calculation, and the energy band labels on which the electrons lie initially are printed out prior to the results being printed out. A table of results is then printed out, suitably labelled, giving the initial and final positions and energies of the hot and promoted electrons, or holes. The excess ionization threshold energy above the conduction band minimum, and the ratio of this excess energy to the energy band gap is also printed out. The values of the first and second derivatives of the four electron states at threshold are then printed out, followed by the type of ionization process.

Throughout the iterative procedure of the position of the hot electron, the positions of the hot and valence electrons are printed out together with the energy difference between the envelope and valence band. Also, there are several error branches within the program, each with its own error message which is printed out if the error branch is encountered. This extra information is printed out to check that the ionization thresholds are determined correctly, and to make the correction of errors, if any, easier.

The computer requirements of the program are; one card reader, two line printer files, which is logically two line printers but physically only one, and a main core store size of 42K bytes.

7.5 The Matrix Elements of the Coulomb Interaction

The matrix element of the coulomb interaction between the initial and final states involved in an impact ionization threshold process is calculated for each threshold position determined. The sizes of these matrix elements are calculated by the two equations described in section 5.2, by the quadruple sum and by the two double sums, both summed over

reciprocal lattice vectors. Before these matrix elements can be calculated, the plane-wave expansions of the wave-functions corresponding to the four electron states involved in the threshold transition have to be calculated by the program written to perform band structure calculations by the Empirical Pseudopotential Method. This program is run, and the coefficients of the plane-wave expansions are written onto a magnetic medium, which are then read in by the computer program which calculates the sizes of the matrix element of the coulomb interaction.

The computer program written to perform the matrix element calculations, reads in the required data for either three or four electron states. If data for only three states is read in, the situation corresponding to the threshold determined by a simple envelope, then the data corresponding to the electron state in which the electrons lie finally is duplicated to represent the data of the fourth electron state. Thus, data corresponding to the four electron states is supplied to the program. The sizes of the matrix element of the coulomb interaction are then calculated, firstly by the two double sums over reciprocal lattice vectors, and then by the quadruple sum over reciprocal lattice vectors, and both values are printed out.

If one, or more, of the energy bands in which the electron states lie is degenerate, then there will be more than one possible way in which the transition may occur. This will result in more than one matrix element, and a range of sizes of matrix elements will then exist, each corresponding to a different combination of the electron states in which the electrons involved lie. All such matrix elements are evaluated at one attempt, by looping back in the program, after the results have been printed out, to the point at which the data concerning the electron states

is read in.

The data for input to the program is submitted through two different media; a small amount through the card reader, and then the majority through either disc storage space or a magnetic tape. The input through the magnetic medium is in the same order and format as the output to the same magnetic medium by the program which performs the band structure calculations. The input through the card reader is in the following order:-

(1) The data set reference number used to input the data from a magnetic medium, which was set up previously.

(2) The number of distinct electron states involved in the threshold process.

(3) The inverse of the screening radius.

(4) The energy band numbers on which the initial and final hot and promoted electron states lie.

The only output is that of the values of the two matrix elements, and the values of the two double sums over reciprocal lattice space occurring in the calculations of the first matrix element.

The computer requirements for the program are; one card reader, one line printer, either disc storage space or a magnetic tape, and a main core store size of 63K bytes.

7.6 The Rate of Increase in the Total Probability

The rate of increase in the number of valence states able to partake in impact ionization for a hot electron just above threshold, is given by equation 6.2.22. This equation is programmed for use on a digital computer to enable a graph to be drawn, one corresponding to each ionization threshold. The computer program written to do this simply draws the required graphs of the proportion of valence band able to partake in impact ionization versus the hot electron energy above threshold.

The program reads in the required data corresponding to an impact ionization threshold, that is the first and second derivatives of all four electron states involved in the process. If the threshold is initiated by a hot hole, then the roles of the valence bands and conduction bands are reversed, which is done by negating all the values of the first and second derivatives. The equation 6.2.22 is then evaluated for a range of values of the excess hot electron, or hole, energy above threshold. This enables the corresponding graph to be drawn over the given range of values of the excess energy, between 0 and 0.25eV. The value of the equation is printed out for the maximum value of the excess energy before the program loops back to read in the data corresponding to another threshold. The graph of this next threshold is drawn on the same figure as the previous graph to enable as many graphs as required to be drawn together.

All the data input to the program is submitted through the card reader, and in the following order:-

(1) The name of the material being considered.

(2) The first derivatives of the initial and final hot electron, or hole, states and the second derivatives of all four hot electron, or hole, states.

The only output is that of the name of the material being considered, followed by the values of the equation 6.2.22 for an excess energy of 0.25eV above threshold, one value for each threshold considered.

The computer requirements for the program are; one card reader, one line printer, one graph plotter and a main core store size of 26K bytes.

8. IMPACT IONIZATION THRESHOLDS FOR SILICON

8.1 Details of the Calculations

The Envelope Method developed in this work is applied to two different band structures of Silicon in a preliminary study [31], the band structures investigated being those of Cohen and Bergstresser [25] and of Stuckel and Euwema [38]. It is thus hoped to obtain useful information concerning the sensitivity of the impact ionization threshold energies to the detailed band structure. Investigations are carried out in the extended zone scheme along the Δ axis for the $\Gamma_{25'}$-X_4 valence band the the Γ_{15}-X_1 and $\Gamma_{2'}$-X_1 conduction bands.

Numerical data from the published energy band diagrams are fitted by suitable polynomial approximations, the analytic expressions used being nowhere in error by more than 0.01eV. The curvatures of the polynomial approximations at their extrema also give correctly the appropriate effective masses of the conduction and valence bands at the energy band extrema. These analytic expressions for the energy bands are then used in the computer program, as described in Chapter 3 and 7, to calculate the impact ionization threshold positions for hot electrons. The computations are carried out until the overall error associated with each threshold energy is less than 0.02eV.

The results of this preliminary study of Silicon are presented in Table 8.1, which contains details of the initial and final states of the hot electron for each threshold position. Details of the initial and final states of the promoted electron are also included where necessary. When this is not given, the information can be readily calculated from the hot electron data. Each threshold energy, E_T, is also expressed in terms of the dimensionless quantity E_T/E_G, where E_G is the band gap for the particular band structure. This facilitates

a more meaningful comparison of the results for band structures with different band gaps.

For comparative purposes, the values of E_T/E_G given by the Franz construction are also calculated for each band structure, the results being presented in Table 8.3. They are obtained by the computer program using a parabolic conduction band based on the appropriate conduction band minimum, and using the polynomial approximations again for the valence bands. The effective mass and the position of the conduction band minimum are taken to be the values appropriate to the band structure considered. Also, the values of E_T/E_G given by the parabolic band approximation are calculated using equation 3.4.18.

This preliminary study was carried out before the Empirical Pseudopotential Method had been programmed for use on the computer, and before it had been decided to calculate the sizes of the matrix element of the coulomb interaction corresponding to each threshold transition. In calculating the sizes of the matrix elements, the opportunity is taken to reproduce the Silicon band structure, of Cohen and Bergstresser only, more accurately by using their form factors in the Empirical Pseudopotential Method. Also, the three principal symmetry directions are investigated and more energy bands are taken into consideration along the Δ axis.

The energy bands investigated along the Δ axis are those of the preliminary study together with the $\Gamma_{25'}$-X_1 valence band. The energy bands investigated along the Λ axis are the Γ_{15}-L_1 and Γ_{15}-L_3 conduction bands and the $\Gamma_{25'}$-L_3 and $\Gamma_{25'}$-L_1 valence bands. The energy bands investigated along the Σ-S axis, that is the Σ axis from Γ to K within the first Brillouin zone and the S axis from K to X along the square face of the adjacent zone, are the Γ_{15}-K_3-X_1

and Γ_{15}-K_1-X_1 conduction bands and the $\Gamma_{25'}$-K_2-X_4, $\Gamma_{25'}$-K_1-X_4 and $\Gamma_{25'}$-K_1-X_1 valence bands. These energy bands are investigated as being able to contain the hot and promoted electrons in an impact ionization threshold transition. In addition to recalculating the threshold positions initiated by hot electrons, threshold positions initiated by hot holes are also calculated.

The energy bands are approximated by suitable Fourier series, in preference to polynomial approximations as a result of the study performed and presented in Chapter 4, and the errors in fitting them to the energy bands are again nowhere greater than 0.01eV. The errors involved in fitting the Γ_{15}-K_3-X_1 conduction band however, are slightly greater than 0.01eV in places, due to the shape of this energy band. The computations are again carried out until the overall error associated with each threshold energy is less than 0.015eV, the slightly greater errors associated with the Γ_{15}-K_3-X_1 conduction band not having a noticeable effect on the overall error.

The result of these improved calculations are presented in Table 8.2, which contains the same details as Table 8.1, plus the sizes of the matrix element of the coulomb interaction as calculated by the two appropriate equations given in Chapter 5. Where one, or more, of the energy bands involved in a threshold transition is degenerate, several values of the matrix element sizes are obtained, but only the largest value is presented. When the coefficients of the plane-wave expansions used in calculating the sizes of the matrix element are determined, the energies of the four electron states involved in the threshold are also calculated. The error in the conservation of energy, ensuring the conservation of wavevector, is thus determined, and this value is also presented in the table for each threshold. Since the error

associated with each threshold energy is less than 0.015eV, the error in the conservation of energy is less than 0.06eV.

The values of E_T/E_G given by the Franz construction and by the parabolic band approximation are again calculated for comparative purposes. The values are calculated by the same procedure as used in the preliminary study, and the results are presented in Table 8.4. In this table, more than one threshold value is given by the Franz construction, along each symmetry axis, due to the extra energy bands being investigated, but the parabolic band approximation still gives only one value along each symmetry axis.

In Tables 8.1 and 8.2, the values of k are measured as a proportion of the distance from Γ to the boundary of the first Brillouin zone, X or L, along the Δ and Λ symmetry axes respectively. Along the Σ-S symmetry axis, the proportional distance is from Γ to the centre of the square face, X, of the adjacent zone. Thus, electron states on the Σ axis within the first Brillouin zone have wavevector k such that $|k| \leqslant 0.75$, and electron states on the S axis on the square face of the adjacent zone have wavevector k such that $0.75 \leqslant |k| \leqslant 1$.

8.2 Discussion of the Results for the Silicon Band Structures

Looking at the results of the preliminary study, presented in Table 8.1, it is seen that the general features for each band structure are similar. The lowest value of E_T/E_G is given by an umklapp process in both cases, with a second umklapp threshold at a higher energy. At an energy between these two umklapp threshold energies, there is a normal threshold which is given by a double envelope solution and corresponds to an intervalley transition. The lowest value of E_T/E_G, in both cases, is also close to the minimum possible energy for any ionization process, which is due to the indirect energy band gap of

Silicon. A detailed comparison of the results shows that there are significant discrepancies between the threshold energies given by the two band structures. Since the value of E_G given by Stuckel and Euwema (1.1eV) is close to that generally accepted for Silicon (see for example, D. Long 'Energy Bands in Semiconductors', p87 [60], and also Kunz [61]), it is believed that the threshold energies given by this band structure are the more reliable.

The values of E_T have also been calculated by Anderson and Crowell [32], using the band structure of Cohen and Bergstresser. A direct comparison of their results with the results presented in Table 8.1 is difficult since Anderson and Crowell adjusted the Cohen and Bergstresser band gap to conform with the commonly accepted value. To confirm the equivalence of the two methods however, the calculations are repeated for the adjusted band structure. Absolute agreement is obtained within the errors of the respective methods, $\pm$ 0.02eV in the present work and $\pm$ 0.2eV for Anderson and Crowell.

It is seen from Table 8.2 that the improved threshold values corresponding to the two umklapp processes of Table 8.1 are in close agreement (E_T/E_G=1.055 and 1.467 respectively) as expected. However, it is surprising to see that there is no longer a normal threshold given by a double envelope, as there was in the preliminary study. This is due to the slight changes in the energy bands concerned, and to the final state of the hot electron being very close to the zone boundary. In the improved calculations this state moves across the zone boundary into another energy band, and thus is not considered as a possible threshold situation. This illustrates the sensitivity of threshold energies to the detailed band structure, particularly when one of the four states involved in the transition is close to the zone boundary.

Table 8.1. Impact Ionization Threshold Results for the Preliminary Study.

When transitions are provided by double envelopes, the figures in parenthesis with the hot electron data refer to the initial and final states of the promoted electron. All such thresholds involve intervalley transitions.

Details of band structure used	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process
Cohen and Bergstresser	0.448	1.64	-0.791	0.82	1.10	U
$\Gamma - \Delta - X$ axis	0.377	1.94	0.963	0.86	1.49	N
(E_G=0.78eV.	-0.074	-0.08	-0.660	1.00		
indirect gap $\Gamma_{25'} \rightarrow \Delta_1$)	0.696	1.97	-0.684	0.96	1.53	U
Stuckel and Euwema	0.286	2.22	-0.850	1.11	1.02	U
$\Gamma - \Delta - X$ axis	0.247	2.34	0.916	1.16	1.13	N
(E_G=1.1eV.	-0.041	-0.02	-0.710	1.16		
indirect gap $\Gamma_{25'} \rightarrow \Delta_1$)	0.706	2.44	-0.674	1.20	1.22	U

Table 8.2. Impact Ionization Threshold Results.

The results are for the band structure of Cohen and Bergstresser. (E_G=0.85eV. indirect gap $\Gamma_{25'} \rightarrow \Delta_1$).

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Δ - X axis Electrons	0.4427	1.745	-0.7868	0.870	1.055	U	0.006	1.3×10^{-12}	0
	0.4424	1.746	-0.7834	0.872	1.056	U	0.007	9.5×10^{-2}	0
	0.7090	2.095	-0.6765	1.013	1.467	U	0.006	6.1×10^{-10}	0
	0.7053	2.117	-0.6643	1.040	1.493	U	0.003	1.6×10^{-11}	0
Γ - Δ - X axis Holes	0.3990	-1.540	-0.0924	-0.145	1.813	N	0.000	2.2×10^{-2}	1.4×10^{-11}
	0.3001	-1.855	-0.1115	-0.206	2.184	N	0.007	7.4×10^{-2}	1.7×10^{-19}
	0.5391	-2.133	-0.1452	-0.334	2.513	N	0.002	1.6×10^{-11}	1.4×10^{-11}
	0.3969	-2.765	-0.2077	-0.612	3.257	N	0.002	7.4×10^{-2}	2.6×10^{-21}
Γ - Λ - L axis Electrons	0.2456	3.867	-0.9097	1.922	3.554	U	0.002	2.5×10^{-2}	0
	0.2548	3.896	-0.8772	1.946	3.589	U	0.001	1.5×10^{-10}	0
	0.6825	4.272	-0.7248	2.090	4.031	U	0.006	7.6×10^{-3}	0
	0.6512	4.308	-0.6835	2.147	4.073	U	0.001	4.5×10^{-12}	0
Γ - Λ - L axis Holes	0.3285	-2.458	-0.1713	-0.146	2.895	N	0.008	5.1×10^{-3}	1.4×10^{-19}
	0.4645	-3.834	0.1837	-0.166	4.515	N	0.006	1.6×10^{-2}	2.8×10^{-10}
Γ-Σ-K-S-X axis Electrons	0.5513	2.620	-0.8440	1.185	2.085	U	0.005	2.1×10^{-15}	0
	0.5018	2.939	-0.7796	1.404	2.462	U	0.019	5.7×10^{-12}	0
	0.4937	2.991	-0.7652	1.467	2.522	U	0.021	2.3×10^{-3}	0
	0.8036	3.526	-0.7776	1.413	3.152	U	0.025	1.5×10^{-15}	0
	0.7680	4.213	-0.6554	1.998	3.962	U	0.014	9.9×10^{-2}	0
	0.7625	4.323	-0.6344	2.115	4.092	U	0.006	3.4×10^{-11}	0
Γ-Σ-K-S-X axis Holes	0.4995	-1.396	-0.1957	-0.156	1.644	U	0.001	7.2×10^{-14}	0
	0.5135	-1.470	-0.2373	-0.244	1.731	U	0.003	2.0×10^{-13}	0
	0.2774	-1.940	-0.2686	-0.335	2.285	U	0.027	2.6×10^{-13}	0
	0.2043	-2.207	-0.2944	-0.429	2.599	U	0.014	1.7×10^{-2}	0
	0.3054	-2.224	-0.3370	-0.606	2.619	U	0.000	8.8×10^{-2}	0
	0.2285	-2.554	-0.3751	-0.770	3.008	U	0.004	4.5×10^{-12}	0

In the improved calculations, there are a further two electron thresholds along the Δ axis, with energies very close to the two threshold energies previously calculated. These correspond to transitions involving an electron in the $\Gamma_{25'}$-X_1 valence band, and thus having threshold energies slightly higher than the transitions involving an electron in the $\Gamma_{25'}$-X_4 valence band. The electron thresholds calculated have an error in the conservation of energy much smaller than the maximum expected error of 0.06eV, all four being in error by less than 0.01eV. There are also four threshold positions which are initiated by hot holes, both valence bands providing the initiating hole for two thresholds. These thresholds correspond to normal processes, and the error in the conservation of energy is less than 0.01eV as in the case for the electron thresholds. The lowest E_T/E_G value for the hole thresholds is much larger than the lowest E_T/E_G value for the electron thresholds, and is also larger than the highest value for the electron thresholds.

By considering only the threshold positions and the values of E_T/E_G, it would appear that the lowest value of E_T/E_G (1.055) should be taken for use in the related theories. However, by looking at the sizes of the matrix element of the coulomb interaction as calculated from equation 5.2.5, it is surprising to see that this would be erroneous, due to the negligibly small size of the matrix element corresponding to this threshold. Indeed, by looking at the sizes of the matrix elements corresponding to the other electron thresholds along the Δ axis, it is surprising to see only one threshold having a significant matrix element size. This threshold is the second lowest (E_T/E_G=1.056), and has a matrix element size of 0.095, which is of the expected order of magnitude. The other three electron thresholds all have matrix element sizes of the order of 10^{-10} or smaller which, while not being mathematically zero, can be considered to be zero.

The lowest hole threshold along the Δ axis, unlike the lowest electron threshold, has a significant matrix element size, and so it would not be erroneous to use this threshold in related theories. However, it would still be erroneous to neglect the effect of the matrix element sizes, as one of the hole thresholds has a negligibly small matrix element size, of the same order of magnitude as that for the lowest electron threshold. The approximate matrix element sizes, as given by equation 5.2.9, are identically zero for all the electron thresholds as expected, since they all correspond to umklapp processes, and hence are not comparable with the proper calculations of equation 5.2.5. The corresponding values for the hole thresholds are all non-zero, but are all negligibly small, thus making only one of comparable size with the proper calculations.

Considering now the thresholds along the Λ axis, it is seen that there are again four electron thresholds, all corresponding to umklapp processes, occurring in pairs with comparable energies, as is the situation for the electron thresholds along the Δ axis. The threshold energies are however, much higher along this axis than along the Δ axis, as is expected from the details of the band structure, the lowest threshold having a value of $E_T/E_G=3.554$. Similarly, the two hole thresholds along the Λ axis, both corresponding to normal processes, have energies much higher than those along the Δ axis. However, the lower of these two thresholds has an energy lower than the lowest electron threshold energy along this axis, which is the opposite situation to that along the Δ axis.

It is again surprising to see that two of the electron thresholds have matrix element sizes which are negligible, although the lowest threshold does have a significant matrix element size. Also, it is perhaps comforting to see that both hole thresholds have significant

matrix element sizes. The approximate matrix element sizes follow the pattern of those corresponding to the thresholds along the Δ axis, those corresponding to the electron thresholds being zero, and those corresponding to the hole thresholds being negligibly small. Thus, none of the approximate matrix element sizes are comparable with the proper calculations of the matrix element sizes. The error in the conservation of energy associated with each threshold is again less than 0.01eV, much smaller than the maximum expected error, as for the thresholds along the Δ axis.

As may be expected from the details of the band structure along the Σ-S axis, there are more threshold positions determined, both for electrons and holes. Also, the lowest electron threshold energy is higher than that along the Δ axis, and lower than that along the Λ axis. However, it is perhaps surprising to see that the lowest hole threshold along this axis is also the lowest hole threshold for this band structure, having a value of $E_T/E_G=1.644$. As for the other two symmetry axes, all the electron thresholds along this axis correspond to umklapp processes, but unlike the other two symmetry axes, the hole thresholds also correspond to umklapp processes and not to normal processes. This is slightly surprising, although some umklapp processes may be expected since the boundary of the first Brillouin zone occurs at $\frac{3}{4}$ of the distance along the axis from Γ.

Once again, it is seen by looking at the sizes of the matrix element, that several of these thresholds can be considered to be unimportant due to the negligibly small sizes of the matrix element. In fact, only two electron thresholds and two hole thresholds have significant matrix element sizes, the lower of these two electron thresholds having an energy substantially higher than the lowest

electron threshold energy. The lowest hole threshold along this axis is now seen to be insignificant, thus making it erroneous to use in related theories. The lowest significant hole threshold along this axis has a very much higher energy than that of the lowest threshold, and being of a comparable energy to the lowest significant electron threshold along this axis. The errors in the conservation of energy are generally greater than those along the other two symmetry axes, but are still considerably smaller than the maximum expected error, being in error by less than 0.03eV.

Had the threshold positions given above been determined without considering the matrix elemant sizes, the lowest electron and hole thresholds would have been taken as having values of E_T/E_G=1.055 and 1.644 respectively. Also, these would have been provided by transitions along two different symmetry axes, the Δ axis and the Σ-S axis respectively. However, by considering also the sizes of the matrix element of the coulomb interaction, neither of these two thresholds can be considered to be significant. The lowest electron and hole thresholds which also have significant matrix element sizes are those for which E_T/E_G=1.056 and 1.813 respectively, both thresholds now occurring along the Δ axis. Thus, these latter two values are those which should be taken for use in related theories, and not the absolute lowest threshold values.

8.3 Comparison with Results from Approximate Band Structure Models

When both the valence band and the conduction band are approximated by parabolae, an expression for E_T is obtained, as presented in Chapter 3. Values of E_T/E_G determined from this expression (equation 3.4.18) for the appropriate values of (k_m,E_m), m_c and m_v given by the polynomial approximations to the two band structures investigated in the preliminary study are shown in Table 8.3. They are referred to as the 'parabolic

band approximation' values. The table also includes the values given by the Franz construction, which corresponds to the removal of the parabolic approximation to the valence band. Finally, the table gives the values obtained by the Envelope Method for the genuine energy bands, and are selected from the many thresholds given in Table 8.1 as being the ones most comparable with the approximate values.

Table 8.3

Comparison of E_T/E_G values for the different band structures

Band structure considered	Parabolic band approximation	Franz construction	Genuine band structure
Cohen & Bergstresser Γ-Δ-X axis (indirect gap)	1.60	1.60	1.53
Stuckel & Euwema Γ-Δ-X axis indirect gap	1.19	1.19	1.22

It is seen from Table 8.3 that there is complete agreement between the values of E_T/E_G given by the Franz construction and those given by the simpler parabolic band approximation. It is also seen that the approximate values are in good agreement with those given by the Envelope Method for the genuine bands. Comparison of Tables 8.1 and 8.3 shows that the approximate models fail to provide the lowest thresholds. In each case the minimum threshold provided by the genuine energy bands is substantially lower and is of the same type, both thresholds corresponding to umklapp processes. Thus the genuine bands give at least one lower threshold of similar importance. In the first example given in the table, for instance, it is seen that in addition

to the umklapp threshold with E_T/E_G=1.53, there is a much lower one with E_T/E_G=1.10. There is also a normal, intervalley threshold with E_T/E_G=1.49.

In Table 8.4, the values of E_T/E_G calculated for the more accurate reproduction of the Cohen and Bergstresser band

Table 8.4

Comparison of E_T/E_G values for different band structure models

Details of thresholds	Parabolic band approximation	Franz construction	Genuine band structure
Γ-Δ-X axis	1.399	1.398	1.467
Electrons	-	1.415	1.493
Γ-Δ-X axis	2.159	2.169	1.813
Holes		3.508	2.513
Γ-Λ-L axis	3.464	3.465	-
Electrons	-	3.466	-
Γ-Λ-L axis	2.404	2.430	-
Holes	-	4.740	-
Γ-Σ-K-S-X axis	2.443	2.296	2.085
Electrons	-	2.831	2.462
	-	2.922	2.522
Γ-Σ-K-S-X axis	1.972	1.946	1.644
Holes	-	2.133	1.731

structure are presented, which correspond to the values given in Table 8.3. It is seen that there are fewer thresholds given by the parabolic band approximation than by the Franz construction, since the parabolic band approximation applies to one valence and one conduction band only. Taking this into consideration, it is seen that there are only minor differences between the corresponding values of E_T/E_G in all thresholds except one, the electron threshold along the Σ-S axis.

It is also seen that the approximate values are only in reasonable agreement with those given by the Envelope Method for the electron thresholds along the Δ axis. All other approximate values of E_T/E_G do not correspond accurately to any genuine thresholds, which is almost certainly due to the departure from the parabolic shape of the energy bands involved. Indeed, it is seen that the approximate models give threshold values along the Λ axis which ought not to exist, since the Γ_{15}-L_1 conduction band and the $\Gamma_{25'}$-$L_{3'}$ valence band are too flat to permit any ionization process to take place.

Comparison of Tables 8.2 and 8.4 shows again that the approximate values of the electron thresholds along the Δ axis fail to provide the lowest threshold, the minimum being provided by a substantially lower value. The approximate values of the hole thresholds along the Δ axis and the electron and hole thresholds along the Σ-S axis do provide the lowest threshold in each case. However, the approximate values are not in agreement with the values given by the Envelope Method, but are considerably higher. The thresholds along the Λ axis are not provided by the Γ_{15}-L_1 conduction band and $\Gamma_{25'}$-$L_{3'}$ valence band since these bands are too flat to permit any ionization process to occur, but are provided by the other two energy bands investigated. The lower of the approximate values for both electrons and holes, which ought not to exist, are lower than the lowest threshold values along this symmetry axis for both electrons and holes, and neither are in reasonable agreement with the values given by the Envelope Method.

9. IMPACT IONIZATION THRESHOLDS FOR GERMANIUM

9.1 Details of the Calculations

The Envelope Method developed in this work and applied to two different band structures of Silicon in a preliminary study [31], is also applied to two different band structures of Germanium in the same preliminary study. The two band structures investigated are those of Cohen and Bergstresser [25] and of Stuckel [39]. As in the study of Silicon, investigations are carried out in the extended zone scheme along the Δ axis for the Γ_2'-X_1 and Γ_{15}-X_1 conduction bands and the Γ_{25}'-X_4 valence band. Additional results are also obtained along the Λ axis for the Γ_2'-L_1 conduction band and the Γ_{25}'-L_3' valence band. Again, it is hoped to obtain useful information concerning the sensitivity of the impact ionization threshold energies to the detailed band structure.

Numerical data from the published energy band diagrams are again fitted by suitable polynomial approximations, the analytic expressions being nowhere in error by more than 0.01eV. Again, the curvatures of the polynomial approximations at their extrema also give correctly the appropriate effective masses of the conduction and valence bands at the energy band extrema. These analytic expressions for the energy bands are then used in the computer program, as described in Chapters 3 and 7, to calculate the impact ionization threshold positions for hot electrons. The computations, as for Silicon, are again carried out until the overall error associated with each threshold energy is less than 0.02eV.

The results of this preliminary study of Germanium are presented in Table 9.1, which contains the details of the initial and final states

of the hot electron for each threshold position, presented in the same manner as were the results for Silicon. For comparative purposes the values of E_T/E_G given by the Franz construction are again calculated for each band structure, the results being presented in Table 9.3. They are obtained in the same manner in which the corresponding results for Silicon were obtained. Also, the values of E_T/E_G given by the parabolic band approximation are calculated using equation 3.4.18.

The opportunity is again taken to reproduce the Germanium band structure, of Cohen and Bergstresser only, more accurately in order to calculate the sizes of the matrix element of the coulomb interaction corresponding to each threshold transition. This is done by using the Cohen and Bergstresser form factors in the Empirical Pseudopotential Method computer program. The three principal symmetry axes are also investigated, and more energy bands are taken into consideration along the Δ and Λ axes.

The energy bands investigated along the Δaxis are those of the preliminary study together with the Γ_{25}'-X_1 valence band, and along the Λ axis are those of the preliminary study together with the Γ_{25}'-L_1 valence band. The energy bands investigated along the Σ-S axis are the Γ_2'-K_3-X_1, Γ_{15}-K_1-X_1, Γ_{15}-K_2-X_3 and Γ_{15}-K_1-X_3 conduction bands and the Γ_{25}'-K_2-X_4, Γ_{25}'-K_1-X_4 and Γ_{25}'-K_1-X_1 valence bands. In addition to recalculating the threshold positions initiated by hot electrons, threshold positions initiated by hot holes are also calculated.

As a result of the study performed and presented in Chapter 4, the energy bands are approximated by suitable Fourier series, in preference to the polynomial approximations used in the preliminary study. The errors involved in fitting the Fourier series to the

energy bands are nowhere greater than 0.01eV for most of the energy bands. However, the Γ_2'-X_1 conduction band along the Δ axis, the Γ_{25}'-K_1-X_1 valence band and the Γ_2'-K_3-X_1 and Γ_2'-K_1-X_1 conduction bands along the Σ-S axis are in error by more than 0.01eV in places, but nowhere in error by more than 0.015eV. Where the errors in approximating to the energy bands are less than 0.01eV, the computations are carried out until the overall error associated with each threshold energy is less than 0.015eV. When the initial state of the hot particle is in the region of an energy band in error by less than 0.015eV, the corresponding computational errors are less than 0.02eV.

The energy band gap of Germanium is generally accepted as being an indirect gap from Γ_{25}' to L_1. In the band structure calculations however, the conduction band minimum does not occur at L as expected, but along the Λ axis away from L. Only one energy value, of those calculated in the Γ_2'-L_1 conduction band, is lower than the energy at L, thus to conform with the accepted position of the energy gap, this one energy value is ignored in the fitting of the Fourier series. Due to this unexpected error and the subsequent action taken, a few of the thresholds, whose final states are in the region of the errors in the Γ_2'-L_1 conduction band, involve larger errors in the conservation of energy than the other thresholds. This larger error does not have a great effect on the accuracy of the threshold energies, the overall errors being less than 0.02eV.

The results of these improved calculations are presented in Table 9.2, and are presented in the same manner as were the results of Silicon in Table 8.2. For comparative purposes, the values of E_T/E_G given by the Franz construction and by the parabolic band approximation

are also calculated, by the same procedures used previously. These results are presented in Table 9.4, in the same manner as were the results of Silicon in Table 8.4.

9.2 Discussion of the Results for the Germanium Band Structures

Both band structures investigated in the preliminary study have three conduction band minima of closely comparable energies, one at Γ, one at L and one along the Γ-Δ-X axis. The ordering by energy of these minima is different for the two band structures considered, and this has a significant influence on the relative order of comparable threshold energies. There are a greater number of thresholds for Germanium than for Silicon, which is a direct consequence of the increased number of conduction band minima, and the possibility of intervalley transitions.

The thresholds, presented in Table 9.1, are tabulated in order of increasing energy for the Cohen and Bergstresser band structure. The thresholds for the Stuckel band structure are tabulated in directly corresponding order, which are seen not to occur in order of increasing energy, indicating just how sensitive each threshold is to the details of the band structure. Also, in some cases there are no directly comparable thresholds, and these are indicated by the blank entries in the table.

The Cohen and Bergstresser band structure along the Λ axis gives two thresholds, both corresponding to umklapp processes, the lower of these being very close to the minimum possible energy for any threshold. Along the Δ axis, the lowest value of E_T/E_G is given by an umklapp process, and is followed closely by a normal process given by a double envelope solution, which corresponds to an intervalley transition, as were the lowest two thresholds along the Δ axis in Silicon. The normal process however, differs from that in Silicon, as it also corresponds

Table 9.1. Impact Ionization Threshold Results for the Preliminary Study.

When transitions are provided by double envelopes, the figures in parenthesis with the hot electron data refer to the initial and final states of the promoted electron. All such thresholds involve intervalley transitions.

Details of band structure used	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process
Cohen and Bergstresser $\Gamma - \Lambda - L$ axis (E_G=0.9eV. indirect gap $\Gamma_{25'} \rightarrow L_1$)	0.148	1.80	-0.962	0.90	1.00	U
	0.375	1.90	-0.888	0.93	1.11	U
Stuckel $\Gamma - \Lambda - L$ axis (E_G=1.2eV. direct gap $\Gamma_{25'} \rightarrow \Gamma_{2'}$)	—	—	—	—	—	—
	—	—	—	—	—	—
Cohen and Bergstresser $\Gamma - \Delta - X$ axis (E_G=0.9eV. indirect gap $\Gamma_{25'} \rightarrow L_1$)	0.401	2.02	-0.799	1.01	1.24	U
	0.131	2.05	0.845	1.02	1.28	N
	-0.034	-0.01	-0.748	1.02		
	0.143	2.18	-0.895	1.08	1.42	U
	0.721	2.24	-0.678	1.09	1.49	U
	0.345	2.29	0.921	1.11	1.54	N
	-0.087	-0.07	-0.663	1.11		
	0.195	2.69	0.021	1.23	1.99	N
	0.238	2.83	0.026	1.25	2.14	N
	—	—	—	—	—	—
	0.434	3.69	0.036	1.29	3.10	N
Stuckel $\Gamma - \Delta - X$ axis (E_G=1.2eV. direct gap $\Gamma_{25'} \rightarrow \Gamma_{2'}$)	—	—	—	—	—	—
	0.343	3.72	0.930	1.85	2.10	N
	-0.048	-0.04	-0.635	1.83		
	0.300	3.50	-0.840	1.75	1.92	U
	0.573	3.48	-0.721	1.73	1.90	U
	—	—	—	—	—	—
	0.211	2.79	0.032	1.24	1.32	N
	0.244	2.91	0.034	1.25	1.42	N
	0.434	3.88	0.977	1.93	2.23	N
	-0.060	-0.06	-0.603	1.89		
	0.473	3.83	0.034	1.25	2.19	N

to an example of an interband process in which the hot electron moves from one conduction band to another. These thresholds occur at higher energies than those along the Λ axis, as may be expected, and there are several more thresholds at higher energies corresponding to both umklapp and normal processes, and including another intervalley transition.

The sensitivity of the thresholds to the detailed band structure is clearly seen for the Stuckel band structure along the Λ axis. Where there are two thresholds given for the Cohen and Bergstresser band structure, there are now no thresholds, as the band structure is too flat to permit any ionization process to take place at all. Along the Δ axis, it also fails to provide the threshold comparable with the lowest threshold for the Cohen and Bergstresser band structure. It does however provide the threshold comparable with the lower normal process corresponding to an intervalley transition in the Cohen and Bergstresser band structure. Although it is of the same type, it has a much higher value of E_T/E_G(2.10 compared with 1.28). At higher energies, there exist comparable thresholds for all but two of the thresholds, but the comparable threshold energies differ considerably. It is interesting to note that the two thresholds, one in each band structure, for which there is no comparable threshold in the other band structure both correspond to intervalley transitions.

Experimental evidence on the band structure of Germanium (see for example, D. Long [60]) indicates an indirect gap of about 0.74eV from $\Gamma_{25'}$ to L_1. This is in general agreement with the Cohen and Bergstresser band structure, and contrasts with the direct gap of 1.20eV from $\Gamma_{25'}$ to Γ_2, given by Stuckel. Accordingly, it is believed that the thresholds given by the Cohen and Bergstresser band structure

are the more reliable. The values of E_T have also been calculated by Anderson and Crowell [32] using the band structure of Cohen and Bergstresser, and absolute agreement is obtained within the errors of the respective methods.

In the improved calculations of the threshold values, the agreement with the corresponding values determined in the preliminary study is not particularly good. The improved threshold values corresponding to the two thresholds along the Λ axis (E_T/E_G=1.023 and 1.026 respectively) and to the two lowest thresholds along the Δ axis (E_T/E_G=1.193 and 1.209 respectively) are in close agreement with those determined in the preliminary study. However, it is surprising to note that the improved threshold values corresponding to all other thresholds determined in the preliminary study differ by significant amounts (having values of E_T/E_G=1.267, 1.374, 1.372, 1.801, 1.883 and 2.789 respectively). This again illustrates the sensitivity of threshold energies to the detailed band structure.

It is seen from Table 9.2 that there are many more thresholds determined than in the preliminary study, due to the extra valence bands being considered. The lowest electron threshold along the Δ axis is still given by an umklapp process, with a normal threshold occurring at a slightly higher energy. At closely comparable energies to these two thresholds there are two other thresholds, of the same type, which occur due to the presence of the second valence band. This situation is expected as it also occurred in the improved calculations for the Silicon band structure. Two thresholds at higher energies (those for which E_T/E_G=1.303 and 1.305) do not correspond to any of the thresholds determined in the preliminary study, which is again due to the details of the band structure. Both these thresholds

Table 9.2. Impact Ionization Threshold Results.

The results are for the band structure of Cohen and Bergstresser. (E_G=0.97eV. indirect gap $\Gamma_{25'} \rightarrow L_1$).

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Δ - X axis Electrons	0.3642	2.130	-0.8211	1.065	1.193	U	0.009	4.8×10^{-12}	0
	0.3637	2.133	-0.8188	1.066	1.195	U	0.011	2.8×10^{-3}	0
	0.1467	2.146	0.8624	1.066	1.209	N	0.023	3.2×10^{-12}	1.2×10^{-22}
	-0.0092	-0.000	-0.7249	1.080					
	0.1471	2.150	0.8665	1.067	1.213	N	0.022	2.2×10^{-1}	6.9×10^{-23}
	-0.0023	-0.002	-0.7217	1.081					
	0.1529	2.202	-0.9084	1.095	1.267	U	0.002	6.7×10^{-12}	0
	0.1543	2.215	-0.9196	1.105	1.280	U	0.003	7.5×10^{-12}	0
	0.7237	2.237	0.7311	1.078	1.303	N	0.027	9.2×10^{-12}	9.1×10^{-18}
	0.0061	-0.000	-0.0013	1.159					
	0.7233	2.240	0.7267	1.079	1.305	N	0.024	1.8×10^{-8}	1.8×10^{-8}
	0.0016	-0.001	-0.0018	1.160					
	0.3266	2.305	0.9636	1.159	1.372	N	0.000	1.8×10^{-11}	1.0×10^{-11}
	-0.0492	-0.036	-0.6862	1.110					
	0.7096	2.306	-0.6770	1.125	1.374	U	0.016	1.0×10^{-11}	0
	0.3218	2.328	0.9949	1.204	1.396	N	0.010	2.0×10^{-11}	1.7×10^{-11}
	-0.0095	-0.008	-0.6826	1.116					
	0.7013	2.346	-0.6565	1.165	1.415	U	0.013	2.8×10^{-12}	0
	0.2105	2.721	0.0246	1.203	1.801	N	0.007	6.3×10^{-10}	6.2×10^{-10}
	0.2326	2.802	0.0260	1.208	1.883	N	0.001	4.8×10^{-12}	6.0×10^{-21}
	0.1799	3.048	0.0601	1.401	2.138	N	0.019	1.4×10^{-10}	1.3×10^{-10}
	0.2758	3.257	0.4780	1.680	2.352	N	0.010	9.9×10^{-12}	6.3×10^{-22}
	0.1764	-0.369	-0.0258	1.208					
	0.3012	3.438	0.3626	2.138	2.539	N	0.010	7.1×10^{-11}	7.2×10^{-11}
	0.0301	-0.070	-0.0313	1.230					
	0.4454	3.681	0.0285	1.218	2.789	N	0.001	1.9×10^{-11}	1.8×10^{-11}
Γ - Δ - X axis Holes	0.1918	-1.385	0.0887	-0.105	1.425	N	0.002	9.0×10^{-2}	2.2×10^{-20}
	0.4416	-1.461	-0.1021	-0.136	1.503	N	0.006	1.1×10^{-11}	7.5×10^{-12}
	0.2559	-1.965	-0.1368	-0.232	2.023	N	0.002	5.8×10^{-2}	1.4×10^{-22}
	0.6012	-2.008	-0.1309	-0.214	2.066	N	0.010	2.1×10^{-11}	8.5×10^{-12}
	0.3591	-2.888	-0.2339	-0.592	2.973	N	0.008	3.6×10^{-2}	5.3×10^{-23}

Table 9.2. continued:-

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_m/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Λ - L axis Electrons	0.1998	1.966	-0.8970	0.983	1.023	U	0.053	5.2×10^{-9}	0
	0.2009	1.968	-0.8993	0.983	1.025	U	0.058	2.9×10^{-2}	0
	0.2822	1.969	-0.8689	0.984	1.026	U	0.009	1.4×10^{-12}	0
	0.2794	1.973	-0.8614	0.985	1.031	U	0.002	8.8×10^{-2}	0
Γ - Λ - L axis Holes	0.7656	-1.005	-0.0449	-0.009	1.034	N	0.005	1.6×10^{-2}	2.0×10^{-13}
	0.1776	-1.221	0.0857	-0.032	1.256	N	0.002	1.0×10^{-1}	7.6×10^{-22}
	0.2178	-1.598	-0.2346	-0.209	1.645	N	0.002	2.5×10^{-2}	4.0×10^{-25}
	0.2618	-2.008	0.0125	-0.001	2.066	N	0.005	6.0×10^{-3}	2.7×10^{-21}
Γ-Σ-K-S-X axis Electrons	0.8662	2.420	0.9630	1.225	1.491	N	0.007	2.1×10^{-16}	3.0×10^{-20}
	0.0943	-0.033	-0.0025	1.162					
	0.1752	2.436	0.0051	1.165	1.507	N	0.023	6.2×10^{-13}	5.5×10^{-13}
	0.8620	2.489	0.8970	1.300	1.561	N	0.011	4.9×10^{-3}	1.2×10^{-17}
	0.0295	-0.023	-0.0055	1.166					
	0.1867	2.496	-0.9581	1.229	1.569	U	0.007	2.3×10^{-12}	0
	0.8601	2.520	0.8720	1.340	1.593	N	0.016	2.8×10^{-10}	2.5×10^{-10}
	0.0055	-0.013	-0.0064	1.167					
	0.2033	2.576	-0.9119	1.279	1.652	U	0.003	1.4×10^{-11}	0
	0.2111	2.611	-0.8968	1.300	1.688	U	0.002	1.7×10^{-2}	0
	0.4537	2.833	-0.8756	1.334	1.916	U	0.001	2.2×10^{-12}	0
	0.3947	2.926	-0.8861	1.434	2.012	U	0.011	7.3×10^{-12}	0
	0.3691	2.931	0.0130	1.184	2.017	N	0.008	2.7×10^{-11}	2.3×10^{-11}
	0.3733	2.932	-0.8176	1.456	2.018	U	0.001	3.7×10^{-3}	0
	0.1777	3.537	0.0696	1.656	2.641	N	0.023	1.4×10^{-10}	1.3×10^{-10}
	0.1844	3.547	0.1408	2.244	2.651	N	0.024	5.5×10^{-10}	5.3×10^{-10}
	-0.0198	-0.067	0.0238	1.236					
	0.7911	3.728	-0.7662	1.617	2.837	U	0.005	2.2×10^{-15}	0
	0.2855	3.819	-0.0151	1.192	2.931	N	0.020	3.2×10^{-16}	5.0×10^{-24}
	0.3973	-0.758	0.6979	1.869					

Table 9.2. continued:-

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_c	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ-Σ-K-S-X axis Electrons (continued)	0.3059	3.904	0.0372	1.333	3.018	N	0.018	5.4×10^{-2}	6.8×10^{-22}
	0.2259	4.012	0.0976	1.931	3.130	N	0.018	4.4×10^{-11}	4.3×10^{-11}
	0.3295	4.026	0.2872	2.825	3.143	N	0.014	8.0×10^{-2}	1.6×10^{-17}
	-0.0355	-0.033	0.0068	1.168					
	0.2322	4.040	-0.0164	1.197	3.158	N	0.001	2.4×10^{-13}	2.0×10^{-22}
	0.4188	-0.848	0.6674	1.995					
	0.3368	4.068	0.3267	2.891	3.187	N	0.017	1.6×10^{-7}	1.6×10^{-7}
	-0.0047	-0.011	0.0054	1.166					
	0.7621	4.235	-0.6579	2.036	3.359	U	0.008	5.4×10^{-2}	0
	0.7554	4.360	-0.6293	2.161	3.487	U	0.012	3.6×10^{-11}	0
	0.1434	4.417	-0.0156	1.194	3.546	N	0.006	2.8×10^{-11}	2.8×10^{-11}
	0.4049	-0.791	0.5639	2.432					
	0.7488	4.485	0.0118	1.180	3.616	N	0.023	7.6×10^{-13}	1.8×10^{-26}
Γ-Σ-K-S-X axis Holes	0.1098	-1.185	0.0541	-0.012	1.220	N	0.002	1.2×10^{-1}	7.4×10^{-23}
	0.2343	-1.261	0.1155	-0.049	1.298	N	0.002	9.1×10^{-13}	3.4×10^{-25}
	0.5673	-1.510	-0.1701	-0.113	1.554	U	0.011	8.3×10^{-15}	0
	0.5770	-1.554	-0.2070	-0.169	1.600	U	0.014	6.9×10^{-13}	0
	0.3171	-1.976	-0.2515	-0.263	2.034	U	0.004	9.3×10^{-13}	0
	0.3429	-2.202	-0.3193	-0.481	2.267	U	0.002	5.5×10^{-2}	0
	0.1938	-2.350	-0.2912	-0.384	2.419	U	0.043	1.5×10^{-2}	0
	0.2150	-2.650	-0.3824	-0.702	2.728	U	0.015	3.3×10^{-13}	0
	0.4359	-2.940	0.0336	-0.005	3.026	N	0.010	1.2×10^{-13}	3.6×10^{-24}
	0.2386	-2.972	-0.0763	-0.022	3.059	N	0.006	5.5×10^{-3}	7.9×10^{-26}
	0.8859	-3.032	0.4345	-0.016	3.121	U	0.002	3.8×10^{-13}	0
	0.8170	-3.468	0.2924	-0.388	3.570	U	0.037	9.1×10^{-13}	0
	0.2830	-3.540	0.0745	-0.021	3.644	N	0.018	4.8×10^{-12}	1.1×10^{-27}
	0.2932	-3.667	0.1159	-0.050	3.774	N	0.007	3.2×10^{-15}	2.6×10^{-22}
	0.2957	-3.698	0.1405	-0.075	3.806	N	0.010	6.8×10^{-2}	7.0×10^{-28}

correspond to intervalley transitions where one of the final states is near the conduction band minimum at Γ and the other is near the conduction band minimum along the Δ axis near X.

The lowest hole threshold along the Δ axis is higher than the lowest electron threshold, and corresponds to a normal process, which is similar to the situation along the Δ axis in Silicon. The error in the conservation of energy corresponding to all the thresholds along the Δ axis are much smaller than the maximum expected error, the largest error being less than 0.03eV.

In Silicon, it was surprising to see that the size of the matrix element of the coulomb interaction corresponding to the lowest electron threshold along the Δ axis was negligibly small. It is also surprising to see that the size of the matrix element corresponding to the lowest electron threshold along the Δ axis in Germanium is similarly negligibly small. It is even more surprising to see that all but two of the electron thresholds along this axis have negligibly small matrix element sizes, while one of those two thresholds has an insignificant matrix element size. The only threshold which has a significant matrix element size is that for which $E_T/E_G = 1.213$, and has a value of 0.22. Thus the situation is similar to that in Silicon, in that it would be erroneous to use the lowest electron threshold in related theories.

The lowest hole threshold, like that in Silicon, has a significant matrix element size, unlike the lowest electron threshold, and so it would not be erroneous to use it in related theories. The second lowest threshold does, however, have a negligibly small matrix element size, as does the second highest threshold, and so the effect of the matrix element sizes cannot be neglected. The approximate matrix element sizes, as given by equation 5.2.9, are all negligibly small, as they were in Silicon, those corresponding to umklapp processes being identically

zero as expected. Thus the significant matrix element sizes are not approximated accurately, and only a few of the negligibly small matrix element sizes are approximated accurately.

All the electron threshold energies along the Λ axis are seen to be close to the minimum possible energy for any threshold, and all correspond to umklapp processes. The lowest hole threshold energy is also close to the minimum possible energy for any threshold, while the other hole thresholds have substantially higher energies. As expected, the lowest electron and hole thresholds along this axis are lower than those along the Δ axis, since the energy band minimum occurs along this axis. The error in the conservation of energy corresponding to all the thresholds, with the exception of the two lowest electron thresholds, is very small, being less than 0.01eV. The larger errors in the conservation of energy corresponding to the two lowest electron thresholds is due to the error in the $\Gamma_2' - L_1$ conduction band mentioned in the previous section.

The situation where some of the thresholds determined have negligibly small matrix element sizes is again repeated, but only for the electron thresholds, two of the thresholds having negligibly small matrix element sizes. One of these negligibly small matrix element sizes corresponds to the lowest electron threshold, and while it would be erroneous to use it in related theories, the correct threshold to use is of almost equal energy. The approximate matrix element sizes follow the pattern of those corresponding to the thresholds along the Δ and Λ axes in Silicon; those corresponding to the electron thresholds being zero, and those corresponding to the hole thresholds being negligibly small, none of them being of comparable sizes to the proper calculations of the matrix element sizes.

There are more threshold positions determined along the Σ-S axis than along the other two axes, as expected from the details of the band structure and the results of Silicon. The lowest electron threshold along the Σ-S axis has an energy higher than the lowest electron threshold along the other two axes, and corresponds to a normal process, in contrast to the other electron thresholds which correspond to umklapp processes. It also corresponds to an intervalley transition where one of the final states is near the conduction band minimum at Γ and the other is near the conduction band minimum at X. This threshold is closely followed by a normal threshold given by a simple envelope solution, and then by another normal threshold of the same type as the lowest threshold. At higher energies there are many more thresholds, corresponding to both umklapp and normal processes, including some more intervalley transitions.

The lowest hole threshold along this axis corresponds to a normal process, and is lower than the lowest hole threshold along the Δ axis, but higher than that along the Λ axis. There are again many more threshold positions at higher energies, corresponding to both umklapp and normal processes, all given by simple envelope solutions. The errors in the conservation of energy of the thresholds are generally greater than those along the other two symmetry axes, but are still within the maximum expected error, being in error by less than 0.025eV for electron thresholds, and by 0.045eV for hole thresholds.

Yet again, it is seen by looking at the sizes of the matrix elements that there are very few thresholds which can be considered to be important due to their significant matrix element sizes. Most of the thresholds along this axis, including the lowest electron threshold, have negligibly small matrix element sizes. Although the lowest significant electron threshold has an energy not much higher than the lowest threshold energy,

the corresponding matrix element size is rather small, having a value of 0.005. The lowest hole threshold however, has a significant matrix element size which is not small, having a value of 0.12.

By selecting, from the many threshold positions determined along all three symmetry axes, the lowest electron and hole thresholds, whether they have significant matrix element sizes or not, it is seen that they both occur along the same axis, the Λ axis, and both have comparable energies, with values of $E_T/E_G = 1.023$ and 1.034 respectively. Since the lowest hole threshold has a significant matrix element size, it is not erroneous to use it in related theories. However, it would be erroneous to use the lowest electron threshold, although the correct threshold to use has an almost equal energy, having a value of $E_T/E_G = 1.031$. Thus the electron and hole thresholds to use in related theories have values of $E_T/E_G = 1.031$ and 1.034 respectively, and both occur along the Λ axis, as may be expected.

9.3 Comparison with Results from Approximate Band Structure Models

It is seen from Table 9.3 that there are only minor differences between the values of E_T/E_G given by the Franz construction and those given by the simpler parabolic band approximation. It is also seen that the approximate values for the Cohen and Bergstresser band structure are in good agreement with those given by the Envelope Method for the genuine bands. The agreement in the Stuckel band structure however is not so good, and it is seen that the approximate models based on the Stuckel band structure give a threshold along the Λ axis which ought not to exist. They also provide much too low a value for E_T/E_G along the Δ axis.

Comparison of Tables 9.1 and 9.3 for the Cohen and Bergstresser band structure shows that the approximate models fail to provide the lowest threshold, as was the case in Silicon, the minimum threshold along each symmetry axis provided by the genuine bands being substantially lower.

The comparison for the Stuckel band structure shows that the approximate models do provide the lowest threshold along the Δ axis but with a much too low a value for E_T/E_G. Thus, for one band structure the approximate band structure models give values of E_T/E_G which are considerably larger than the lowest value for the genuine band structure, and in the other band structure they give values of E_T/E_G which are considerably smaller than the lowest value for the genuine band structure.

Table 9.3

Comparison of E_T/E_G values for the different band structures

Band structure considered	Parabolic band approximation	Franz construction	Genuine band structure
Cohen & Bergstresser Γ-Λ-L axis (indirect gap)	1.06	1.11	1.11
Stuckel Γ-Λ-L axis (direct gap)	1.09	1.07	—
Cohen & Bergstresser Γ-Δ-X axis (indirect gap)	1.52	1.53	1.49
Stuckel Γ-Δ-X axis (direct gap)	1.20	1.17	1.32

It is seen from Table 9.4 that there are minor differences between the corresponding values of E_T/E_G given by the Franz construction and those given by the simpler parabolic band approximation only for the electron thresholds, the corresponding values for the hole thresholds differing substantially. It is also seen that the approximate values

given by the Franz construction are only in reasonable agreement with those given by the Envelope Method for three threshold positions, all along the Σ-S axis; the lowest electron threshold and the two lowest hole thresholds given by the Franz construction. The approximate values given by the parabolic band approximation are only in reasonable agreement for the electron threshold along the Σ-S axis, all other approximate values of E_T/E_G differing substantially from the corresponding values given by the Envelope Method. The approximate models

Table 9.4

Comparison of E_T/E_G values for different band structure models

Details of thresholds	Parabolic band approximation	Franz construction	Genuine band structure
Γ-Δ-X axis	1.566	1.559	1.374
Electrons	-	1.609	1.415
Γ-Δ-X axis	2.089	1.859	1.503
Holes	-	2.167	-
	-	3.023	2.066
Γ-Λ-L axis	1.356	1.348	1.026
Electrons	-	1.484	1.031
Γ-Λ-L axis	1.316	1.229	1.034
Holes	-	2.296	-
Γ-Σ-K-S-X axis	1.431	1.422	1.507
Electrons	-	1.589	-
	-	2.028	-
Γ-Σ-K-S-X axis	-	1.591	1.554
Holes	-	1.651	1.600
	2.546	2.351	-

also give some thresholds which ought not to exist, as did those in the preliminary study for the Stuckel band structure, and those along the Λ axis in Silicon.

Comparison of Tables 9.2 and 9.4 shows again that the approximate

values of E_T/E_G corresponding to all the thresholds, except the hole threshold along the Λ axis, fail to provide the lowest thresholds. Of those not provided, the hole threshold along the Δ axis and the electron thresholds along the Λ and the Σ-S axes are provided by slightly lower values, while the other two thresholds are provided by substantially lower values. However, the approximate values of the hole threshold along the Δ axis and of the electron and hole thresholds along the Λ axis are not in agreement with the values given by the Envelope Method, but are considerably higher. Thus the only threshold given by the approximate band structure models which is in reasonable agreement with the lowest threshold given by the Envelope Method is the electron threshold along the Σ-S axis, all other values being considerably higher than the lowest threshold values given by the Envelope Method. The Γ_2'-K_3-X_1 conduction band, and the Γ_{25}'-X_4, Γ_{25}'-L_3, and Γ_{25}'-K_2-X_4 valence bands are all too flat to provide all the thresholds given by the Franz construction, the lower thresholds only being provided.

10. IMPACT IONIZATION THRESHOLDS FOR 3C SILICON-CARBIDE

10.1 Details of the Calculations

The band structure used in the investigation of impact ionization thresholds in 3C Silicon-Carbide is reproduced from the data of Hemstreet and Fong [29]. As a result of the pilot study carried out, which was presented in Chapter 2, the effect of the non-local, angular-momentum-dependent potential term in the pseudopotential analysis is neglected. As in the investigations for Silicon and Germanium, the three principal symmetry directions are investigated, in an extended zone scheme, for impact ionization threshold initiated by hot electrons and by hot holes.

The energy bands investigated along the Δ axis are the Γ_1-X_1 and Γ_{15}-X_3 conduction bands and the Γ_{15}-X_5 and Γ_{15}-X_3 valence bands. The energy bands investigated along the Λ axis are the Γ_1-L_1, Γ_{15}-L_3 and Γ_{15}-L_1 conduction bands and the Γ_{15}-L_3 and Γ_{15}-L_1 valence bands. Those investigated along the Σ-S axis are the Γ_1-K_1-X_1, Γ_{15}-K_1-X_3, Γ_{15}-K_2-X_5 and Γ_{15}-K_1-X_5 conduction bands and the Γ_{15}-K_2-X_5, Γ_{15}-K_1-X_5 and Γ_{15}-K_1-X_3 valence bands. The numerical values of these energy bands are fitted by suitable Fourier series and, as in the cases for Silicon and Germanium, the analytic expressions used are, in general, nowhere in error by more than 0.01eV. However, the Γ_1-X_1 and Γ_1-K_1-X_1 conduction bands are, in parts, slightly less accurate. Where the threshold positions do not involve electron states in the inaccurate parts of these two conduction bands, the computations are carried out until the overall error associated with each threshold energy is less than 0.015eV.

The energy values on the Γ_1-X_1 conduction band calculated by the

Empirical Pseudopotential Method do not accurately reproduce the expected energy values as calculated by Hemstreet and Fong. The conduction band minimum does not occur at X as is expected, but along the Δ axis at a proportional distance of about 0.05 from the zone boundary with an energy of 2.31eV. Also the "smoothness" of the energy band is not reproduced, deviations from the expected values occurring at proportional distances of between 0.6 and 0.7 from the centre of the first Brillouin zone. This unexpected deviation causes greater errors in the curve fitting than are normally expected, and consequently some of the thresholds along this axis are subject to considerable errors in the conservation of energy.

Due to the shape of the Γ_1-K_1-X_1 conduction band, the errors in fitting the appropriate analytic expression are greater than 0.01eV in places, but are less than 0.015eV. The computations of threshold positions involving electron states in the slightly less accurate parts of this conduction band are carried out until the overall error associated with each threshold is less than 0.02eV. The results of the calculations of threshold positions are presented in Table 10.1, and are presented in the same manner as were the results of Tables 8.2 and 9.2 for Silicon and Germanium respectively. The values of E_T/E_G given by the Franz construction and by the parabolic band approximation are also calculated for comparative purposes, by the same procedure used previously, and the results are presented in Table 10.2, and are presented in the same manner as were the results of Tables 8.4 and 9.4.

10.2 Discussion of the Results

It is seen from Table 10.1 that along the Δ axis there are four electron thresholds, all corresponding to umklapp processes, and occurring in pairs with comparable energies. This is similar to the

Table 10.1. Impact Ionization Threshold Results.

The results are for the band structure of Hemstreet and Fong. (E_G=2.31eV. indirect gap $\Gamma_{15} \rightarrow \Delta_1$).

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Δ - X axis Electrons	0.3491	4.926	-0.8476	2.443	1.131	U	0.005	1.9×10^{-2}	0
	0.3460	4.944	-0.8375	2.462	1.139	U	0.008	3.8×10^{-3}	0
	0.7493	5.984	-0.6487	2.970	1.589	U	0.085	2.7×10^{-2}	0
	0.7450	6.007	-0.6414	2.986	1.599	U	0.099	1.3×10^{-2}	0
Γ - Δ - X axis Holes	0.5745	-2.997	-0.0928	-0.175	1.297	N	0.019	1.3	1.3
	0.3646	-3.692	-0.1278	-0.317	1.597	N	0.035	1.2×10^{-1}	2.5×10^{-4}
	0.5492	-6.359	-0.1195	-0.282	2.751	N	0.003	7.6×10^{-2}	9.3×10^{-3}
Γ - Λ - L axis Electrons	0.2536	8.876	-0.8910	4.434	2.841	U	0.006	2.7	0
	0.2541	8.883	-0.8756	4.441	2.844	U	0.004	9.9×10^{-2}	0
Γ - Λ - L axis Holes	0.4328	-4.907	-0.1233	-0.092	3.123	N	0.001	4.7×10^{-1}	1.3×10^{-2}
	0.6040	-7.262	0.6526	-1.288	3.142	N	0.003	2.9	9.2×10^{-2}
	0.6066	-7.295	0.2542	-0.344	3.156	N	0.002	3.9×10^{-2}	3.1×10^{-2}
	0.6143	-7.392	0.2840	-0.419	3.199	N	0.008	1.8×10^{-2}	1.3×10^{-4}
	0.6424	-7.739	-0.0901	-0.047	3.349	N	0.009	1.5×10^{-1}	1.1×10^{-2}
Γ-Σ-K-S-X axis Electrons	0.5051	6.103	-0.9423	2.542	1.641	U	0.000	2.0	0
	0.3709	6.814	-0.8904	2.991	1.948	U	0.009	2.0×10^{-1}	0
	0.3274	6.862	-0.8542	3.346	1.969	U	0.019	5.4×10^{-2}	0
	0.7966	8.089	-0.9486	2.505	2.500	U	0.004	7.8×10^{-2}	0
	0.7352	9.837	-0.6872	4.687	3.256	U	0.001	8.7×10^{-2}	0
	0.7285	10.054	-0.6506	4.968	3.350	U	0.002	1.6×10^{-4}	0
	0.7261	10.129	0.9139	2.764	3.383	U	0.004	3.0×10^{-1}	0
Γ-Σ-K-S-X axis Holes	0.6315	-2.684	-0.1732	-0.148	1.161	U	0.028	8.8×10^{-2}	0
	0.3667	-3.470	-0.2998	-0.524	1.501	U	0.040	3.9×10^{-2}	0
	0.2457	-4.058	-0.3500	-0.768	1.756	U	0.013	1.7×10^{-1}	0
	0.7973	-5.502	-0.0986	-0.056	2.380	N	0.016	2.1×10^{-1}	5.9×10^{-5}
	0.6103	-5.766	-0.1836	-0.174	2.495	U	0.017	1.6×10^{-2}	0
	0.3609	-6.265	0.1678	-0.136	2.711	N	0.003	3.2×10^{-1}	3.5×10^{-1}
	0.3725	-6.481	-0.2985	-0.519	2.804	U	0.040	1.6×10^{-2}	0
	0.3899	-6.798	0.1607	-0.122	2.941	N	0.012	3.4×10^{-2}	3.9×10^{-2}
	0.3936	-6.864	0.1767	-0.156	2.970	N	0.013	3.7×10^{-3}	1.3×10^{-3}
	0.3952	-6.892	0.1914	-0.196	2.982	N	0.014	4.4×10^{-3}	1.4×10^{-2}

situation in Silicon, and may be expected due to the similarities in the band structure along this axis. There are also three hole thresholds, all corresponding to normal processes, the lowest of which is substantially higher than the lowest electron threshold. As expected, the thresholds along this axis are determined fairly accurately, with the exception of those thresholds involving electron states in the inaccurate part of the Γ_1-X_1 conduction band, the error in the conservation of energy being less than 0.02eV for each threshold.

The two highest electron thresholds, which are considerably higher than the lowest two thresholds, involve electron states in the inaccurate part of the Γ_1-X_1 conduction band, as does the second lowest hole threshold, which is also considerably higher than the lowest hole threshold. The consequent errors in the conservation of energy are thus not unexpected, being 0.085eV, 0.099eV and 0.035eV respectively for the electron and hole thresholds. The errors in the initial positions of the hot electrons are not thought to be too large, as the electron states lie in the Γ_{15}-X_3 conduction band, and it is the final states of the electrons which are in error. The error for the hole threshold is mainly due to the error in the initial position of the promoted hole, and thus the error in the initial position of the hot hole is not thought to be very large.

The energy gap along the Λ axis in Silicon is much larger than that along the Δ axis, and consequently the threshold energies along the Λ axis were much higher than those along the Δ axis. The energy gap along the Λ axis in 3C Silicon-Carbide is similarly much larger than that along the Δ axis, and thus it is expected that the threshold energies along this axis will be much higher than those along the Δ axis. This is seen to be the case for both electron and hole thresholds.

Only two electron thresholds are determined, both corresponding to umklapp processes and both of comparable energies, while all the hole thresholds correspond to normal processes. The same situation concerning the type of threshold processes also occurs along the Δ axis, and also along the two corresponding axes in Silicon. Another comparison which can be made with Silicon is that while the lowest electron threshold along the Δ axis is lower than the lowest hole threshold, the opposite situation occurs along the Λ axis, the lowest hole threshold being lower than the lowest electron threshold.

Since the minimum energy gap occurs very close to X along the Δ axis, the energy gap along the Σ-S axis is of a comparable energy, and it may be expected that the lowest thresholds along this axis will be of comparable energies to those along the Δ axis. However, the lowest electron threshold along this axis is much higher than that along the Δ axis, and in contrast to this, the lowest hole threshold along this axis is substantially lower than that along the Δ axis, and is consequently also much lower than the lowest electron threshold along the Σ-S axis. As expected, all the electron thresholds along the Σ-S axis correspond to umklapp processes, but of the hole thresholds, some correspond to umklapp processes, including the lowest threshold, while others correspond to normal processes.

The errors in the conservation of energy corresponding to all the thresholds along the Λ axis are much smaller than the maximum expected error, the largest error being less than 0.01eV. The electron thresholds along the Σ-S axis also have corresponding errors in the conservation of energy much less than the maximum expected error, but a few being larger than those along the Λ axis, although being less than

0.025eV. Those corresponding to the hole thresholds along the Σ-S axis are generally larger than the others, but are all less than 0.04eV.

By considering only the threshold positions and the values of E_T/E_G, it would appear that the lowest electron threshold occurs along the Δ axis with a value of $E_T/E_G=1.131$, and that the lowest hole threshold occurs along the Σ-S axis with a value of $E_T/E_G=1.161$. In Silicon and Germanium, the corresponding situation was seen to be erroneous due to matrix element sizes being negligibly small, and was rather surprising. However, by looking at the matrix element sizes of the lowest electron and hole thresholds in 3C Silicon-Carbide, it would not have been erroneous, since both are of a significant size, and have values of 0.019 and 0.088 respectively. Indeed, by looking at the matrix element sizes corresponding to all the thresholds determined, the situations which occurred in Silicon and Germanium, in which very few matrix element sizes were significant, are seen not to be repeated. All the thresholds have corresponding matrix element sizes which are significant, which was the situation expected before the results of Silicon were obtained.

The lowest electron and hole thresholds given are thus the correct values to use in related theories, unlike the situations for Silicon and Germanium (for electrons only). Also, the largest matrix element sizes are considerably larger than the largest sizes in Silicon and Germanium, and it is surprising that they are even greater than unity. Of the approximate matrix element sizes which do not correspond to umklapp processes, and are consequently non-zero, only a few are in reasonable agreement with the proper calculations.

Thus the lowest electron threshold occurs along the Δ axis while

the lowest hole threshold occurs along the Σ-S axis, which is in contrast to the results of Silicon and Germanium for which both the lowest electron and hole thresholds occurred along the same symmetry axis. As in Germanium, both the thresholds are of comparable energies, and they also have comparable matrix element sizes. It is noted that all the threshold positions are determined by simple envelope solutions and that no double envelope solutions, which correspond to intervalley transitions, exist. This is due to the absence of multiple conduction band minima of comparable energies in the band structure, there being only the one distinct conduction band minimum.

10.3 Comparison with Results from Approximate Band Structure Models

The results presented in Table 10.2 show that there are minor differences between the corresponding values of E_T/E_G given by the Franz construction and those given by the simpler parabolic band approximation only for the electron thresholds along the Δ and Λ axes and for the hole threshold along the Σ-S axis. The corresponding values for the other thresholds differ substantially, the values given by the parabolic band approximation being higher than those given by the Franz construction. The approximate values of E_T/E_G are nowhere in agreement with the corresponding values given by the Envelope

Table 10.2

Comparison of E_T/E_G values for different band structure models

Details of Thresholds	Parabolic band approximation	Franz construction	Genuine band structure
Γ-Δ-X axis	1.683	1.651	1.131
Electrons	-	1.778	1.139
Γ-Δ-X axis	1.751	1.527	1.297
Holes	-	2.766	-
Γ-Λ-L axis	2.889	2.883	-
Electrons	-	2.924	-
Γ-Λ-L axis	2.850	1.934	-
Holes	-	3.266	-
Γ-Σ-K-S-X axis	2.620	2.182	1.641
Electrons	-	3.269	1.948
	-	3.998	1.969
Γ-Σ-K-S-X axis	1.554	1.534	1.161
Holes	-	2.558	-

Method for the genuine bands, all corresponding values being substantially higher. Indeed, as was the situation along the Λ axis in Silicon, the lowest lying conduction band and the highest lying valence band along the Λ axis (the Γ_1-L_1 and Γ_{15}-L_3 bands) are too flat to permit any ionization process to take place at all, and the approximate band structure models give threshold values for these bands which ought not to exist.

Along the Δ axis, the Γ_{15}-X_5 valence band is too flat to provide more than one hole threshold position, but the Franz construction provides a second hole threshold which thus ought not to exist. The same situation also occurs along the Σ-S axis, for which the Γ_{15}-K_2-X_5 valence band

is too flat to provide more than one hole threshold position. Thus, there are only values corresponding to the approximate band structure models given by the Envelope Method for the electron thresholds and the lower hole threshold along the Δ and the Σ-S axis. The lowest electron and hole thresholds along these axes are also provided by the approximate band structure models, as is seen by comparing Tables 10.1 and 10.2. However, the approximate values of E_T/E_G are substantially higher than the values given by the Envelope Method.

11. IMPACT IONIZATION THRESHOLDS FOR GALLIUM-PHOSPHIDE

11.1 Details of the Calculations

The band structure used in the investigations of impact ionization thresholds in Gallium-Phosphide is reproduced from the data of Walter and Cohen [26], and as in the investigations of the previous semiconductors, the three principal symmetry axes are investigated in an extended zone scheme. Again, impact ionization thresholds initiated by both hot electrons and hot holes are thus determined by the Envelope Method. The energy bands investigated are the Γ_1-X_1, Γ_{15}-X_3 and Γ_{15}-X_5 conduction bands and the Γ_{15}-X_5 and Γ_{15}-X_3 valence bands along the Δ axis, the Γ_1-L_1, Γ_{15}-L_3 and Γ_{15}-L_2 conduction bands and the Γ_{15}-L_3 and Γ_{15}-L_2 valence bands along the Λ axis, and the Γ_1-K_1-X_1, Γ_{15}-K_1-X_3, Γ_{15}-K_2-X_5 and Γ_{15}-K_1-X_5 conduction bands and the Γ_{15}-K_2-X_5, Γ_{15}-K_1-X_5 and Γ_{15}-K_1-X_3 valence bands along the Σ-S axis.

As in the previous investigations, the numerical values of these energy bands are fitted by suitable Fourier series, the analytic expressions used being nowhere in error by more than 0.01eV for all the energy bands except the Γ_{15}-X_5 and the Γ_{15}-K_1-X_3 conduction bands. The shapes of these two conduction bands are similar to some of the conduction bands in Germanium and 3C Silicon-Carbide which were fitted slightly less accurately, and consequently it is not unexpected that these energy bands are also in error by slightly greater than 0.01eV in places. They are however, in error by less than 0.015eV, but this slightly greater error has no noticeable effect on the overall accuracy of the energies of the thresholds which involve electron states in the Γ_{15}-X_5 conduction band. Thus, the computations are carried out until the overall error associated with each threshold energy is less than

0.015eV for all thresholds not involving electron states in the inaccurate parts of the Γ_{15}-K_1-X_3 conduction band. For the very few threshold positions which do involve electron states in the inaccurate parts of this conduction band, the threshold energies are in error by less than 0.02eV.

The results of the calculations of the threshold positions are presented in Table 11.1, and are presented in the same manner as were the results of Tables 8.2, 9.2 and 10.1 for Silicon, Germanium and 3C Silicon-Carbide respectively. The values of E_T/E_G given by the Franz construction and by the parabolic band approximation are also calculated, as before, for comparative purposes, and by the same procedures as used previously. The results of these approximate threshold values are presented in Table 11.2, in the same manner as were the results of Tables 8.4, 9.4 and 10.2.

11.2 Discussion of the Results

In the band structure of Germanium, there are three conduction band minima of comparable energies, and there are also three in the band structure of Gallium-Phosphide, one at Γ, one at X and one at L, the lowest of these occurring at X, unlike that in Germanium. As a consequence of these multiple conduction band minima and the possibility of intervalley transitions, there are a greater number of thresholds for Gallium-Phosphide, as there are for Germanium, than there are for Silicon and 3C Silicon-Carbide, as can be seen from Table 11.1.

Of the electron thresholds along the Δ axis, the lowest two are of comparable energies, the lowest having a value of E_T/E_G=1.118, and both corresponding to umklapp processes. All the other electron thresholds correspond to normal processes and occur at substantially higher energies than the lowest two thresholds. Two of these normal thresholds also

Table 11.1. Impact Ionization Threshold Results.

The results are for the band structure of Walter and Cohen. (E_G=?.24eV. indirect gap $\Gamma_{15}\rightarrow X_1$).

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Δ - X axis Electrons	0.5157	4.736	-0.7754	2.346	1.119	U	0.002	6.3×10^{-2}	0
	0.5116	4.762	-0.7541	2.374	1.130	U	0.004	9.0×10^{-3}	0
	0.1873	5.796	0.0298	2.823	1.592	N	0.012	2.4×10^{-2}	3.4×10^{-21}
	0.2651	6.268	0.5555	2.809	1.803	N	0.017	4.4×10^{-4}	1.6×10^{-22}
	0.2422	-0.493	-0.0482	2.877					
	0.2973	6.507	0.4322	3.391	1.910	N	0.001	5.5×10^{-14}	2.1×10^{-22}
	0.0704	-0.172	-0.0645	2.944					
	0.3484	6.904	0.1103	3.203	2.088	N	0.009	1.2×10^{-12}	9.6×10^{-13}
	0.4954	8.198	0.2664	4.091	2.666	N	0.025	5.5×10^{-3}	4.2×10^{-19}
Γ - Δ - X axis Holes	0.8669	-2.263	-0.0107	-0.002	1.012	N	0.000	1.6×10^{-2}	1.7×10^{-2}
	0.4198	-2.835	-0.1304	-0.157	1.268	N	0.000	1.3×10^{-1}	6.0×10^{-12}
	0.4986	-3.460	-0.1769	-0.284	1.547	N	0.004	3.6×10^{-2}	2.4×10^{-12}
	0.5739	-4.019	0.2603	-0.561	1.797	N	0.018	3.5×10^{-1}	7.0×10^{-14}
Γ - Λ - L axis Electrons	0.9037	5.723	-0.7688	2.639	1.560	U	0.002	1.7×10^{-2}	0
	0.2830	5.763	0.0229	2.804	1.577	N	0.007	5.8×10^{-15}	2.0×10^{-15}
	0.7937	5.811	-0.6175	2.891	1.599	U	0.001	3.0×10^{-14}	0
	0.3509	5.827	-0.0234	2.805	1.606	N	0.006	2.9×10^{-3}	2.1×10^{-22}
	0.4086	-0.403	0.7829	2.619					
	0.3013	5.894	-0.0248	2.807	1.636	N	0.002	1.8×10^{-11}	1.8×10^{-11}
	0.4422	-0.447	0.7683	2.640					
	0.4431	5.987	0.5198	3.106	1.677	N	0.002	6.4×10^{-10}	6.4×10^{-10}
	0.0334	-0.040	-0.0433	2.841					
	0.3424	6.216	0.4127	3.348	1.780	N	0.001	1.3	1.3
	0.0308	-0.034	-0.0395	2.834					
	0.4103	6.822	0.1793	3.365	2.051	N	0.014	3.1×10^{-1}	3.1×10^{-1}
	0.4424	7.135	0.3079	3.524	2.191	N	0.003	2.1×10^{-12}	1.9×10^{-12}

Table 11.1. continued:-

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Λ - L axis Holes	0.4013	-2.818	-0.2089	-0.121	1.360	N	0.006	2.9×10^{-1}	2.4×10^{-13}
	0.4263	-3.031	0.2039	-0.116	1.356	N	0.002	9.2×10^{-1}	6.9×10^{-12}
	0.4995	-3.625	0.1211	-0.047	1.621	N	0.011	7.5×10^{-1}	5.8×10^{-13}
	0.6149	-4.482	0.4768	-0.497	2.004	N	0.004	8.8×10^{-1}	9.9×10^{-13}
	0.8462	-5.707	-0.0585	-0.012	2.552	N	0.002	2.0×10^{-1}	5.5×10^{-2}
Γ-Σ-K-S-X axis Electrons	0.7784	5.144	0.9878	2.247	1.300	N	0.018	9.9×10^{-2}	1.0×10^{-1}
	0.2019	-0.103	-0.0075	2.794					
	0.7555	5.539	0.8415	2.640	1.477	N	0.016	2.7×10^{-1}	2.3×10^{-1}
	0.0678	-0.085	-0.0182	2.814					
	0.1014	5.602	0.0024	2.790	1.505	N	0.000	3.6×10^{-14}	3.1×10^{-14}
	0.7480	5.672	0.7969	2.786	1.537	U	0.014	1.1	0
	0.0224	-0.045	-0.0265	2.841					
	0.2194	5.694	0.0076	2.794	1.546	N	0.002	3.3×10^{-2}	5.0×10^{-26}
	0.1218	5.765	0.0220	2.826	1.578	N	0.001	2.3×10^{-2}	1.8×10^{-8}
	0.7332	5.947	-0.8240	2.691	1.659	U	0.011	1.3×10^{-3}	0
	0.4444	6.009	0.8313	2.669	1.687	U	0.008	1.9×10^{-4}	0
	0.3667	-0.520	-0.0202	2.820					
	0.1772	6.323	0.0648	3.067	1.828	N	0.001	5.4×10^{-12}	7.5×10^{-12}
	0.3496	6.332	-0.0270	2.843	1.832	U	0.012	1.9×10^{-2}	0
	0.4166	-0.689	0.7932	2.800					
	0.5100	6.530	0.6615	3.379	1.921	N	0.012	9.4×10^{-2}	8.2×10^{-2}
	0.1164	-0.272	-0.0351	2.879					
	0.5109	6.538	0.0287	2.850	1.924	N	0.016	4.5×10^{-14}	4.2×10^{-14}
	0.6985	6.583	-0.7043	3.177	1.944	U	0.007	1.2×10^{-2}	0
	0.5287	6.675	0.5969	3.703	1.985	N	0.015	9.7×10^{-1}	6.9×10^{-2}
	0.0308	-0.082	-0.0374	2.890					

Table 11.1. continued:-

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_m/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ-Σ-K-S-X axis Electrons (continued)	0.6882	6.746	-0.6702	3.337	2.017	U	0.012	3.4×10^{-3}	0
	0.2263	6.829	-0.0290	2.851	2.054	U	0.013	6.5×10^{-13}	0
	0.5148	-1.085	0.7701	2.893					
	0.4312	6.925	0.5951	3.712	2.097	N	0.001	1.1×10^{-3}	7.4×10^{-13}
	0.1264	-0.322	-0.0375	2.891					
	0.4503	7.073	0.5112	4.136	2.163	N	0.011	2.1×10^{-12}	2.4×10^{-12}
	0.0276	-0.067	-0.0333	2.870					
	0.6510	7.128	0.0289	2.851	2.188	N	0.018	2.3×10^{-3}	2.7×10^{-6}
	0.3675	7.250	0.5097	4.144	2.243	N	0.003	9.4×10^{-4}	1.3×10^{-3}
	0.1091	-0.237	-0.0331	2.869					
	0.3816	7.269	0.0739	3.137	2.251	N	0.003	2.9×10^{-3}	1.6×10^{-12}
	0.3905	7.284	0.4093	4.479	2.258	N	0.002	2.1	2.1
	0.0086	-0.007	-0.0102	2.798					
Γ-Σ-K-S-X axis Holes	0.8950	-2.257	-0.0511	-0.008	1.009	N	0.004	1.1×10^{-3}	1.7×10^{-3}
	0.9133	-2.638	-0.0423	-0.006	1.180	N	0.015	2.3×10^{-3}	8.6×10^{-15}
	0.4368	-2.717	-0.2652	-0.215	1.215	U	0.020	5.2×10^{-3}	0
	0.2740	-2.874	0.1347	-0.041	1.285	N	0.003	3.8×10^{-3}	1.7×10^{-28}
	0.4750	-3.017	-0.2527	-0.185	1.349	U	0.005	5.1×10^{-3}	0
	0.4873	-3.108	0.2375	-0.154	1.390	N	0.001	1.4×10^{-2}	4.5×10^{-13}
	0.2946	-3.117	-0.2712	-0.232	1.394	U	0.035	4.6×10^{-2}	0
	0.3252	-3.457	-0.3240	-0.395	1.546	U	0.003	7.6×10^{-2}	0
	0.7273	-3.784	0.3538	-0.483	1.692	N	0.030	1.2×10^{-2}	5.6×10^{-4}
	0.4429	-4.511	0.0353	-0.004	2.017	N	0.004	1.7×10^{-2}	2.4×10^{-14}
	0.5722	-5.206	0.1812	-0.082	2.328	N	0.022	2.6×10^{-1}	9.3×10^{-13}
	0.8166	-5.672	0.2879	-0.281	2.537	U	0.054	9.9×10^{-3}	0

correspond to intervalley transitions, being determined by double envelope solutions, for which the final state of the promoted electron is near the conduction band minimum at Γ, and the final state of the hot electron is near the conduction band minimum at X. The hole thresholds along this axis all correspond to normal processes, as is expected from the results of the semiconductors investigated previously. The lowest threshold energy is close to the minimum possible energy for any threshold, and is also lower than the lowest electron threshold along this axis. The errors in the conservation of energy corresponding to all the threshold positions along this axis are, as may be expected from the previous results, much smaller than the maximum expected error, being less than 0.02eV for all the thresholds except one. This threshold is the highest electron threshold determined, and the slightly larger error of 0.025eV associated with it is due to the inaccuracies of the Γ_{15}-X_5 conduction band, in which the hot electron state initially lies.

In Silicon, several of the thresholds are insignificant due to the negligible sizes of the matrix elements corresponding to those thresholds, and in Germanium, most of the thresholds are insignificant due to the same reason. However, in 3C Silicon-Carbide all the thresholds have matrix element sizes which are significant. The situation for Gallium-Phosphide is slightly different from these situations, in that most of the thresholds have corresponding matrix element sizes which are significant. The lowest electron and hole thresholds both have corresponding matrix element sizes which are significant, thus it would not have been erroneous to have taken them for use in related theories without considering the matrix element sizes. Indeed, there are only

two insignificant matrix element sizes corresponding to thresholds along this axis, both electron thresholds with large values of E_T/E_G. The matrix element sizes given by the approximate equation 5.2.9 are seen to be in agreement with the proper calculations of equation 5.2.5 for only two thresholds, the lowest hole threshold and one of the higher electron thresholds, of which that corresponding to the lowest hole threshold is the only one which is of a significant size.

The energy gap along the Λ axis is slightly higher than that along the Δ axis, and consequently the threshold energies are also higher along the Λ axis. The lowest electron threshold is provided by an umklapp process, and has a considerably higher energy than that of the lowest electron threshold along the Δ axis. At higher energies there are several more thresholds, only one of which corresponds to an umklapp process, all the others corresponding to normal processes including a few intervalley transitions given by double envelope solutions. The lowest hole threshold is also higher along this axis than that along the Δ axis, and similar to the situation along the Δ axis, is lower than the lowest electron threshold. Once again, all the hole thresholds along the Λ axis correspond to normal processes. The errors in the conservation of energy corresponding to all the thresholds along this axis are again very small, the error being nowhere greater than 0.015eV, and only exceeding 0.01eV for two thresholds, one electron and one hole threshold.

As along the Δ axis, the lowest electron and hole thresholds along the Λ axis both have corresponding matrix element sizes which are significant. However, the sizes of the matrix elements corresponding to the thresholds along this axis tend to follow the situation along

the Λ axis in Germanium, in that several of those corresponding to electron thresholds are of an insignificant size (about half) while all those corresponding to hole thresholds are of a significant size. In contrast to the situation in the semiconductors investigated previously, the approximate sizes of the matrix elements corresponding to the electron thresholds are nearly all in agreement with the proper calculations, only those corresponding to the two umklapp processes and to one normal process differing substantially. The situation for the hole thresholds however, is similar to those previously, none of the approximate matrix element sizes being in agreement with the proper calculations.

The lowest of the conduction band minima in Gallium-Phosphide occurs at X, and hence the energy gap along the Σ-S axis is the same as that along the Δ axis. The lowest thresholds along these two symmetry axes may therefore be expected to be of comparable energies. This is not the case for the electron thresholds, the lowest threshold along the Σ-S axis being substantially higher than that along the Δ axis. This threshold is provided by a normal, intervalley transition, which is in contrast to the situation along the other two symmetry axes, but similar to the situation along the Σ-S axis in Germanium, which is a direct consequence of the multiple conduction band minima of comparable energies. At higher energies there are several umklapp processes, and also some more normal processes, some of which are also intervalley transitions. It is interesting to note that there are also a few umklapp processes which also correspond to intervalley transitions, a direct consequence of the zone boundary being $\frac{3}{4}$ of the proportional distance along this axis from Γ.

The lowest hole threshold, as expected, does have an energy comparable to the energy of the lowest hole threshold along the Δ axis, being very slightly lower. This lowest threshold is provided by a normal process, and the situation for the thresholds at higher energies is similar to that in Germanium and 3C Silicon-Carbide, there being some umklapp processes and some normal processes. As may be expected from the errors in the conservation of energy corresponding to the thresholds along the Σ-S axis in Silicon, Germanium and 3C Silicon-Carbide, those along the Σ-S axis in Gallium-Phosphide are slightly greater than those along the other two axes. The errors associated with all the electron threshold energies however, are still very small, being nowhere greater than 0.02eV, while the errors associated with all but the highest hole threshold are nowhere greater than 0.035eV. The error in the conservation of energy of 0.054eV associated with the highest hole threshold, although less than the maximum expected error, is due to the inaccuracies in the Γ_{15}-K_1-X_3 conduction band.

Similar to the situation along the Δ and Λ axes, only a few of the matrix element sizes corresponding to the electron thresholds along the Σ-S axis can be considered to be insignificant, while those corresponding to the hole thresholds along this axis are all significant. Also, the lowest electron threshold has a significant size of matrix element, as does the lowest hole threshold. Again, the matrix element sizes given by the approximate equation are in agreement with the proper calculations corresponding to several of the electron thresholds, but to only the lowest hole threshold. The proper calculations of the matrix element sizes corresponding to the majority of the thresholds along the Σ-S axis are not approximated accurately by the approximate

calculations.

Thus, from the many thresholds determined along all three symmetry axes, the lowest electron threshold is provided by an umklapp process along the Δ axis with a value of $E_T/E_G=1.118$, and the lowest hole threshold is provided by a normal process along the Σ-S axis with a value of $E_T/E_G=1.009$. However, the lowest hole threshold along the Δ axis is only very slightly higher than that along the Σ-S axis, having a value of $E_T/E_G=1.012$. Since the matrix element sizes corresponding to all three of these thresholds are significant, they can also be considered to be the lowest significant thresholds. The hole threshold along the Δ axis, although it has a very slightly greater energy than that along the Σ-S axis, has a matrix element size greater than that along the Σ-S axis by an order of magnitude. Thus, the lowest hole threshold along the Δ axis should be used in related theories in preference to that along the Σ-S axis, and hence the lowest threshold for both electrons and holes occurs along the Δ axis.

The values of E_T have also been calculated by Anderson and Crowell [32], but using the band structure of Cohen and Bergstresser [25]. The band structure of Walter and Cohen [26], which is reproduced and used in the present work, differs slightly from that of Cohen and Bergstresser and is regarded as being the more accurate. The results presented in this work are thus compared with the results of Anderson and Crowell, and very good agreement is obtained within the errors of the respective methods. The lack of absolute agreement is almost certainly due to the slight differences in the two band structures investigated.

11.3 Comparison with Results from Approximate Band Structure Models

It is seen from Table 11.2 that there are minor differences between the corresponding values of E_T/E_G given by the Franz construction and those given by the parabolic band approximation for all the electron thresholds, but for only the hole threshold along the Σ-S axis. The corresponding values for the hole thresholds along the Δ and Λ axes are not in good agreement, the values given by the parabolic band approximation being substantially higher than those given by the Franz construction. It is also seen that the approximate values of E_T/E_G given by both the approximate band structure models are only in agreement with the value given by the Envelope Method for the lowest hole threshold along the Σ-S axis. In fact, there is only one other threshold given by the Envelope Method for which the approximate values may be compared, the lowest hole threshold along the Δ axis. This situation also occurred in the investigations of Silicon, Germanium and 3C Silicon-Carbide, but is much more pronounced here.

Comparison of Tables 11.1 and 11.2 shows that the two threshold values given by the Envelope Method for which approximate threshold values also exist, correspond to the lowest hole thresholds along the Δ and the Σ-S axes. Thus, the lowest hole threshold along the Σ-S axis is also given accurately by both the approximate band structure models,

Table 11.2

Comparison of E_T/E_G values for different band structure models

Details of Thresholds	Parabolic band approximation	Franz construction	Genuine band structure
Γ-Δ-X axis	1.096	1.094	-
Electrons	-	1.107	-
Γ-Δ-X axis	1.299	1.219	1.012
Holes	-	1.527	-
Γ-Λ-L axis	1.342	1.341	-
Electrons	-	1.357	-
Γ-Λ-L axis	1.245	1.128	-
Holes	-	2.583	-
Γ-Σ-K-S-X axis	1.808	1.769	-
Electrons	-	2.649	-
	-	3.288	-
Γ-Σ-K-S-X axis	1.063	1.061	1.009
Holes	-	1.212	-
	-	2.426	-

although by slightly higher values, while that along the Δ axis is given by approximate values which are considerably higher than the value given by the Envelope method.

The Γ_1-X_1 conduction band along the Δ axis is too flat to permit any ionization process to take place at all, and thus the E_T/E_G values given by the approximate band structure models ought not to exist. The same situation also applies to the Γ_1-L_1 conduction band along the Λ axis, the Γ_1-K_1-X_1 conduction band along the Σ-S axis and the Γ_{15}-L_3 valence band along the Λ axis. The Γ_{15}-X_5 valence band along the Δ axis is only able to provide one hole threshold position, and

thus only one E_T/E_G value given by the approximate band structure models ought to exist, as is the Γ_{15}-K_2-X_5 valence band along the Σ-S axis.

12. IMPACT IONIZATION THRESHOLDS FOR GALLIUM-ARSENIDE

12.1 Details of the Calculations

As in the investigation of impact ionization thresholds in Gallium-Phosphide, the band structure used in the investigation of impact ionization thresholds in Gallium-Arsenide is reproduced from the data of Walter and Cohen [26], and the investigations are carried out along the three principal symmetry axes in an extended zone scheme. Again, impact ionization thresholds initiated by both hot electrons and hot holes are determined by the Envelope Method. The energy bands investigated are the Γ_1-X_1 and Γ_{15}-X_3 conduction bands and the Γ_{15}-X_5 and Γ_{15}-X_3 valence bands along the Δ axis, the Γ_1-L_1, Γ_{15}-L_3 and Γ_{15}-L_2 conduction bands and the Γ_{15}-L_3 and Γ_{15}-L_2 valence bands along the Λ axis, and the Γ_1-K_1-X_1, Γ_{15}-K_1-X_3, Γ_{15}-K_2-X_5 and Γ_{15}-K_1-X_5 conduction bands and the Γ_{15}-K_2-X_5, Γ_{15}-K_1-X_5 and Γ_{15}-K_1-X_3 valence bands along the Σ-S axis.

The numerical values of these energy bands are fitted by suitable Fourier series, the analytic expressions used being nowhere in error by more than 0.01eV for all the energy bands except the Γ_1-X_1, Γ_{15}-X_3 and Γ_{15}-K_1-X_3 conduction bands. The shapes of these three conduction bands are similar to some of the conduction bands in Gallium-Phosphide, and also in Germanium and 3C Silicon-Carbide, which were fitted less accurately. Consequently, the errors involved in fitting the analytic expressions to these conduction bands are greater than 0.01eV in places, as may be expected. The errors in the Γ_1-X_1 and Γ_{15}-X_3 conduction bands are only greater than 0.01eV in the region between about 0.3 and 0.4 of the proportional distance along the Δ axis measured from Γ. The errors in this region vary considerably, but the analytic expressions fitted are nowhere in error by more than 0.03eV, and as a result of these inaccuracies,

a few threshold positions, for both electrons and holes, are determined less accurately.

The errors in the Γ_{15}-K_1-X_3 conduction band are only slightly greater than 0.01eV in places, and are nowhere in error by more than 0.015eV. Thus, for threshold positions not involving electron states in the inaccurate parts of these three conduction bands, the computations are carried out until the overall error associated with each threshold energy is less than 0.015eV. For the very few thresholds which involve electron states in the inaccurate parts of the Γ_{15}-K_1-X_3 conduction band, the threshold energies are in error by less than 0.02eV, while the threshold energies of those few thresholds involving electron states in the inaccurate parts of the Γ_1-X_1 or Γ_{15}-X_3 conduction bands are in error by less than 0.035eV.

The results of the calculations of the threshold positions are presented in Table 12.1, and are presented in the same manner as were the results of Tables 8.2 and 9.2 for Silicon and Germanium. As usual, the values of E_T/E_G given by the Franz construction and by the parabolic band approximation are also calculated, again for comparative purposes, by the same procedures used previously. The results of these approximate threshold values are presented in Table 12.2 in the same manner as were the results of Tables 8.4 and 9.4.

12.2 Discussion of the Results

The previous semiconductors investigated all have indirect band gaps, unlike the band gap in Gallium-Arsenide, which is a direct gap from Γ_{15}-Γ_1. However, similar to the band structures of Germanium and Gallium-Phosphide, there are a further two conduction band minima of comparable energies, one at L and one at X. Consequently there are a similar number of impact ionization threshold positions in Gallium-Arsenide as there are in Germanium and Gallium-Phosphide,

Table 12.1. Impact Ionization Threshold Results.

The results are for the band structure of Walter and Cohen. (E_G=1.45eV. direct gap $\Gamma_{15} \rightarrow \Gamma_1$).

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
$\Gamma - \Delta - X$ axis Electrons	0.2498	3.280	0.0234	1.480	1.266	N	0.044	6.1×10^{-12}	6.0×10^{-12}
	0.3270	3.546	0.0349	1.520	1.450	N	0.065	2.0×10^{-12}	3.7×10^{-12}
	0.6603	3.625	0.7276	2.144	1.505	N	0.014	1.7×10^{-10}	1.7×10^{-10}
	0.0584	-0.030	-0.0089	1.451					
	0.6395	3.736	0.6819	2.233	1.581	N	0.009	6.9×10^{-4}	6.9×10^{-4}
	0.0216	-0.030	-0.0208	1.473					
	0.2423	3.929	0.0758	1.762	1.715	N	0.031	7.6×10^{-2}	7.0×10^{-2}
	0.3782	4.197	0.5906	2.435	1.900	N	0.036	7.9×10^{-13}	1.3×10^{-12}
	0.1911	-0.287	-0.0213	1.475					
	0.5208	4.249	0.0269	1.491	1.936	N	0.008	7.1×10^{-13}	7.3×10^{-13}
	0.5153	4.263	-0.7471	2.131	1.945	U	0.018	1.4×10^{-13}	0
	0.5134	4.268	-0.7445	2.132	1.948	U	0.017	5.6×10^{-6}	0
	0.4005	4.301	-0.7799	2.144	1.971	U	0.009	2.3×10^{-3}	0
	0.4047	4.314	-0.7947	2.154	1.980	U	0.052	8.2×10^{-2}	0
	0.1658	4.335	-0.9084	2.167	1.995	U	0.006	2.6×10^{-2}	0
	0.1637	4.343	-0.9165	2.170	2.001	U	0.001	2.8×10^{-2}	0
	0.4305	4.358	0.4739	2.852	2.011	N	0.007	6.8×10^{-7}	3.4×10^{-11}
	0.0221	-0.031	-0.0213	1.475					
$\Gamma - \Delta - X$ axis Holes	0.2473	-1.703	0.1155	-0.120	1.177	N	0.004	6.4×10^{-2}	2.1×10^{-11}
	0.3594	-2.664	-0.1684	-0.233	1.340	N	0.009	3.6×10^{-2}	1.2×10^{-13}
	0.4463	-3.386	-0.2085	-0.335	2.339	N	0.022	3.1×10^{-2}	5.1×10^{-13}
	0.4894	-3.733	0.0949	-0.081	2.579	N	0.032	3.4×10^{-2}	4.9×10^{-12}
	0.5165	-3.945	0.1076	-0.105	2.726	N	0.040	4.9×10^{-1}	1.7×10^{-11}
$\Gamma - \Lambda - L$ axis Electrons	0.3914	5.471	0.3925	2.544	2.780	N	0.005	3.1×10^{-3}	4.5×10^{-20}
	0.3646	5.503	0.3952	2.541	2.802	N	0.003	3.5×10^{-11}	3.4×10^{-11}
	0.9856	5.575	0.2693	2.563	2.852	N	0.026	6.7×10^{-2}	2.6×10^{-13}

Table 12.1. continued:-

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ - Λ - L axis Holes	0.2359	-1.534	0.1148	-0.041	1.060	N	0.005	1.6	1.2×10^{-14}
	0.3102	-2.192	-0.2219	-0.131	1.514	N	0.003	4.9×10^{-2}	7.2×10^{-14}
	0.3565	-2.593	0.0235	-0.001	1.791	N	0.007	1.0×10^{-1}	2.1×10^{-14}
	0.6972	-5.113	0.2464	-0.163	3.533	N	0.001	1.4×10^{-1}	1.2×10^{-12}
	0.7623	-5.461	0.3092	-0.262	3.773	N	0.001	9.0×10^{-2}	5.2×10^{-2}
	0.7947	-5.612	-0.0524	-0.009	3.878	N	0.003	7.4×10^{-1}	1.8×10^{-2}
Γ-Σ-K-S-X axis Electrons	0.1916	3.015	0.0046	1.457	1.083	N	0.009	6.4×10^{-16}	8.1×10^{-16}
	0.8473	3.716	0.9877	2.205	1.567	N	0.016	1.8×10^{-2}	1.9×10^{-2}
	0.1364	-0.054	-0.0040	1.457					
	0.4797	3.769	0.0171	1.488	1.604	N	0.002	2.2×10^{-12}	1.4×10^{-12}
	0.8353	3.913	0.8968	2.404	1.703	N	0.020	3.7	3.7
	0.0533	-0.047	-0.0082	1.462					
	0.8318	3.970	0.8495	2.488	1.743	N	0.023	2.8	3.9×10^{-1}
	0.0085	-0.017	-0.0092	1.465					
	0.2025	4.505	-0.9858	2.206	2.112	U	0.022	3.1×10^{-4}	0
	0.2304	4.528	0.0891	2.108	2.128	N	0.015	9.0×10^{-2}	8.6×10^{-2}
	0.1377	4.560	-0.9628	2.246	2.150	U	0.028	1.2×10^{-2}	0
	0.1182	4.598	-0.9463	2.287	2.177	U	0.004	3.1×10^{-2}	0
	0.3125	4.666	-0.9664	2.238	2.224	U	0.004	7.7×10^{-4}	0
	0.7805	4.788	0.0062	1.459	2.308	N	0.021	1.6×10^{-5}	2.0×10^{-5}
	0.3511	4.843	-0.0151	1.481	2.346	U	0.001	3.5×10^{-14}	0
	0.4044	-0.642	0.7706	2.720					
	0.4009	5.127	-0.8335	2.524	2.542	U	0.034	5.3×10^{-4}	0
	0.4233	5.254	0.5223	3.649	2.630	N	0.002	1.6×10^{-4}	1.9×10^{-5}
	0.0843	-0.126	-0.0147	1.479					

Table 12.1. continued:-

Details of thresholds	Initial state of hot electron k_h	E_h	Final state of hot electron k_i	E_i	E_T/E_G	Type of process	Error in the conservation of energy	Matrix element from equation 5.2.5	Matrix element from equation 5.2.9
Γ-Σ-K-S-X axis Electrons (continued)	0.2253	5.292	-0.0158	1.483	2.656	N	0.014	1.8×10^{-2}	1.2×10^{-15}
	0.4888	-0.951	0.7299	2.858					
	0.4303	5.296	0.4343	3.831	2.659	N	0.021	9.2×10^{-1}	9.2×10^{-1}
	0.0019	-0.010	-0.0021	1.455					
	0.4412	5.367	-0.7862	2.667	2.708	U	0.027	5.8×10^{-3}	0
	0.1558	5.548	-0.0178	1.492	2.833	N	0.011	8.0×10^{-4}	4.1×10^{-4}
	0.5307	-1.105	0.7043	2.953					
	0.7303	5.661	-0.8237	2.550	2.911	U	0.016	2.0×10^{-4}	0
	0.7050	6.142	-0.6983	2.977	3.244	U	0.019	2.9×10^{-3}	0
	0.7017	6.201	-0.0098	1.466	3.284	N	0.006	4.3×10^{-2}	1.6×10^{-2}
	0.6950	6.315	-0.6615	3.133	3.363	U	0.003	8.1×10^{-4}	0
Γ-Σ-K-S-X axis Holes	0.1455	-1.485	0.0720	-0.015	1.026	N	0.002	3.0×10^{-7}	1.7×10^{-27}
	0.3227	-1.610	0.1591	-0.076	1.112	N	0.013	2.7×10^{-2}	2.5×10^{-12}
	0.9552	-2.201	-0.0213	-0.001	1.521	N	0.009	3.9×10^{-6}	7.7×10^{-6}
	0.8978	-2.463	-0.0498	-0.008	1.701	N	0.016	4.3×10^{-4}	1.6×10^{-16}
	0.4661	-2.640	-0.2488	-0.198	1.824	U	0.004	2.5×10^{-3}	0
	0.8362	-2.808	0.4102	-0.663	1.940	U	0.019	8.0×10^{-3}	0
	0.4971	-2.833	-0.2422	-0.184	1.957	U	0.005	2.3×10^{-3}	0
	0.2718	-3.089	-0.2729	-0.261	2.134	U	0.036	7.9×10^{-4}	0
	0.2978	-3.398	-0.3390	-0.459	2.348	U	0.014	7.5×10^{-4}	0
	0.3380	-3.846	-0.0458	-0.007	2.657	N	0.000	8.6×10^{-4}	6.4×10^{-18}
	0.4112	-4.564	0.1026	-0.029	3.153	N	0.002	2.6×10^{-2}	6.8×10^{-15}
	0.4771	-5.116	0.1939	-0.113	3.535	N	0.001	1.4×10^{-3}	7.2×10^{-3}
	0.4820	-5.153	0.2264	-0.155	3.560	N	0.002	5.1×10^{-3}	6.5×10^{-14}

due to the multiple conduction band minima and the possibility of intervalley transitions.

Along the Δ axis in the semiconductors investigated previously, the lowest two electron threshold were seen to have comparable energies, the lowest threshold promoting an electron from the higher lying valence band and the higher threshold promoting an electron from the lower lying valence band. The situation along the Δ axis in Gallium-Arsenide is in contrast to these results, the lowest threshold being substantially lower than the second lowest threshold, having values of E_T/E_G = 1.266 and 1.450 respectively. This is due to the Γ_1-X_1 conduction band being too flat to permit impact ionization processes to take place which involve electrons in the Γ_{15}-X_3 valence band. Both these thresholds are provided by normal processes, and the final states of both the electrons, for both thresholds, are near the conduction band minimum at Γ, as may be expected. At higher energies there are, as usual, several more thresholds provided by both umklapp and normal processes, and including a few intervalley transitions.

The hole thresholds along the Δ axis all correspond to normal processes, as is expected, and similar to the situation in Gallium-Phosphide, the lowest hole threshold is lower than the lowest electron threshold along this axis. The next lowest hole threshold however, has a very much larger energy than the lowest (E_T/E_G=1.840 compared with E_T/E_G=1.177 for the lowest threshold). The errors in the conservation of energy associated with the thresholds along this axis are not particularly good, the large errors occurring due to the inaccuracies involved in fitting the analytic expressions to the two conduction bands. The large errors in the conservation of energy associated with the two lowest electron thresholds are due to the initial states of the hot electrons being in the inaccurate part

of the Γ_1-X_1 conduction band. Two other electron thresholds, at much higher energies, also have large errors in the conservation of energy associated with them due to the initial hot electron states being in the inaccurate part of the Γ_{15}-X_3 conduction band. The errors in the conservation of energy associated with the hole thresholds are much smaller than those for the electron thresholds, and are all less than the maximum expected error.

There are surprisingly few electron thresholds along the Λ axis (only three) considering the similarities in the Gallium-Phosphide and Gallium-Arsenide band structures. These thresholds also occur at very high energies, all having energies larger than all the thresholds determined, both electron and hole thresholds, along the Δ axis. These large energies are due to the Γ_1-L_1 conduction band being too flat to permit any ionization process to take place at all, the hot electron in each threshold being provided by either the Γ_{15}-L_3 or the Γ_{15}-L_2 conduction band.

The lowest hole threshold along this axis has a value of E_T/E_G=1.060, close to the minimum possible energy for any threshold, and lower than the lowest hole threshold along the Δ axis. Similar to the situation along the Δ axis, the next lowest hole threshold along the Λ axis occurs at an energy considerably higher than that of the lowest hole threshold. Once again, all the hole thresholds along this axis are provided by normal processes. The errors in the conservation of energy corresponding to all the threshold along this axis, except the highest electron threshold, are very small indeed, as may be expected, none being greater than 0.01eV. The error corresponding to the highest electron threshold, although larger than all the others along this axis, is still smaller than the maximum expected error.

Considering now the thresholds along the Σ-S axis, the situation is seen to be similar to that along the Σ-S axis in both Germanium and Gallium-Phosphide as expected. There are many electron thresholds, provided by both umklapp and normal processes, some of which also correspond to intervalley transitions, and there are also many hole thresholds, again provided by both umklapp and normal processes. The lowest electron threshold along this axis has an energy only slightly greater than that of the lowest hole threshold along the Λ axis, and lower than the energies of the lowest electron thresholds along the Δ and Λ axes. Similar to the lowest electron threshold along the Δ axis, this lowest threshold is provided by a normal process in which the final electron states are both near the conduction band minimum at Γ. Also, the next lowest electron threshold has an energy considerably higher than that of the lowest threshold.

The lowest hole threshold along the Σ-S axis is close to the minimum possible energy for any threshold, as is that along the Λ axis, and is even lower than the lowest hole threshold along the Λ axis. This lowest threshold is provided by a normal process, as is the next lowest threshold which has an energy not much greater than that of the lowest threshold, in contrast to the situation along the other two symmetry axes. However, the third lowest hole threshold does have an energy considerably higher than those of the two lowest thresholds.

The errors in the conservation of energy corresponding to the thresholds along the Σ-S axis are generally smaller than those along the Δ axis, in contrast to the situations in the other semiconductors

investigated, but generally greater than those long the Λ axis, in keeping with the situations in the other semiconductors investigated. Most of the electron thresholds have a corresponding error in the conservation of energy less than 0.02eV, with only a few in error by more than this. Only one threshold is in error by more than 0.03eV, but is still less than the maximum expected error, and is due to the inaccuracies in the Γ_{15}-K_1-X_3 conduction band. All but one hole threshold have corresponding errors in the conservation of energy less than 0.02eV, while the one less accurate threshold, in error by 0.036eV, is again due to the inaccuracies in the Γ_{15}-K_1-X_3 conduction band.

The situation concerning the sizes of the matrix element of the coulomb interaction is similar to that in Gallium-Phosphide, in that most of the thresholds have corresponding matrix element sizes which are significant. However, in contrast to the situation in Gallium-Phosphide, the lowest electron threshold along the Δ axis has an insignificant matrix element size, the lowest significant threshold not occurring until a very much higher energy is achieved, the threshold having a value of E_T/E_G=1.715. The lowest hole threshold along the Δ axis does have a significant matrix element size, in keeping with the situation in Gallium-Phosphide.

The lowest electron and lowest hole thresholds along the Λ axis both have corresponding matrix element sizes which are significant, similar to the situation along the Λ axis in Gallium-Phosphide. Indeed, the matrix element size corresponding to the lowest hole threshold is greater than unity, a situation which was first noticed in the results of 3C Silicon-Carbide. Along this axis there is only

one threshold, an electron threshold, which has a corresponding matrix element size which is insignificant.

Again, there are only a few thresholds along the Σ-S axis which can be considered to have insignificant matrix element sizes. However, unlike the situations along the other two symmetry axes, both the lowest electron and the lowest hole thresholds have corresponding matrix element sizes which are insignificant. The second lowest electron threshold along the Σ-S axis is the lowest which also has a significant matrix element size, but it has a very much higher energy. Also, a further two electron thresholds, at even higher energies, are seen to have corresponding matrix element sizes greater than unity, as did the lowest hole threshold along the Λ axis. The lowest significant hole threshold, unlike the lowest significant electron threshold, occurs at an energy not much greater than that of the lowest hole threshold.

The approximate calculations of the matrix element sizes are in general better than those in the other semiconductors investigated. Most of the electron thresholds along the Δ axis which are provided by normal processes have corresponding matrix element sizes which are also approximated accurately by the matrix element sizes calculated from equation 5.2.9. This also applies to the electron thresholds along the Σ-S axis which are provided by normal processes. However, the situation for the electron thresholds along the Λ axis, and for all the hole thresholds are more in keeping with the situations in the other semiconductors investigated, in that very few of the matrix element sizes corresponding to these thresholds are approximated accurately by the equation 5.2.9.

From the many thresholds determined along all three principal symmetry axes, the lowest electron threshold has a value of E_T/E_G=1.083 and is provided by a normal process along the Σ-S axis, while the lowest hole threshold has a value of E_T/E_G=1.026, slightly lower than the lowest electron threshold, and is also provided by a normal process along the Σ-S axis. However, it would be erroneous to use these values in related theories as both of them have corresponding matrix element sizes which are insignificant. It is necessary to consider the matrix element sizes corresponding to the thresholds before attempting to use them in related theories. By doing this, it is seen that the lowest electron threshold which has a significant matrix element size does not occur until a very much higher energy than the overall lowest electron threshold. It has a value of E_T/E_G=1.567, and is provided by a normal process along the Σ-S axis, and also corresponds to an intervalley transition.

The lowest hole threshold which has a significant matrix element size occurs at a comparable energy to that of the overall lowest hole threshold, being only slightly higher with a value of E_T/E_G=1.060. However, this threshold is not provided along the same axis, but along the Λ axis, thus being in contrast to the expected situation which occurred in Silicon, Germanium and Gallium-Phosphide. That is, the lowest electron and hole thresholds which are significant are provided along different symmetry axes, the situation which also occurred in 3C Silicon-Carbide. Even though the lowest significant hole threshold occurs along the Λ axis, the lowest significant hole thresholds along the other two symmetry axes occur at only slightly higher energies, although they do have significantly smaller matrix element sizes.

The values of E_T have also been calculated by Anderson and Crowell [32], but using the band structure of Cohen and Bergstresser [25]. As for Gallium-Phosphide, the band structure of Walter and Cohen which is reproduced and used in the present work, differs slightly from that of Cohen and Bergstresser, and is regarded as being the more accurate. The results presented in this work are thus compared with the results of Anderson and Crowell, and with the exception of the electron thresholds along the Λ axis, very good agreement is obtained within the errors of the respective methods. The lack of agreement in the electron thresholds is almost certainly due to the slight differences in the two band structures investigated.

12.3 Comparison with Results from Approximate Band Structure Models

As expected from the results of the approximate band structure models for the semiconductors investigated previously, there is not complete agreement between the values of E_T/E_G given by the parabolic band approximation and the corresponding values given by the Franz construction. There are however, only minor differences between the corresponding values for all the thresholds except the hole threshold along the Δ axis, for which the value of E_T/E_G given by the parabolic band approximation is substantially higher than that given by the Franz construction. Similar to the situation in Gallium-Phosphide, there are only two threshold positions determined by the Envelope method for which the approximate values may be compared, the lowest electron thresholds along the Δ and the Σ-S axes. Of these two thresholds, only that along the Σ-S axis is approximated accurately, the approximate values corresponding to the threshold along the Δ axis being substantially lower.

Comparison of Tables 12.1 and 12.2 shows that

Table 12.2

Comparison of E_T/E_G values for different band structure models

Details of Thresholds	Parabolic band approximation	Franz construction	Genuine band structure
Γ-Δ-X axis	1.124	1.127	1.266
Electrons	-	1.465	-
Γ-Δ-X axis	1.876	1.607	-
Holes	-	1.873	-
	-	2.091	-
	-	3.103	-
Γ-Λ-L axis	1.073	1.057	-
Electrons	-	1.491	-
Γ-Λ-L axis	1.933	1.929	-
Holes	-	3.884	-
Γ-Σ-K-S-X axis	1.046	1.034	1.083
Electrons	-	1.148	-
	-	1.521	-
Γ-Σ-K-S-X axis	-	1.603	-
Holes	-	1.754	-
	1.981	1.975	-
	-	4.488	-

the two threshold values given by the Envelope method, for which approximate threshold values also exist, correspond to the lowest electron thresholds along the Δ and Σ-S axes. Thus the lowest electron threshold along the Σ-S axis is also given accurately by both the approximate band structure models, while that along the Δ axis is given by approximate values which are substantially lower than the true values.

The Γ_{15}-X_5 valence band along the Δ axis is too flat to permit any ionization process initiated by hot holes to take place at all,

and thus the values of E_T/E_G given by the approximate band structure models ought not to exist. The same situation also applies to the Γ_1-L_1 conduction band and the Γ_{15}-L_3 valence band along the Λ axis, as it does for the corresponding energy bands along the Λ axis in Silicon, 3C Silicon-Carbide and Gallium-Phosphide. The Γ_{15}-K_2-X_5 valence band along the Σ-S axis is also too flat to permit any ionization process to take place at all, and again, the values of E_T/E_G given by the approximate band structure models ought not to exist. The Γ_1-X_1 conduction band along the Δ axis is only able to provide one electron threshold position, and thus only one value of E_T/E_G given by the approximate band structure models ought to exist. This situation also applies to the Γ_1-K_1-X_1 conduction band along the Σ-S axis.

13. RESUMÉ OF RESULTS

13.1 Summary of Work Done

The aim of this thesis has been to investigate the relation between the detailed band structure and impact ionization for a number of semiconductors. This research project was prompted by the lack of knowledge concerning the impact ionization threshold energies for realistic band structures, since at that time the only calculations of impact ionization threshold energies had been made using approximate band structure models. These assumed the conduction band to be parabolic, and most of them also assumed the valence band to be parabolic.

In order to perform calculations for realistic band structures, it was necessary to obtain accurately the details of the band structures investigated. This was done by reproducing the band structure calculations of previous workers, which were made by the Empirical Pseudopotential Method (EPM). In the EPM calculations of 3C SiC, the band structure was reproduced from the calculations of Hemstreet and Fong [28,29], for which a nonlocal, angular-momentum-dependent potential term was added to the local potential term. As a result of a pilot study into the effect of this nonlocal term, an error in the calculations of Hemstreet and Fong was revealed, and it was shown that the nonlocal term had a negligible effect on the band structure, contrary to the results of Hemstreet and Fong.

Once the realistic band structures of the various semiconductors had been calculated, the impact ionization threshold energies were then determined. For the purpose of calculating these threshold energies, a graphical method was developed in this work, referred to as the Envelope Method. The results of these calculations were

presented in Chapters 8 to 12, together with the appropriate values given by two approximate band structure models which previous workers had used in their calculations. These models were; the method developed by Franz [10] as a generalization of the method introduced by Tewordt [9], and the familiar parabolic band, or effective mass, approximation.

Since the impact ionization threshold energy appears as a parameter in the total probability of a transition, it was decided to calculate two other quantities appearing in the total transition probability. One of these was the relative importance of the density of states of the energy bands involved in a transition for hot electron energies just above threshold, for thresholds having comparable energies. However, it was concluded that for Si, this quantity was not important in determining the relative importance of thresholds having comparable energies, and was therefore not calculated for the thresholds of the other semiconductors investigated.

The second quantity appearing in the transition probability was the matrix element of the coulomb interaction between the electron states involved in a transition. This calculation was performed by employing the theory of Beattie and Landsberg [17], in which the states involved in the transition are described by orthonormal, one-electron functions. Furthermore, these one-electron functions were expanded as a finite series of plane-waves, the coefficients of which were calculated by means of the EPM. In the calculations of the sizes of the matrix element, only the term which is referred to as the 'direct term' was evaluated; no evaluation was made of the term which is referred to as the 'exchange term'. The results of these calculations were presented in Chapters 8 to 12, some of which were very

surprising, as a number of threshold positions had corresponding matrix element sizes which were insignificant.

13.2 Accuracy of Band Structures in Calculating the Lowest Threshold Energies

In the chapters dealing with the results, not only were the threshold values given by the Envelope Method for the realistic band structures presented, but also the corresponding values given by the two approximate band structure models. It was seen that while the values given by the two approximate models were, in general, comparable with each other, the comparison with the corresponding values given by the Envelope Method were very poor indeed. Since the lowest threshold of each type (electron and hole) is of greatest importance, they are summarized in Table 13.1, in which the lowest values given by the approximate band structure models are compared with the lowest values given by the Envelope Method.

It is seen that in the majority of situations, the lowest approximate values are not directly comparable to the lowest values given by the Envelope Method, but correspond to different threshold positions. Even for those approximate values which are directly comparable to the lowest values given by the Envelope Method, the threshold energies are in general considerably different, and are only comparable in two cases. However, there are a few other threshold energies given by the approximate band structure models which are of comparable energies to those given by the Envelope Method, although the threshold positions are not directly comparable.

From these results it is clear that, of the 30 lowest threshold positions determined in the various semiconductors, since only two of these are also given accurately by the approximate band structure

Table 13.1

Summary of the comparison between approximate band structure models and genuine bands in the calculation of threshold energies

Material and process	Γ-Δ-X axis		Γ-Λ-L axis		Γ-Σ-K-S-X axis	
	Parabolic band approx.	Franz construction	Parabolic band approx.	Franz construction	Parabolic band approx.	Franz construction
Si Electrons	NH	NH	NL	NL	H	H
Holes	H	H	NL	NL	H	H
Ge Electrons	NH	NH	NH	NH	NE	NE
Holes	H	NH	H	H	NH	NH
SiC Electrons	H	H	NE	NE	H	H
Holes	H	H	NH	NL	H	H
GaP Electrons	NE	NE	NL	NL	NH	NH
Holes	H	H	NE	NL	E	E
GaAs Electrons	L	L	NL	NL	E	E
Holes	NH	NH	NH	NH	NH	NH

All refer to the lowest threshold in each case.

Key to table:
- N: Approximate value corresponds to different threshold
- E: Approximate value comparable to true value
- L: Approximate value lower than true value
- H: Approximate value higher than true value

models, it is dangerous to rely on the values given by approximate band structure models. To obtain reliable values for impact ionization threshold energies it is necessary to use the genuine band structures of the semiconductors being investigated. Also, from the results of the preliminary studies of Si and Ge, and from the comparison with the corresponding improved calculations, it was seen that the threshold values are sensitive to the precise details of the band structure. Hence, the realistic band structures used should be in substantial agreement with experimental data to ensure the greatest possible accuracy.

In the light of this last remark, it should be mentioned that the accuracies of the band structures investigated in the present work vary considerably. In the improved calculations of the threshold positions in Si the band structure was reproduced from the data of Cohen and Bergstresser [25], although it was known that it did not agree with the generally accepted band structure, the energy band gap being too small. This band structure was reproduced to enable the matrix elements of the coulomb interaction to be calculated by the method previously described, which uses the wavefunctions of the pseudopotential method which are readily available. While this was the best pseudopotential band structure calculation available at the time, the matrix elements could not have been calculated as readily if, say, the Stuckel and Euwema [38] band structure had been used.

The band structures of Ge, GaP and GaAs were all reproduced from accurate band structure calculations which were in substantial agreement with available experimental data. However, there is an uncertainty about the band structure calculations of 3C SiC, as not a great amount of experimental data is available. Also, the band structure calcula-

tions of Hemstreet and Fong [28,29], using an extra, nonlocal potential term in the pseudopotential analysis, were found to be in error. Even without the effect of this extra term, there were discrepancies between the band structure calculations of Hemstreet and Fong and those in the present work, for which their data was used. The reliability and accuracy of this band structure is consequently in doubt, as are the accuracies of the resulting threshold energies.

13.3 Relative Significance of the Lowest Thresholds

Even though the accuracies of the Si and SiC band structures investigated are in doubt, the threshold energies calculated from them should still provide some useful information; especially the lowest threshold energies for both electrons and holes. The size of the matrix element of the coulomb interation corresponding to each threshold was seen to be an important factor in determining the relative significance of threshold energies. It was surprising to see that many of the threshold positions determined had corresponding matrix element sizes which were insignificant, including some which were the lowest threshold positions determined. This resulted in the lowest significant threshold position occurring at a higher energy, and in a few instances at a much higher energy.

Had the sizes of the matrix element not been considered, this would have led to some incorrect threshold energies being used in related theories, such as in the calculation of the impact ionization coefficients for electrons and holes. The lowest threshold energy within each semiconductor, together with the lowest threshold energy which has a corresponding significant matrix elment size, and the principal symmetry axis on which each threshold occurs is thus summarized

in Table 13.2.

Both the lowest electron and hole thresholds in Si have corresponding matrix element sizes which are insignificant, but the lowest significant electron threshold occurs at very nearly the same energy, while the lowest significant hole threshold occurs at a much higher energy. Also, the lowest significant hole threshold occurs along a different symmetry axis than does the absolute lowest threshold; the same symmetry axis along which the lowest electron threshold occurs. The situation for the lowest electron threshold in Ge is

Table 13.2

Lowest Impact Ionization Thresholds

Material and Process	Lowest Threshold E_T/E_G	Lowest Significant Threshold E_T/E_G	Axis on which lowest significant threshold occurs
Si. Electron	1.055	1.056	Γ-Δ-X
Hole	1.644 *	1.813	Γ-Δ-X
Ge. Electron	1.023	1.031	Γ-Λ-L
Hole	1.034	1.034	Γ-Λ-L
SiC. Electron	1.131	1.131	Γ-Δ-X
Hole	1.161	1.161	Γ-Σ-K-S-X
GaP. Electron	1.118	1.118	Γ-Δ-X
Hole	1.009 *	1.012	Γ-Δ-X
GaAs. Electron	1.083	1.567	Γ-Σ-K-S-X
Hole	1.026 *	1.060	Γ-Λ-L

* Occurs along a different symmetry axis (the Γ-Σ-K-S-X axis).

similar to that in Si; namely the lowest significant threshold occurring at very nearly the same energy as the absolute lowest threshold. The lowest hole threshold however is also the lowest significant hole threshold, and also occurs along the same symmetry axis as does the lowest eleetron threshold.

All the threshold positions in 3C SiC were seen to have corresponding matrix element sizes which were significant, and consequently both the absolute lowest electron and hole thresholds are also the lowest significant thresholds. However, these thresholds occur along different symmetry axes, unlike the situation in Si and Ge. The lowest significant electron threshold in GaP is also the absolute lowest threshold, while the lowest significant hole threshold has an energy very nearly equal to that of the absolute lowest threshold, but occurs along a different symmetry axis. Thus the lowest significant electron and hole thresholds are provided along the same symmetry axis, a similar situation to that in Si.

In GaAs, both the lowest electron and hole thresholds are insignificant, as they were in Si, although the lowest significant electron threshold occurs at a very much higher energy. The lowest significant hole threshold however, has an energy very nearly equal to that of the absolute lowest threshold, but occurs along a different symmetry axis. Thus the lowest significant electron and hole thresholds occur along different symmetry axes, as was the situation in 3C SiC.

It is indicated from these results that it would be unwise to use any impact ionization threshold energies in related theories, such as in the calculation of the impact ionization coefficients, without first calculating the corresponding size of the matrix element of the coulomb interaction. Indeed, in a few cases it would have been erroneous if the

relative significance of the threshold positions had not been considered, and would have led to totally misleading results.

The method used to calculate the sizes of the matrix element of the coulomb interaction assumes that the wavefunctions of the electron states involved in a transition can be expanded as a finite series of plane-waves. It is recognised that this is an approximation, and that more accurate and reliable calculations are possible and should perhaps be used. However, it is hoped that the results presented in this thesis prove useful, and serve as a basis for further calculations of the matrix elements by using more reliable and exact models.

13.4 Application to Impact Ionization Coefficients

Having calculated the impact ionization threshold energies of hot electrons and holes in a number of semiconductors, and also determined the lowest significant threshold energies, they can now be used to determine which of the two processes will be the more important; impact ionization by electrons or by holes. This can be done by calculating the total probability of transition, in which the threshold energy appears as the lower limit of the integration over energy, or by calculating the impact ionization coefficient, which has a negative exponential dependence on the threshold energy. The impact ionization coefficients for electrons, α, and for holes, β, will also depend upon the total transition probability, and thus an idea of the relative magnitudes of α and β, within a particular semiconductor, can be obtained by considering the lowest threshold energies and the corresponding matrix element sizes.

Looking at Table 13.2, it is seen that the lowest significant electron threshold for Si has an energy considerably lower than that of the lowest significant hole threshold, and thus it is expected that

$\alpha > \beta$, which is in agreement with experimental evidence (see for example, Miller [7]). In Ge the two thresholds are of almost equal energies, and so it may be expected that $\alpha \approx \beta$ in this case. However, by considering also the corresponding matrix element sizes of these two thresholds, it is expected that $\alpha \lesssim \beta$, as is indicated by Miller [6].

The situation in Ge also applies to 3C SiC, in that the lowest significant electron and hole thresholds have nearly equal energies. However, the corresponding matrix element of the hole threshold is slightly larger than that of the electron threshold, and thus it is expected that $\alpha \lesssim \beta$. The thresholds in GaP are also of comparable energies, although the electron threshold is the slightly higher of the two, but the matrix element corresponding to the electron threshold is also larger than that corresponding to the hole threshold, and consequently it is expected that $\alpha \approx \beta$. In Ga As, the lowest significant electron threshold energy is considerably larger than that of the lowest significant hole threshold, the opposite situation to that in Si, and thus it is expected that $\alpha < \beta$ as is reported experimentally by Stillman et. al [62].

These impact ionization coefficients depend upon the impact ionization threshold energies, which in turn depend upon the direction in which the electric field is set up in the semiconductor. For example, if the electric field is set up in the direction of one particular symmetry axis, then the electrons and holes will gain energy from the field by moving in the direction of the field, along that particular symmetry axis. However, it was seen from the results presented in Chapters 8 to 12 that the lowest significant threshold energies along each of the principal symmetry axes differ considerably.

Under the assumptions made by Shockley [2] for small electric fields, the electron and hole concentration peaks in the direction of the electric field, and that only those electrons avoiding collisions with the lattice will gain sufficient energy to partake in impact ionization. Thus in this situation, the direction in which the electric field is set up in the semiconductor will have a significant effect upon the impact ionization coefficients. The same situation does not necessarily apply under the assumptions made by Wolff [1] for large electric fields, where the electrons may be scattered and may assume a distribution which is almost spherically symmetric.

14. Conclusions, Recommendations and some Ideas for Extending the Work

By comparing the impact ionization threshold positions calculated accurately by the Envelope Method for the realistic band structure with the corresponding values calculated from two approximate band structure models, for a number of semiconductors, it has been shown that the approximate band structure models give unreliable threshold values. Consequently, it is necessary to calculate impact ionization threshold positions accurately, by making full use of the details of the band structure, and that it is dangerous to rely on values given by approximate band structure models. Also, the threshold positions are sensitive to the precise details of the band structure, and thus the band structures used should be in substantial agreement with experimental data.

Having determined the impact ionization threshold positions, it was then shown that it is necessary to calculate the sizes of the matrix element of the coulomb interaction associated with each threshold position to establish which threshold positions are significant. It was surprising to see that several impact ionization threshold positions had corresponding matrix element sizes which were insignificant, including some which were the lowest threshold for the particular band structure. Thus, having determined the significant impact ionization threshold positions, these may then be used in the related theories, such as calculating the total transition probabilities or the impact ionization coefficients of electrons and holes.

The realistic band structures used in the investigation of impact ionization threshold positions were calculated by the Empirical Pseudopotential Method. In the calculation of the band structure of 3C SiC by Hemstreet and Fong [28,29], a nonlocal, angular-momentum-

dependent potential was added to the local, spherically symmetric potential in the pseudopotential analysis. However, in the present work an error in the calculations of Hemstreet and Fong was revealed, and that the nonlocal potential had a negligible effect on the band structure, contrary to the results of Hemstreet and Fong.

The series expansion in plane-waves of the wavefunctions used in calculating the sizes of the matrix element of the coulomb interaction could have been taken to be two different lengths. The first was a short, or basis, expansion involving only the plane-waves which were treated exactly in the perturbation theory employed in calculating the band structure, while the second was a full expansion which also involved all the plane-waves included through the perturbation theory. In a pilot study it was shown that the improved accuracy obtained by using the full expansion rather than the basis expansion was very small, and that the extra computer time used was vast and would have been prohibitive for a large number of calculations.

The calculation of the sizes of the matrix element of the coulomb interaction involved a quadruple sum over reciprocal lattice vectors. However, under the assumptions made by Beattie and Landsberg [17], this quadruple sum could be approximated by the product of two double sums over reciprocal lattice vectors. The matrix element sizes were thus calculated by both the quadruple sum and the two double sums and presented in the results of Chapters 8 to 12. It was shown that the approximate matrix element sizes were not in good agreement with the true matrix element sizes, and that it would be dangerous to rely on matrix element sizes calculated by using the approximating assumptions of Beattie and Landsberg.

In trying to determine the relative importance of impact ionization

threshold positions with almost equal energies, the rate of increase in the number of electron states available for impact ionization for hot electron energies just above threshold was investigated. However, it was concluded that for Si this factor proved to be unimportant, and thus was not considered in the investigations of the other semiconductors.

It was mentioned in the introduction to Chapter 3 that the Envelope Method developed in the present work has a few disadvantages compared with the method developed by Anderson and Crowell [32]. One of these disadvantages was that the final states of both electrons were restricted to lie in the lowest lying conduction band, or the final states of both holes in the highest lying valence band, whereas this was not a restriction of the method developed by Anderson and Crowell. However, the method developed here could be extended to the generality of the method of Anderson and Crowell, in that the final states of both electrons, or holes, need not lie in the same energy band. In doing this more impact ionization threshold positions would be determined, some of which may have energies not much greater than the lowest threshold energy, and may also prove to be more significant when the corresponding matrix element sizes are calculated.

While the impact ionization threshold positions of a number of semiconductors have been calculated in the present work, there are a great many more semiconductors to which the methods employed here can be applied. Thus the work presented here can be extended to calculate accurate impact ionization threshold positions in other semiconductors for which accurate, realistic band structures are available, and these can then replace the approximate values now being used.

The impact ionization threshold energies and corresponding sizes of the matrix element of the coulomb interaction presented in this work can be used to evaluate the total transition probability associated with each significant, and important, threshold position. Thus the relative importance of impact ionization, and also Auger Recombination, to processes involving phonons or traps can be determined. The impact ionization coefficients can also be calculated for both electrons and holes, which can then be used, for example, to calculate the avalanche breakdown conditions in a reverse biased p-n junction.

REFERENCES

[1] P.A. Wolff, Phys. Rev. 95, 1415 (1954).

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

[3] G.A. Baraff, Phys. Rev. 128, 2507, (1962).

[4] L.V. Keldysh, Sov. Phys. JETP. 21, 1135, (1965).

[5] C.A. Lee, R.A. Logan, R.L. Batdorf, J.J. Kleimack and W. Wiegmann, Phys. Rev. 134, A761 (1964).

[6] S.L. Miller, Phys. Rev. 99, 1234 (1955).

[7] S.L. Miller, Phys. Rev. 105, 1246 (1957).

[8] Y. Okuto and C.R. Crowell, Phys. Rev. B6, 3076 (1972).

[9] L. Tewordt, Z. Phys. 138, 499 (1954).

[10] W. Franz, Handb. d. Phys. 17, 190 [Berlin: Springer-Verlag] (1956).

[11] D.L. Dexter, Proc. Int. Conf. on Semiconductor Physics, Prague [New York: Academic Press] p122-6 (1960).

[12] E. Antoncik, Czech. J. Phys. 7, 674 (1957).

[13] E. Antoncik, Czech. J. Phys. 8, 492 (1958).

[14] A.R. Beattie, J. Phys. Chem. Solids, 23, 1049 (1962)

[15] R.J. Hodgkinson, Proc. Phys. Soc. 82, 58 (1963).

[16] R.J. Hodgkinson, Proc. Phys. Soc. 82, 1010 (1963).

[17] A.R. Beattie and P.T. Landsberg, Proc. Roy. Soc. A249, 16 (1958).

[18] J.R. Hauser, J. Appl. Phys. 37, 507 (1966).

[19] L. Huldt, Phys. Stat. Sol. (a) 8, 173 (1971).

[20] D.L. Camphausen and C.J. Hearn, Phys. Stat. Sol. (b) 50, K139 (1972).

[21] E.O. Kane, J. Phys. Chem. Solids 1, 249 (1957).

[22] S. Mahadevan, S.M. Hardas and G. Suryan, Phys. Stat. Sol. (a) 8, 335 (1971).

[23] G.C. Fletcher, The Electron Band Theory of Solids. p67ff, [London: North-Holland] (1971).

[24] D. Brust, Phys. Rev. 134, A1337 (1964).

[25] M.L. Cohen and T.K. Bergstresser, Phys. Rev. 141, 789 (1966).

[26] J.P. Walter and M.L. Cohen, Phys. Rev. 183, 763 (1969).

[27] H.G. Junginger and W. van Haeringen, Phys. Stat. Sol. 37, 709 (1970).

[28] L.A. Hemstreet, Jr. and C.Y. Fong, Solid State Commun. 9, 643 (1971).

[29] L.A. Hemstreet, Jr. and C.Y. Fong, Phys. Rev. B6, 1464 (1972).

[30] L.A. Hemstreet, Jr., C.Y. Fong and M.L. Cohen, Phys. Rev. B2, 2054 (1970).

[31] R.A. Ballinger, K.G. Major and J.R. Mallinson, J. Phys. C: Solid State Phys. 6, 2573 (1973).

[32] C.L. Anderson and C.R. Crowell, Phys. Rev. B5, 2267 (1972).

[33] P.T. Landsberg, Solid State Theory: Methods and Applications. p266ff, [London: Wiley-Interscience] (1969).

[34] A.R. Beattie and P.T. Landsberg, Proc. Roy. Soc. A258, 486 (1960).

[35] P.T. Landsberg, Lectures in Theoretical Physics. 8A, p313ff, [Boulder: Colorado Press] (1966).

[36] A.R. Beattie and G. Smith, Phys. Stat. Sol. 19, 577 (1967).

[37] S. Ahmad and W.S. Khokley, J. Phys. Chem. Solids 28, 2499 (1967).

[38] D.J. Stukel and R.N. Euwema, Phys. Rev. B1, 1635 (1970).

[39] D.J. Stukel, Phys. Rev. B3, 3347 (1971).

[40] J. Korringa, Physica 13, 392 (1947).

[41] W. Kohn and N. Rostoker, Phys. Rev. 94, 1111 (1954).

[42] J.C. Slater, Phys. Rev. 51, 846 (1937).

[43] B. Segall, Phys. Rev. 105, 108 (1957).

[44] C. Herring, Phys. Rev. 57, 1169 (1940).

[45] J.C. Phillips, Phys. Rev. 112, 685 (1958).

[46] J.C. Phillips and L. Kleinman, Phys. Rev. 116, 287 (1959).

[47] L. Kleinman and J.C. Phillips, Phys. Rev. 118, 1153 (1960).

[48] M.H. Cohen and V. Heine, Phys. Rev. 122, 1821 (1961).

[49] B.J. Austin, V. Heine and L.J. Sham, Phys. Rev. 127, 276 (1962).

[50] D. Brust, J.C. Phillips and F. Bassani, Phys. Rev. Lett. 9, 94 (1962).

[51] D. Brust, M.L. Cohen and J.C. Phillips, Phys. Rev. Lett. 9, 389 (1962).

[52] F. Bassani and D. Brust, Phys. Rev. 131, 1524 (1963).

[53] M.J.G. Lee and L.M. Falicov, Proc. Roy. Soc. A 304, 319 (1968).

[54] P.O. Löwdin, J. Chem. Phys. 19, 1396 (1951).

[55] J.H. Wilkinson, The Algebraic Eigenvalue Problem. pp290-299, pp515-521, [Oxford: Clarendon Press] (1965).

[56] M.A. Efroymson, Mathematical Methods for Digital Computers. p191ff, [London: Wiley] (1960) [edited by A. Ralson and H.S. Wilf].

[57] G.B. Wetherill, Elementary Statistical Methods. pp150-152, [London: Chapman and Hall] (1972).

[58] I.C.L., System 4 Statistics Scheme Manual. [I.C.L.]

[59] F.B. Hildebrand, Introduction to Numerical Analysis. Second Edition. p414ff, London: McGraw-Hill (1974).

[60] D. Long, Energy Bands in Semiconductors, p87, [New York: Wiley] (1968).

[61] A.B. Kunz, Phys. Rev. Lett. 27, 567 (1971).

[62] G.E. Stillman, C.M. Wolfe, J.A. Rossi and A.G. Foyt, Appl. Phys. Lett. 24, 471 (1974).