THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/73492

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Theoretical studies of wavepacket propagation in semiconductor quantum well structures

Stephen Collins

University of Warwick

PhD

1986

Abstract

In this thesis a heuristic expression for the current through a GaAs/GaAlAs heterostructure is derived. This expression is shown to give rise to agreement between experiment and theory.

The expression itself is derived within the effective mass formalism, which is discussed to show that its use will not generate large errors. This conclusion is contrary to previous work which will be shown to be in error due to misunderstandings concerning effective mass theory.

To justify the approach used to obtain the tunnel current expression the behaviour of a wavepacket incident upon a square potential barrier is studied. The study shows that the wavepacket traverses the potential sufficiently rapidly to allow scattering to be neglected, and that the total transmission probability can be calculated from the solution of the time independent Schrodinger equation.

The current expression is reduced to a one dimensional integral by assuming parabolic conduction bands, position independent mass and a thermalised electron distribution. The resulting expression is different from the usual Tsu-Esaki formula, a difference which can be seen to arise because the Tsu-Esaki formula does not account for the different velocities on each side of the barrier.

The final stage, before any comparison is made to experimental results, is to show that the numerical technique of Vigneron and Lambin is more accurate than the WKB technique.

A comparison of experimental results and the results of the numerical integration of the current density expression shows that they can only be reconciled if a resistance or diode is assumed in series with the tunnel barrier. This fitting parameter is then shown to be sufficient for good fits to be obtained between experiment and theory for the first time.

CONTENTS

List of Illustrations

Acknowledgements

There are many people who have helped me to reach the stage at which I am able to submit this thesis. Some of them I would like to thank here:-

My wife and parents for the sacrifices which allowed me to continue my education.

Dr J R Barker for inspiration and guidance.

My colleagues at Warwick, and more recently RSRE, for their co-operation and friendship. Special thanks to Dr D Lowe for his constant willingness to discuss my problems and for reading this thesis.

Professors P W McMillan and A J Forty for use of the facilities of the Physics Department and the SERC for financial support.

Finally Mrs C Edge for her cheerful dedication to the typing of this thesis.

Declaration

This thesis contains an account of my own independent research work performed in the Department of Physics at the University of Warwick between October 1982 and October 1985 under the general supervision of Dr John R Barker.

Some of the work has previously been published as follows:-

1. "Theory of Transient Quantum Transport in Heterostructures"
   J R Barker, S Collins, D Lowe and S Murray.
   Proc 17th Int Conf Phys Semicon, San Francisco (1984)
   Ed J D Chadi and W A Harrison
   Springer-Verlag, New York (1985)

2. "On the accuracy of the effective mass approximation for scattering at heterojunctions"
   S Collins, D Lowe and J R Barker
   J Phys C 18 L637 (1985)

3. "Quantum Theory of hot electron tunnelling in microstructures"
   J R Barker
   An invited paper at 4th Int Conf on Hot Electrons in Semicon
   Physica 134B 22 (1985)

TO ALISON

INTRODUCTION

Motivation

The advances in the growth technologies for compound semiconductors in the past decade have made possible the controlled, reproducible fabrication of structures with features on a nanometre scale (1). This has motivated investigations into structures which exploit Quantum Mechanical effects, with a general aim of manufacturing devices controlled by Quantum Mechanics.

The structures which are of interest in this thesis were first suggested by Tsu and Esaki (2). They showed that the transmission resonances, predicted as early as 1952 (3), could be exploited to give a resonant current. However, the results at that time were not sufficiently encouraging to sustain interest, with no resonance being observed at room temperature.

The advances in fabrication technology since this early work have led to a renewed interest based on the expectation that the improved sample fabrication now possible would lead to room temperature resonance. The first observation of room temperature resonance was reported by Sollner et al (4). These resonances, however, could only be detected in the first derivative of the current. It was nearly two more years before Shewchuk et al (5) observed a resonance at room temperature.

At present, a stage has been reached at which a resonant current has been observed at room temperature, but, several unanswered questions remain: Why is the peak to valley ratio so diasppointingly small? Why don't the

current voltage characteristics show the expected symmetry with respect to the direction of the bias? Why did the data obtained by Shewchuk et al contain a discontinuity?

The aim of this thesis is to provide a theoretical model to address these questions and to provide a method of predicting the characteristics of other systems which may be proposed in the future. To be widely useful restrictions must be placed on the type of model, it must be easy to implement, flexible and quick.

A simple current voltage expression is therefore required and it must be shown that it can at least give a good estimate of the magnitude and shape of the observed characteristic.

Outline of the Thesis

The expression which is to be used will not be derived until Chapter 4. Before then several points implicitly assumed in the derivation of the expression will be justified.

In the first chapter a simple one dimensional Schrodinger equation (S.E.) for the wave function in a Gallium Aluminium Arsenide (GaAlAs) system, with an Aluminium mole fraction which varies in one dimension will be given. The equation arises from the work of Bastard (6) which will be discussed to show that the assumptions made can be justified and to obtain criteria for the applicability of the equation.

The available values for the parameters needed to model both electrons and holes using the one dimensional equation will be discussed. It will be

shown that there is considerable doubt concerning the accuracy of some of the values, doubts which mean that the results of methods not requiring knowledge of the hole masses will be favoured when deciding the best experimentally determined value of the conduction band discontinuity. The last section of the chapter will consider previous criticism of the simple equation to show that misunderstandings caused overestimates of the errors involved in using the equation.

The next two chapters are concerned with the results of the one dimensional time dependent equation which arises from the simple model. In chapter two, it is shown that the solution to the time independent equation arises naturally in the solution of the time dependent equation. This justifies the later use in the current expression of the results of the time independent equation, even though the current is carried by electrons whose behaviour can only be predicted using the time dependent equation.

In the next chapter, Chapter 3, the time required by a wavepacket to traverse a potential barrier, the traversal time, will be considered. It will be shown that, if the wavepacket has a sufficiently well defined momentum, the traversal time is related to the derivative of the phase difference between the incident and transmitted waves with respect to the wavevector. The main point of this chapter will be to demonstrate that the tunnelling of an electron through a potential barrier is a rapid process, typically taking of the order of ten femtoseconds. This is faster than the mean free time which is of the order of a picosecond so that the probability of an electron being scattered during tunnelling is assumed to be negligible, allowing scattering to be neglected in the derivation of the tunnelling current.

At the beginning of Chapter 4 the current expression itself is derived. The expression shows that to be able to calculate the current across a general potential a method of calculating the transmission coefficient of a general potential is required. The remainder of the chapter is devoted to demonstrating that a numerical method is much more accurate than the popular WKB method, and can be extended to calculate traversal time and the bandstructure of a superlattice.

In the final chapter the current expression itself is used. Initially the integrand is approximated to demonstrate that the expression can be used to qualitatively explain the observations of Collins et al (7), and to predict the behaviour of systems with two layers of GaAlAs separated by one layer of GaAs. The remainder of the chapter is concerned with the numerical evaluation of the current expression for systems which have either been fabricated or proposed for fabrication. The model developed will be shown to fit the observations to an accuracy which is sufficient for the model to be useful in predicting the behaviour of proposed devices. The lack of agreement for one system can be explained as a failure of one of the assumptions which is included in the model to simplify the integration over the electron distribution. This failure does not imply that the model should not be used but indicates that care is needed in applying the model to ensure that the conditions needed for the model to be applicable do occur.

The overall conclusion of the thesis is that the current density expression developed in section 4.2 is based on a justifiable heuristic approach. The expression itself is accurate enough to be used in assessing and understanding devices if the electrons in the contacts are in thermal equilibrium.

BACKGROUND, A SIMPLE MODEL FOR A HETEROSTRUCTURE

1.1 Introduction

The aim of this chapter is to outline a simple model obtained by Bastard which allows a semiconductor with position dependent composition to be modelled using an equivalent electrostatic potential. This considerable simplification is vital in obtaining a simple picture of resonant tunnelling and a method for calculating the current.

The chapter itself is organised into three distinct sections. In the first section Bastard's work is outlined to show that the assumptions used to extract a simple model equation from the equation representing the bandstructure are justifiable.

The second section is devoted to a discussion of the properties of GaAs, AlAs and GaAlAs which enter the model as parameters. This discussion will outline the reasons behind the values for the parameters which will be used later and to show that the materials are not sufficiently well characterised to allow experimental verification of the simple model.

With experimental verification excluded, two theoretical tests of the accuracy of the model are considered. The authors of the original work are critical of the simple model, however it will be shown that the criticism arises from misunderstandings concerning the model. The overall conclusion is that an accuracy of better than ten percent is expected.

The chapter concludes with a short summary section.

1.2 Theoretical model for a heterojunction

In this section Bastard's work will be outlined to justify the simple representation of heterostructures used later in the thesis. When considering electron transport in semiconductor structure, one of the fundamental properties is the distribution of allowed electron states, the bandstructure. The most accurate available method of obtaining a bandstructure is to represent the ions and their associated bound electrons by a pseudopotential and expand the wave function of the remaining electrons using an appropriate basis set. Since the unknown electron distribution gives rise to a contribution to the potential, the wave function must be iterated until a self-consistent ground state is obtained.

This procedure is not ideal. The large amount of computer facilities required and its sophistication restrict its present use to specialist groups working on small systems. It is also difficult to gain any insight into the processes governing the formation of the bandstructures making trends difficult to identify and requiring a study of each individual system of interest.

The restrictions of the pseudopotential method mean that although it is the most fundamental available technique its usefulness is limited to studies of individual systems. These studies indicate the approximations which can be used to obtain a method simple enough to be used to study many systems.

Such a calculation has been performed by Pickett et al (8), on a (1,1,0)

AlAs-GaAs interface. The most significant result for this thesis is that the self-consistent pseudopotential for the interface shows only small differences from the bulk pseudopotentials used as the initial condition. These small differences are localised within ~10Å of the As layer which formed the interface.

The fact that this calculation showed that in GaAs-AlAs heterojunctions, the heterojunction is a region of sharp transition between apparently bulk materials has led to several models in which a heterojunction is represented by juxta positioning bulk bandstructures (6, 9, 10, 11,).

In the first of these, Bastard used the two band Kane representation of a bulk bandstructure (12) to obtain a simple model for heterostructures. In this representation the conduction and valence band states have the symmetry of the s and p atomic orbitals respectively. This led Bastard to denote the conduction (valence) band states as S (P) and the conduction band minimum (valence band maximum) as V_s (V_p).

The wave function in a material, A, was expanded as a sum over all the bands, of the Bloch cell periodic part of the wave function at the zone centre u_j^A , multiplied by a slowly varying envelope function F_j^A ,

$$\psi^A = \sum_j F_j^A \; u_j^A$$

If this wave function represents an electron near to the conduction band minimum only two of the F's are significant, corresponding to the two spin

states representing the conduction band minimum. These two functions, labelled F_1 and F_2 by Bastard, are coupled by eigenvalue equations. Bastard used several assumptions to obtain a simple set of coupled equations for F_1 and F_2 and then decouple them to obtain a simple S.E. to model a heterostructure. These assumptions are:

(i) The momentum matrix element, π , is equal in both materials

(ii) The spatial localisation of any potential perturbation is small compared to the scale of variation of the envelope

(iii) The only affect of the potential perturbation at the interface is to shift the bandstructures relative to each other

(iv) The transverse wavelength is greater than the thickness of all the layers.

Assumption (i) is justified by the recent work of Merian and Bhattecharjee (13) which shows that although π is composition dependent the total variation is only a small percentage of the mean value.

The work of Pickett et al (8) showed that the potential perturbation is localised to a few atomic planes. The perturbation is therefore on the same spatial scale as the cell periodic component u_j , which by definition, is a much shorter scale than any variation in F. Assumption (ii) is therefore justified.

Pickett's work can also be used in an attempt to justify assumption (iii), since the small potential perturbation calculated by Pickett is unlikely

to mix states from different bands.

The requirement for the last assumption arises from a need to decouple the continuity conditions which must be imposed upon the functions. The coupled continuity conditions are the continuity of

$$\frac{1}{E_A + E - V_P} \left[-\sqrt{2}\, \frac{dF_1}{dz} - ik_+ F_2 \right]$$

and

$$\frac{1}{E_A + E - V_P} \left[-\sqrt{2}\, \frac{dF_2}{dz} + ik_- F_1 \right]$$

where E_A is the bandgap of material A and

$$k_\pm = (k_x \pm ik_y)/\sqrt{2}$$

These equations are decoupled if the second term in each bracket can be neglected. Since the derivative can be approximated by the envelope function divided by the layer thickness, the second terms are negligible if the electrons wavelength is much greater than the layer thickness. That is when assumption (iv) is valid. The approximate nature of the requirement for assumption (iv) means that it is not a strict condition. However, it does indicate that care is needed at high temperatures for thin barriers, for example a 50 Å barrier requires that the transverse energy E_T is such that

$$E_T < 22.6 \text{ meV}$$

which cast doubt on the use of the model at room temperature.

The last assumption decouples the equations produced using the other three assumptions, so that if all four are valid, the equation for one of the

two envelope functions is (6)

$$\left[V_S(z) + \frac{\pi^2}{3} P_- \frac{1}{E_A+E-V_P(z)} P_+ + \frac{\pi^2}{3} P_+ \frac{1}{E_A+E-V_P(z)} P_- + \right.$$

$$\left. \frac{2\pi^2}{3} P_z \frac{1}{E_A+E-V_P(z)} P_z \right] \quad F = EF \qquad (1.2.1)$$

where $P_\pm = (P_x \pm iP_y)//2$. All the variables in equation (1.2.1) are functions of z alone, therefore F is separable. The translational symmetry of the Hamiltonian perpendicular to z means that the envelope function must take the form

$$F(x,y,z) = f(z) \exp (i(k_x x + k_y y))$$

Defining the constant of motion

$$E_t = \frac{\hbar^2(k_x^2 + k_y^2)}{2m_A}$$

and writing the total energy as the sum of two components

$$E = E_t + E_z$$

the equation for f(z) is

$$\left(\frac{-\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m(z)} \frac{\partial}{\partial z} + V_s(z) - E_t \left(\frac{1-M_A}{m(z)} \right) - E_z \right) f(z) = 0 \qquad (1.2.2)$$

where the mass m(z) is defined as

$$m(z) = \frac{3}{4\pi^2} (E_A + E - V_P(z)) \qquad (1.2.3)$$

A further simplification is obtained if the energy is assumed to be in the small range such that for all z

$$E_A + E - V_p(z) \approx E_A + V_s(z) - V_p(z)$$

so that m(z) can be replaced by the conduction band minimum mass. This approximation is only valid for electrons near the conduction band minimum for all compositions present, which restricts both the energy and composition range for which the simplified equation with an energy independent mass is valid.

Despite the limitations imposed by this further simplification, the simplicity achieved means that it is the simplified model which will be used throughout the remainder of the thesis so that in further discussion

$$m(z) = \frac{3}{4\pi^2}\,(E_A + V_s(z) - V_p(z)) \qquad (1.2.4)$$

The model has so far excluded any external potential such as would arise from an electrostatic field. To extend the model to be useful in GaAs/GaAlAs devices these must be included. Altarelli (70) has shown that if an applied potential is included in the k.p. model the potential is added to the diagonal elements of the Hamiltonian matrix so that an applied voltage V(z) can be simply included in Bastards model to give

$$\left[\frac{-\hbar^2}{2}\,\frac{\partial}{\partial z}\,\frac{1}{m(z)}\,\frac{\partial}{\partial z} + V_s(z) + V(z) - E_t\left(\frac{1-m_A}{m(z)}\right) - E_z\right] f(z) = 0$$

This shows that the conduction band edge can be simply interpreted as being equivalent to an electrostatic potential.

A model has therefore been obtained in which the envelope function obeys equation (1.2.2) with the mass m(z) defined by equation (1.2.4).

The position dependence of both $V_S(z)$ and m(z) in equation (1.2.2) shows that a heterojunction can be represented by a step potential, whilst the Kronig Penney model can be used to represent a superlattice. These are two of a class of problems in which the behaviour of the potential changes discontinuously. In these problems it is usual to solve the wave functions on each side of the discontinuity and then match the wavefunctions using continuity conditions. The exact nature of the continuity conditions is important in section 1.4 and so they will be discussed here.

The continuity conditions obtained by Bastard, after decoupling, require the continuity of

$$f \ , \quad \frac{1}{m(z)} \ \frac{\partial f}{\partial z}$$

where m(z) is defined by (1.2.3) but can be approximated by (1.2.4). These conditions arise from the Hamiltonian directly and must therefore be the correct continuity conditions. A conclusion supported by Morrow and Brownstein (69).

1.3 Properties of GaAs, AlAs and GaAlAs

In this section the known properties of GaAs, AlAs and GaAlAs required for the two band model obtained in section 1.2 will be discussed. This

will show that attempts at experimental verification of (1.2.2) are premature, and provide the best estimates of the required parameters for use later.

All three materials form a Zinc Blende structure, with one sublattice in GaAlAs containing both Ga and Al ions. This crystal structure leads to a first Brillouin zone with several points of high symmetry amongst which $\Gamma(0,0,0)$, $X(1,0,0)$ and $L(\frac{1}{2},\frac{1}{2},\frac{1}{2})$, representing local conduction band minima, are important.

To begin, consider the best characterised of the three materials GaAs, a direct bandgap material with both the conduction band minimum and valence band maximum occurring at the Γ point.

Although GaAs is well characterised, there are still uncertainties about several of the accepted values, especially the hole masses. These uncertainties arise because the valence band maximum is formed by two degenerate anisotropic bands, denoted as the heavy hole and light hole bands. The degeneracy means that experiments yield signals from two bands, whilst the anisotropy can lead to a wide variation in the experimental values obtained for the same parameter. The evidence agrees that the [1,1,1] direction is the heaviest in both bands, however the degree of anisotropy differs from experimental values of 25% or 50% (14) to a theoretical value of approximately 100% (15). These difficulties mean that the values listed in Table 1.1 for the hole masses are concensus isotropic values for anisotropic parameters. A list of parameters for AlAs is more difficult to compile. AlAs is an indirect material with the conduction band minimum at the X point lying below that at the Γ point. This makes the bandgap of interest, the direct bandgap,

TABLE 1.1

Property	GaAs	AlAs	$Ga_{1-x}Al_xAs$
E_g^Γ	1.519[b]	3.11[d]	1.519 + 1.247 x (x<0.45) 1.519 + 1.247 x + 1.147 $(x-0.45)^2$ (x>0.45)
E_g^L	1.815[b]	2.35[c]	1.815 + 0.535 x
E_g^X	1.981[b]	2.17[a]	1.981 + 0.189 x
m_l^Γ	0.067[b]	0.15[a]	0.067 (1 + 1.24 x)
m_{lh}	0.082[b]	0.15[a]	0.082 (1 + 0.83 x)
m_{hh}	0.51[b]	0.76[a]	0.51 (1 + 0.49 x)

(a) after S Adachi, J Appl Phys 58 R1 (1985)

(b) after J S Blakemore, J Appl Phys 53 R123 (1982)

(c) after H C Casey Jr and M B Panish, Heterostructure Lasers, Part A, Academic Press (1978)

(d) after B Monemar, Phys Rev B 8 57 11 (1973)
(lowered slightly to fit in with the GaAlAs formula)

difficult to measure.

Several attempts have been made to measure the direct bandgap of AlAs. The only low temperature measurement seems to be due to Monemar et al (16) and gave a value of 3.13 eV at 4K. All the other measurements were at room temperature and gave values of 3.03 eV (16), 2.95 eV (17) and 3.14 eV (18). The high value given by Yim (18) can be attributed to the technique used. From an absorption spectra, the bandgap was calculated by extrapolating the linear part of the absorption against photon energy curve to zero absorption. This was done without first subtracting off the indirect contribution, a failure which can lead to an overestimate.

The conclusion is therefore that the bandgap at absolute zero is approximately 3.13 eV, with a decrease to approximately 3 eV at room temperature. The value given in Table 1.1 is slightly less than 3.13 and is chosen to fit in with the expression given for the bandgap of GaAlAs.

The electron mass required for the model is the Γ valley mass, a parameter whose measurement is made more difficult by the indirect nature of the material. The only measurement, by Monemar et al (16), gave an estimated value, from an examination of the exitonic spectrum of AlAs of

$$m = 0.113 \times 9.109 \times 10^{-31} \text{ kg}$$

The lack of further experimental data means that consideration should be given to theoretical values.

A self-consistent orthogonalised plane wave calculation performed by

Stukel and Euwema (19), with only the lattice parameter as input, gave

$$m = 0.105 \times 9.109 \times 10^{-31} \text{ kg}$$

which is in general agreement with the value obtained from a k.p. calculation by Braunstein and Kane (20)

$$m = 0.110 \times 9.109 \times 10^{-31} \text{ kg}$$

The value given by Braunstein and Kane for the conduction band mass in GaAs was 15% higher than the currently accepted value. This can be used to give an estimated error for the AlAs value of

$$m = (0.110 \pm 0.017) \times 9.109 \times 10^{-31} \text{ kg} \qquad (1.3.1)$$

with the expectation that the true value is at the lower end of the range. Although the value tabulated is the widely accepted value it would appear that this is an overestimate and a better value would be (1.3.1), which would make the light hole mass greater than the conduction band mass, as in GaAs.

It appears that the hole masses in AlAs have never been measured. The tabulated values are the accepted values which appear to originate in the work of Lawaetz (80). In this early paper the then accepted values for Si and Ge parameters were used to estimate the properties of compound semiconductors. It should be stressed that the values are estimates, which Lawaetz hints contain large errors, since, to obtain any value at all, the γ parameters used had to be quoted to a greater accuracy than the method justified. To estimate these errors consider the GaAs values. The

heavy hole mass is in error by approximately twenty percent and the light hole mass is in error by approximately ten percent. This leads to estimates for AlAs of

$m_l = 0.15 \pm 0.015 \quad (\times 9.109\ 10^{-31}\text{kg})$

$m_h = 0.76 \pm 0.15 \quad (\times 9.109\ 10^{-31}\text{kg})$

The more recent CPA calculations of Chen and Shur (15) gave the values for $Ga_{0,1}\ Al_{0.9}As$ as

$m_l[1,1,1] = 0.137 \quad (\times 9.109\ 10^{-31}\text{kg})$

$m_l[1,0,0] = 0.103 \quad (\times 9.109\ 10^{-31}\text{kg})$

$m_h[1,1,1] = 1.053 \quad (\times 9.109\ 10^{-31}\text{kg})$

$m_h[1,0,0] = 0.435 \quad (\times 9.109\ 10^{-31}\text{kg})$

Whilst these can only be used as estimates for the AlAs masses they do show the high degree of anisotropy in the hole masses and could be used to support an argument that the tabulated light hole mass is an upper bound, being equal to the highest anisotropic value.

It must be concluded that the masses given for AlAs in Table 1.1 are only the currently accepted estimates and further characterisation work is needed.

Despite the uncertainty in the values for some of the properties of the binary compounds estimates are required for the properties of tertiary alloys. A simplistic approach to obtain the GaAlAs parameters is to interpolate between the values for GaAs and AlAs. Then, for an alloy with

an Al molefraction x, a parameter p is estimated as

$$p(x) = (1-x)p(o) + x\,p(1)$$

A lack of experimental results means that this approach is used at present to obtain all the effective masses in GaAlAs, although another approach is suggested by k.p. theory (20), which gives the conduction band and light hole band masses as

$$\frac{1}{m_c} = 1 + \frac{2P^2}{3\hbar^2}\left(\frac{2}{E_g} + \frac{1}{E_g+\Delta_o}\right)$$

$$\frac{1}{m_\ell} = 1 - \frac{4P^2}{3\hbar^2 E_g}$$

For GaAs and AlAs, the spin split off energy Δ_o is much less than the bandgap E_g, so that the conduction band mass can be written as

$$\frac{1}{m_c} = 1 + \frac{2P^2}{\hbar^2 E_g}$$

Then, since both expressions are dominated by the second term and the matrix element P has been shown to be approximately constant to a first approximation the masses will have the same compositional dependence as Eg.

There are several studies of the variation of bandgap with composition. Obtaining a conclusive answer is difficult, a point clearly demonstrated when different techniques obtained different results from the same samples (21).

The tabulated formula is the one which appears to be the most widely accepted and seems to originate in unpublished work (22). The formula has a discontinuous change at an Al molefraction very close to that associated with the change from a direct to an indirect bandgap. This has lead Chen and Sher, whose CPA calculation showed no such behaviour, to claim that the discontinuity is caused by experimental error in determining the direct bandgap in an indirect material.

A different viewpoint can be taken (23) by considering the more usual expression for the dependence of the bandgap on composition

$$E_g(x) = E_g(o) + (E_g(1) - E_g(o) - c)x + cx^2$$

If the non-linearity, measured by the bowing parameter c, is mainly caused by alloy scattering, then c itself may be composition dependent. A conclusion supported by the CPA calculations which show a bowing parameter which is dependent on x. The tabulated formula would then be considered as a simple fit, with composition independent bowing parameter, to experimental data which shows composition dependent bowing and would therefore be expected to change as further measurements improve the known data.

The most important property of a GaAlAs system is the manner in which the bandstructures of two layers with different compositions align at a heterojunction. The usual method of expressing this alignment is to quote the difference in the energies of the Γ point conduction band minima across the heterojunction as a fraction of the total difference in the direct bandgaps, Q. There is an implicit assumption that this is constant for all systems, which means that doped heterojunctions, in which charge

transfers to the low bandgap material, are excluded and that the results of Waldrop et al (24) will eventually be shown to be due to an uncontrolled variable, which caused apparently identical heterojuntions to give rise to different discontinuities.

Until recently it was widely accepted that the value for Q was 0.85, a result obtained by Dingle (25). More recently, results have tended to show this to be an overestimate.

The "Dingle Rule" was obtained from the absorption spectra of quantum wells by matching a simple model for the expected electron and hole energy levels to the observed transitions. This or the related luminescence experiment have proved very popular, with the more recent work by Miller et al (26) on parabolic and square quantum wells giving

$$Q = 0.57$$

and the results from square wells, by Dawson et al (27)

$$Q = 0.75$$

Although this technique is widely used the three papers(25,26,27) demonstrate some of the difficulties which arise. All the authors assume that the simple model derived in the previous section is correct. The different authors use different connection formulae to obtain the eigenstates, a difference which leads to significantly different energy levels for the same parameters. The uncertainty in the hole masses means that Miller et al are justified in using the hole mass as well as Q as an adjustable parameter, and the uncertainty in growth allows Dawson et al to vary the quantum well width.

The conclusion is that, by assuming the validity of the simple model and the values of some of the parameters, this technique is assuming too much. Consequently it is not a good technique for obtaining Q unless sufficient data is obtained to allow Q and all the masses to be variables and still have sufficient data to prove the validity of the model. A technique is therefore required which assumes neither the validity of the simple model or the values of the effective masses.

Such a technique is C-V profiling which can be used to obtain Q. Since its original use by Kroemer (28) the technique has been improved by Watanabe et al (29). After extracting the value for Q, it has been shown that the result can be checked by modelling the measured carrier concentration, demonstrating the measurements sensitivity to Q. The other advantage of the C-V technique is that by using n and p type doped GaAs it is possible to obtain the conduction band and valence band discontinuities separately. The fact that Watanabe et al obtained results which exactly added up to the bandgap difference was used as an indication of the accuracy of the technique.

The result obtained using this technique was

$$Q = 0.62$$

a result which was supported recently by Heiblum et al (30) from an examination of the photocurrent across a sample containing a heterojunction.

In the past, agreement with this result has also been claimed from a hole mobility experiment (31). A model of the hole mobility in the two dimensional electron gas next to the heterojunction was used to extract a

difference in hole band maxima of 220 meV, the total bandgap difference being obtained from a photoluminescence measurement. The samples Al molefraction, 0.5, meant that the material was indirect and the measured value, 550 meV, is much closer to an estimate of the indirect gap, 562 meV, than the estimated direct gap, 626 meV. The conclusion is that Wang et al measured the indirect gap and the true value of Q, using the estimated direct gap is

$$Q = 0.67 \pm 0.05$$

which is in good agreement with the analysis they produced of the results of a similar experiment by Stormer

$$Q = 0.69 \pm 0.04$$

Despite this it would appear that the best estimate of Q presently available is Q = 0.62, a conclusion which is supported by the recent theoretical model purposed by Tersoff (32), a model experimentally verified by Margaritondo (33).

One final point which should be made is that although the bandgaps themselves are temperature dependent, there is some evidence (34) that the temperature dependence of the bandgap is independent of the composition. This means that the bandgap difference across a heterojunction is temperature independent, with the consequence that the potentials used to represent the GaAlAs structures in Chapter 5 are temperature independent.

1.4 Comparison between the simple model and other models

A direct comparison between the results of a tight binding calculation and those of the simple model from section 1.2 for the transmission coefficient for a single layer has been performed by Osbourn (35). The layers studied were chosen to be sufficiently thick that the transmission coefficient from the simple model could, to a good approximation, be written as

$$T = A \exp(-2\kappa d)$$

where d is the layer thickness.

Using this form Osbourn obtained the value of κ which would be required to fit the tight binding calculation. The required value was found to be smaller than the value which Osbourn claimed arose from the simple model, the difference increasing with Al molefraction. The conclusion was therefore reached that the simple model was inaccurate.

In the above analysis the decay constant was defined as

$$\kappa = \left[\frac{2m(V-E)}{\hbar^2} \right]^{\frac{1}{2}}$$

as would be expected, however, the potential V was taken to be the total bandgap difference rather than the conduction band difference. Osbourn therefore overestimates the value of κ which would arise in the simple model, a fact which could explain the discrepancy observed. The compositional dependence of the disagreement then arises because the overestimate increaseas with composition. This misunderstanding invalidates Osbourn's conclusion.

A more recent comparison was performed by Marsh and Inkson (9,36). The comparison, based on empirical pseudopotential calculations of the phaseshift suffered by a wavefunction reflected from a heterojunction, concluded that the simple model was very inaccurate and could not be expected to give any quantitative predictions of the properties of systems containing heterojunctions. As with Osbourn this conclusion can be shown to be based on a misunderstanding concerning the simple model (37).

The expression used by Marsh and Inkson for the phaseshift calculated using the simple model was

$$\phi = \arctan\left[\frac{2k\kappa}{(\kappa^2-k^2)}\right] + \pi \qquad (1.4.1)$$

This expression arises when the wavefunction is matched across the heterojunction using the continuity conditions which require the continuity of

$$f \ , \ \frac{\partial f}{\partial z}$$

For comparison the result obtained using the continuity relations which arise naturally in the model, which require the continuity of

$$f \ , \ \frac{1}{m}\frac{\partial f}{\partial z}$$

is

$$\phi = \arctan\left[\frac{2\kappa k}{m_1 m_2\left(\left(\frac{\kappa}{m_2}\right)^2-\left(\frac{k}{m_1}\right)^2\right)}\right] + \pi \qquad (1.4.2)$$

where m_1 and m_2 are the masses in GaAs and GaAlAs respectively.

To compare the results of the pseudopotential calculation and the correct formula (1.4.2), a value for the potential V is required which is unfortunately not given. However the expression used, (1.4.1), is a

function of V/E and so a fit to the curve obtained from (1.4.1) will give a unique value for V.

In figure (1.1) the results obtained for an Al molefraction of 0.25 are shown. The given curve from expression (1.4.1) is shown as curve B which was fitted (curve C) to obtain the potential V. This value was then inserted into the correct formula (1.4.2) to give an improved, but unsatisfactory result (curve D). This remaining disagreement can be reduced further if the value of V extracted from curve B is considered.

The fitting procedure shows that a potential corresponding to more than 1.5 times the accepted bandgap difference between the two materials was used in (1.4.1), an error which may in part be due to following Osbourn's definition of the decay constant.

To use the simple model properly a value for the conduction band discontinuity is required which is as consistent as possible with the results of the pseudopotential calculations. To do this the conduction band discontinuity was taken to be the accepted bandgap difference minus the valence band discontinuity ΔE_v given by Marsh and Inkson

$$V = 1.247x - \Delta E_v \quad eV$$

This is not the exact value which should be used, since the pseudopotential method doesn't give the exact bandstructure, but, if the pseudopotentials do give a good model of a real heterojunction, this must be a good approximation. The result of using this value of potential in the correct expression is shown as curve E. The agreement is encouraging, especially when it is recognised that this is not an optimum fit, but the

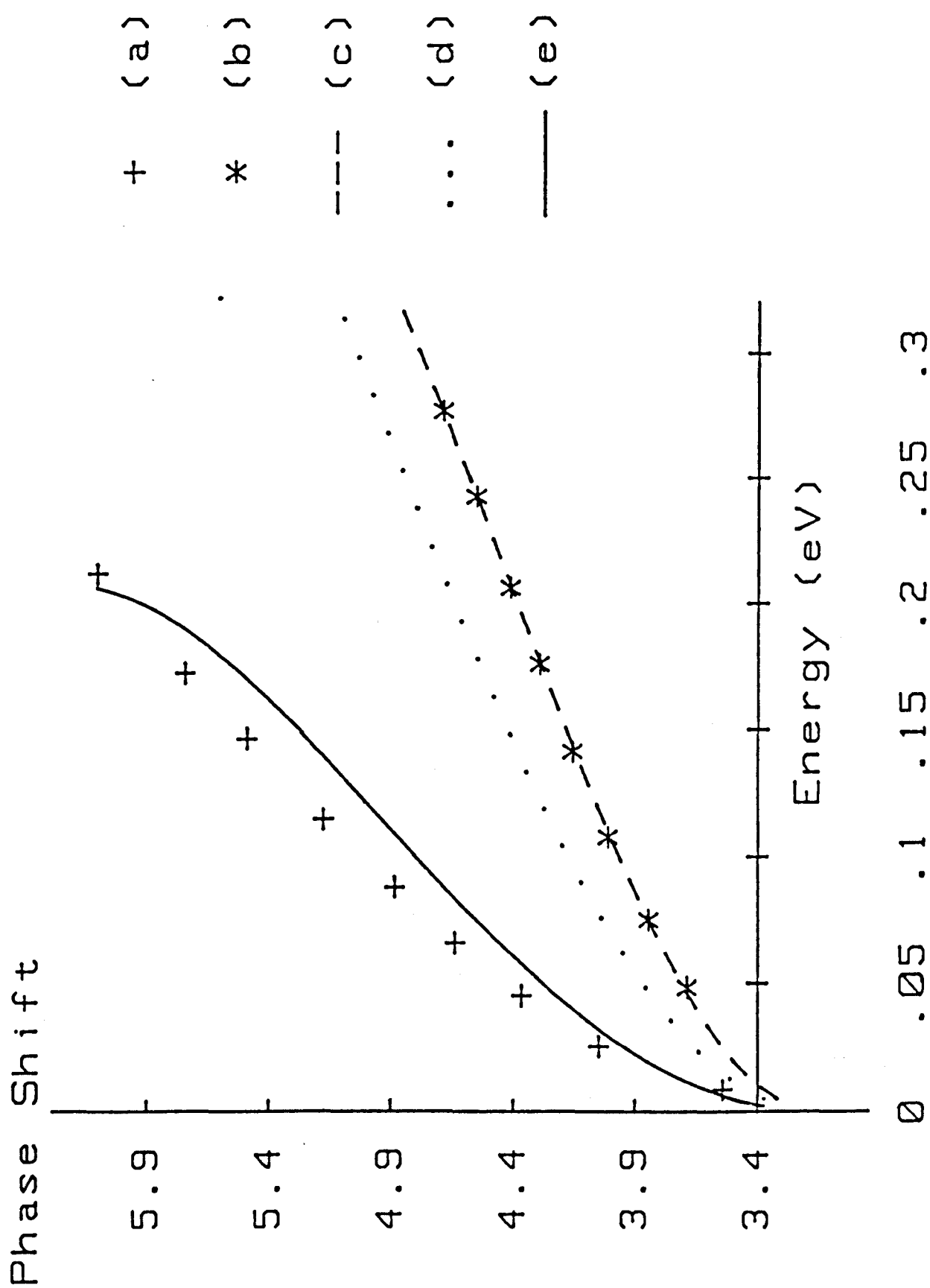

Figure 1.1 Comparison of phase shifts calculated from pseudopotentials and the simple model
(a) Pseudopotential result
(b) Marsh and Inkson's result from (1.4.1)
(c) Fit to (b) to extract the potential
(d) Results of (1.4.2) using the potential from (c)
(e) Results of (1.4.2) using more realistic potential

result of using the most accurate parameters available.

More recently Marsh and Inkson (38) extended the calculations to evaluate the eigenenergies of a single quantum well. The eigenvalue equation used for a well of length L was

$$kL + \alpha(k) = m\pi$$

In this expression $\alpha(k)$ was taken to be the phaseshift between the incident and reflected components, which in this case was taken to have two components, one from the envelope function and one from the cell periodic part of the wavefunction. The simple model cannot reproduce $\alpha(k)$ because it contains no means of calculating the contribution from the cell periodic component. This would appear to be a severe criticism until expression (1.4.3) is considered in detail.

Condition (1.4.3) is exactly the same as that which would be obtained using a square well model for the Quantum well. The square well is the potential in the S.E. for the envelope function alone and not the total wavefunction. Therefore, the condition which gives rise to (1.4.3) should only be applied to the envelope function and not the total wave function, giving rise to the eigenvalue equation

$$kL + \phi(k) = m\pi$$

as originally defined by Marsh and Inkson (9), where it was also stated that the cell periodic contribution to α was expected to be small.

In two of the crystallographic directions studied by Marsh and Inkson the difference between α and ϕ does appear to be small. However, in the

(1,0,0), direction chosen to calculate the eigenenergies of a Quantum well, the difference is large. In this direction $\emptyset$ is almost constant over the entire energy range except for a resonance, but α is a smooth function very similar to the other results. The smoothing of the resonant feature by the cell periodic component suggests that the strange behaviour observed in the (1,0,0) direction is due to the manner in which the phase of the total wave function is subdivided between the cell periodic function and the envelope function. If this is correct, then as with the other crystallographic directions, the difference between $\emptyset$ and α can be assumed to be small.

Despite these misunderstandings, the work of Marsh and Inkson remains the best from which to obtain an estimate of the errors involved in using the simple model. The conclusion is therefore that in the worst possible case, a GaAs/AlAs heterostructures, an error of up to ten percent can be expected, with smaller errors in GaAs/GaAlAs heterostructures due to the lower conduction band discontinuity.

1.5 Summary

In this chapter it has been demonstrated that Bastard used justifiable approximations to produce a simple one dimensional model S.E. for the conduction band states in systems with varying material composition.

The model obtained by Bastard requires several input parameters before it can be used to predict the behaviour of GaAlAs systems, and the accepted values for the parameters were tabulated. Although there is an accepted value for all the parameters doubts about the accuracy of several of these

were expressed, doubts which indicate a need for further characterisation of AlAs and bring into question the use of superlattice absorption/emission experiments to determine the conduction band discontinuity. These doubts cast upon the results of superlattice experiments meant that the results of such experiments were ignored and an accepted value for Q of

$$Q = 0.62$$

was taken from a C-V profiling experiment.

The results of the final section showed that criticisms of the simple model by Osbourn and Marsh and Inkson were based on misunderstandings of the simple model. The work of Marsh and Inkson was corrected to show that their estimate of errors of 10% for an AlAs/GaAs heterojunction was good enough to be taken as a guide to the order of magnitude of the errors.

A point has now been reached at which a simple S.E. can be used with some confidence to model a GaAlAs/GaAs heterostructure.

THE SOLUTION OF A SIMPLE TIME DEPENDENT PROBLEM

2.1 Introduction

The last chapter demonstrated that a simple one dimensional model could be used for systems in which the Aluminium concentration in GaAlAs varied. The tunnelling current through such a system arises from the motion of electrons, motion which must be treated as a time dependent problem. In this chapter it will be demonstrated that a simple formula, which involves only the solution to the time independent S.E., can be used to calculate tunnelling currents. This is important later in deriving a current voltage expression which only requires the solution of the time independent equation.

In the first three sections the basis sets for two simple model potentials and the expressions for the transmitted and reflected wave functions will be obtained. The expressions for the wave functions can be used to give a simple picture of the processes involved in tunnelling, and show how the solution to the time independent S.E. arises naturally in a time dependent problem. The solution for the double barrier potential shows that the transmission coefficient has a resonance, a property which can, hopefully, be exploited and is the major reason for interest in these systems.

In the fourth section a general numerical technique is outlined which can be used to complement the analytic solution. The major use of this technique here is to confirm the accuracy of the preceeding analysis. This is confirmed in the fifth section before a short summary of the conclusions is given.

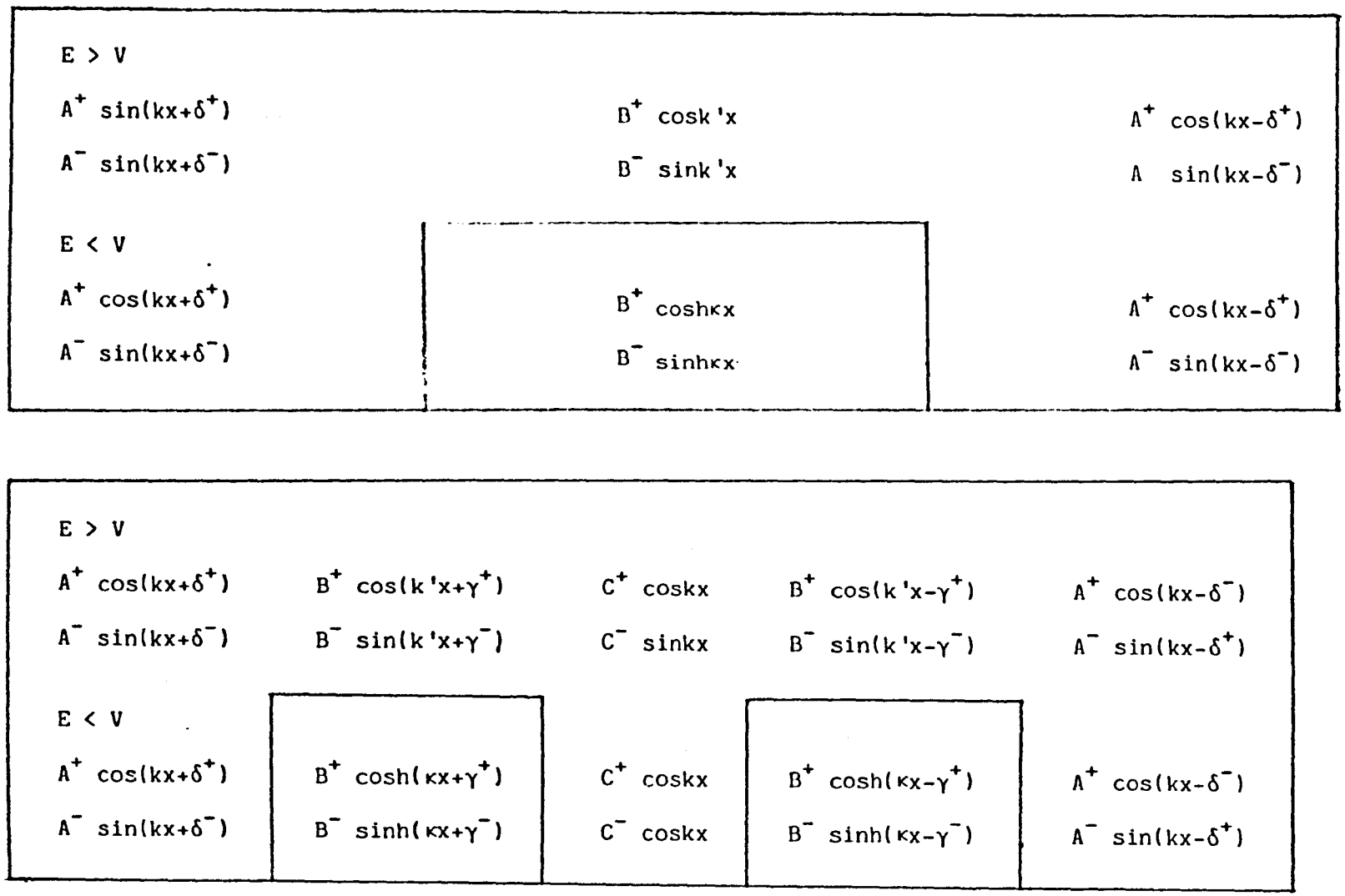

Figure 2.1 Basis sets for one and two square barriers

2.2 Basis set for a square barrier potential

Before attempting to solve the problem of a wavepacket incident upon a square barrier, a complete orthonormal basis set is required. The basis set are the eigenstates for the Hamiltonian, shown in figure (2.1), which, because of the symmetry of the potential, can be split into odd and even parity states, each set having one state at each energy.

The orthogonality of the basis set can be demonstrated in two simple stages.

(1) Consider any two non-degenerate states ψ_1 and ψ_2. It is easily shown (39) that these two states must be orthogonal.

$$H\psi_1 = E_1\psi_1 \qquad \text{and} \qquad H\psi_2 = E_2\psi_2$$

$$\langle \psi_1 | H | \psi_2 \rangle - \langle \psi_1 | H | \psi_2 \rangle = 0$$

$$(E_1 - E_2) \langle \psi_1 | \psi_2 \rangle = 0$$

The states being non-degenerate means $E_1 \neq E_2$ so that

$$\langle \psi_1 \left| \psi_2 \right. \rangle = 0$$

and the states are orthogonal.

(2) With all the non degenerate states orthogonalised it remains to orthogonalise the degenerate pairs. The orthogonality of degenerate pairs can be simply demonstrated using the parity of the states, this property being the motivation for the choice of the co-ordinates.

With the orthogonality of the states demonstrated they must be normalised. The normalisation process is the same for all states and will be

demonstrated for an even parity state with an energy less than that equivalent to the barrier height. The normalisation integral is

$$\langle\psi|\psi\rangle = 2\left[\int_0^a dx\ B^2\ \cosh^2 \kappa x + \int_a^\infty A^2\ \cos^2(kx-\delta^+)\right]$$

where the parity of the state has been used to simplify the expression. The next step is to write the integrals in the form

$$\langle\psi|\psi\rangle = 2\left[\int_0^a dx\ (B^2\cosh^2\kappa x - A^2\cos^2(kx-\delta^+)) + \int_0^\infty dx\ A^2\cos^2(kx-\delta^+)\right]$$

The first integral can be easily performed to give a contribution

$$\frac{a(B^2-A^2)}{2} + \frac{B^2\sinh 2\kappa a}{4\kappa} - \frac{A^2[\sin 2(\kappa a-\delta^+) - \sin 2\delta^+]}{4k}$$

The second integral is more difficult, direct evaluation giving infinity. Therefore consider

$$\lim_{k\to k'} \int_0^\infty dx A(k)\ A(k')\ \cos(kx-\delta^+(k))\ \cos\ (k'x-\delta^+(k'))$$

This can be evaluated split into two contributions.

$$\lim_{k\to k'} \frac{A(k)}{2}\ A(k') \int_0^\infty dx \left\{ \cos\ ((k-k')x - (\delta^+(k)-\delta^+(k'))) + \right.$$

$$\left. \cos\ ((k+k')x - (\delta^+(k) + \delta^+(k'))\right\}$$

Each one of these can then be evaluated by the same technique. Consider the first term and let $(k-k')y = (k-k')x - (\delta^+(k) - \delta^+(k'))$.

Then the first integral becomes

$$\int_0^\infty dy \cos((k-k')y) + \int_{\frac{\Delta\delta}{\Delta k}}^0 dy \cos((k-k')y)$$

where $\Delta\delta/\Delta k = (\delta^+(k) - \delta^+(k'))/(k-k')$

The first integral must be integrated using

$$\int_0^\infty e^{ikx}\, dx = \pi\, \delta(k) + i\, \frac{(1 - \delta_{k,0})}{k}$$

to give

$$\int_0^\infty dx \cos\left((k-k')x - (\delta^+(k) - \delta^+(k'))\right) =$$

$$\pi\, \delta(k-k') + \frac{1}{(k-k')} \sin\left(-(\delta^+(k) - \delta^+(k')) / (k-k')\right)$$

The problem that arises in normalising the basis set can now be seen. The free space contribution has given rise to the $\delta(k-k')$ term, whilst the remainder of the wave function has given rise to a finite term. These two terms require different schemes of normalisation. The δ function term requires A to be such that

$$\langle \psi(k) \mid \psi(k') \rangle = \delta(k-k')$$

whilst the other term requires A such that

$$\langle\psi(k) \mid \psi(k)\rangle = 1$$

These two conditions are mutally exclusive.

The only way to proceed is to consider the wave function. The free space component to the wave function occupies all but the central region of space. That is all of space except for a small region much less than a micron in length. It is therefore reasonable to expect the free space component of the wave function to dominate, therefore, it would seem reasonable to normalise the basis set in the same manner as a plane wave. For this reason all the contributions to the integral, $\langle\psi|\psi\rangle$, except for the $\delta(k-k')$ contribution are neglected and the wave function δ function normalised.

This means that the approximation

$$\langle\psi|\psi\rangle = A^2 \pi \delta(k-k')$$

will be used so that A(k) is a constant with a value of $1/\sqrt{\pi}$. With the states normalised the parameters δ must be defined to characterise the states, this is done using the continuity relations

$$\delta^+ = ka + \arctan(\kappa \tanh \kappa a/k) \qquad E < V$$

$$\delta^+ = ka - \arctan(k' \tan k' a/k) \qquad E > V$$

$$\delta^- = ka - \arctan(k \tanh \kappa a/\kappa) \qquad E < V$$

$$\delta^- = ka - \arctan(k \tan k' a/k') \qquad E > V$$

where $k' = \left(\frac{2m(E-V)}{h^2}\right)^{\frac{1}{2}}$ $\qquad \kappa = \left(\frac{2m(V-E)}{h^2}\right)^{\frac{1}{2}}$

2.3 Expansion of the initial condition

Having obtained a basis set in the previous section. The next step is to expand the initial condition in terms of the basis set so that the wavepacket can be evolved by multiplying each basis function by the phase factor exp (-iωt) and re-integrating.

There are two possible restrictions which could be imposed to study tunnelling. The wavepacket can be such that there is either no probability of finding an electron on one side of the barrier or no probability of finding an electron with an energy greater than the barrier height. Two conditions which are made mutually exclusive by the uncertainty relation connecting wavevector and space. The aim of the thesis is to study current voltage characteristics due to electrons traversing a barrier system. This means that the first of these conditions is more applicable, and is therefore adopted.

The initial condition is such that the wavepacket is restricted at t=0 to one side of the barrier, Region I, so that to reach the other side, Region III, the barrier must be traversed. This is unlike the other condition where there is always some probability of finding an electron in Region III at t=0. The expansion coefficients of the initial condition for the odd and even states are

$$c(k) = \int dx' \quad \psi(x'o)\, A^{+}(k) \cos (kx + \delta^{+})$$

$$s(k) = \int dx' \quad \psi(x'o)\, A^{-}(k) \sin (kx + \delta^{-})$$

These can be written as linear combinations of the fourier coefficients of the initial condition

$$f(k) = \int dx' \; e^{ikx'} \; \psi(x'o)$$

$$c(k) = A^{+}(k) \left[e^{i\delta^{+}} \; f(k) + e^{-i\delta^{+}} \; f(-k) \right]/2$$

$$s(k) = A^{-}(k) \left[e^{i\delta^{-}} \; f(k) - e^{-i\delta^{-}} \; f(-k) \right]/2$$

The wave function in Region III is then the sum of two integrals

$$\psi(x,t) = \int_{o}^{\infty} dk A^{+}(k) \; c(k) \; \cos \; (kx-\delta^{+}) \; e^{-i\omega_k t}$$
$$+ \int_{o}^{\infty} dk \; A^{-}(k) \; s(k) \; \sin \; (kx-\delta^{-}) \; e^{-i\omega_k t}$$

Using the result, from the previous section, that both the normalisation constants A^{+} and A^{-} are to a very good approximate given by $1/\sqrt{\pi}$ the two terms can be combined. The expression is further simplified by noting that the δ's are both odd functions of k to give

$$\psi(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \; f(k) \; \cos \; (\delta^{+}-\delta^{-}) \; \exp \left[i(kx + \delta^{+} + \delta^{-} - \omega_k t) \right]$$

A similar procedure for the wave function in Region I gives

$$\psi(x,t) = \frac{1}{2\pi} \left[\int_{-\infty}^{\infty} dk \; f(k) \; \exp \; (-ikx-i\omega t) \right.$$
$$\left. + \; i \int_{-\infty}^{\infty} dk \; f(k) \; \sin \; (\delta^{-}-\delta^{+}) \; \exp \; (i(kx-\omega t+\delta^{+}+\delta^{-})) \right] \qquad (2.2.2)$$

If it is remembered that the wave function in free space would be

$$\psi(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \; f(k) \; \exp \left[-i(kx-\omega t) \right]$$

these result have a simple interpretation. The wave function in Region

III can then be identified as a free propagating wavepacket, with each fourier component modified by a prefactor which can be identified as the transmission coefficient for that fourier component. The wave function in Region I has two components. The first is a free evolving wavepacket, whilst the other is a wavepacket moving in the opposite direction with each fourier component premultiplied by a reflection coefficient r(k).

This interpretation means that;

(i) the condition $t^2 + r^2 = 1$ is automatically satisfied.

(ii) the phase shifts of the reflected and transmitted fourier components differ by $\pi/2$

(iii) because both the phase angles δ tend to zero as the barrier width tends to zero, the solution to the freely evolving wavepacket is easily recovered.

The most important result arises when the expressions $\cos^2(\delta^+-\delta^-)$ and $(\delta^++\delta^-)$ are expanded

$$\cos^2(\delta^+-\delta^-) = \frac{4\kappa^2k^2}{4\kappa^2k^2 + (\kappa^2-k^2)^2\sinh 2\kappa a}$$

$$(\delta^++\delta^-) = 2\kappa a + (\kappa^2-k^2)\tanh \kappa a/\kappa k$$

and found to be identical to the transmission probability and phaseshift which would be obtained by solving the time independent S.E. with a unit incident flux in Region I. Although this result has been obtained for a simple potential there appears to be no stage in the analysis which would

prevent the extension of this result to other potentials. The conclusion is therefore reached that for a general potential the wave function in Regions I and III can be obtained by a simple three stage process;

(i) fourier transform the initial condition

(ii) solve the time independent S.E. for unit incident flux to obtain t(k) and r(k)

(iii) integrate the expressions (2.3.1) and (2.3.2)

It is now possible to give a general description of the interaction of a wavepacket and a potential. Firstly consider Region III, causality requires that the wave function in this region must be zero until the incident wavepacket has reached the front of the potential. A requirement fulfilled by subtle phase cancellation in (2.3.1) which makes the integral zero until sometime after the impact of the wavepacket upon the potential. If the transmission coefficient and phaseshift are both smooth functions of wavevector, the transmitted wavepacket has only slightly modified fourier components when compared to the initial condition. Then the transmitted wavepacket resembles the incident wavepacket reduced in magnitude.

In Region I there are two contributions. The first is a freely evolving wavepacket which moves as if the potential were absent, and, if evaluated alone, would show a wavepacket which evolved freely out of Region I. The second represents the reflected wavepacket. Its similar phase structure to the transmitted wavepacket leads to the expectation that it is also zero until the incident wavepacket has impacted with the potential. Again modified fourier components are expected to give rise to a reflected wavepacket which resembles a reduced incident wavepacket moving in the

opposite direction. It is a combination of these two terms that gives rise to an interference structure seen in numerical simulations (40). A result clearly shown in figure (2.2), where both contributions are plotted separately and then combined.

The overall picture is that a wavepacket freely evolving until it impacts with the barrier. At about this time interference appears in Region I near the potential and a transmitted wavepacket begins to emerge from the back of the potential. Both the interference and transmitted wavepacket grow until finally a reflected wavepacket emerges from the interference, travelling backwards, and a transmitted wavepacket moves through Region III.

This interpretation can be confirmed by integration of expressions (2.3.1) and (2.3.2). The complex nature of the functional dependence of the phaseshifts δ^+ and δ^- on wavevector means that the integration must be performed numerically. The initial condition was assumed to be a Gaussian restricted to one side of the potential so that

$$f(k) = \int_{-\infty}^{-a} dx \exp\left(-(x-x_o)^2/\sigma^2 + i(k_o-k)x\right)$$

This can be evaluated analytically using standard integrals (71) to give

$$f(k) = \frac{\sqrt{\pi}\sigma}{2}\left[1 - \Phi\left(((x_o+a/\sigma) - i(k-k_o)\sigma/2\right)\right]$$

$$\exp\left(-i(k-k_o)x_o - (k-k_o)^2\sigma^2/4\right) \qquad (2.3.3)$$

where $\Phi(\alpha + i\beta)$ is the error function of complex argument. This expression was simplified by an appropriate choice of initial condition.

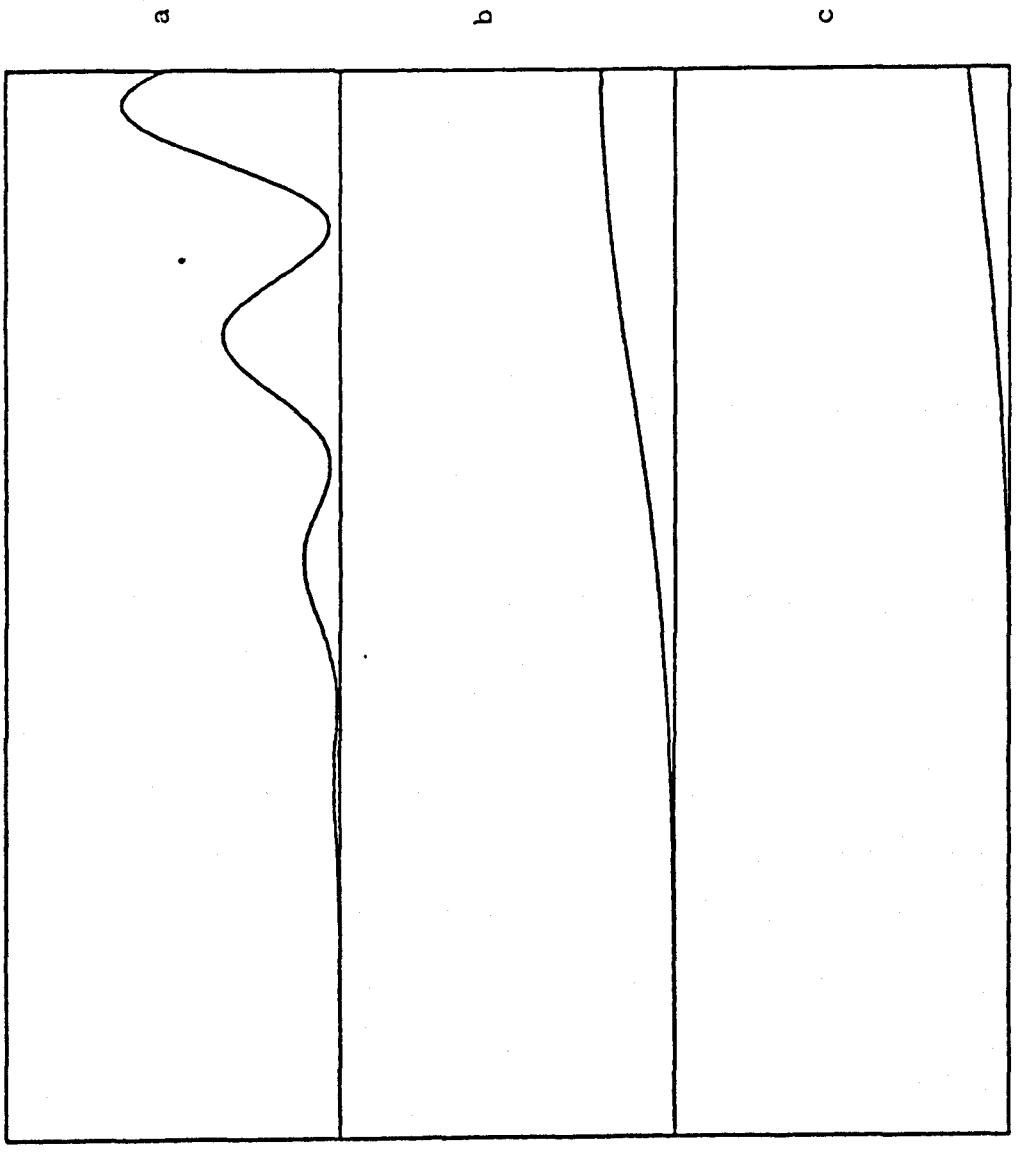

Figure 2.2 Evaluation of expression (2.3.2)
(a) both contributions
(b) incident wavepacket contribution
(c) reflected wavepacket contribution

Consider the error function (72)

$$\phi(\alpha + i\beta) = \phi(\alpha) + e^{-\alpha^2} f(\alpha, \beta)$$

An initial condition with a peak three standard deviations from the front of the barrier, gives α a value of three, a value large enough for the β dependence of the error function to be negligible.

The fourier coefficients of the initial condition are then approximately Gaussian even though the spatial distribution is non Gaussian.

Simpson's Rule was used to evaluate the integral numerically. To obtain a good result a representative sample is required. If the number of points in the sample is to be kept to an acceptable level, the initial position and the time of evaluation t must be restricted to prevent the exponential term

$$i(k_o - k)x_o + i\omega_k t$$

from oscillating too rapidly.

The same analysis that was used to produce (2.3.3) can be used to show that restricting the sample to a range within three standard deviations of the mean wavevector will produce no noticeable affect on the spatial distribution. This restriction means that the maximum value of $k-k_o$ is $3\sigma_k$. Then limiting the number of oscillations in the exponential over all the wavevector range to n gives rise to the condition

$$3\sigma_k x_o < n\pi$$

The initial condition has already been chosen to peak three standard deviations from the front of the barrier, and with the relationship between σ_k and σ_x this leads to

$$a < (n\pi - 18)\,\sigma_x$$

This result means that n must be greater than six and is proportional to the ratio of the width, 2a, to the standard deviation. This condition indicates that care is needed to integrate narrow wavepackets and wide barriers, since these give rise to the greatest number of oscillations in the exponential.

The time dependent term in the exponential leads to a condition on the time at which the integration can be evaluated. For the time dependent term to change by 2π over the energy range in a wavepacket ΔE

$$t < 2\pi\hbar/\Delta E$$

which for a wavepacket with an energy spread of 40 meV give

$$t < 10^{-13}\ s$$

Since the minimum time of interest is the time required for the wavepacket to reach the potential this condition limits the techniques usefulness for wide low energy wavepackets, which need a long time to reach the potential, giving rise to a highly oscillatory term.

Despite these limitations the technique was used to confirm the picture of interaction of a wavepacket and potential, and to confirm that the multiple peaks observed when the wavepacket first arrives at the barrier are due to interference between the incident and reflected wavepackets.

2.4 Double barrier results

The preceeding analysis can be applied to a system of two identical barriers, a system which again allows odd and even parity states to be defined.

The states are defined in figure (2.1) and the parameters obtained by matching are, for energies less than that equivalent to the barrier height

$$\gamma^{+} = \frac{\kappa b}{2} + \tanh^{-1}(k \tan(kb/2)/\kappa)$$

$$\delta^{+} = ka + \tan^{-1}(\kappa \tanh(\kappa a-\gamma^{+})/k)$$

$$\gamma^{-} = \frac{\kappa b}{2} - \tanh^{-1}(\kappa \tan.(kb/2)/k)$$

$$\delta^{-} = ka - \tan^{-1}(k \tanh(\kappa a-\gamma^{-})/\kappa)$$

whilst, for the remainder of the energy range the parameters are

$$\gamma^{+} = \frac{k'b}{2} - \tan^{-1}(k \tan(kb/2)/k')$$

$$\delta^{+} = ka - \tan^{-1}(k' \tan(k'a-\gamma^{+})/k)$$

$$\gamma^{-} = k'b - \tan^{-1}(k' \tan(kb/2)/k)$$

$$\delta^{-} = ka - \tan^{-1}(k \tan(k'a-\gamma^{-})/k')$$

Again the transmission coefficient is given by $\cos^2(\delta^{+}-\delta^{-})$, which for energies less than the barrier equivalent, reaches unity for k such that

$$\cos kb \cosh \kappa c + \left(\frac{\kappa^2-k^2}{2\kappa k}\right) \sinh \kappa c \sin kb = 0 \qquad (2.4.1)$$

This is in exact agreement with the results which will be obtained in Chapter 4, where it will be shown that for a system of N barriers these transmission resonances occur at

$$\text{coskb cosh } \kappa c + \left(\frac{\kappa^2-k^2}{2\kappa k}\right) \text{sinh } \kappa c \text{ sinkb} = \cos\left(\frac{r\pi}{N}\right)$$

The result, (2.4.1), is then recovered because n is restricted to be less than N, so that for two barriers it can only be unity.

It is these transmission resonances which can hopefully be exploited in double layer systems.

2.5 General numerical method

In the previous sections the solution to a simple time dependent problem was given, however the method used has restrictions;

(1) it is limited to the potentials for which an analytic transmission coefficient can be obtained.

(2) the integration of the final expression restricts the initial condition and time of evolution that can be treated.

(3) the potential is time independent.

(4) it is based on an approximately normalised basis set.

It is therefore useful to develop a complementary technique which has none of these restrictions, but, may have some of its own. In this section a general numerical technique will be outlined to be the complement to the

analytic technique. The first step is to obtain an approximate discrete equation which can be used by the computer to simulate the S.E. Since the general approach has already been discussed by Goldberg et al (40) the present discussion will be restricted to a generalisation to a time dependent potential.

Consider the Taylor expansion of the wavefunction to the first two orders

$$\psi(t + \Delta t/2) = \psi(t) + \left(\frac{\partial \psi}{\partial t}\right)_t \frac{\Delta t}{2}$$

$$\psi(t + \Delta t/2) = \psi(t) - \left(\frac{\partial \psi}{\partial t}\right)_{t+\Delta t} \frac{\Delta t}{2}$$

which means that to be consistent

$$\psi(t) + \left(\frac{\partial \psi}{\partial t}\right)_t \frac{\Delta t}{2} = \psi(t+\Delta t) - \left(\frac{\partial \psi}{\partial t}\right)_{t+\Delta t} \frac{\Delta t}{2}$$

The general form of the time dependent S.E. can be used to give

$$\left(1 - iH(x,t) \frac{\Delta t}{2}\right)\psi(x,t) = \left(1 + iH(x,t+\Delta t) \frac{\Delta t}{2}\right)\psi(x,t+\Delta t)$$

a form obtained by Goldberg et al using the Cayley form of the evolution operator. This relationship can be used to obtain $\psi(t+\Delta t)$ from $\psi(t)$ if the potential at $t+\Delta t$ is known. If the time dependence of the potential is not known, as would arise if the potential contained a Poisson correction dependent upon the wave function, the substitution

$$H(x,t+\Delta t)\ \psi(x,t+\Delta t) \approx H(x,t)\ \psi(x,t+\Delta t)$$

could be used.

A Taylor expansion of the Hamiltonian shows that for the substitution to

cause minimum error

$$H(x,t)\ \psi(x,t+\Delta t) >> \frac{\partial H(x,t)}{\partial t} \quad \Delta t \quad \psi(x,t+\Delta t)$$

A condition which must be added to those discussed by Goldberg et al.

The equation with a discrete timestep is then

$$\left(1 - iH(x,t)\ \frac{\Delta t}{2}\right)\psi(x,t) = \left(1 + iH(x,t)\ \frac{\Delta t}{2}\right)\psi(x,t+\Delta t)$$

from which an approximate, with a discrete spatial step, can be obtained as described by Goldberg et al (40).

The method was implemented using a fixed spatial step of 1 Å with the timestep fixed so that the parameter λ defined by Goldberg et al was unity. A condition which uses the largest possible timestep for a fixed spatial step.

2.6 Comparison of the analytic and numerical results

In the previous sections a general method for solving the time independent S.E. was developed and a simple problem solved analytically. A position has now been reached where the accuracy of the general method can be compared for the simple problem using the analytical solution.

The simple analytically solvable problem is a Gaussian wavepacket incident upon a square barrier. Any comparison between the analytic solution and

the general method requires the definition of a square barrier on a lattice. The discontinuous nature of the potential leads to some ambiguity as to the number of points required to define the potential. Vigernon and Lambin have studied the time independent S.E. for a square barrier and chose the discrete form for a square barrier of length L and height V on a lattice spacing Dx.

$N = L/Dx$

$V(0) = V(N) = V/2$

$V(n) = V \qquad 0<n<N$

The end points 0 and N are separated by L and can be considered the bounds of the square barrier. This technique of discretising a square barrier was also found to give the best results for the time dependent problem, a result which probably arises because the same procedure is used in both cases to obtain a discrete form of the spatial derivative.

For the time dependent problem the best method of comparing the results of the two techniques for evaluating the wave function, is to plot the wave function at a given time as a function of position. The best time at which to evaluate the wave function for comparison is whilst the wave function is still interacting with the potential barrier. A time chosen because it is when the interference between incident and reflected components occurs, allowing the height and separation of several features to be compared. The comparison is shown in figure (2.3), where it can be seen that the general method does reproduce the wave function very accurately, if the correct method is used to represent the potential.

Another comparison which can be made is the expected transmission

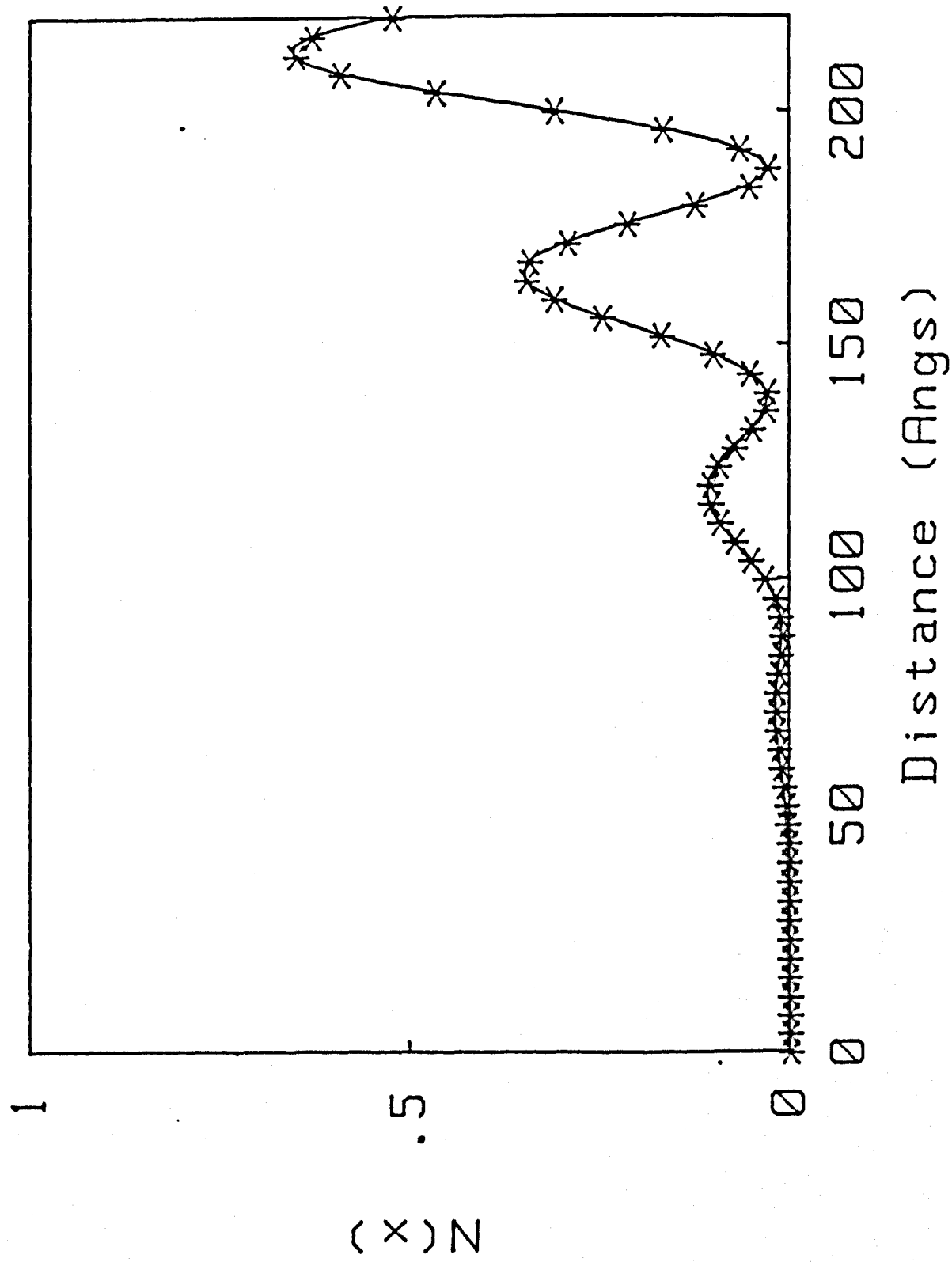

Figure 2.3 Comparison of the numerical results (*) and the analytic results (—).

probability when there is no longer any wave function in the barrier. Analytically the wave function in the region beyond the barrier is

$$\psi(x) = \int_{-\infty}^{\infty} f(k)t(k) \exp i(kx-\omega_k t)\, dk$$

If the barrier edge is at b then the probability of being beyond the barrier, P, is

$$P = \frac{\int_b^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdkdk' t(k)t(k')f(k)\exp\,(i(k-k')x-(\omega_k-\omega_{k'})t)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdkdk' f(k)f(k')e^{i(k-k')x}}$$

To simplify this formula assume $b\to\infty$, an assumption which must add to the integration a region over which the integrand integrates to zero, if evaluated at a sufficiently large time, t. Then the expression reduces to the form

$$P = \frac{\int_{-\infty}^{\infty} dk\ |t(k)|^2\ |f(k)|^2}{\int_{-\infty}^{\infty} dk\ |f(k)|^2} \qquad (2.6.1)$$

A well known result whose derivation shows that f(k) must be such that the transmitted pulse is localised at a sufficiently large time well away from the barrier, to allow the step $b \to \infty$.

It is possible to allow the numerical technique to iterate until the wavepacket has left the region around the barrier and then integrate in the region beyond the barrier and check the result with the formula (2.6.1). As an example consider a potential which is two 50 Å, 0.23 V

barriers, separated by 50 Å. The results plotted in figure (2.4) shows that the numerical technique agrees to within 0.5% with the value given by (2.6.1), and that the finite width of the wavepacket smooths the features in the transmission coefficient especially the very sharp resonance.

Another interesting result arises when studying the interaction of a wavepacket and a double barrier system. When a wavepacket is incident at approximately the same energy as a transmission resonance two pulses are reflected. The fact that this only occurs at the energy of the transmission resonance shows that the phenomena must be due to the zero in the reflection coefficient. The momentum distribution has two distinct negative momentum peaks and it can be shown that these two peaks each give rise to a spatial peak. This explains the observation that as time evolves one peak, with higher momentum is found to pass through the other.

2.7 Summary

The analytic solution to a simple time dependent S.E. has shown how the transmission coefficient, defined by solving the time independent S.E. with unit incident flux, naturally arises in the solution of a time dependent problem.

This result was confirmed, when it was demonstrated that the results obtained agreed with the numerical solution to a discrete form of the time dependent S.E. This numerical technique was also used to prove that a simple formulae for the total transmission probability is correct. A result which is important because the same analyses which can be used to derive expression (2.6.1) will be implicity used to derive the current voltage relationship in Chapter 4.

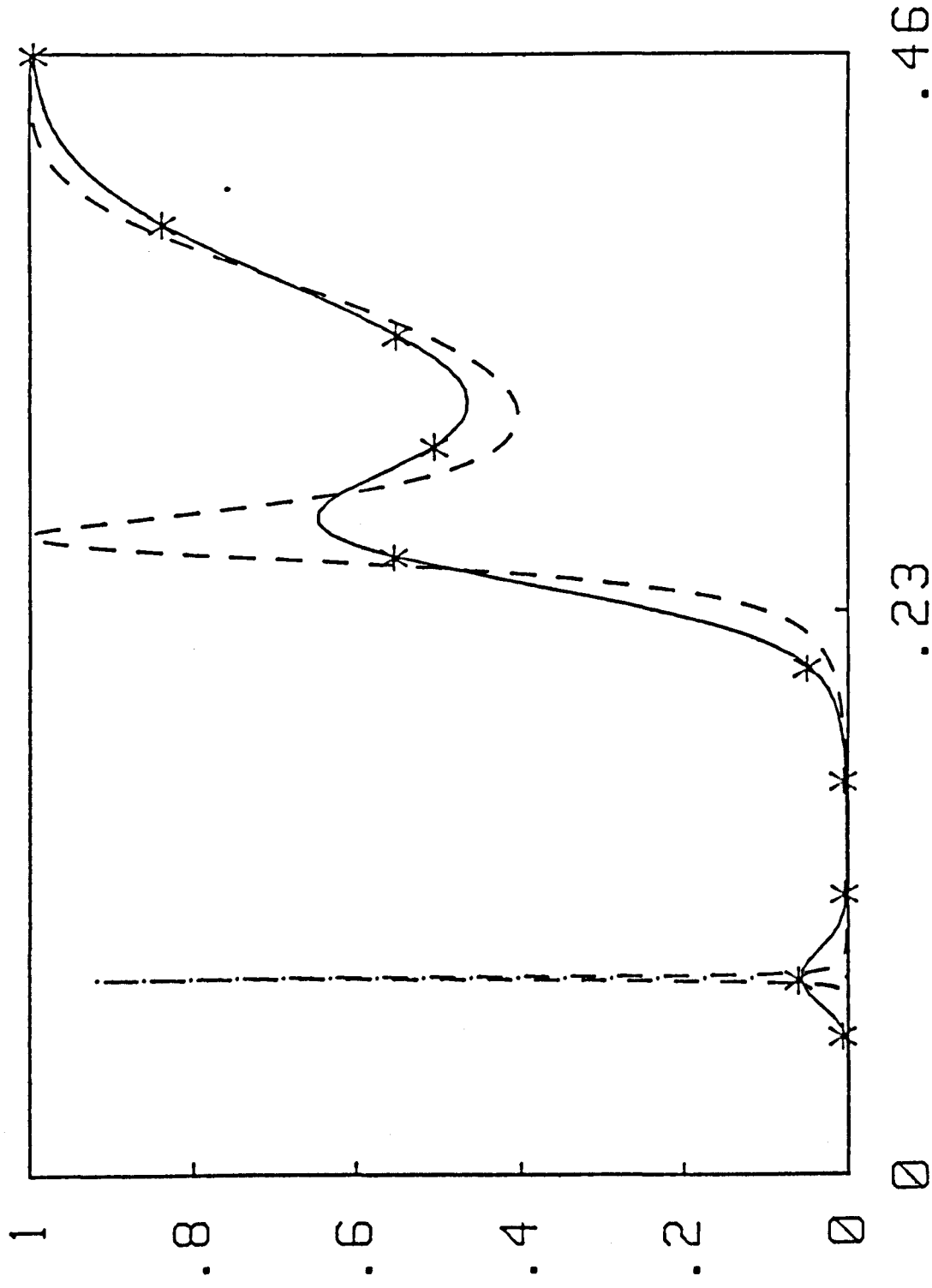

Figure 2.4 Calculation of transmission probability for Gaussian wavepacket, standard deviation 400Å, incident upon a double barrier system, two 50Å, 0.23 eV barriers separated by 50Å

--- Transmission coefficient

—— Transmission probability from (2.6.1)

xxx Numerical transmission probabilities

... Position of resonance, 79.9 meV

A DISCUSSION OF THE TRAVERSAL TIME OF A BARRIER

3.1 Introduction

In this chapter the time required for a pulse to traverse a system will be studied using the results of the previous chapter. This is done to establish an estimate of the time required for an electron to traverse a system and hence the importance of scattering processes during tunnelling.

In the first section the analytic result for the transmitted wavepacket is used to relate the traversal time across the system and the derivative of the phaseshift with respect to wavevector. The second section is used to discuss the square barrier results obtained by Stevens and Buttiker, results which contradict the result obtained in the first section. It will be shown that the analysis used by each author contains an error.

Since the thesis has a specific interest in double barrier systems these will be considered in the third section. It will be shown that, the transmission resonance is associated with an increase in the time required to tunnel, resonant tunnelling and unassisted hopping are different dynamical pictures of the same process, and that the picture proposed by Ricco and Azbel is contrary to the solution of the time dependent S.E.

Finally a primitive Monte Carlo model of tunnelling will give rise to a new interpretation of the result obtained in the first section.

3.2 Derivation of the traversal time expression

McColl (41) was the first to consider the time taken for a pulse to

traverse a barrier, using an analysis similar to that produced later by Hartman (43). The initial condition was restricted to one side of the barrier and the conclusion was reached that the wave function on the other side became significant when the initial pulse reached the front of the barrier and then later decreased. Unfortunately this statement has been misinterpreted as a statement that the transmitted pulse peak appears at the time that the incident pulse impinges upon the front of the barrier implying a zero tunnelling time.

The phase time result is a text book result, first obtained by Wigner (42) by considering a wavepacket made up of two components. The analysis was refined by Hartman (43) to a consideration of a full Gaussian wavepacket incident upon a square barrier. Hartman's final step was based on a stationary phase argument, which is usually used when the integrand involves a parameter which can be made large, forcing the integrand to oscillate violently except at the value where the oscillation is damped.

To avoid using stationary phase consider the Wigner function for the wavepacket in the region beyond the barrier. From the previous chapter the fourier components of the wavepacket in this region are

$$\psi(p,t) = f(p)\, t(p)\, \exp(i\omega_p t)$$

where $\omega_p = \hbar^2 p^2/2m$

Defining the Wigner function as an integral over k space gives

$$F(q,p,t) = \int \frac{dQ}{2\pi\hbar}\; t^*(p+Q/2)\; t(p-Q/2)\; f(p+Q/2)\; f(p-Q/2)$$

$$\exp\left[\, i(qQ+hptQ/m)\,\right]$$

If the initial condition is assumed to be the truncated Gaussian of the previous chapter then, to a good approximation, the fourier spectrum is Gaussian, and

$$f^*(p+Q/2)\ f(p-Q/2) = \exp\left[i(q_oQ-(p-p_o)^2/2\sigma p^2-Q^2/2\sigma p^2\right]$$

where σp is the standard deviation in the Fourier spectrum. The Wigner function is then

$$F(q,p,t) = \frac{1}{2\pi\hbar}\ \exp\left[-(p-p_o)^2/2\sigma p^2\right]$$
$$\int_{-\infty}^{\infty} dQ t^*(p+Q/2)t(p-Q/2)\exp\left[-Q^2/2\sigma p^2+iQ(q-q_o)-\frac{hpt}{m}\right]$$

If the initial Gaussian in momentum space is very localised, the pre-exponential Gaussian means that all values of p other than p_o are negligible so that

$$F\ (q,p,t) \approx F(q,p_o,t)$$

This assumption also means that the integrand itself is restricted so that only the region near $Q\approx 0$ contributed to the integral.

To proceed further an approximation must be made to account for the behaviour of the transmission coefficient t(p) must be used. The simplest approximation is to assume that the Gaussian is so narrow that t(p) is constant over the range of integration. The wavepacket is then found to traverse the barrier in the time taken for it to traverse the same width of free space and the only affect of the barrier is to reduce the

magnitude of the wavepacket by the transmission coefficient evaluated at the peak momentum.

To obtain the next approximation to the wavepacket the variation in the phaseshift $\phi(p)$ must be taken into account.

To proceed write the transmission coefficient in the form

$$t(p) = \exp(\ln t(p)+i\phi(p))$$

Then

$$t^*(p_o+Q/2)t(p_o-Q/2) = \exp\left[\ln t(p_o-Q/2) + \ln t(p_o+Q/2) + i\,\phi(p_o+Q/2)-i\,\phi(p_o-Q/2)\right]$$

because only a small range of Q contributes to the integral this expression can be Taylor expanded about Q=0 to give

$$t(p_o+Q/2)t(p_o-Q/2) = \exp\left[\ln(t(p_o)+t'Q/2)+\ln(t(p_o)-t'Q/2) + iQ\phi'\right] \qquad (3.2.1)$$

where $t' = \left(\frac{\partial t}{\partial p}\right)_{p_o}$ and $\phi' = \left(\frac{\partial \phi}{\partial p}\right)_{p_o}$

If the transmission coefficient is slowly varying so that

$$t(p_o) \gg t'Q/2 \qquad (3.2.2)$$

for all values of Q which contribute to the integral, expression (3.2.1) can be reduced to

$$t^*(p_o+Q/2)\ t(p_o-Q/2) = |t(p_o)|^2 \exp\left[iQ\phi'\right]$$

The Wigner function is then

$$F(q,p,t) \approx \frac{|t(p_o)|^2}{2\pi\hbar} \exp\left[-(p-p_o)^2/2\sigma p^2\right] \int_{-\infty}^{\infty} dQ \exp\left[-Q^2/2\sigma p^2 + iQ(q-q_o - \frac{\hbar p_o t}{m} - \phi')\right]$$

and the integration can be identified as a Gaussian centred on

$$q - q_o - \phi' - \hbar p_o t/m$$

which can be written as

$$q - q_o - \hbar p_o(t - m\phi'/\hbar p_o)/m$$

This expression shows that the transmitted wavepacket has the same form as the incident wavepacket, but reduced in magnitude by $|t(p_o)|^2$ and, compared to the pulse in the absence of the potential, it is delayed by a time τ where

$$\tau = \frac{m}{\hbar p_o}\left(\frac{\partial \phi}{\partial p}\right)_{p_o} \qquad (3.2.3)$$

This is a general result, but, caution is required in its use, its exact interpretation depends on the phase referencing of the plane waves matched across the potential. The interpretation given here being only valid for the phase referencing used. If another definition for the phase of the plane waves is used a similar analysis must be undertaken to find the exact interpretation of (3.2.3) for the phaseshift defined.

Of the two assumptions used in the derivation of (3.2.3) the first was that the wavepacket was localised in momentum about $\hbar p_o$, a condition which is expected if a function of p_o is to be defined. The second assumption was that the transmission coefficient was slowly varying over the range of wavevectors present in the wavepacket. These conditions

combined quantitatively in (3.2.2)

$$|t(p_o)| > \sigma_k \left(\frac{\partial |t(p)|}{\partial p}\right)_{p_o}$$

A condition which shows that if the transmission coefficient is rapidly varying the analysis is only valid for wavepackets with small standard deviations.

The earlier work by Hartman has been criticised because it required that the transmitted peak position was correlated with the time the peak of the incident pulse arrived at the front of the barrier. This caused concern, when later numerical results published by Goldberg et al (40) showed several peaks at the front of the barrier. The understanding of wavepacket behaviour obtained in the last chapter shows that these peaks are caused by the interference of the incident and reflected pulses. Despite this interference, it is possible to identify a wavepacket which impinges on the barrier at the expected time, justifying Hartman's analysis.

3.3 Single square barriers

The analysis in the previous section can be used on any one dimensional potential. There are specific results for square barriers obtained by other authors which contradict the general expression. These are now considered to clarify the situation.

Stevens (44) has obtained an expression for the velocity of an electron of energy E in a semi-infinite step potential of height V. If this expression is extended to be applicable to barriers of finite width, a, it

gives a traversal time Υ

$$\Upsilon = \frac{\dot{a}}{\sqrt{2(V-E)/m}} \qquad (3.3.1)$$

The wavepackets used for this analysis were specially chosen to allow complex analysis to be used in the barrier region and the wave function in the barrier was written as

$$\psi(x) = \frac{2k}{k+\kappa} \exp\left[-\kappa x\right]$$

The analysis used to obtain the phase-time expression requires a region of free space beyond the barrier. The wave function at the back of a barrier of width, a, would be

$$\psi(a) = \frac{2k}{k+\kappa} \exp\left[-\kappa a\right]$$

To match this wave function, the wave function beyond the barrier must be of the form

$$\psi(x) = t(k) \exp\left[ik(x-a)\right]$$

This is a different system of phase referencing than that used in interpreting equation (2.2.3), so that a similar analysis must be used to obtain a peak position x at time t such that

$$x - a - x_o - hk_o\left(\frac{t-m\phi'}{hk_o}\right)/m = 0$$

The time for the peak to reach the back of the barrier at a is then

$$t = \frac{-mx_o}{hk_o} + \frac{m}{hk_o}\left(\frac{\partial\phi}{\partial k}\right)_{k_o}$$

The first term is the time required to reach the front of the barrier (x_o being negative) so the time to cross the barrier is the derivative of the phase of t. By matching amplitudes at a, there are insufficient variables to match derivatives, t is found to be given by

$$t = \frac{2k}{k+\kappa} \exp\left[-\kappa a\right]$$

This appears to have no phase, but if a phase is defined as

$$\phi = i\kappa a$$

then the traversal time is

$$t = \frac{a}{(2(V-E)/m)^{\frac{1}{2}}}$$

The Stevens result is therefore recovered by using the phasetime analysis on the wave functions used by Stevens. This shows that the Stevens result is exact for the step potential for which the wave function used is exact, however, the extension to finite potentials can only be an approximation because of the approximate wave function used.

Buttiker (45) has derived an expression for the traversal time across a square barrier by considering a barrier containing a small magnetic field. The geometry is such that the potential is a function of y, the field is applied in the z direction and the beam of incident electron are spin polarised in the x direction.

The B field then has two affects;

(1) It adds a spin dependent term to the decay constant, the decay constant becoming a function of the z component of the spin. This polarises the spin parallel to the B field, by selectively reflecting the electrons with spin anti-parallel to the field.

(2) The magnetic field causes Larmor precession in the plane normal to the B field. This Larmor precession giving rise to a y component of spin.

The transmitted electrons are therefore no longer polarised along the x

direction but have components in all three directions, given by Buttiker as

$$s_z = \frac{\hbar}{2}\left(\frac{T_+ - T_-}{T_+ + T_-}\right)$$

$$s_y = -\hbar \sin(\Delta\Phi_+ - \Delta\Phi_-)\ \frac{(T_+T_-)^{1/2}}{T_+ + T_-}$$

$$s_x = \hbar \cos(\Delta\Phi_+ - \Delta\Phi_-)\ \frac{(T_+T_-)^{1/2}}{T_+ + T_-}$$

where $T_+(T_-)$ and $\Delta\Phi_+(\Delta\Phi_-)$ are the transmission probabilities and phaseshifts of the electrons with spin components parallel (anti-parallel) to the magnetic field.

If the small field limit is taken, a Taylor expansion for T gives

$$\langle s_z \rangle = \frac{\hbar}{2}\ \omega_L \cdot T_z$$

and for the trigonometric functions gives

$$\langle s_y \rangle = -\frac{\hbar}{2}\,\omega_L\ T_y$$

$$\langle s_x \rangle = \frac{\hbar}{2}\ (1-(\omega_L T_x)^2/2)$$

where

$$T_z = -\frac{m}{\hbar\kappa}\frac{\partial}{\partial\kappa}\ \ell n\ T^{1/2}$$

and

$$T_y = -\frac{m}{\hbar\kappa}\frac{\partial\Delta\Phi}{\partial\kappa}$$

Buttiker defines T_x using the conservation of spin

$$\langle s_x^2 \rangle + \langle s_y \rangle^2 + \langle s_z \rangle^2 = \hbar^2/4 \qquad (3.3.1)$$

to get

$$T_x = (T_y^2 + T_z^2)^{1/2}$$

Then

$$T_x = \frac{m}{h\kappa} \frac{1}{D} \frac{\partial D}{\partial \kappa} \qquad (3.3.2)$$

with D defined in terms of the phaseshift $\Delta\Phi$ and transmission probability T as

$$D = T^{\frac{1}{2}} \exp\left[i(\Delta\Phi - k d) \right]$$

However, the Taylor expansions used in defining T_x mean that the condition (3.3.1) can no longer be applied to the T's because of the neglected terms in the Taylor series. The application of condition (3.3.1) is equivalent to adding all the neglecting terms from $\langle s_z \rangle$ and $\langle s_y \rangle$ into $\langle s_x \rangle$. To be consistent with its own definition T_x must be defined as

$$T_x = T_y = \frac{-m}{h\kappa} \frac{\partial \Delta\Phi}{\partial \kappa}$$

and not the expression given by Buttiker. This conclusion is supported by Buttiker whose analysis of a magnetic field in free space gives rise to expressions for $\langle s_x \rangle$ and $\langle s_y \rangle$ which imply that $T_x = T_y$

The fact that a time can be defined is not sufficient, it must be shown to be physically relevant. To demonstrate the relevance of the time defined in (3.3.2), Buttiker used an analogy to another experiment.

The change in spin polarisation $\langle s_z \rangle$ was considered to be due to electrons changing spin states, a process considered to be analogous to the changes in energy level population produced by an oscillating potential. The actual mechanism for spin polarisation is not electrons changing state but preferential selection of one state compared with the other. A better

analogy would therefore be with electrons of two slightly different kinetic energies incident upon a barrier. The higher energy state would be preferentially transmitted, leading to an increase in the average energy of the transmitted beam compared to the incident beam. This selective transmission could not be used to define a physically relevant time.

The conclusion is that Buttiker used the wrong analogy to demonstrate the physical relevance of the time τ, defined in (3.2.2). A better analogy casts doubt on the physical relevance of τ.

The final conclusion must be that both Stevens and Buttiker produced flawed derivations of traversal times. By considering a step potential Stevens neglected a component of the wave function which is non negligible in a square barrier, an omission which is equivalent to the WKB approximation to the transmission coefficient.

Buttiker has derived (3.3.2), which is identified as the traversal time. In the derivation condition (3.3.1) was misused, whilst in attempting to demonstrate the relevance of (3.3.2) an incorrect analogy was used.

3.4 Numerical results

In this section the expression obtained for the time required to traverse a square barrier (3.2.3) will be compared to the results obtained by numerically integrating (2.3.1).

A comparison has been made previously by Schnupp (46), using the method of

Goldberg et al (40). It was found that for the one situation studied the peak in the wavepacket appeared much later than the time predicted by expression (3.2.3).

There are two sources of doubt concerning the validity of this isolated result;

(1) Experience of comparing the results of the numerical technique with the analytic expression has shown that the discrete potential needed to represent a square barrier is not obvious. Since Schnupp had no independent check on his results it is probable that the discrete potential used was an incorrect representation of a square barrier.

(2) The other source of doubt is the applicability of the phasetime expression to the system being studied. These doubts cannot be expressed quantitatively because of a lack of information. However, the barrier is 15 Å thick with the spread of the Gaussian wavepacket being of the same order. This localisation of the spatial distribution will de-localise the momentum distribution, whilst the derivation of the phasetime expression requires a momentum distribution localised at a specific value.

It would therefore be acceptable to conclude that the disagreement observed by Schnupp was due to either a bad representation of the potential, a system for which the phasetime expression cannot be expected to be valid or both.

Schnupp's work shows that to compare the analytic expression (3.2.3) and numerical results isolated values are insufficient, the functional dependence of the numerical results must be examined. For this reason the

energy dependence of the numerical results was obtained for several potentials.

The definition of traversal time used to obtain the numerical results was the same as that used in obtaining the phasetime expression, which required the correlation of the transmitted peak position and the initial peak position. To evaluate the transmitted wavepacket, to find the peak, expression (2.3.1) was used to avoid any doubt as to the validity of the representation of the potential. Since it was desirable not to assume the applicability of the phasetime result, the expression for the transmitted wavepacket was evaluated at a time when the peak was well away from the potential. This allowed the wave function to be evaluated over a large region which ensured that the peak of the wave function was found. The traversal time was then obtained assuming that the peak had been moving with the incident velocity. Although this assumption gives only a small error in the velocity of the transmitted peak, the fact that the traversal time was small compared to the total time meant that large errors were produced in the traversal time.

The accuracy of the numerical results was therefore improved by using the first result as an estimate for the time at which to evaluate the wave function much closer to the potential. This method was found to give good results which confirmed the validity of the phasetime. Once a few results had confirmed that the phasetime result was at least a good estimate, the phasetime expression itself was used to give the estimate for the search close to the potential.

The results obtained for two systems are shown in figures (3.1) and (3.2), with the analytic results of Buttiker and the dwell time being plotted for

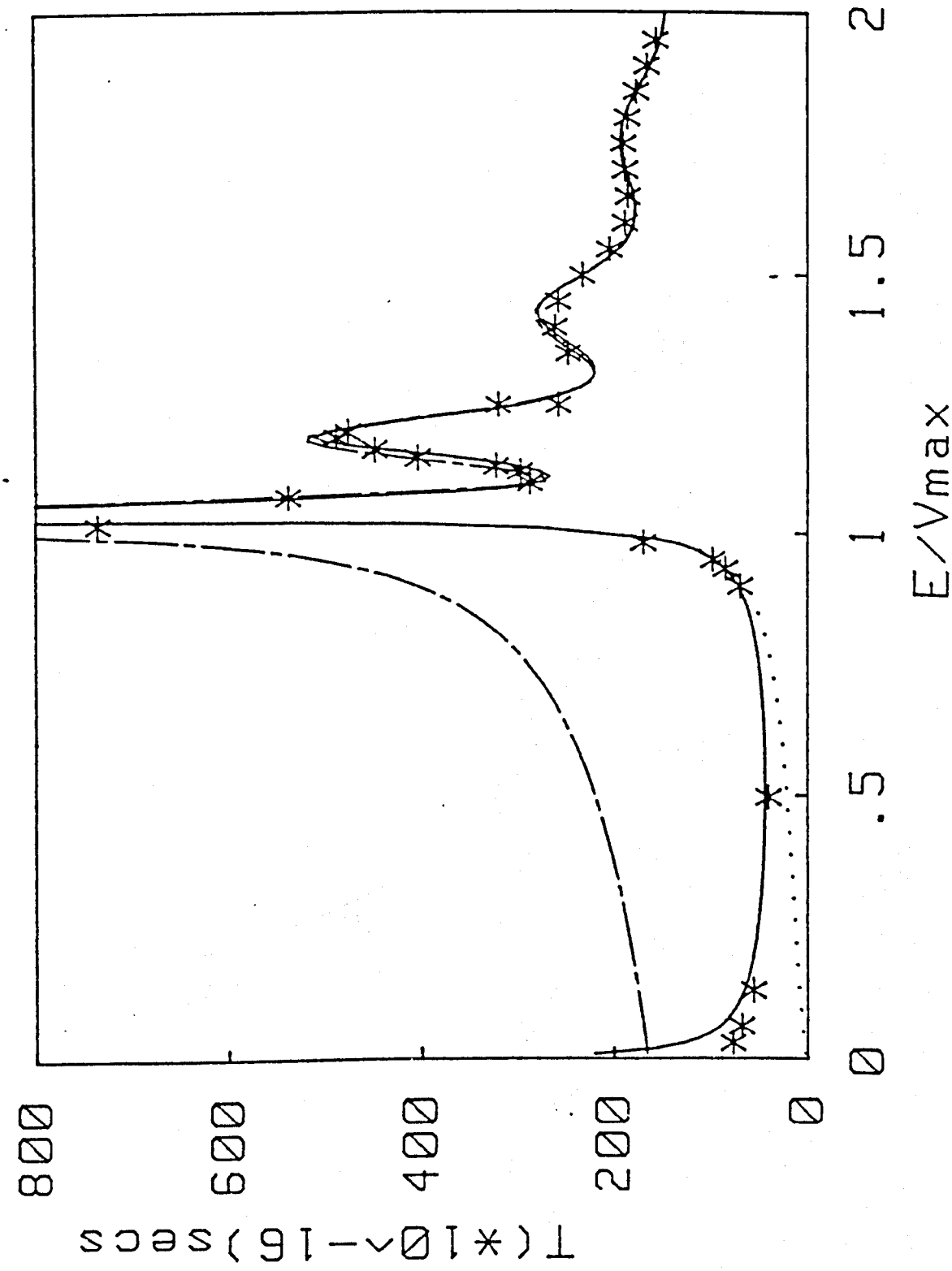

Figure 3.1 Characteristic times for a 200Å, 0.3 V square barrier

-.- Buttiker's traversal time

—— Phasetime

... dwell time

xxx Numerical results for a Gaussian wavepacket with a 0.3 μm standard deviation

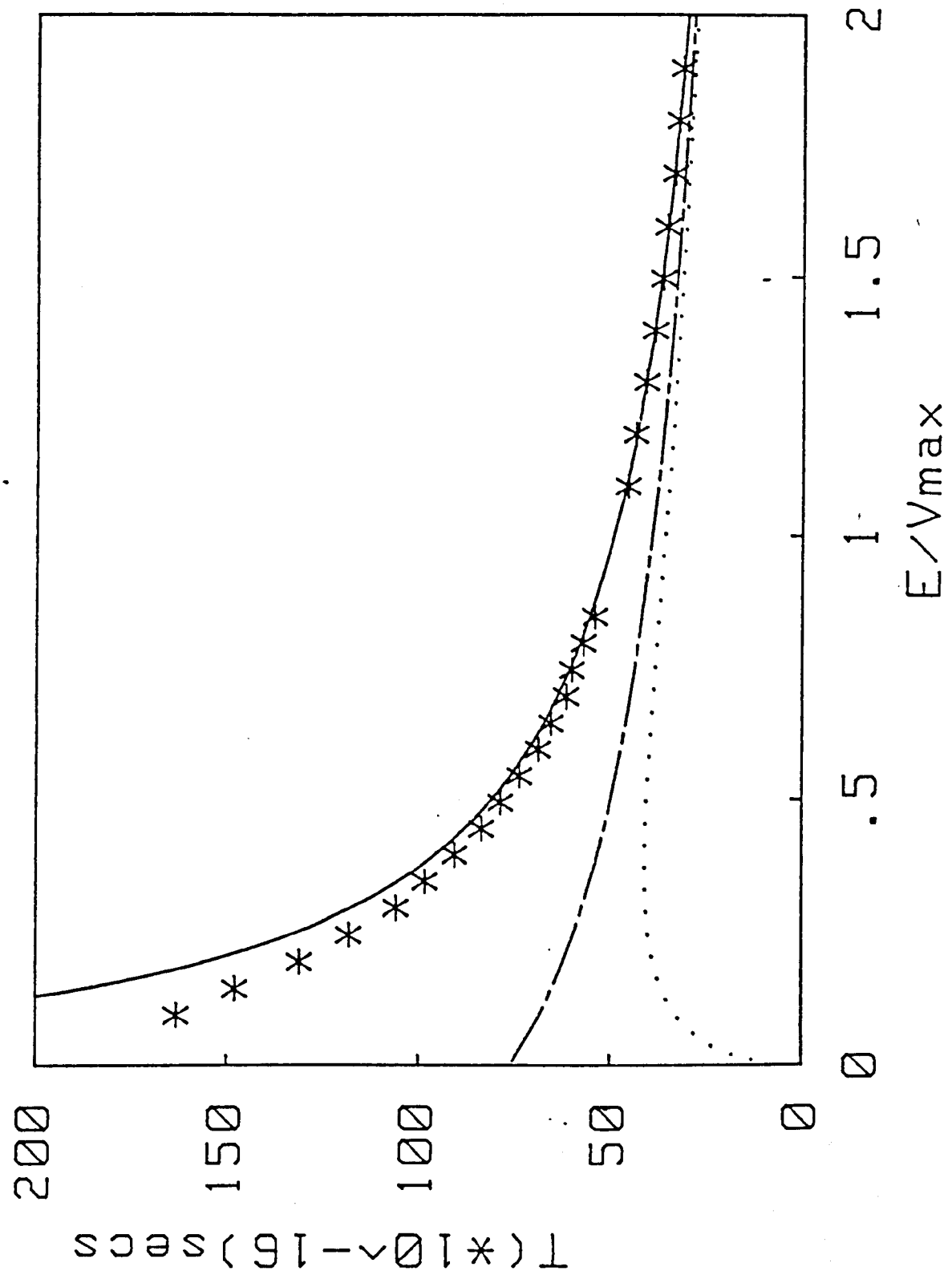

Figure 3.2 Characteristic times for a 25Å, 0.1 V square barrier
-.- Buttiker's result
—— Phasetime
... dwell time
xxx Numerical results for a Gaussian wavepacket with a 0.3 μm standard deviation

comparison. Several points emerge;

(1) All the results tend to the classical traversal time

$$t = a\sqrt{m/2(E-V)}$$

above the barrier.

(2) The results show that as the energy increases the agreement between the numerical results and the analytic expression also increases. This is due to condition (3.2.2). The rate of change of the transmission coefficient is inversely proportional to the energy. Therefore for the fixed standard deviation, 0.3 μm was used, the higher the energy the better the expected agreement, as observed.

(3) Tunnelling is a fast process. This means that scattering during tunnelling is expected to be rare except for ion scattering and phonon emission. These two processes may still occur because the probability of ion scattering is related to ionic position, whilst the high energies which can be gained from the field means phonon emission may occur very rapidly.

3.5 Double barrier systems

The transmission coefficients show a resonance which could be exploited if resonant tunnelling is a sufficiently fast process to allow high frequency operation. To determine the time taken for resonant tunnelling consider the wavefunction beyond the barrier

$$\psi(x,t) = \int dk\ t(k)f(k)\ \exp\left[ikx-i\omega t\right]$$

where the transmission coefficient at resonance is

$$t(k) = \frac{1}{(E(k) - E_r) + i\Gamma/2}$$

Messiah (39) has shown that, by assuming a parabolic E-k relationship and a fourier distribution which is approximately constant over the range of the resonance, the transmitted wave function is

$$\psi(x,t) \approx k_r \quad f(Er) \exp\left[i(k_r x - \omega_r t) - \Gamma(t-x/v_r)/2\hbar\right] \qquad (3.5.1)$$

If this is evaluated near to the barrier it shows that the wave function in the barrier is decaying exponentially with a decay constant for the probability density of $\hbar/\Gamma$. This situation can be simulated using the general numerical technique discussed in section 2.5. The result, shown in figure (3.3), is the time dependence of the total wave function in the barrier when a Gaussian wavepacket with a standard deviation of 120 Å is incident upon two 0.23 eV, 50 Å barriers separated by 50 Å at the resonant energy of 80 meV.

The evolution of this system showed that after approximately 80 fs there is a stationary peak in the centre of the barriers and only 0.18% of the wave function had been transmitted. The reflected pulse emerged after 140 fs to leave a quasi-bound state in the barriers and a transmission probability of 0.37%. After this time the quasi-bound state decays, half of the decaying probability going to each side of the barrier. Since the quasi-bound state contained approximately 3% of the wave function after 140 fs this means that the total transmission probability which the simulation predicted was approximately 1.9%. This compared favourably with the theoreticl value of 2%, especially when it is noted that the simultation had to be extended by deleting the wavepacket reflected of the

boundaries of the simulation, to prevent interference from the boundaries.

A close examination of the results shown in figure (3.3) shows an erratic transient followed by a smooth decay. The transient is caused by the penetration of the impacting wave function, with the initial rapid decay being caused by the withdrawal of this penetrated wave function as it is reflected. The smooth decay can be very satisfactorily fitted using the decay constant corresponding to the width of the resonance, 0.65 ps, and fitting the amplitude to a point well away from the transient, conclusively verifying (3.5.1).

Another situation arises if the momentum distribution is considered to be very narrow compared to the resonance. Then the phasetime expression obtained in section 3.1 can be used to give a traversal time

$$T = \hbar \frac{\partial \Phi}{\partial E}$$

The phaseshift at resonance is

$$\Phi = \tan^{-1} \frac{\Gamma}{2(E-E_r)}$$

giving a final expression

$$T = \frac{2\hbar}{\Gamma} \quad \frac{(\Gamma/2)^2}{(E-E_r)^2 + (\Gamma/2)^2}$$

which peaks at resonance to $2\hbar/\Gamma$, twice the decay constant, and is Lorentzian in shape.

A simulation of this situation for the previously used potential would be prohibitively lengthy. To show that a different dynamical picture emerges consider a Gaussian wavepacket with a spatial standard deviation σ_x, which has a sharp fourier spectrum compared to the resonance

$$\sigma_x > 85 \text{ nm}$$

In the traversal time the impacting pulse has travelled a distance of approximately 100 nm. This means that as the peak reaches the front of the barrier the point which was initially one standard deviation in front of the peak is just about to emerge. Therefore at no time will it be possible to define a reflected pulse, a quasi-bound state and little transmission. The transmitted and reflected pulses will emerge almost together, a picture which is confirmed by simulations on other system.

Two dynamical picture emerge of resonant tunnelling. If the momentum distribution of the wavepacket is large compared to the width of the resonance, the fourier components of the quasi-bound state are all present in the incident pulse and the quasi-bound state is formed. A situation is identifiable in which a quasi-bound state is isolated in the barrier region and this decays exponentially, with equal probability in each direction. This situation can be described in hopping terminology. The incident electron has a small probability of hopping into the energy level of the well from which it can either hop back or forward. If it hops forward it contributes to the current across the barrier.

If the momentum distribution is narrow compared to the resonance then no quasi-bound state is formed. The dynamic picture is one in which the transmitted and reflected pulse emerge at approximately the same time and a hopping description cannot be used.

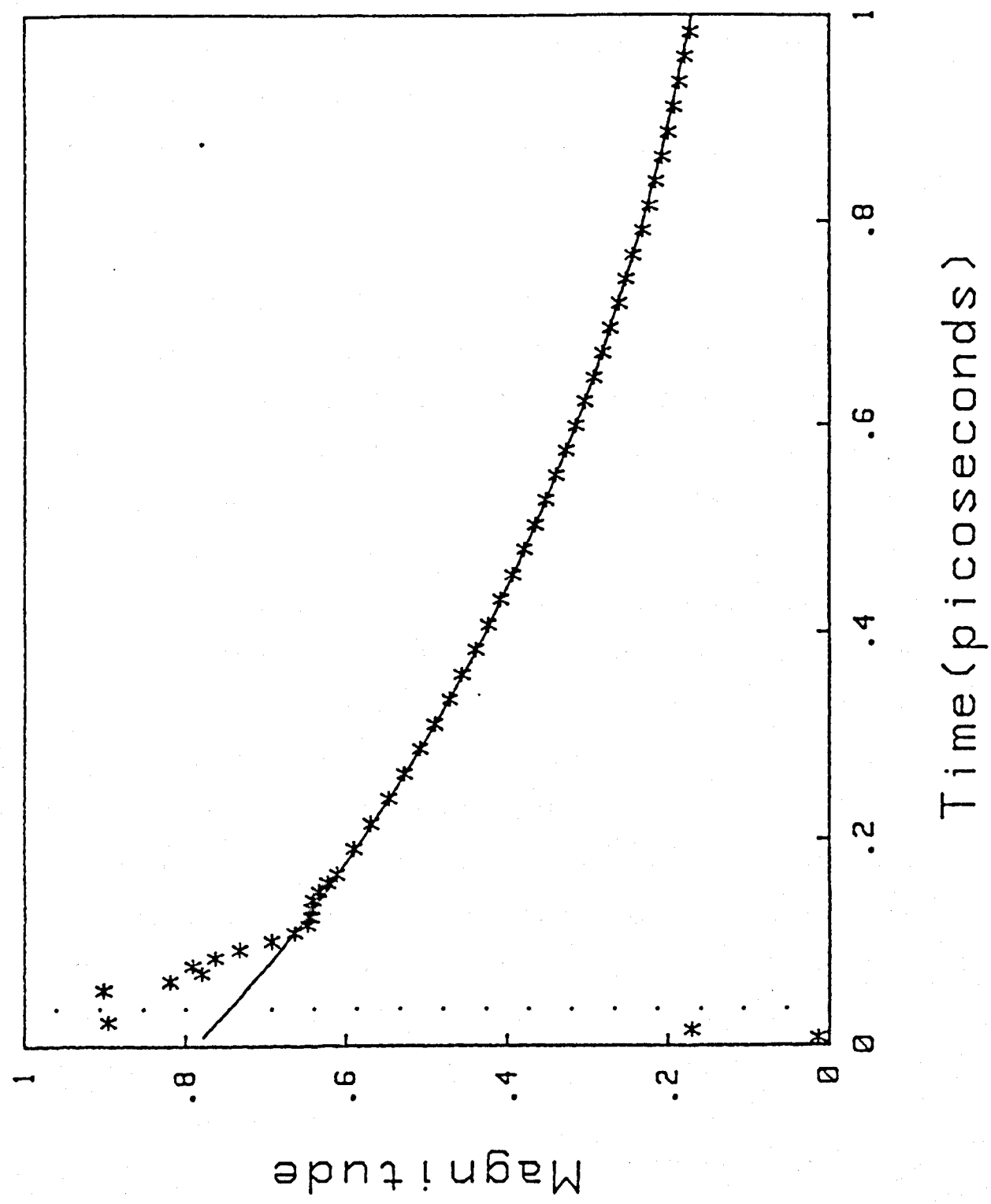

Figure 3.3 Magnitude of the wave function in a double barrier system as a function of time

x Numerical results

— Exponential decay with a decay constant of 0.65 ps

.. Time at which wavepacket peaks impinged on the barrier

In the above the momentum distribution has been considered to be the variable. In real systems it is the width of the energy level which is most easily adjusted. The above discussion would then lead to the expectation that for a fixed electron environment, momentum distribution, it is the narrow energy levels which give rise to conduction via a "hopping" mechanism.

These results contradict the time dependence of resonant tunnelling outlined by Ricco and Azbel (66). This picture arose from a consideration of the solution to the time independent S.E. and not from a solution of the time dependent equation. The time independent equation shows that the condition for a transmission resonance is identical to the condition for a quasi-bound state. The large occupation probability associated with the quasi-bound state is therefore assumed to be critical in allowing resonant tunnelling. If this is true and the quasi-bound state is initially empty, the first electrons cannot resonantly tunnel. Only a few lucky electrons are transmitted by the first barrier into the central region, where they are trapped by the same process that forms the quasi-bound state. The number density in the barrier gradually increases with time until it is so great that the leakage through the second barrier gives rise to a current equal to the incident current. Current continuity then demand that no reflection is taking place, and resonant transmission has been established.

This picture has been used by Luryi (77) to study the time required to set up a resonance. The approach used was to consider one contact and the central well as two plates of a capacitor being charged by the leakage current across a barrier. This gives rise to an RC time constant

$$t = \frac{2\pi\epsilon}{\alpha c k} \exp\left[2kd\right]$$

where k is the electron wavevector, ϵ and d are the permittivity and thickness of the barrier and α and c are universal constants. This gave an estimated time before resonance could occur which set the maximum possible frequency of operation of resonant tunnelling much less than the frequency observed by Sollner et al. The conclusion was therefore reached that the peak observed by Sollner was due to a mechanism other than resonant tunnelling.

The results of the time dependent equation show that the Ricco and Azbel picture is not consistent with the solution of the time dependent equation. Resonance tunnelling has been observed for initial conditions which were zero in the barrier. In resonant tunnelling, with a large transmission probability, the transmitted wavepacket emerges at most a few picoseconds after the incident pulse impacts upon the barrier, in agreement with the frequency performance obtained by Sollner et al.

The Ricco and Azbel picture of resonant tunnelling requires the first part of the wavepacket to establish the conditions for resonance. A wavepacket would therefore suffer a large amount of reflection before the conditions for resonance were established, a point which is impossible to reconcile with the total transmission probability of

$$T = \int dk |f(k)|^2 |t(k)|^2$$

which implies that the whole wavepacket interacts with a system which is characterised by a resonant transmission coefficient, whilst Ricco and

Azbel require different components of the wavepacket to interact with the double barrier system or isolated single barriers depending upon the conditions at the time of impact.

Another approach to the characteristic times of a resonant system has been taken by Stevens (48). Stevens studied a model system consisting of one square barrier thickness b a distance a from a step potential, a system chosen as the simplest which showed the required resonant transmission coefficient. If the energy of the incident pulse is much less than the barrier height and very close to the energy of the lowest resonance then the pole due to the resonance gives a contribution which decays in a characteristic time

$$T = \frac{m^2 a^4 V}{2\pi^3 \hbar^3} \exp[\, 2b(2mV/\hbar^2)^{1/2}] \qquad (3.5.1)$$

During the derivation of the above expression Stevens used the approximations

$$ka = \pi$$

and

$$\kappa^2 = 2mV/\hbar^2$$

If these are used in (3.5.1), then it can be written

$$T = \frac{\kappa^2}{4k^2} \frac{ma}{\hbar k_o} \exp\,[2\kappa b]$$

This can be further simplified if the transmission coefficient for a square barrier is used

$$T = \frac{4\kappa^2 k^2}{4\kappa^2 k^2 + (\kappa^2 + k^2)^2 \sinh^2 \kappa b}$$

which with the approximation $\kappa > k$, $\kappa b > 1$ becomes

$$T = \frac{16k^2}{\kappa^2 \exp[2\kappa b]}$$

The expression for τ is then

$$\tau = \frac{2m\dot{a}}{\hbar k_o} \quad \frac{2}{T}$$

The first part of the expression is recognised as the time for the electron/wavepacket to travel a distance of 2a and the inverse of the transmission coefficient is the average number of "attempts" needed for an electron to escape. The factor of 2 arises because the time produced by the number of bounces multiplied by the time between bounces is characteristic of $|\psi|^2$, so that considering ψ as Stevens does produces a factor of 2.

The Stevens time can therefore be identified not as a traversal time, but as an escape time from the region between the barrier and the step.

3.6 Semi-classical model

One possible use of the traversal time expression is to generate a semi-classical model of tunnelling which could reproduce some of the dynamics of the process. The wavepacket will be represented by a distribution of particles which evolve classically, their distribution at time t modelling the probability distribution at t.

Consider a wavepacket incident upon a square barrier. The simplest possible model would be to replace the barrier by a partial reflector at one point and delay the transmitted and reflected components by the time required to establish the correct peak positions. This approach cannot be

used, because, the analysis used to produce the traversal time expression cannot be used on the reflected wavepacket, so that there is no analytic expression for the delay which would need to be imposed on the reflected component.

The next simplest model replaces the barrier by a partial reflector on each side. This model reproduced the three regions of space present in the full system, and the symmetry of the potential means that each reflector is identical.

The model assumes that each particle evolves classically up to the first reflector. Those reflected move away at the incident velocity whilst those which are transmitted move in a finite time T to the other reflector. Here they are again transmitted or reflected, the transmitted component moving away with the classical velocity and the reflected component returning the first reflector. If the individual reflectors have a reflection probability r then the total transmission and reflection probabilities T and R are

$$T = \frac{1-r}{1+r}$$

$$R = \frac{2r}{1+r}$$

The parameter r being chosen so that T is equal to the transmission probability for the plane wave representing the same momentum.

Then the average time $< T >$ taken for a particle which is eventually transmitted to cross the barrier is

$$<T> = \frac{T(1+r^2)}{(1-r^2)} \qquad (3.6.1)$$

This particle model will not be able to reproduce the oscillations on the incident side of the barrier which are caused by interference or the decay of the wave function in the barrier. The model must therefore be judged on its ability to reproduce the transmitted and reflected wavepackets.

The sample used to represent the initial condition used 900 points spread evenly over the region within two standard deviations of the mean position or momentum, each point being the initial condition for 50 particles. Each particle was added into the final distribution with a weight proportional to the probability of the initial position.

Initially the phasetime expression (3.2.3) was used to give the value of τ directly, however the results produced were unsatisfactory. Much better results were then obtained by using the phasetime expression to define the average time in expression (3.6.1) and hence τ. The results obtained using this definition of τ for a minimum Gaussian wavepacket with a standard deviation of 300 Å and an energy of 0.15 eV incident upon a 50 Å, 0.2 eV barrier are shown in figure (3.4). The result is promising with a model transmission probability of 0.1639 compared to 0.1640 obtained by solving the time dependent S.E. and the wavepackets peak positions reproduced to a few nanometers.

The model could be improved. The use of two reflectors always produces a peak between the two, a feature which does not necessarily appear, and fails to reproduce the decay of the wave function. To be more accurate the model must be modified. Probably the best approach would be to model the barrier as a region in which the particle has a finite probability of being reflected whilst travelling between two points. This would reproduce the decaying wave function, however there are difficulties in

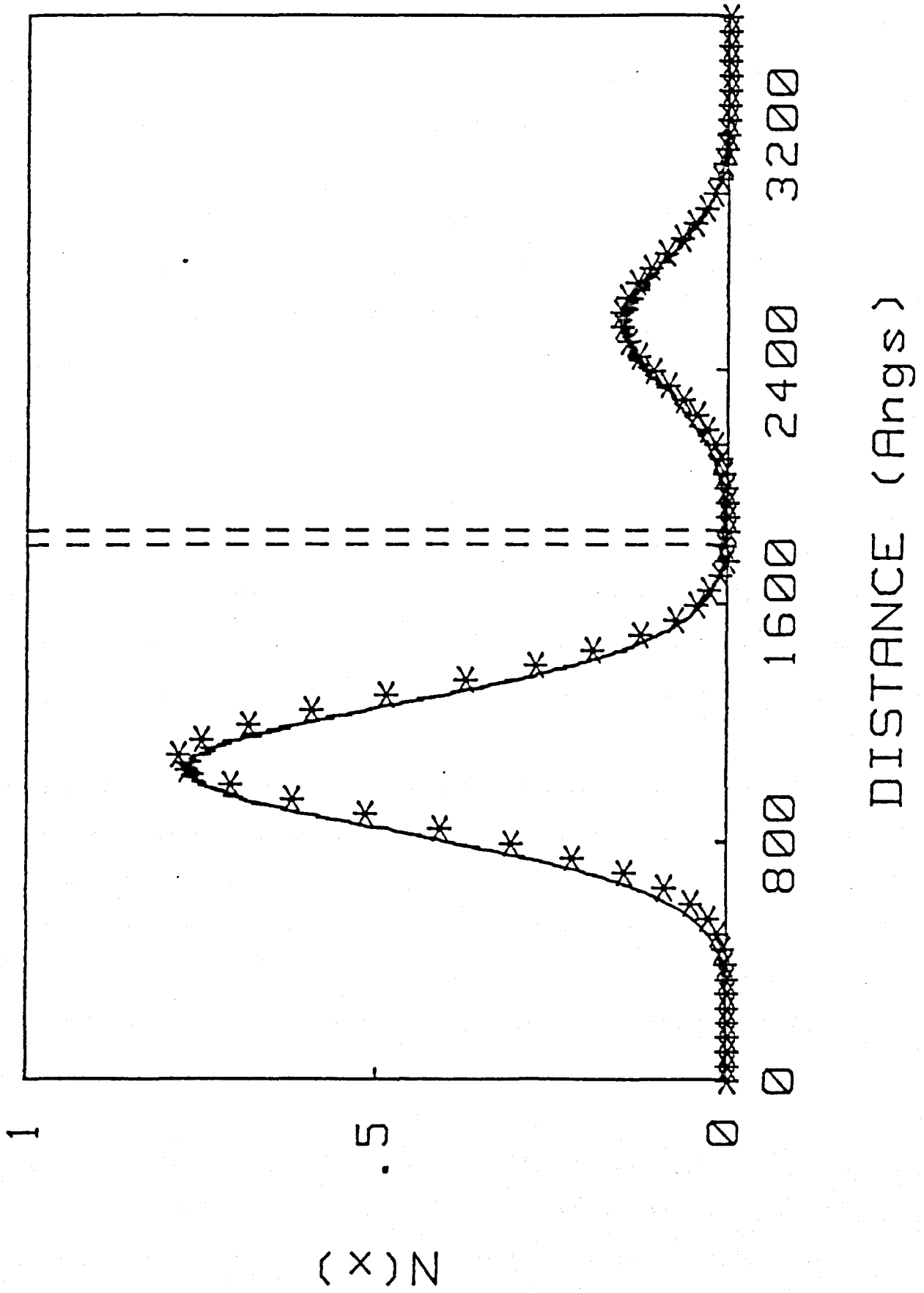

Figure 3.4 Comparison of the semiclassical model (—) and the numerical solution to the Schrodinger Equation (*) for a Gaussian pulse, σ_x = 300Å, with an energy 0.15 eV incident upon a 50Å, 0.2 eV square barrier.

assigning the time taken to travel between two points, a difficulty which may be overcome using the Stevens result (3.3.1).

Despite these difficulties the outstanding result of this simplistic model is an indication that the result of the phasetime expression should be interpreted as the average time required for an electron to be transmitted.

3.7 Summary

In this chapter it has been shown that the traversal time across a barrier is related to the derivative of the phaseshift. The exact relationship depends on the phase referencing of the plane waves used in the definition of Φ, so that this should also be defined when stating the relationship.

The result obtained by Stevens for a square barrier has been shown to be exact for a step potential. It cannot be extended to a square barrier because of the neglected second component of the wave function in the barrier, which, although small, is important. In contrast the result obtained by Buttiker has been shown to rest on an incorrect analogy which means that a physically meaningless time is purposed as the traversal time.

In a study of double barrier systems, it was found that hopping and resonant tunnelling are the same process described by different dynamical pictures. The transmission resonances were found to be associated with resonances in the traversal time, with a traversal time inversely proportional to the resonance width. The characteristic time produced by Stevens from the analysis of a system with a quasi-bound state was found

to be the semiclassical time obtained by multiplying the classical time between collisions with the confining barrier with inverse of the transmission probability. The picture proposed by Ricco and Azbel for time dependent resonant tunnelling was shown to be contrary to the solution of the time dependent S.E. This contradiction cast considerable doubt on the conclusion of Luryi that resonant tunnelling could not be the mechanism causing the resonant current observed by Sollner et al.

A simple Monte Carlo model indicated that the traversal time should be considered as an average time that a transmitted electron spends in the barrier.

The most important result is that tunnelling is a very rapid process measured in femto seconds. The probability of scattering during traversal is therefore negligible. The only affect of the scattering processes usually associated with electron transport, will therefore be to shape the electron distribution in the contacts. The only exceptions to this rule could be ionised impurity scattering from ions in the barrier and phonon emission due to the high energies gained from the field. The first is expected to cause difficulties for systems which are doped in the barrier region whilst the second is neglected.

THE TUNNEL CURRENT AND THE CALCULATION OF THE TRANSMISSION COEFFICIENT

4.1 Introduction

In the previous two chapters it has been demonstrated that the behaviour of a wavepacket incident upon a barrier can be predicted by considering the fourier components of the initial condition and, that a wavepacket traverses a barrier in a time which is short compared to the mean free time. Both these results are used in the first section of this chapter to derive a heuristic current density expression, the first result being used to justify considering plane waves incident upon the potential and the second the neglect of scattering events. The result of this first section is a current density expression which shows that for any current voltage characteristics to be calculated a general method of calculating transmission coefficients is required.

Two general methods are discussed in the following two sections. The first method is the analytic WKB method, which is sufficiently well known that only an outline of particularly relevant results will be given. The second method is a numerical technique first used by Vignernon and Lambin. This method will be outlined in some detail, because the results of section 4.6 show that it is far more accurate than the WKB method and is therefore adopted for the current voltage described in Chapter 5.

Before this conclusion is reached some exact analytic results will be obtained against which to compare the accuracy of the two general methods. The results are obtained for a system of two square barriers and, since this is a good model for fabricaaed layer systems (4,5), the analysis will be extended to include the ability to predict the dependence of the

resonant energy and resonance width on any change of system parameter.

After the accuracy of the numerical technique has been demonstrated the last two sections of the chapter extend the same technique to calculate the phase difference between the incident and transmitted waves, and the bandstructure of a superlattice. The first of these will give an ability to calculate the traversal time for a general potential, whilst the second confirms the accuracy of the simple model derived in Chapter 1 and can be used to demonstrate that hopping may be the dominant transport mechanism at high temperatures for superlattices with long periods.

4.2 Expression for the tunnel current

In this section a heuristic expression for the tunnel current through a potential will be derived. The results of section 3.2 showed that the behaviour of a wavepacket could be calculated by considering the fourier components of the initial condition independently before integrating to obtain the transmitted wavepacket. This approach is adopted in this section. Consider the situation shown in figure (4.1) where the 1D potential implies that the plane waves can be specified by a total energy E, and the wavevector in the plane normal to the z axis, k_t.

The simple single particle picture adopted here means that the current across the system can be decomposed into a component due to electrons incident upon the potential from each contact. Current continuity then means that the total current and each contribution can be calculated at any point. This is used to simplify the derivation, by allowing the evaluation of the contributions from the collector and emitter at points A and B respectively.

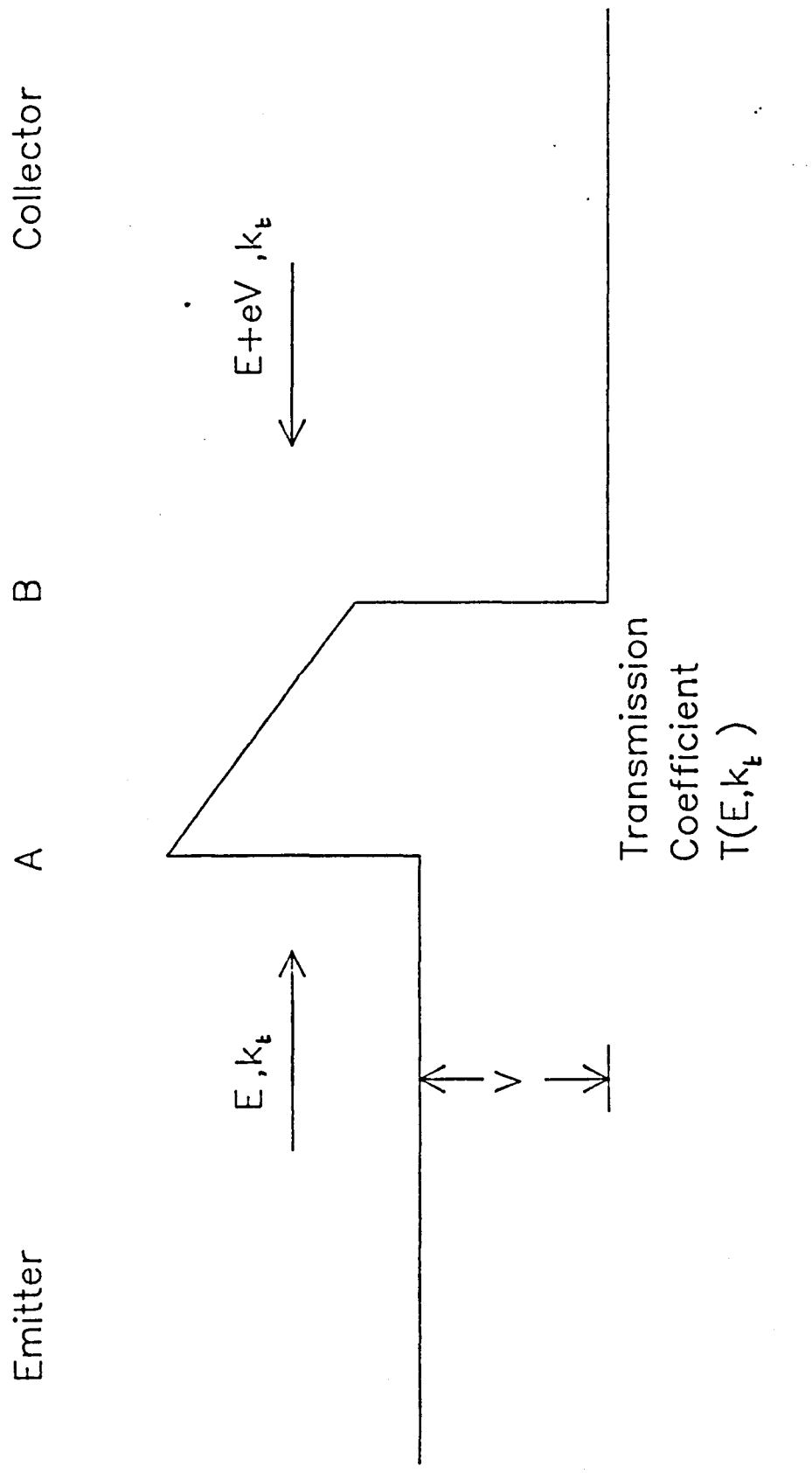

Figure 4.1 The system considered in the derivation of the heuristic current density expression.

First consider the contribution from plane waves incident from the emitter. Plane waves of energy E and transverse wavevector k_t have a statistical weight $f_e(E,k_t)$ and give rise to a total incident current from a unit volume of k space of

$$j_i = \frac{2}{(2\pi)^3} f_e(E,k_t) \frac{e}{\hbar} \left(\frac{\partial E}{\partial k_z}\right)_{E,k_t}$$

with the factor 2 accounting for spin and $(2\pi)^3$ for the density of states.

To contribute to the transmitted current the plane wave must not only couple to a state on the other side of the potential, but, because the electrons are fermions, that state must also be empty. If the transmission coefficient is defined as the ratio of transmitted to incident current, the transmitted current from this state is

$$j_1 = \frac{2e}{(2\pi)^3\hbar} \left(\frac{\partial E}{\partial k_z}\right)_{E,k_t} (1-f_c(E+eV,k_t))f_e(E,k_t)T(E,k_t) \qquad (4.2.1)$$

A similar analysis can be used to obtain the contribution from the plane waves incident from the collector

$$j_2 = \frac{2e}{(2\pi)^3\hbar} \left(\frac{\partial E}{\partial k_z}\right)_{E+eV,k_t} (1-f_e(E,k_t))\, f_c(E+eV,k_t)T(E,k_t) \qquad (4.2.2)$$

The total current can now be obtained by taking the difference between the two contributions and integrating over wavevector. To combine the contributions in this way the transmission coefficient in both contributions must be the same, this means the transmission coefficient defined as a ratio of currents must be used (39). After performing the

polar integration the total current density J is given by

$$J = \frac{2e}{(2\pi)^2\hbar}\int_0^\infty\int_0^\infty k_t dk_t dk_z T(E,k_t)\left[\left(\frac{\partial E}{\partial k_z}\right)_{E,k_t} f_e(E,k_t)(1-f_c(E+eV,k_t))\right.$$

$$\left. -\left(\frac{\partial E}{\partial k_z}\right)_{E+eV,k_t} f_c(E+eV,k_t)(1-f_e(E,k_t))\right] \qquad (4.2.3)$$

This is not the usual expression which is used, and has become known as the Tsu-Esaki Formula (2), which is

$$J_{TE} = \frac{2e}{(2\pi)^2\hbar}\int_0^\infty\int_0^\infty k_t dk_t dk_z\ T(E,k_t)\left(\frac{\partial E}{\partial k_z}\right)_{E,k_t}(f_e(E,k_t)-f_c(E+eV,k_t))$$

This can be recovered by assuming that the two derivatives are equal, an assumption which appears to slip into the derivation of J_{TE} when (4.2.2) is written down as analogous to (4.2.1) without realising that the incident velocities are different.

To enable comparisons to be made between the current expression (4.2.3) and the more usual form re-write (4.2.3) as

$$J = J_{TE} + \Delta J \qquad (4.2.4)$$

where

$$\Delta J = \frac{2e}{(2\pi)^2\hbar}\int_0^\infty\int_0^\infty k_t dk_t dk_z\ T(E,k_t)f_c(E+eV,k_t)(1-f_e(E,k_t))$$

$$\left[\left(\frac{\partial E}{\partial k_z}\right)_{E+eV,k_t} - \left(\frac{\partial E}{\partial k_z}\right)_{E}\right]$$

The expression for ΔJ shows that this correction term tends to zero for both high and low voltages. At high voltages the term $f_c(E+eV,k_t)$ must

tend to zero as V tends to infinity to maintain a finite energy density whilst at low voltages the two derivatives become equal.

To simplify the current expression three assumptions can be used. The first two, that the conduction bands are parabolic, and that the mass is position independent, make both the transmission coefficient and the derivative independent of k_t, leaving all the dependence on k_t in the occupation probabilities. The third assumption, that the occupation probabilities are Fermi functions, allows the k_t integral to be performed leaving a one dimensional integral.

If the Fermi distributions in the emitter (collector) are characterised by Fermi energies $\mu_e(\mu_c)$ and electron temperatures $T_e(T_c)$ the transverse integral can be performed and the remaining integral changed to an integral over longitudinal energy E_ℓ to give

$$J_{TE} = \frac{2mk}{(2\pi)^2\hbar^3} \int_0^\infty dE_\ell T(E_\ell) \left[T_e \ln(1+\exp(\beta_e(\mu_e-E_\ell))) \right.$$
$$\left. -T_c \ln(1+\exp(\beta_c(\mu_c-E_\ell-eV))) \right]$$

This expression reduces to the more usual form when the electron temperatures T_e and T_c are equal so that

$$J_{TE} = \frac{2mkT_e}{(2\pi)^2\hbar^3} \int_0^\infty dE_\ell \; T(E_\ell) \; \ln\left[\frac{1+\exp(\beta(\mu_c-E_\ell))}{1+\exp(\beta(\mu_c-E_\ell-eV))} \right] \qquad (4.2.5)$$

where

$$\beta = \frac{1}{kT}$$

Using the same assumptions and assuming equal electron tempertures, the correction term can be simplified to give

$$\Delta J = \frac{2mkT}{(2\pi)^2\hbar^3(1-\exp[\beta(\mu_e+eV-\mu_c)])} \int_0^\infty dE_\ell \; T(E_\ell) \left[\frac{(E_\ell+eV)^{\frac{1}{2}}-(E_\ell)^{\frac{1}{2}}}{(E_\ell)^{\frac{1}{2}}} \right] \ln \left[\frac{1+\exp(\beta(\mu_c-E_\ell))}{1+\exp(\beta(\mu_c-E_\ell-eV))} \right] \quad (4.2.6)$$

A form which shows that the correction is small if $V > kT/e$ (where kT/e is of the order of tens of millivolts).

The final full expression for the current density which is a combination of (4.2.5) and (4.2.6) contains the function

$$D(E_\ell,\mu_c,\mu_c,T) = \ln \left[\frac{1+\exp(\beta(\mu_e-E_\ell))}{1+\exp(\beta(\mu_c-E_\ell-eV))} \right]$$

which is usually termed the supply function and $T(E_\ell)$ the transmission coefficient calculated as the ratio of currents.

The final requirement before any current calculations can be performed is for a method of calculating the transmission coefficient of a general potential.

4.3 General analytic technique

The requirement for a general method for calculating transmission coefficients could be fulfilled using the WKB approximation. Because a full description of this approximation can be found in many standard

texts only the results of special interest, that is systems of more than one potential barrier, will be discussed.

Bohm (3) used the WKB approximation to study the transmission coefficient for a system of two identical barriers and obtained a transmission coefficient.

$$T = \left[1 + (4\Theta^2 - \frac{1}{4\Theta^2}) \sin^2 \left[\tfrac{1}{2}(\pi - J/\hbar) \right] /4 \right]^{-1}$$

The transmission coefficient is equal to unity if

$$J = (N+\tfrac{1}{2})\hbar$$

a condition identical to that for a metastable state of the system.

The metastable state can be considered as being due to coherent interference of the wave function with itself as it is reflected from the inside of each component of the system. The resonance is then considered to be due to the constructive interference of the incident wave function as it penetrates into the region between the barriers with the wave function already present. This constructive interference leads to an increase in the wave function in the potential system until the current transmitted through the back of the system is equal to the incident current. This is the picture used by Azbel to discuss the time dependence of resonant tunnelling, as discussed in Chapter 3.

A system of two dissimilar barriers has been treated by Hrach (74). The transmission coefficient has the form

$$T = 1/ (1+s^2+\sigma^2)$$

This shows that for two dissimilar barriers the transmission coefficient

is unity if;

(i) $S^2 = 0$, a condition which is fulfilled only if the barriers have the same transmission probability for the energy under consideration.

(ii) $\sigma^2 = 0$ the same condition as obtained by Bohm.

The final result of interest is the WKB result for a system of N square barriers obtained by Pshenichnov (47).

$$T = [\cos^2 N\Phi + \sin^2 N\Phi \sin^2(J/2\hbar)(\theta + 1/4\theta)^2/\sin^2\Phi]^{-1}$$

where $\cos \Phi = \left(\theta + \frac{1}{4\theta}\right) \cos (J/2\hbar)$

This will be used to check the accuracy of the WKB approximation.

4.4 General numerical technique

In this section a detailed description of the numerical technique for calculating the transmission coefficient for a general potential is given for two reasons;

(1) Unlike the WKB technique, the numerical technique is not widely known and has only previously been discussed in the paper in which it was first developed (49).

(2) Presuming the conclusion of section 4.6 the numerical technique will be used in the remainder of the thesis.

The technique is based on the definition of the wave function on a lattice, with the wave function on the j^{th} point being designated as $\psi(j)$. A discrete equation is required which emulates the behaviour of the S.E.

The most general form of the time independent S.E. is

$$p(z)\,\frac{\partial^2\psi}{\partial z^2} + q(z)\,\frac{\partial\psi}{\partial z} + r(z)\,\psi = 0$$

The equivalent difference equation, for a lattice with spacing Δ, is obtained by replacing the differentials by equivalent differences obtained from a Taylor expansion about the point j. The result is the difference

$$\left(\frac{p(j)}{\Delta^2} + \frac{q(j)}{2\Delta}\right)\psi(j+1) + \left(r(j)-\frac{2p(j)}{\Delta^2}\right)\psi(j) + \left(\frac{p(j)}{\Delta^2}-\frac{q(j)}{2\Delta}\right)\psi(j-1) = 0$$

where p(j), q(j) and r(j) are the functions of p,q and r evaluated at point j.

For convenience this equation can be written in the form

$$l(j)\,\psi(j+1) + m(j)\,\psi(j) + n(j)\,\psi(j-1) = 0$$

which, for a region of constant coefficients, can be reduced to a quadratic equation by assuming a wave function of the form

$$\psi(j) = u^j$$

The solution of the quadratic equation then shows that u must take one of two forms

$$u^{\pm} = \frac{-m\pm\sqrt{m^2-4ln}}{2l}$$

The significance of the two forms can be found by inserting them into the expression for the current in the discrete system (49),

$$J(j+\tfrac{1}{2}) = -2\,\left|\frac{\psi(j+1)}{\Delta}\right|^2 \mathrm{Im}(\psi(j)/\psi(j+1))$$

where Im(a) is the imaginary part of a. Each form is found to carry current in one direction, so that u^+ and u^- can be identified as the discrete equivalents of plane waves.

The transmission coefficient can now be defined by considering the situation shown in figure (4.2), a situation equivalent to that for the differential equation when a plane wave, amplitude A, is incident upon the potential.

The reflection coefficient can be defined as

$$R = \frac{|B|^2}{|A|^2}$$

and an expression for R obtained by considering the quantity

$$N(-1) = \frac{\psi(-1)}{\psi(0)} = \frac{Au_1^{+}+Bu_1^{-}}{A+B} \qquad (4.4.1)$$

which can be rearranged to give

$$R = \left| \frac{N(-1)-u_1^{-}}{N(-1)-u_1^{+}} \right|^2$$

The calculation of R is thus reduced to the calculation of the quantity N(-1), which contains all the information concerning the potential.

Figure 4.2 The analogy between the solutions of the discrete and differential equations.

To calculate N(-1) the difference equation (4.4.1) can be arranged to give

$$N(j-1) = \frac{-m(j)}{n(j)} - \frac{\ell(j)}{n(j)\,N(j)}$$

An equation which can be solved using the continued fraction techniques described by Wall (68), with the boundary condition

$$N(j) = u_j^+ \qquad j > J$$

The transmission coefficient is then given by

$$|T|^2 = 1 - |R|^2 \tag{4.4.3}$$

which implies that $|T|^2$ is the transmission coefficient defined as the ratio of transmitted and incident currents rather than the ratio of amplitudes. This can be seen by considering a potential which includes a potential difference across the barrier. The potential difference means that conservation of energy requires a unit incident state, wavevector k to connect to a transmitted state, wavevector k' amplitude t. The continuity of current demands that

$$(1 - |R|^2)k = |t|^2 k' \tag{4.4.4}$$

The definition of $|t|^2$ means that it is the transmission coefficient defined as a ratio of amplitudes and a comparison of (4.4.3) and (4.4.4) shows that

$$|T|^2 = \frac{|t|^2 k'}{k}$$

The definition of $|t|^2$ and the proportionality of velocity to wavevector means that the right hand side of (4.4.3) is a ratio of the transmitted to incident current. As stated, $|T|^2$ is the transmission coefficient which must be used in current calculations.

4.5 Simple analytic model

In the previous two sections two techniques have been described to obtain the transmission coefficient for a general potential. The choice of which technique to employ depends on the accuracy of the techniques which can only be ascertained by comparison with exact results.

In this section the exact transmission coefficient for a system of square barriers will be obtained. This choice of 'test' system was made because the square barrier is one of the simplest potentials for which to solve the S.E., and, using the model in Chapter 1 the square barrier is a model for a layer of material.

The most effective method for solving the S.E. for a system of identical barriers is to obtain the transfer matrix for a single barrier, a method used by Headings (50). The wave function between the r th and (r+1)th barrier was written as

$$\psi = A_r \exp(-ik(z-ra-rb)) + B_r \exp(ik(z-ra-rb))$$

with the transfer matrix defined as

$$\begin{pmatrix} A_{r-1} \\ B_{r-1} \end{pmatrix} = T \begin{pmatrix} A_r \\ B_r \end{pmatrix}$$

This definition of T and phase referencing for the wave function leads to matrix elements for T .

$$t_{11} = \frac{im_1m_2}{2k\kappa}\left[\left(\left(\frac{k}{m_1}\right)^2 - \left(\frac{\kappa}{m_2}\right)^2\right)\sinh\kappa a - 2i\kappa k\cosh\kappa a/m_1m_2\right]\exp(ikb)$$

$$t_{21} = \frac{im_1m_2}{2\kappa k}\left[\left(\frac{k}{m_1}\right)^2 + \left(\frac{\kappa}{m_2}\right)^2\right]\sinh(\kappa a)\exp(ikb)$$

These are sufficient to determine the whole matrix since the reflection symmetry of the potential leads to the relationships

$$t_{11}^{*} = t_{22}$$

$$t_{12}^{*} = t_{21}$$

The transmission coefficient can now be calculated by multiplying together the individual transfer matrices. The diagonalistion of the matrices used in the multiplication leads to results in terms of the parameter defined by

$$\cos\Phi = \cosh\kappa a\cos kb - \frac{\left(\left(\frac{k}{m_1}\right)^2 - \left(\frac{k}{m_2}\right)^2\right)\sinh\kappa a\sin kb}{2\kappa k/m_1m_2} \qquad (4.5.1)$$

for the energies such that

$$\left|\cosh\kappa a\cos kb - \frac{\left(\left(\frac{k}{m_1}\right)^2 - \left(\frac{k}{m_2}\right)^2\right)\sinh\kappa a\sin kb}{2k\kappa/m_1m_2}\right| \leqslant 1 \qquad (4.5.2)$$

The transmission coefficient is then (50)

$$T = \frac{\sin\emptyset}{\sin N\emptyset\sin\emptyset + \sin N\emptyset(\cos\emptyset - T_{22})}$$

from which it can be seen that under the condition

$$\Phi = n\pi/N \tag{4.5.3}$$

the transmission coefficient is unity.

A combination of (4.5.1) and (4.5.3) leads to the eigenvalue equations for the position of these transmission resonances

$$\cosh\kappa a \cos kb - \frac{\left(\left(\frac{k}{m_1}\right)^2 - \left(\frac{\kappa}{m_2}\right)^2\right)}{2\kappa k/m_1 m_2} \sinh\kappa a \sin kb = \cos(n\pi/N)$$

where n is restricted to values such that

$$1 \leqslant n < N$$

This eigenvalue condition only applies in the energy intervals for which condition (4.5.2) holds, and these intervals are separated by others in which no resonance can occur. As the number of barriers increases so does the limit on n and more resonances occur in each energy interval. As N tends to infinity the number of resonances becomes so great that over the whole of the energy range for which (4.5.2) applies the transmission coefficient is unity, and the condition for resonance tends to the condition

$$\left| \cosh\kappa a \cos kb - \frac{\left(\left(\frac{k_1}{m_1}\right)^2 - \left(\frac{\kappa}{m_2}\right)^2\right)}{2\kappa k/m_1 m_2} \sinh\kappa a \sin kb \right| < 1$$

This is the condition for the conduction bands obtained by the Kronig

Penney analysis and shows that transmission resonances are the primitive conduction bands of a finite system.

Consider a system of two square barriers in more detail, with especial consideration to two important properties; the resonance position and width.

The importance of the position of the resonance will be seen in section 5.5 where it will be shown that to a good approximation the voltage at which the current peaks, V_r , is related to the position of the transmission resonance $E_r(o)$ by

$$V_r = 2E_r(0)/e$$

Since V_r is one of the system properties over which close control will be required, a simple formula for predicting the change of system parameter necessary to produce a required change in resonant voltage is desirable. To obtain such a simple formula consider the condition for resonance written in the general form

$$f\ (E_r, V, a, b) = \text{constant}$$

If one of the parameters, p, is varied by a small amount dp the Taylor expansion to the lowest order gives a change in resonant energy dE_r

$$dE_r = -\left(\frac{\partial f/\partial p}{\partial f/\partial E_r}\right) dp \qquad (4.5.4)$$

Considering the three system parameters, barrier height, V, barrier width, a, and barrier separation, b, the derivatives are:

$$\frac{\partial f}{\partial V} = \frac{\partial \kappa}{\partial V}\left[a\left(\sinh \kappa a \cos kb - \frac{((k/m_1)^2 - (\kappa/m_2)^2)}{2\kappa k/m_1 m_2}\cosh \kappa a \sin kb\right) + \frac{((k/m_1)^2 + (\kappa/m_2)^2)}{2\kappa^2 k/m_1 m_2}\sinh \kappa a \sin kb\right]$$

$$\frac{\partial f}{\partial a} = \kappa \cos kb \sinh \kappa a - \frac{((k/m_1)^2 - (\kappa/m_2)^2)}{2k/m_1 m_2}\cosh \kappa a \sin kb$$

$$\frac{\partial f}{\partial b} = -k \sin kb \cosh \kappa a + \frac{((\kappa/m_2)^2 - (k/m_1)^2)}{2\kappa/m_1 m_2}\sinh \kappa a \cos kb$$

$$\frac{\partial f}{\partial E_r} = \frac{\partial k}{\partial E}\left[b\left(-\cosh \kappa a \sin kb - \frac{((k/m_1)^2 - (\kappa/m_2)^2)}{2\kappa k/m_1 m_2}\sinh \kappa a \cos kb\right) - \frac{((k/m_1)^2 + (\kappa/m_2)^2)}{2\kappa\, k^2/m_1 m_2}\sinh \kappa a \sin kb\right] + \frac{\partial \kappa}{\partial E}\left[a\left(\sinh \kappa a \cos kb - \frac{((k/m_1)^2 - (\kappa/m_2)^2)}{2\kappa\, k/m_1 m_2}\cosh \kappa a \sin kb\right) + \frac{((k/m_1)^2 + (\kappa/m_2)^2)}{2\kappa^2 k/m_1 m_2}\sinh \kappa a \sin kb\right]$$

where

$$\frac{\partial \kappa}{dV} = \frac{2m_2}{\hbar^2 \kappa}$$

$$\frac{\partial k}{\partial E} = \frac{2m_1}{\hbar^2 k}$$

$$\frac{\partial \kappa}{dE} = -\frac{\partial \kappa}{dV}$$

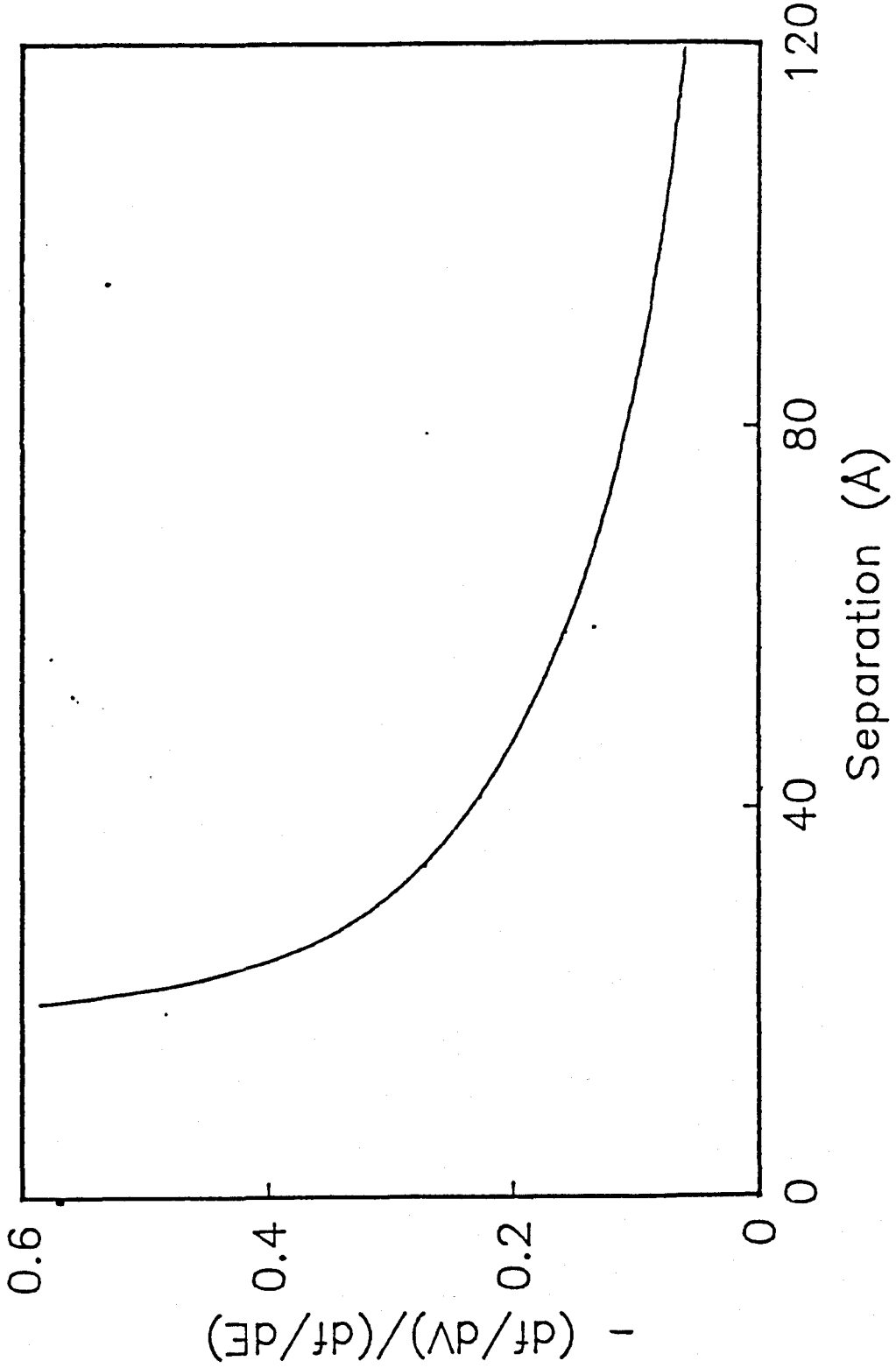

Figure 4.3a -(∂f/∂V)/(∂f/∂E) for a system of two 200 mV, 25Å barriers as a function of their separation, showing an increase in the resonant energy as the potential barrier increases.

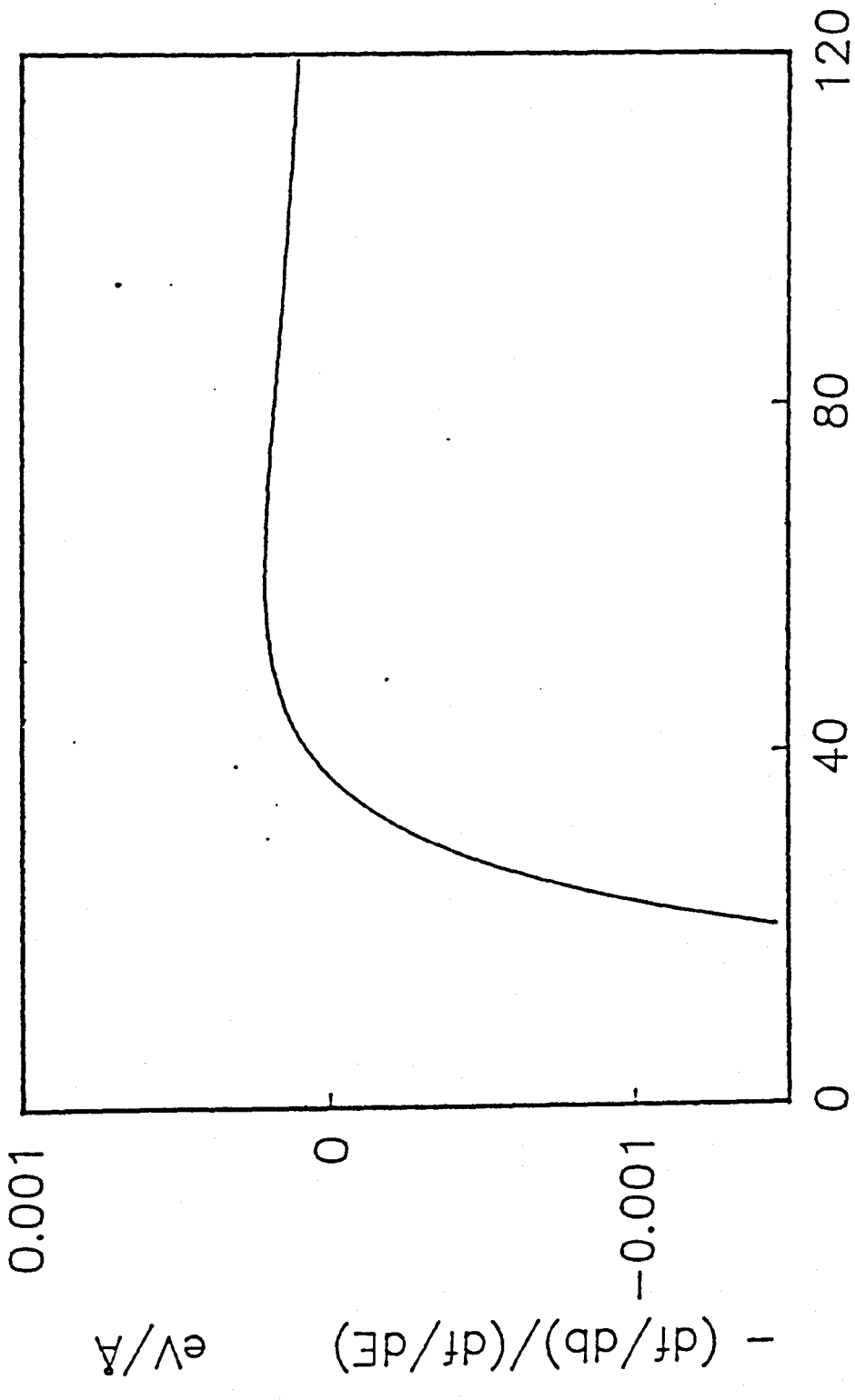

Figure 4.3b -(∂f/∂b)/(∂f/∂E) for two 200 mV, 25Å barriers as a function of their separation.

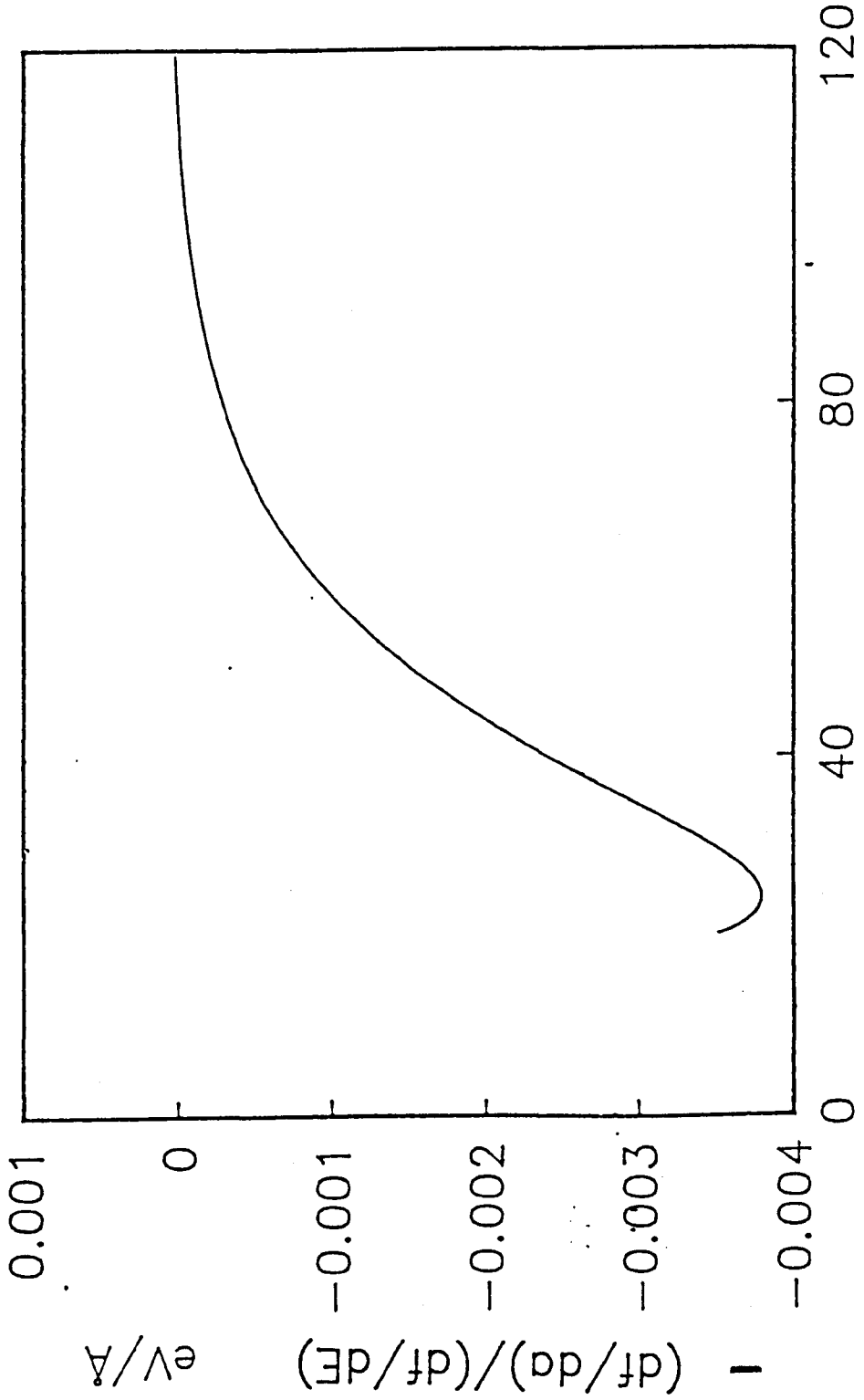

Figure 4.3c $-(\partial f/\partial a)/(\partial f/\partial E)$ for two 200 mV, 25Å barriers as a function of their separation. (The y axis is in ratio of eV/Å)

These derivatives can be used in equation (4.5.4) to study the affect of varying the three parameters. Unfortunately the complex nature of the derivatives means that they have to be evaluated numerically. As shown in figure (4.3), the numerical results show another consequence of the complex nature of the derivatives, only one simple dependence emerges. The resonant energy is proportional to the barrier height V. The dependence on the barrier width and separation, a and b, are too complex for any simple rules to be proposed. With no simple rules the only method of predicting the change needed to produce a required resonant energy is to evaluate (4.5.4) for each system separately. This method is used in Chapter 5 to check if an error in fabrication is the cause of disagreement between observed and predicted behaviour.

It would be reasonable to expect that the amount of current which will flow through a resonance to be proportional to its width. It would therefore be desirable to predict the affect of any parameter change on the width of the resonance and hence the magnitude of the current. Consider the magnitude of the transmission coefficient for a double barrier potential

$$|T|^2 = |T_1|^2 \sin^2\emptyset \; / \; (|R_1|^2 \sin^2 N\emptyset + |T_1|^2 \sin^2\emptyset)$$

This can be expanded about the resonant energy, assuming that the single barrier transmission coefficient is constant, to give

$$|T|^2 = \frac{\Gamma^2}{(E-E_r)^2 + \Gamma^2} \qquad (4.5.5)$$

a Lorentzian peak with a width Γ

$$\Gamma = \left| \frac{|T_1| \sin \phi_r}{|R_1| \; D \; N} \right|$$

where D is the derivative

$$D = \left(\frac{d\cos\phi}{dE} \right)_{E_r}$$

A Taylor expansion around the resonance energy shows that the Lorentzian shape is not surprising, for small changes ΔE

$$T(E_r + \Delta E) = T(E_r) + \left(\frac{\partial T}{\partial E} \right)_{E_r} \Delta E + \tfrac{1}{2} \left(\frac{\partial^2 T}{\partial E^2} \right)_{E_r} \Delta E^2 + \ldots.$$

The definition of E_r as the position of a maximum in T means that the second term is zero leaving an expansion with the same small ΔE limit as the Lorentzian. A Lorentzian peak close to resonance is therefore not surprising. The surprise is that the only resonances not found to be Lorentzian all the way down to the background value of T, many orders of magnitude below the maximum, are those resonances which are wide enough or at large enough energy for the condition on the transmission coefficient for the single barrier to be invalid. The majority of resonances are therefore Lorentzian with a predictable width, so that the affect of any change can be calculated.

The conclusion is that simple equations (4.5.4) and (4.5.5) can be used to qualitatively predict how the major features of a double barrier system, the resonant voltage and current, will behave when a small change of system parameter occurs.

4.6 Comparison of general techniques

To enable a choice to be made of which of the two general techniques to use in calculating the current expression, the results from the general techniques must be compared to the exact results for the test system.

Before any comparison the numerical technique must be studied in more detail. As outlined in Chapter 2 representing a discontinuous potential on a lattice gives rise to some ambiguity. In the original paper, Vignernon and Lambin used a method involving two points at half the barrier height. This method was found to give the best results over a wide range of square barriers, with a lattice spacing of 1Å, if the two half points were taken as the positions of the discontinuities. Using this method the numerical technique was capable of an accuracy of better than 0.1% for energies up to twice the barrier height, as shown in figure (4.4). Another important factor which was found to determine the accuracy of the calculation was the precision of the calculation. It was found that changing the variables accuracy to 4 bytes instead of the more usual 8 bytes increased the error by a factor of approximately five.

The potential representation method already described was sufficient for the parabolic Hamiltonian. However, for any Hamiltonian containing derivatives of the system parameters the accuracy decreased. The loss of accuracy is caused by the errors involved in defining the derivatives using n point difference schemes. Accuracy can be retained if the potential is defined in a way which allows the derivatives to be defined analytically.

This is done by representing a step potential by the inverse tangent

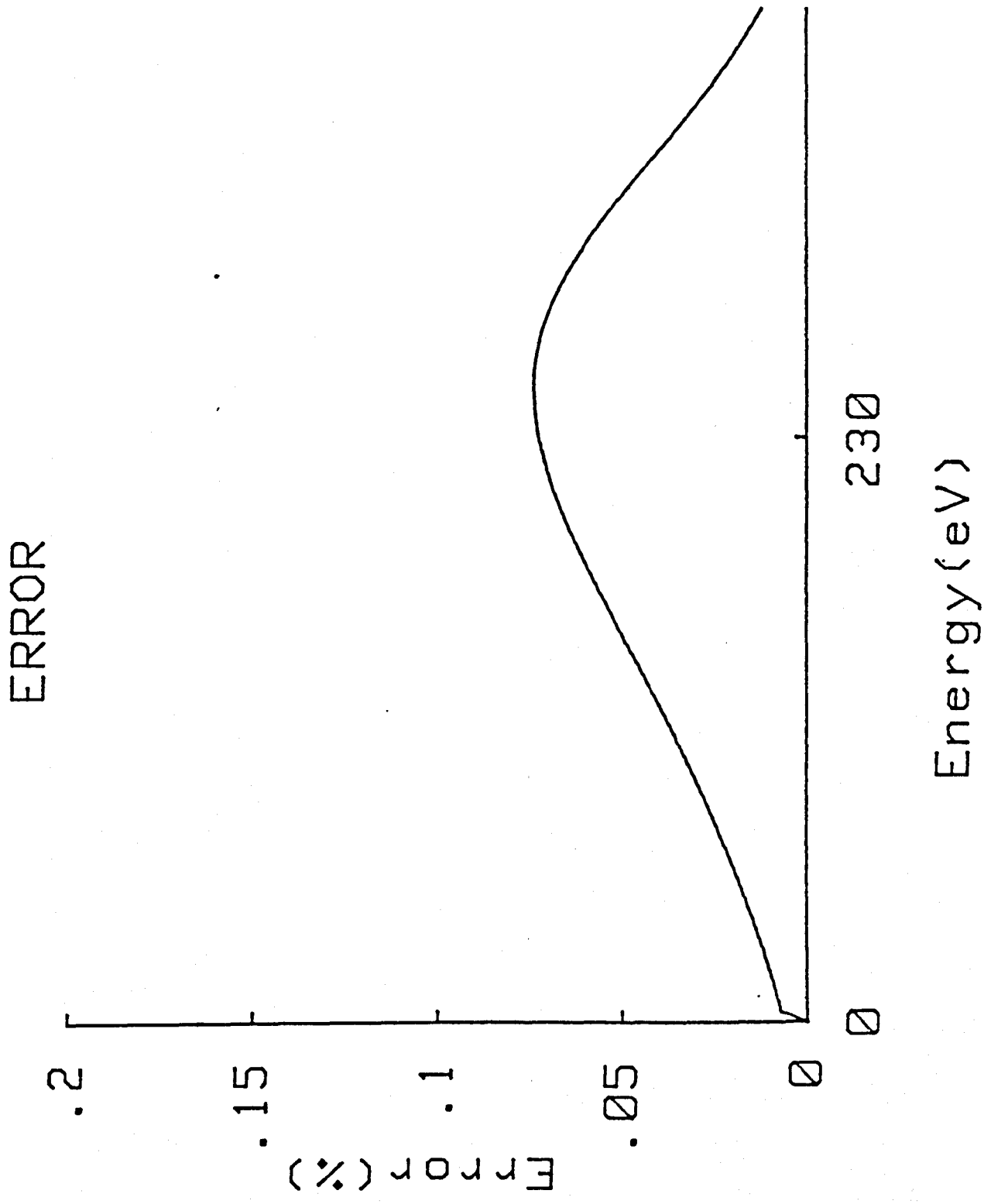

Figure 4.4 The percentage error in the transmission coefficient of 50Å, 0.23 V square barrier calculated using the numerical method of Vignernon and Lambin.

function

$$V(x) = V_o \text{ arc tan } \left(\frac{(x+\epsilon-a)\ \tan(1)}{2\epsilon} \right)$$

over the range $(a-\epsilon < x < a+\epsilon)$. This function has the desired property of changing from 0 to V_o over the range of definition and in the limit as ϵ tends to zero it is a step. The derivatives can then be expressed analytically

$$\frac{dV}{dx} = \frac{2V_o \epsilon \tan(1)}{(2\epsilon)^2 + ((x + \epsilon - a) \tan(1))^2}$$

$$\frac{d^2V}{dx^2} = \frac{-4V_o \epsilon (\tan(1))^3 (x + \epsilon - a)}{((2\epsilon)^2 + ((x + \epsilon - a) \tan(1))^2)^2}$$

Using this method of representing a discontinuous potential the numerical technique can be extended to non-parabolic Hamiltonians, without a loss of accuracy.

The comparison between the two general techniques, shown in figure (4.5), is for a system which is representative of those to be considered. The transmission resonance is the property of the system which it is hoped to exploit and is therefore of major importance. The WKB technique predicts the existence of the resonance, but, the approximations used in treating transmission through the individual barriers give rise to an incorrect resonant energy. The numerical technique is much more accurate, the exact and numerical results being indistinguishable. This accuracy is not limited to the one system shown and the greater accuracy demonstrated means that the numerical technique will be used throughout the remainder

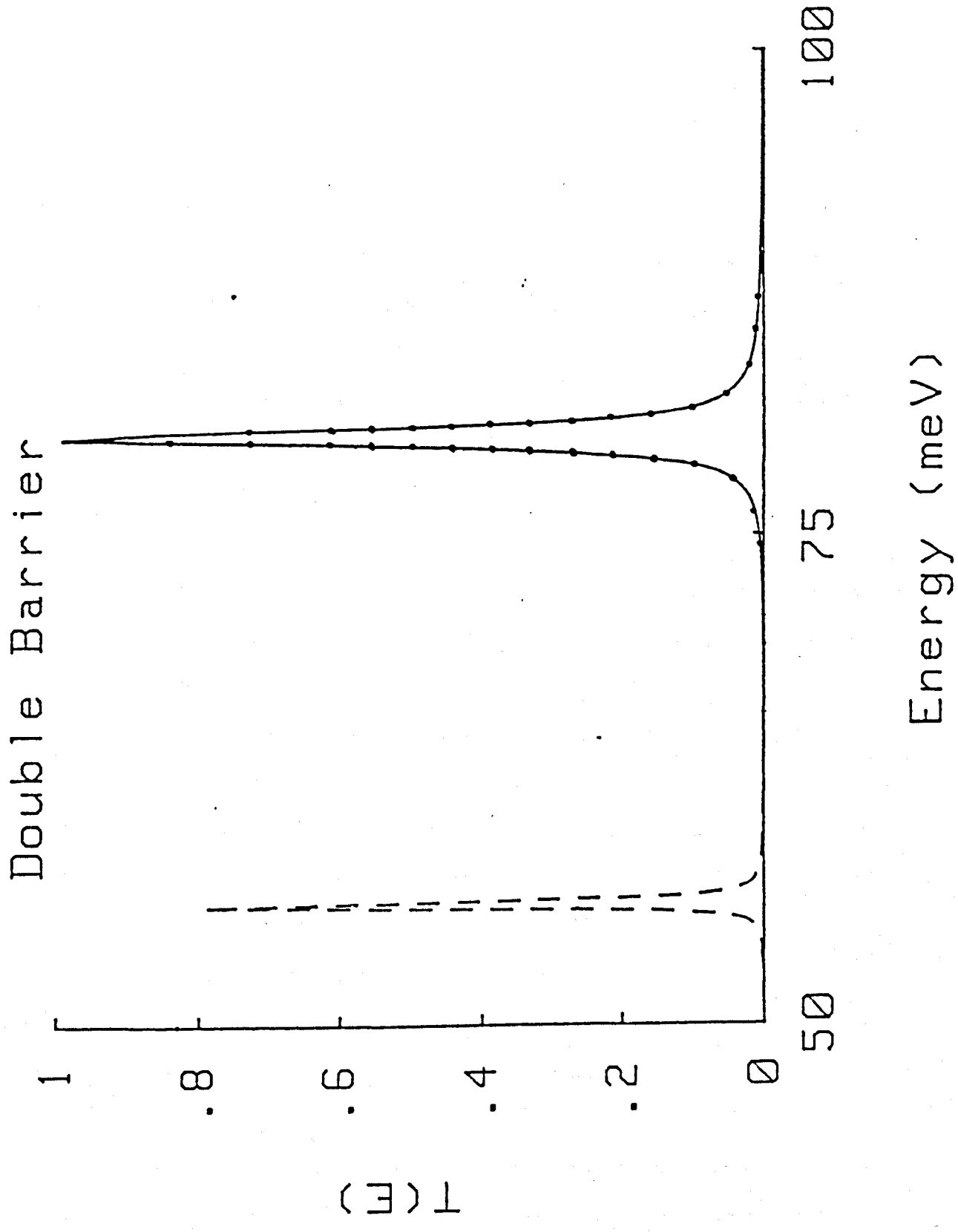

Figure 4.5 A comparison of the exact (•) numerical (—) and WKB (--) transmission coefficients of a double barrier system.

of the thesis.

Before proceeding with the current calculation, it is informative to study the transmission coefficient for two barriers. The systems which are to be modelled are two terminal devices with the voltage applied between the emitter and collector. Some of this voltage is dropped across the layers forming the barriers and in the model described in Chapter 1 this will be represented as a simple linear addition to the potential representing the conduction band minimum.

As the potential across the barriers is increased the resonant energy and height are both found to decrease. The resonant energy appears to be fixed with respect to the energy of the centre of the well between the two barriers. Since the energy reference is the emitter, as the applied potential is increased the potential of the centre of the well decreases, with a consequent decrease in resonant energy. This leads to a simple rule for the resonant energy, independent of the method of distributing the potential between the barriers and the well, as a function of the voltage, V, applied across the potential system

$$E_r(V) = E_r(0) - eV/2$$

For voltages greater than $2E_r(o)/e$ the resonance disappears.

The observed decrease in resonance height is due to the asymmetry introduced by the applied field. WKB showed that resonance to unity is only possible if the barriers have the same transmission probability. As the field increases the asymmetry increases with the consequent decrease in the transmission resonance.

The affect of barrier asymmetry has been investigated by Kane (51) For

simplicity the asymmetry studied was not a double barrier system with an applied field, but two barriers of different thickness a_1 and a_2. Kane wrote the ratio of transmitted and incident currents j_t and j_i as

$$\frac{j_t}{j_i} = P \left[f_1 \exp(\kappa_1 a_1 + \kappa_2 a_2) + f_2 \exp(\kappa_1 a_1 - \kappa_2 a_2) + f_3 \exp(\kappa_2 a_2 - \kappa_1 a_1) + f_4 \exp(-\kappa_1 a_1 - \kappa_2 a_2) \right]^{-2} \qquad (4.6.1)$$

where P is a prefactor and f's are phase factors.

If the transmission coefficients for individual barriers are approximated to be

$$T = \exp[-2\kappa a]$$

then each term in the denominator can be interpreted as a product or ratio of the transmission coefficients of the individual barriers.

The denominator will be dominated by the first term over most of the energy range and the transmission coefficient is then proportional to the product of the individual transmission coefficients.

$$\frac{j_t}{j_i} = P \exp[-2(\kappa_1 a_1 + k_2 a_2)]$$

At specific energies when the phasefactor f_1 is zero, the denominator is dominated by the second or third term, depending upon which barrier is least transmitting. For a system with the first barrier least transmitting the transmission coefficient is then

$$\frac{j_t}{j_i} = P \exp[-2(\kappa_1 a_1 - \kappa_2 a_2)]$$

and the transmission coefficient is proportional to the ratio of the transmission coefficients of the individual barriers.

A picture therefore emerges in which the transmission coefficient is proportional to the product of the transmission coefficients of the individual barriers over most of the energy range. However, at specific energies it becomes proportional to their ratio, with a consequent resonant increase.

This widely used result can be compared to the results of the numerical technique. Figure (4.6) shows the ratio of the individual barrier transmission coefficients divided by the resonant transmission coefficient plotted as a function of the ratio of the widths of the barriers. It can be seen that when the two barriers are sufficiently different

$$\frac{j_t}{j_i} = 4 \frac{T_1}{T_2}$$

which can be understood if the prefactor P is examined in detail. In Kane's notation

$$P = \frac{2^8 k_1 \kappa_1^2 k_2^2 \kappa_2^2 k_3}{(k_1^2 + \kappa_1^2)(\kappa_1^2 + k_2^2)(k_2^2 + \kappa_2^2)(\kappa_2^2 + k_3^2)}$$

where the subscript denotes the regions. To a good approximation the wavevectors and decay constants are equal so that

$$P \approx 2^4$$

The phasefactor for resonance is

$$f_2 = \left[\exp\left[i\left(\phi_1 - \phi_3 + \phi_4\right)\right] - \exp\left[i\left(-\phi_1 + \phi_3 - \phi_4\right)\right] \right] \exp\left[i(\phi_2 + \phi_5)\right]$$

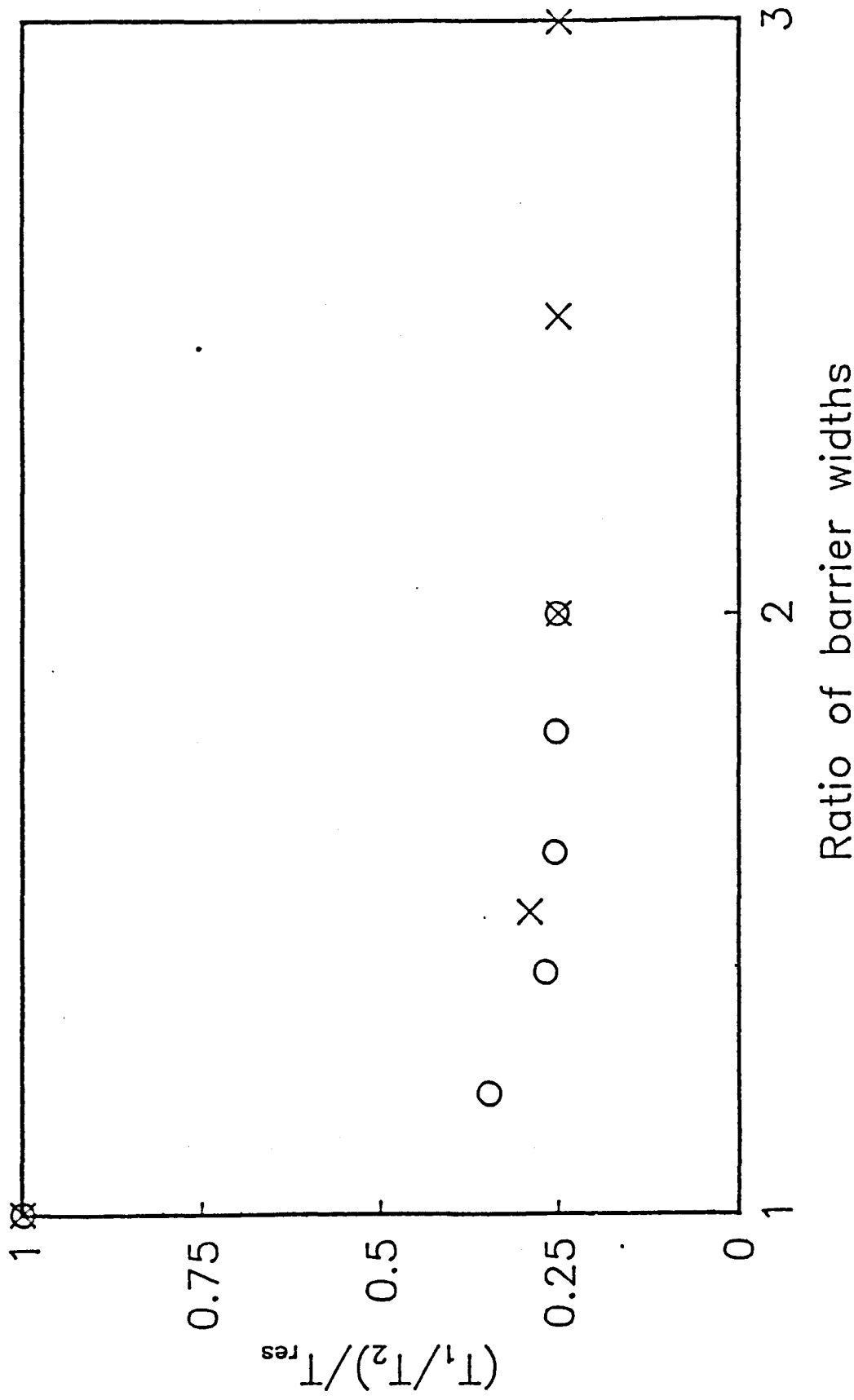

Figure 4.6 The ratio of the transmission coefficients of two barriers, calculated at the resonance energy, divided by the transmission coefficient at resonance plotted against the ratio of the widths of the two barriers
X Fixed barrier 5 nm
O Fixed barrier 10 nm
All barriers 0.23 V separated by 5 nm

where

$$\emptyset_1 = k_2 w_3$$

$$\emptyset_2 = \text{arc tan } \kappa_1/k_1$$

$$\emptyset_3 = \text{arc tan } \kappa_1/k_2$$

$$\emptyset_4 = \text{arc tan } \kappa_2/k_2$$

$$\emptyset_5 = \text{arc tan } \kappa_2/k_3$$

The condition for resonance reduces this expansion to

$$f_2 = 2 \sin (2\emptyset_4) \exp [i(\emptyset_2 + \emptyset_5)]$$

If the other contributions to the bracket in (4.6.1) are neglected and all the decay constants and wavevectors are taken to be approximately equal the bracket reduces to

$$[\ ]^2 = 4 \exp [2(\kappa_1 a_1 - \kappa_2 a_2)]$$

and ratio of currents becomes

$$\frac{j_t}{j_i} = 4 \exp [2(\kappa_1 a_1 - \kappa_2 a_2)] \qquad (4.6.2)$$

Figure (4.6) can now be explained. The region before the plateau is caused by the third term of the denominator in (4.6.1) which contributes until the barriers are sufficiently different. After this point all but the most significant term can be neglected and equation (4.6.2) is valid.

The final result is that to a good approximation, for systems with sufficiently different transmission coefficients, the transmission coefficient at resonance is four times the ratio of the transmission coefficients of the individual barriers.

4.7 Numerical calculation of the phase shift

The numerical technique chosen to calculate the transmission coefficient can be extended to calculate the phase difference between the incident and transmitted waves. This can then be used to calculate the traverse time as defined in section 3.2.

The wave function in Region III is defined as

$$\psi(J + 1 + p) = U_{III}^{+p} \qquad p > J$$

which is the solution to the discrete equation equivalent to the solution to the continuous solution

$$\psi(x) = \exp [ik(x-L)] \qquad x > L = (J+1)\ \Delta$$

Similarly the solution in Region I is the equivalent of

$$\psi(x) = A \exp [ikx] + B \exp [-ikx]$$

Since the usual definition of a transmission coefficient is for a unit incident flux the transmission and reflection coefficient can be identified as

$$t = \frac{\exp [ikL]}{A}$$

$$r = \frac{B}{A}$$

where A and B are functions of k. The calculation of the phase difference between transmitted and incident waves, obtained from t, requires the calculation of A.

During the calculation of $| T |^2$ the quantity N(-1) was defined and

calculated, where N(-1) is given by

$$N(-1) = \frac{Au_1^+ + Bu_1^-}{A + B}$$

which can be rearranged to give

$$A = B \frac{(u_1^+ - N(-1))}{(N(-1) - u_1^-)}$$

This condition can be inserted into the equation for the wave function at the origin

$$\psi(0) = A + B$$

Then the definition of the factors N(p)

$$\psi(p) = N(p)\ \psi(p-1)$$

and the initial condition

$$\psi(J+1) = 1$$

can be used to give the final equation for A

$$A = (u_1^+ - N(-1)) \prod_{o}^{J} N(p) \,/\, (u_1^+ - u_1^-) \qquad (4.7.1)$$

Since the factors N(p) can be calculated from the program used to calculate $|T|^2$ this method of calculating A is easily implemented and efficient, however, it is very inaccurate. The product in expression (4.7.1) means that the percentage error in the final result is the sum of the errors in the N s. The result is that this direct method is not sufficiently accurate, and another method of calculating the phase difference is required.

This other method was based on a general result for one dimensional potentials. Messiah (39) has shown that if a given potential is

considered with flux incident separately from the left and right then

$$\tau_L = \tau_R$$

$$p_L - \tau_R = \pi - p_R + \tau_R$$

where τ and p are the phase angles of t and r, the subscripts denoting the side from which the flux was incident. These two results combine to give

$$\tau = (p_L + p_R - \pi)/2 \qquad (4.7.2)$$

and reduce the calculation of the required phase difference to the calculation of p_L and p_R, which is equivalent to two calculations of R.

The expression for N(-1) can be rearranged to give

$$R = \frac{N(-1) - u_1^-}{u_1^+ - N(-1)}$$

an expression which is expected to be accurate to within a few percent since it involves the same quantities as the calculation of $|T|^2$ which is accurate to better than one percent.

The values p_L and p_R were then calculated from (4.7.2) using an inverse tangent routine, which must be extended beyond the $-\pi/2$ to $\pi/2$ range used by the computer library routine in order to ensure that τ is a smooth function.

In comparing the numerical results with the analytic results for a single square barrier a correction is needed to the numerical results. This accounts for the region of free space which must be included in the definition of the potential for the numerical technique to be used. The

resulting comparison is shown in figure (4.7), where the functional dependence of both sets of results is shown rather than the percentage error because of the zero in the function. An estimate of the error over the range shown would be approximately 0.5%.

To extend the method to resonant systems care must be taken. At resonance the reflection coefficient is zero, and this must be prevented from changing the range in which the inverse tangent function operates.

Once this precaution has been taken the results for the double barrier are as accurate as those for the single barrier. The numerical results show the rapid change of approximately π predicted by scattering theory (39). This change occurs over a range of a few halfwidths of the resonance and, through the traversal time expression (3.2.3), leads to each transmission resonance being associated with a resonance in the traversal time.

Once the functional dependence of the phase difference has been calculated by this numerical method the traversal time for any general potential can be obtained, the accuracy being limited by the wavevector step size used in the calculation of the derivative.

4.8 Simple numerical superlattice bandstructure calculation

The numerical technique used to obtain the transmission coefficient can be extended to calculate the E-k relationship for a superlattice. Vignernon and Lambin have shown (52) that if a wave function of energy E is to be periodic over a length L, represented by a n discrete points

$$(A(n + 1) - B(n - 1))^2 - 4 \leqslant 0$$

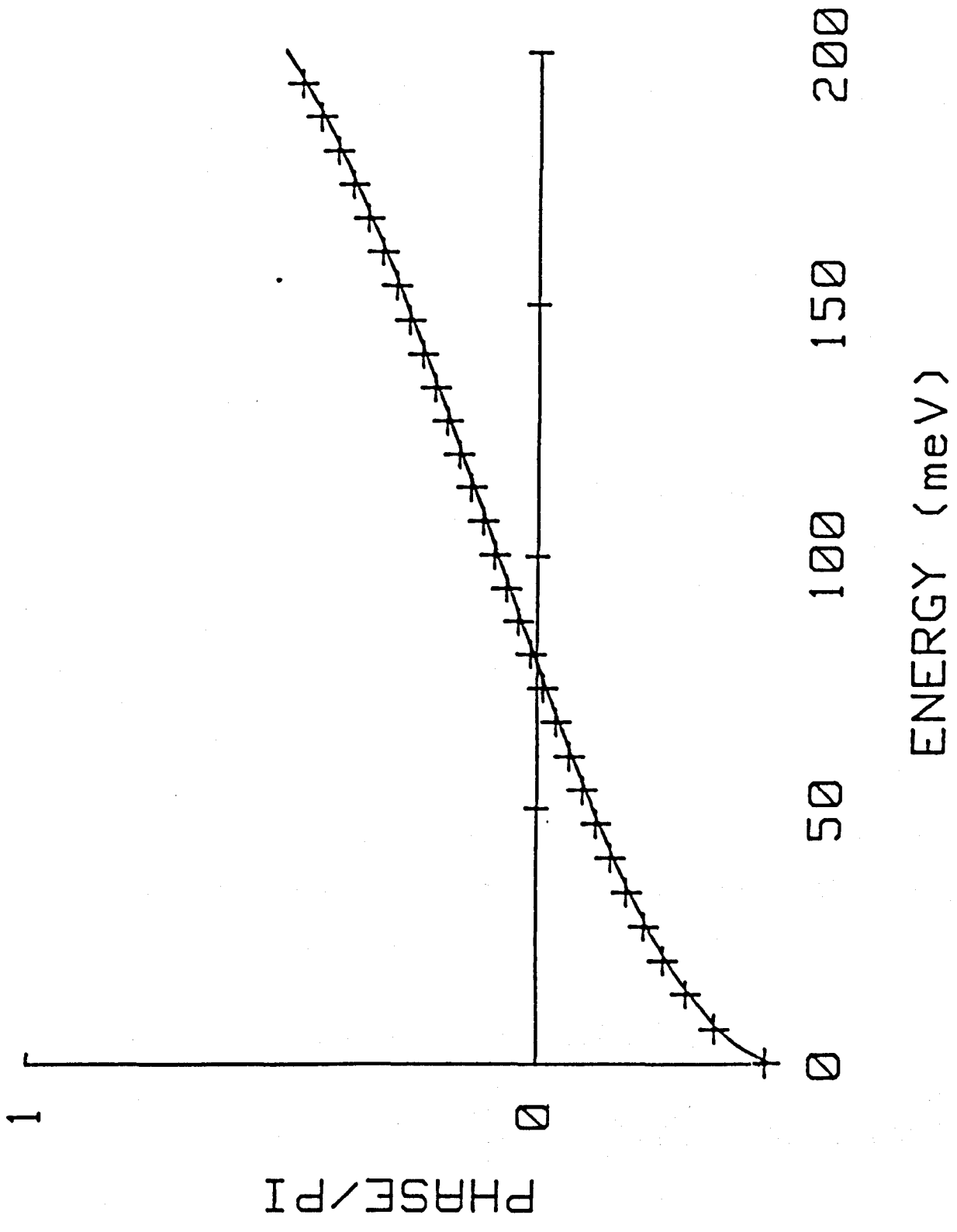

Figure 4.7 A comparison of the exact (—) and numerical phaseshifts (X) for a 10 nm, 23% Al square barrier.

The coefficients being obtained from the recurrence relationships

$$A(p) = b(p)\,A(p-1) - a(p-1)\,A(p-2)$$

$$B(p) = b(p+1)\,B(p-1) - a(p)\,B(p-2)$$

with the initial conditions

$$A(0) = 1 \qquad A(1) = b(1)$$

$$B(0) = 1 \qquad B(1) = b(2)$$

with a(p) and b(p) being defined by Wall (68) from the continued fraction form of the S.E. (4.4.2). The wavevector k corresponding to an allowed energy E can then be obtained using

$$\cos kL = (A(n+1) - B(n-1))\,/2$$

so that a full E-k relationship can be obtained.

Once the full dispersion relationship is known the effective mass and density of states can be calculated. The effective mass, being defined as

$$m^{*} = (\hbar^2/\partial^2 E/\partial k^2)_E$$

can be calculated for a system with uniform energy steps ΔE at E(p) using

$$\frac{d^2E}{dk^2} = \left[\frac{2\,(2k(p) - k(p+1) - k(p-1))\,\Delta E}{(k(p+1) - k(p-1))\,(k(p) - k(p-1))\,(k(p+1) - k(p))} \right]$$

where k(p) is the wavevector corresponding to E(p). The density of states can be calculated from

$$N([E(p) + E(p+1)]/2) = \frac{(k(p+1) - k(p))}{\Delta E}$$

This method can be used to compare the superlattice bandstructure calculated using the simple model described in section 1.2 with the results of other methods such as the calculations by Ivanov and Pollman (79).

The general results of this comparison are in agreement with the conclusions of section 1.4 and Ivanov and Pollman. The density of states showed the expected pole at each extremum, the masses agreed to within approximately 10% with those of Ivanov and Pollman, and although no figures are given by Ivanov and Pollman, the positions of the bands also appear to agree to within 10%.

As well as supporting the accuracy of the model, this method can be used to discuss the transport mechanisms which may be expected to dominate in a superlattice.

Although an electron cannot be realistically represented by a Gaussian wavepacket, the Gaussian wavepacket does possess the major properties required, being both localised in position and momentum. Consider a Gaussian wavepacket with a momentum distribution $f(k,k_o)$, then

$$\psi(x,t) = \int dk\ f(k,k_o)\ \exp\ [i(kx-\omega_k t)]$$

To propagate as a free particle between scattering events the dispersion relationship

$$E = \frac{hk^2}{2m^*}$$

must be a good approximation to the bandstructure over the range of wavevectors for which $f(k,k_o)$ is significant, that is within a few

standard deviations of the mean.

The results of the simple superlattice calculation indicate that the bandstructure is parabolic over approximately 10% of the wavevector range, so that if the electron is to evolve as a free particle between collisions

$$\sigma_k < \frac{\pi}{nL}$$

where n is of the order of 25. This condition is equivalent to a condition for the standard deviation in space σ_x

$$\sigma_x > \frac{nL}{\pi}$$

This indicates that to evolve as a free particle an electron must be spread over tens of lattice periods.

At the other extreme is a wavepacket made up of contributions from all the states in one band. Then

$$\sigma_k > \frac{\pi}{2L}$$

which is equivalent to

$$\sigma_x < \frac{2L}{\pi}$$

a wavepacket localised in one period, which would be expected to contribute to the conduction mechanism by a hopping process.

The result is an indication that a hopping process will dominate conduction if the electron is localised to one period by scattering mechanisms or by the admixture of all the states in one band. This last

condition implies that for narrow bands at high temperature, when it can be expected that all states have equal probability of occupation, the dominant transport mechanism is expected to be hopping with a free electron-like behaviour observed at lower temeratures.

4.9 Summary

Although the last two sections are interesting in demonstrating the versatility of the numerical technique the important results for the major theme of this thesis were obtained in sections 4.2 and 4.6.

In the first of these sections the current density expression (4.2.4), which will be used in Chapter 5 to calculate current voltage characteristics, was obtained. This expression demonstrated a requirement for a method of calculating the transmission coefficient of a general potential, a requirement which will be filled by the general numerical method outlined in section 4.4.

This choice of method was the result of the comparison in section 4.6, where it was demonstrated that the numerical method was accurate to within 0.5% over all the energy range whilst the alternative WKB method gave an incorrect peak position.

The other result of significance from section 4.6 was the voltage dependence of the resonant energy

$$E_r(V) = E_r(o) - eV/2$$

which gives rise to a good approximation to the resonant voltage V

$$V_r \qquad 2E_r(o)/e$$

This in turn means that the analytic results of section 4.5 can be used to help engineer double barrier structures with a desired resonant voltage and current.

CURRENT VOLTAGE CHARACTERISTICS

5.1 Introduction

This chapter is devoted to discussing the current voltage characteristics of different systems. Although in most of the chapter the results of numerical integration of the expression for the current density (4.2.4) are discussed the first two sections are qualitative discussions of the temperature and voltage dependence of the current using model transmission coefficients. The first section also shows that (4.2.4) can be used to understand the qualitative features observed in a study where insufficient data was given to justify a full quantitative analysis. This approach could be extended to predicting the behaviour of systems before fabrication, and gives some confidence that the quantitative predictions will be of some value.

In the third section the numerical technique used to integrate the current density expression, in the presence of the sharp integrand structure caused by the resonance in the transmission coefficient, is discussed. This prepares the way for the first direct comparison between the predictions of the simple model and the observed current-voltage characteristics.

Although most of the parameters are taken from the design specifications, there are still two features which must be assumed - the conduction band discontinuity and the distribution of the applied field. The results of section 1.3 indicated that a conduction band discontinuity of 62% of the total bandgap discontinuity should be assumed. Because of the remaining uncertainty values of 60% and 65% are used to give some indication of the

sensitivity of the final result to any error in the accepted value. The model does not include an attempt to solve Poisson's equation. This means that two model field configurations will be assumed, leading to a total of four current voltage characteristics to be compared to the observed results.

The following two sections, 5.5 and 5.6, show that the lack of agreement between these calculations and the observed results cannot be explained as differences between the design and the fabrication of the device, but, is due to a resistance in series with the double barriers. Any remaining disagreement is then shown to be attributable to the experimental technique, although the asymmetry observed in the experimental data means that a detailed fit is not undertaken because of the lack of accurate control over the system parameters implied by the asymmetries.

In section 5.7 studies of the affects of various asymmetries show that the major cause of asymmetry in the peak voltage is the series device. Preliminary calculations indicate that deliberately introduced asymmetries, designed to enhance the peak to valley ratio for a specific resonance, do not produce a large effect.

In the following three sections, systems which have been designed or fabricated are considered. These calculations demonstrate that the model is useful both in explaining observed data and predicting the behaviour of suggested designs. This latter use is probably the major use to which this model should be put, enabling the design of systems with desired properties without fabrication, a suggested design being modelled in one day.

In the final section before the summary, results of the one failure to obtain any sort of agreement between predicted and observed behaviour are presented. This section demonstrates that although the model appears to be genererally applicable caution is still required in its use.

The chapter ends with a short summary of results.

5.2 Qualitative study of single barrier systems

A heuristic current voltage expression using a simple model for a GaAlAs system was obtained in Chapter 4. The first term in this expression is that used by other authors, to which a correction term was added which accounted for the velocities difference in the two contacts due to the applied potential.

The simple analytic discussion in this section is simplified if only one of the terms is considered. Numerical integration of the full expression has shown that for the systems of interest the second term represents a maximum of ten percent of the total. Therefore, in the remainder of this section only the dominant contribution will be considered, so that

$$J \approx \frac{2mkT_e}{4\pi^2\hbar^3} \int_0^\infty dE \;\; T(E,V)\; D(E,V,\mu_e,\mu_c)$$

Using model transmission coefficients it is possible to reach some general conclusions concerning the voltage and temperature dependence of the current density.

Consider the supply function

$$D(E,V,\mu_e,\mu_c) = \ln\left[\frac{1 + \exp((\mu_e-E)/kT)}{1 + \exp((\mu_c-E-eV)/kT)}\right]$$

For simplicity assume that $\mu_e = \mu_c$, which allows the same approximations to be used for both the numerator and denominator at voltages small compared to kT/e. The supply function can then be approximated as

$$D \approx \frac{eV}{kT} . \qquad 0 < E < \mu$$

$$D \approx \exp((\mu-E)/kT\ [1-\exp(-eV/kT)] \qquad E > \mu$$

for all the energy range except for a small region of a few kT around the Fermi energy. The current density can then be approximated as

$$J \propto T\theta(\mu) \int_{\cdot}^{\mu} dE\ T(E,V)\ \frac{eV}{kT} +$$

$$T\int_{\mu\theta(\mu)}^{\infty} dE\ T(E,V)\ \exp\ ((\mu-E)/kT)\ \ [1-\exp(-eV/kT)]$$

Two behaviours can be identified. For a high Fermi energy (doping) the first term can dominate because of the exponential damping of the second term. For low voltages, the voltage dependence of the transmission coefficient can be neglected to give a current density J

$$J \propto \frac{eV}{k} \int_{\cdot}^{\mu} dE\ T(E,0)$$

Figure (5.3) shows that at high doping the Fermi level is only a weak function of temperature, so that J is Ohmic and only a weak function of temperature.

The other behaviour arises for low doping. The Fermi level is then less than zero and since the energy must be greater than zero the first term no longer exists. The current density is then

$$J \propto T(1-\exp(-eV/kT))\ \exp\ (\mu/kT)$$

$$\int_{\cdot}^{\infty} dE\ T(E,V)\ \exp\ (-E/kT)$$

The voltage has already been assumed to be small compared to kT/e, to approximate the supply function, so that the current density is approximately

$$J \propto \frac{eV}{k} \cdot \exp(\mu/kT) \int_{\cdot}^{\infty} dE\ T(E,V)\ \exp(-E/kT)$$

Again the low voltage current density is ohmic. However, it is no longer a weak function of temperature, both the integrand and prefactor being exponential in reciprocal temperature. These, combined with the temperature dependence of the Fermi level, mean that the current density at low voltage is temperature sensitive.

This type of analysis can be used to consider the observed characteristics of square barrier systems. Collins et al studied the current voltage characteristics of p and n doped GaAlAs barries (7).

For the p type layers it was found that the zero bias resistance was inversely proportional to temperature from 300 K to 77 K. Below 77 K the characteristics remained temperature independent down to 4.2 K. As the barrier thickness increased for 60Å to 120Å the zero bias resistance was found to increase by 2.5 orders of magnitude.

For n type layers the zero bias resistance was found to be inversely proportional to temperature down to 30 K after which the characteristics became temperature independent. The current for a 60Å barrier was found to be an order of magnitude greater than for the p type 60Å barrier and the zero bias resistance increased by less than one order of magnitude as the barrier thickness increased to 120Å.

Since the material parameters given were not measured but projected from the growth parameters no quantitative analysis is possible, however, qualitative agreement can be obtained.

Assuming that the only affect of the barrier doping is to give rise to an electrostatic contribution to the barrier potential, the only difference between the n and p type systems is the barrier potential V , which is greater for the p type systems. Using the approximate supply function gives

$$J \propto \frac{eV}{k}\ \theta(\mu) \int_0^{\mu} dE\ T(E,V) + T \int_{\mu\theta(\mu)}^{\infty} dE\ \exp((\mu-E)/kT) \left[1-\exp(-eV/kT)\right]$$

The next step is to assume a transmission coefficient which models the behaviour of the actual transmission coefficient. The outstanding feature of the transmission coefficient of a square barrier is the transition from low to high transmission probability at an energy approximately corresponding to the barrier height. The major voltage dependence, at least for small voltages, is a shift in this edge. These features lead to a model transmission coefficient

$$T(E,V) = T_L(V_b-E-eV/2) + T_H(E-V_b+eV/2)$$

where $T_L < T_H$ and V is the applied voltage.

Since the barrier height is much greater than the Fermi energy

$$J \propto \theta(\mu) \int_0^{\mu} dE T_L\ \theta(V_b-E-eV/2)\ \frac{eV}{k}$$

$$+ T \int_{\mu\theta(\mu)}^{\infty} dE \left\{ [T_L\ \theta(V_b-E-eV/2) + T_H(E-V_b+eV/2)] \right.$$

$$\left. \exp\ ((\mu-E)/kT)(1-\exp(-eV/kT)) \right\}$$

which integrates to give

$$J \propto \theta(\mu)\ \frac{eV}{k}\ T_L\ \mu + \frac{1}{k}\ [\ 1-\exp(-eV/kT)\ ]\ \exp(\mu/kT)$$

$$[T_L\ \left\{ \exp(-(V_b-eV/2)/kT) - \exp(-\mu\theta(\mu)/kT) \right\}$$

$$+\ T_H\ \exp((-V_b+eV/2)/kT)]$$

Since T_H is several orders of magnitude greater than T_L the last term is dominated by the T_H contribution and the term which dominates the temperature dependence of the current density is

$$\exp(-V_b/kT)$$

This predicts that the system with the higher barrier will give rise to a temperature independent current at a higher temperature, as observed.

The observed behaviour can be seen to occur because the temperature dependent contribution to the current arises from electrons which have a high transmission probability. These electrons must have an energy greater than the barrier height, therefore the higher the barrier the higher the temperature at which this contribution can be neglected, and the current remain temperature independent.

The dependence of the second term on the barrier potential shows that the current decreases with an increase in the barrier height. This explains the observation that the lower n type barrier gives rise to larger current densities than the equivalent p type barrier. Again the reason is the contribution from electrons with an energy greater than that equivalent to the barrier potential.

The final observation to be explained is the dependence upon barrier doping, of the sensitivity of the zero bias resistance to variations in barrier width. To explain this observation consider the conductivity given by

$$\frac{\partial J}{\partial V} \propto \theta(\mu)\,\frac{e}{k}\int_{\cdot}^{\cdot} dE\; T(E,V) + \frac{e}{k}\;\exp\,(-eV/kT)\int_{\mu\propto\mu}^{\infty} dE\; T(E,V)\;\exp((\mu-E)/kT)$$

The zero bias resistance is proportional to the inverse of the conductivity evaluated at zero applied voltage. At zero volts the temperature dependence of the second term means that the zero bias resistance is inversely proportional to the temperature. When the second term becomes negligible the resistance becomes temperature independent as observed, and is given by

$$\rho \propto \left[\int_{0}^{\mu} dE\; T(E,0)\;\right]^{-1}$$

A study of the dependence of the resistance upon the barrier width requires a model transmission coefficient which itself contains a width dependence. Since the Fermi energy is less than the barrier height a WKB approximate to the transmission coefficient can be used as the model. Neglecting the prefactor

$$T(E,0) \propto \exp(-2a(2m(V_b-E)/h^2)^{1/2})$$

which can be integrated to give

$$\rho \propto \left[\frac{1}{a^2}\left\{(1+2a(2m(V_b-\mu)/\hbar^2)^{1/2})\;\exp(-2a(2m(V_b-\mu)/\hbar^2)^{1/2})\right.\right.$$

$$\left.\left. - (1+(2a\,V_b/\hbar^2)^{1/2})\;\exp\,(-2a(2mV_b/\hbar^2)^{1/2})\right\}\right]$$

The overall expression is simplified by assuming that μ is sufficiently small compared to V_b for the expression to be dominated by the first term of the Taylor expansion in which μ is treated as a small change in V_b then

$$\rho \propto \frac{1}{\mu} \exp(2a(2mV_b/\hbar^2)^{1/2})$$

An expression which agrees with the observations that the resistivity increases with width, a, and that the dependence of the magnitude of the increase upon the barrier doping V_b.

5.3 Qualitative study of double barrier systems

In the previous section results were obtained for a system with a monotonically increasing transmission coefficient, a class of systems which does not include resonant tunnelling structures. To investigate the behaviour of a system with a transmission coefficient which contains a resonance the model transmission coefficient

$$T(E,V) = f(V)\ \delta(E_r - eV/2 - E) + T_L \theta(E - V_m + eV/2)$$

$$+ T_H\ \theta(V_m - E - eV/2)$$

can be used. The first term represents a resonance with a voltage dependent amplitude and position, whilst the remaining terms represent the smooth background. Since the main system of interest is a double barrier system, the voltage dependence of the peak position is chosen to agree with that of double barriers and the smooth background has an edge at the barrier height minus the average applied voltage.

In general the barrier height is much greater than the Fermi energy so

that for voltages such that

$$V_m - eV/2 > \mu$$

the high energy approximate to the supply function can be used to evaluate the T_H contribution.

The contribution from the low transmission probability region is less than

$$J \propto T_L T(V_m - eV/2) \ln\left[\frac{1+\exp(\mu/kT)}{1+\exp((\mu-eV)/kT)}\right]$$

which is made negligible by T_L, so that the current can be approximated by

$$I \propto T\, f(V) \ln\left[\frac{1+\exp((\mu-E_r+eV/2)/kT)}{1+\exp((\mu-E_r-eV/2)/kT)}\right]$$

$$+ T^2 k T_H \exp((\mu+eV/2-V_m)/kT)(1-\exp(-eV/kT))$$

For small voltages such that

$$\mu + eV/2 - V_m < 0$$

the second contribution is exponentially damped and the current is further reduced to

$$I \propto T\, f(V) \ln\left[\frac{1+\exp((\mu-E_r-eV/2)/kT)}{1+\exp((\mu-E_r+eV/2)/kT)}\right]$$

Resonance then occurs when the transmission peak is close to the maximum in the supply function, which is when

$$V \approx 2E_r/e \qquad (5.3.1)$$

and the current at resonance, I_p, is

$$I_p \propto T\, f(2E_r/e) \ln\left[\frac{1+\exp(\mu/kT)}{1+\exp(\mu-2E_r)/kT}\right]$$

The resonance energy E_r is usually much greater than the Fermi level μ so that the denominator in the logarithmic function is negligible leaving

$$Ip \quad \propto \quad T\, f(2E_r/e) \ln\,(1 + \exp(\mu/kT))$$

This shows that the temperature dependence of the current resonance depends on the doping density. For high doping density and low temperature, I_p has the same temperature dependence as the Fermi level μ

$$I_p \quad \propto \quad f(2E_r/e)\mu/k$$

which from figure (5.3) means that it is approximately temperature independent. In contrast at low doping and high temperature, I_p is very temperature sensitive

$$I_p \quad \propto \quad T\, f(2E_r/e)\, \exp\,(\mu/kT)$$

At a voltage slightly greater than the resonant value, where the current voltage characteristics have a minimum or valley, condition (5.3.1) gives

$$I_v \quad \propto \quad T_H\, kT^2 \exp((\mu+E_r-V_m)/kT)(1-\exp(-2E_r/kT))$$

$$+ \quad T_L T\,(V_m-2E_r)\, \ln\left[\frac{1+\exp(\mu)kT)}{1+\exp((\mu-2E_r)/kT)}\right] \qquad (5.3.2)$$

in which the last term represents an estimate of the contribution from the electrons with low transmission probability.

This expresssion shows that two distinct behaviours can be identified. For systems consisting of high barriers with a low resonant energy, E_r, and Fermi level, μ, the first contribution is exponentially damped, especially at low temperatures. Under these conditions the first contribution becomes negligible and I_v demonstrates a temperature

dependence

$$I_v \propto T_L T\,(V_m - 2E_r)\,\ln\left[\frac{1+\exp(\mu/kT)}{1+\exp(\mu-2E_r)/kT}\right]$$

The other behaviour arises for systems with low barriers and high Fermi levels. Then the first term is no longer exponentially damped and the dependence of the two terms on the transmission coefficient means that the first dominates. This leads to a more sensitive temperature dependence

$$I_v \propto T_H T^2 \exp((\mu+E_r-V_m)/kT)(1-\exp(-2E_r/kT))$$

An important property of the resonant systems is the peak to valley ratio. Again two distinct behaviours can be identified, depending upon the dominant term in (5.3.2). For the systems with the high barriers

$$\frac{I_p}{I_v} \approx \frac{f(2E_r/e)}{T_L(V_m-2E_r)} \qquad (5.3.3)$$

This expression was obtained for a model transmission coefficient with special properties, care is therefore needed when considering its consequences. The expression indicates that the peak to valley ratio can be increased by decreasing the minimum transmission probability T_L. This conclusion may be valid for some systems, however, in a realistic double barrier system a decrease in T_L could be accompanied by a decrease in the transmission resonance width. This decrease in resonance width could decrease the resonant current sufficiently for the current ratio to become proportional to T_L, contrary to (5.3.3). Despite these difficulties a conclusion which may be drawn from (5.3.3) is that the peak to valley ratio for these systems is temperature independent.

This behaviour is in contrast to the behaviour for systems with low barriers. Then the peak to valley ratio is

$$\frac{I_p}{I_v} \approx \frac{f(2E_r/e)\ln\left[\frac{1+\exp(\mu/kT)}{1+\exp((\mu-2E_r)/kT)}\right]}{T_H T \exp\left((\mu+E_r-V_m)/kT\right)(1-\exp(-2E_r/kT))}$$

Again this expression must be interpreted with care. The barrier height is usually much greater than kT so the dominant term will be

$$\exp - (\mu + E_r - V_m)/kT$$

which shows that the ratio will be very temperature dependent.

This qualitative analysis has shown that for highly doped contacts at low temperatures the peak current is temperature independent, whilst at low doping the peak current is temperature sensitive. These results are confirmed in figures (5.1) and (5.2), which shows the temperature dependence of the peak current for a double barrier system with contact Fermi levels at 0 meV and 52.8 meV respectively. The system with the high doping density has a temperature independent peak current until 200 K whilst the low doping density gives rise to a temperature independent peak current below approximately 50 K. These figures also show the temperature dependence of the current at voltages less than and greater than the resonant value. These show how the temperature independence of the peak current for systems with highly doped contacts coupled with the temperature dependence of the current at voltages greater than the resonant value means that the peak disappears at a lower temperature for the highly doped contacts.

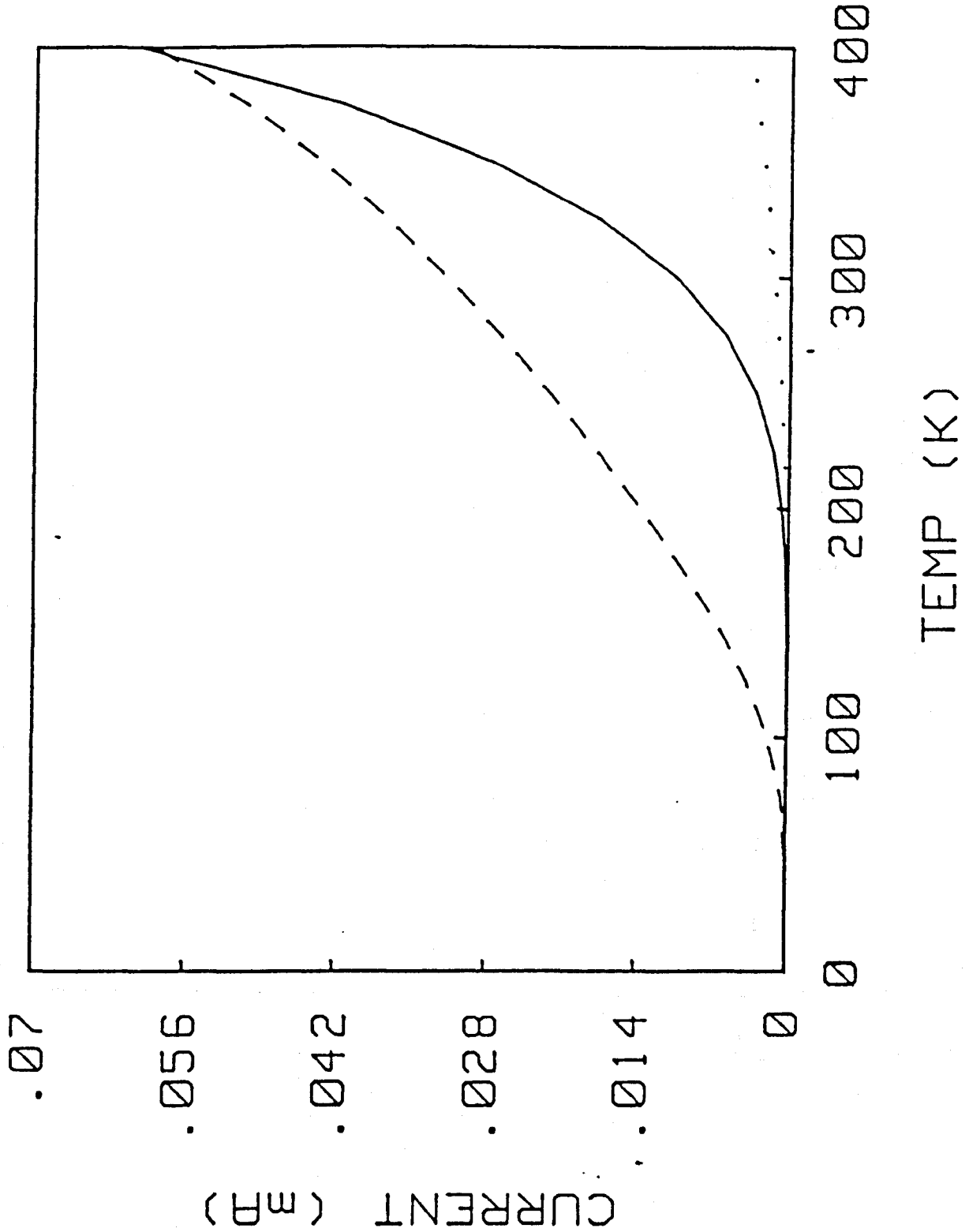

Figure 5.1 The temperature dependence of the current at three voltages for a double barrier with a Fermi level at 0 meV
(—) after resonance, 300 mV
(--) near resonance, 135 mV
(...) before resonance, 10 mV

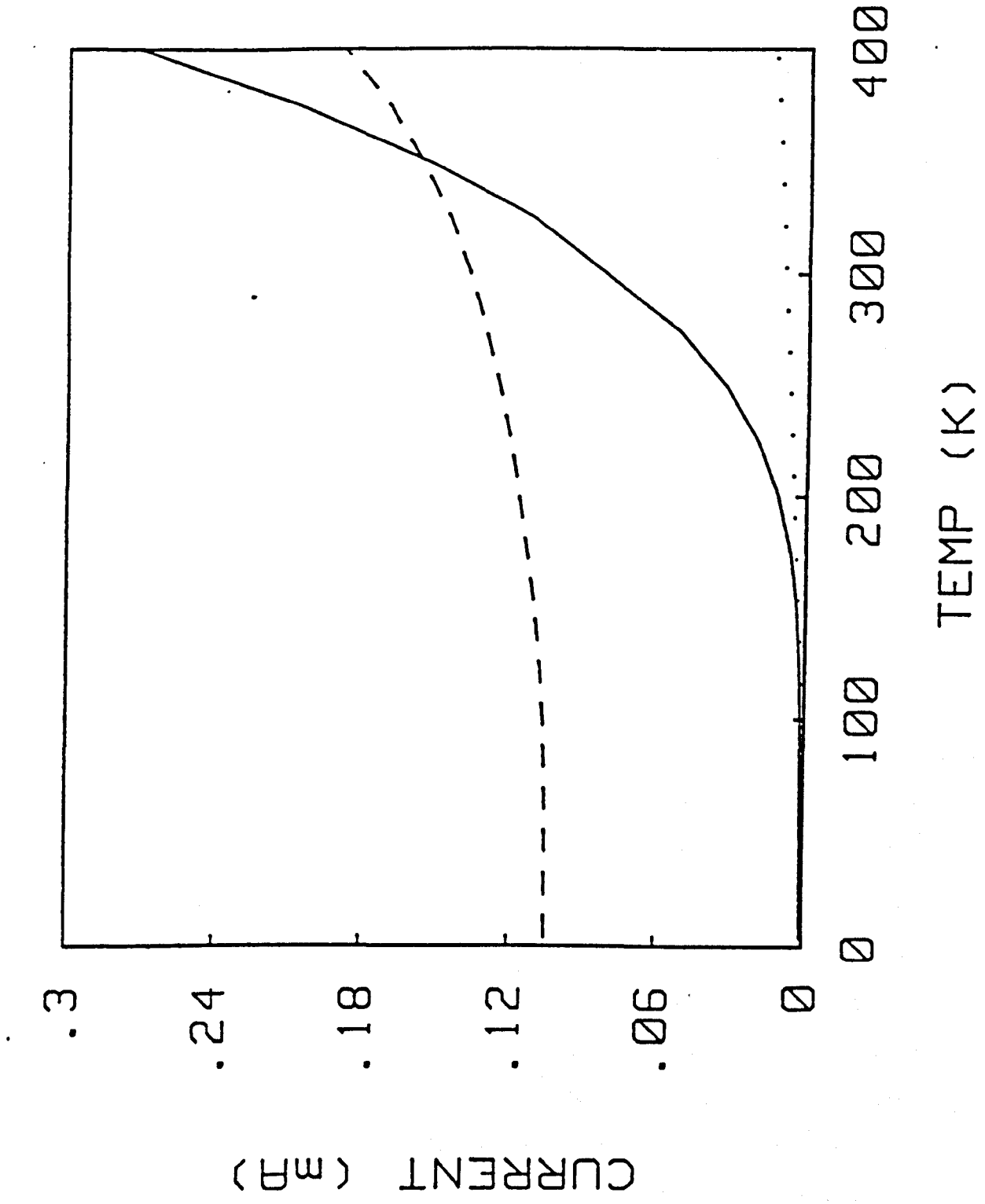

Figure 5.2 The temperature dependence of three voltages for a double barrier system with a contact Fermi energy of 52.8 meV
(—) after resonance, 300 mV
(--) near resonance, 135 mV
(..) before resonance, 10 mV

5.4 Numerical integration technique and the Fermi level

The remainder of this chapter will be devoted to discussing results obtained by numerically integrating the current density expression (4.2.4). Before any numerical integration the integrand must be considered to ensure that the integration is performed accurately.

The supply function D is a smooth function of energy, as is the function which occurs in the second integrand in (4.2.4)

$$(1 + eV/\Sigma)^{\frac{1}{2}} - 1$$

Unlike these smooth functions which cause no difficulty, the transmission coefficient may contain sharp resonant structure. To perform an accurate numerical integration in the presence of such sharp structure, without excessive use of computer resources, the range of integration was divided. Each division was then integrated using Simpson's rule with particular care taken in any region containing a resonance.

To enable a division of the range in this manner a method of finding the resonance positions and widths was required. Since it was desirable to use the same method for any potential a general method was adopted in which the transmission coefficient was evaluated to locate the peaks before any integration was performed. The algorithm used also included any sharp edges with the resonances to be integrated carefully. The accuracy of the general program was checked against the results of a program which used the special properties of double barrier systems to find resonances.

Once the range of integration had been divided, Simpson's rule was used on

each region. The use of a constant number of steps meant that the divisions containing resonances, being smaller, were integrated more carefully.

The final requirement before current voltage characteristics can be calculated is for a method for obtaining the Fermi level from the given doping density. The dopants are all one species and the current expression (4.2.4) assumes a parabolic band. This means that a simple balance equation can be used, Blakemore (78). For a parabolic band with a Fermi level η kT, the number of electrons n_e is

$$n_e = N_c F_{1/2}(\eta)$$

where $F_{1/2}(\eta)$ is the half order Fermi-Dirac Integral and $N_c = 2(m^*kT/2\pi\hbar^2)^{1/2}$. The number of ionised donors from a dopant density N_D with a donor level εkT below the conduction band and degeneracy β^{-1} is

$$N = \frac{N_D}{1+\beta^{-1}\exp(\varepsilon+\eta)}$$

The equilibrium condition used to find the Fermi level is therefore

$$N_c F_{1/2}(\eta) = \frac{N_D}{1+\beta^{-1}\exp(\varepsilon+\eta)} \qquad (5.4.1)$$

This condition was solved for η by the method of bisection with the Fermi-Dirac integral evaluated by two methods;

(1) Direct evaluation by Simpson's Rule, which is simple but lengthy.

(2) The use of approximate analytic expressions, correct to within two percent over the whole energy range.

$$\frac{\exp(\eta)}{1+0.27\exp(\eta)} \qquad \eta < 1$$

$$4(\eta^2+1.7)^{3/4}/3\pi^{1/2} \qquad \eta > 1$$

The accuracy of the approximate expressions is not sufficient to guarantee that their use in solving (5.4.1) will lead to accurate Fermi levels. However, using the first method to check the results of the second it was found that the calculated electron densities agreed to within a few percent. The anâlytic approximates were therefore used for simplicity.

The results shown in figure (5.3) for the temperature dependence of the Fermi level for systems with high and low doping density, show that high doping gives rise to a Fermi level which is almost temperature independent whilst the Fermi level at low doping is very temperature sensitive, a result which has been used in sections 5.2 and 5.3.

5.5 Double barrier systems

Several authors have observed the current voltage characteristics of systems containing more than one GaAlAs layer (53, 54, 55, 4, 5). In the model used in this thesis, there is no attempt to solve Poisson's equation. The application of an external voltage to a two terminal device is therefore assumed to give rise to a linear voltage drop across the system separating the contacts. This assumption restricts the use of the model to simple systems. There is some evidence that this assumption is not valid for systems containing many GaAlAs layers (75), therefore the present discussion is limited to systems containing only two GaAlAs layers. The data which will be considered in detail is amongst the most recently published data, (56).

In this paper all the data that can be expected is given. However, there still remains two details needed in the model about which no information

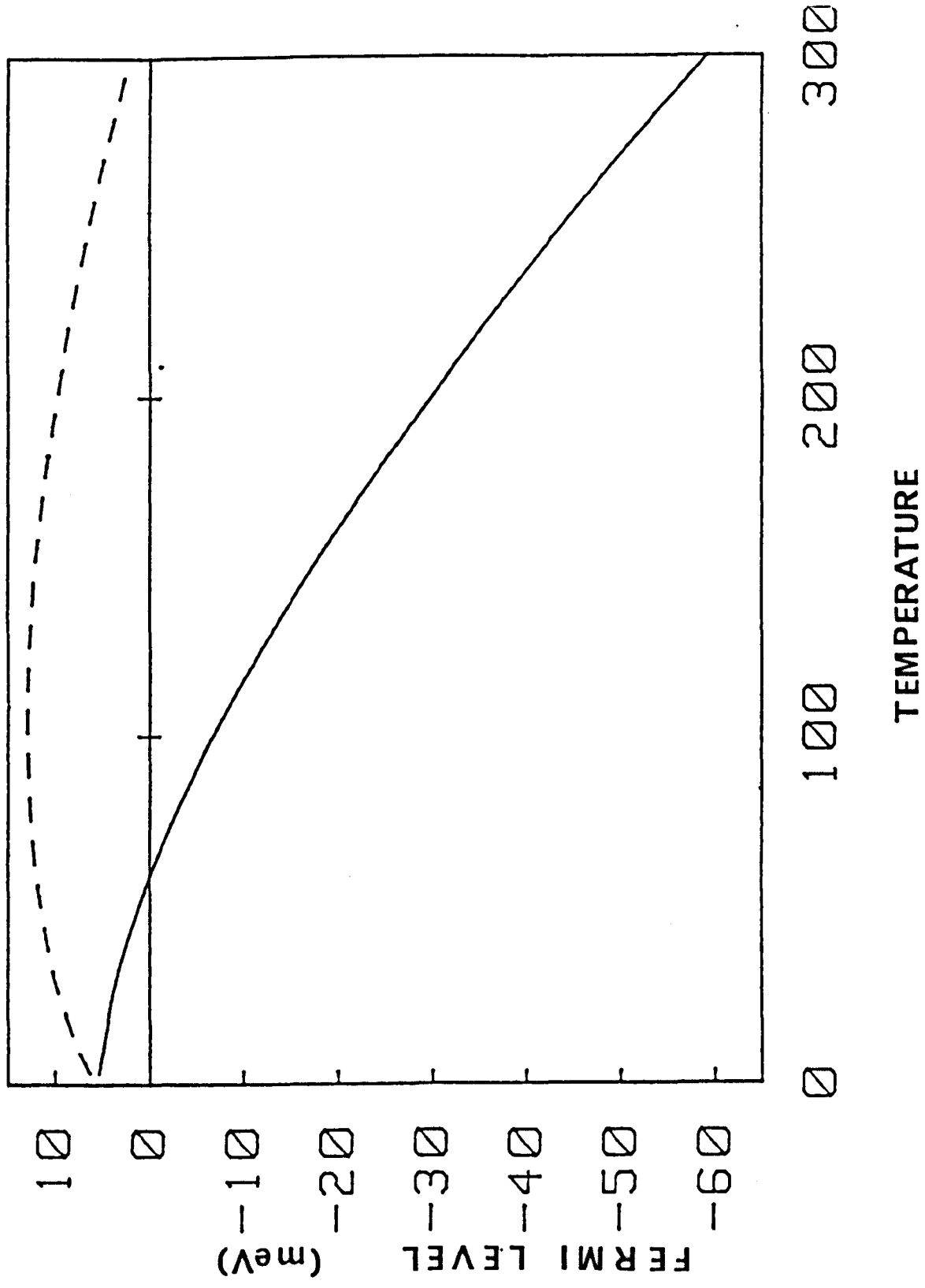

Figure 5.3 Temperature dependence of the Fermi level for two dopings
(—) Donor density, 5×10^{16} cm^{-3}
(--) Donor density, 10^{18} cm^{-3}

could be obtained, the distribution of the applied potential and the bandgap discontinuity. The system described by Sollner was therefore chosen for detailed modelling to fix these final details. The parameters given by Sollner and used in the modelling are given in Table (5.1).

As discussed in Chapter 1 the accepted value for the barrier height which should be used to model a $Ga_{1-x}Al_xAs$ layer is

$$V = 0.62(1.247x) \quad eV \quad (x < 0.45)$$

There is some uncertainty in this parameter, so that it was decided to use 0.60 or 0.65 instead of 0.62 to assess the affect of any uncertainty.

Since the contacts are designed to be conducting it was assumed that all the applied voltage is dropped across the two layers of GaAlAs and the separating layer of GaAs. There are many possible ways in which the potential drop could be distributed amongst the three layers. Rather than model all possibilities it was decided to model two extremes. In one, each AlGaAs layer supports the same field but the GaAs none, whilst in the other the potential is distributed so that each layer supports the same field. The first of these models would be consistent with the GaAs layer retaining a sufficiently high electron population to be considered conducting, whilst, the second is consistent with the thin GaAs layer being denuded of electrons by the action of the applied voltage, the electrons moving into the collector.

The two conduction band discontinuity formulae and the two potential distributions give rise to four model current voltage characteristics to be considered. The outstanding feature of the characteristics which can be compared are the position and height of resonance. These are listed

TABLE 5.1

Property	Value
Doping density of contacts	$10^{18} cm^{-3}$
Fermi level	12.96 meV
Electron temperature	100 K
Barrier width	50 Å
Barrier separation	50 Å
Area of device	$2.25\pi \mu m^2$
Aluminium fraction	35%
Electron mass	$0.067 \times 9.109 \times 10^{-31}$ kg

in Table 5.2, and a typical result shown in figure (5.4).

A comparison of the predicted resonant voltages to the observed value is disappointing, with all the values being only one half that observed.

One conclusion from these results could be that the model with its various assumptions, is too simple. A conclusion which would require the adoption of much more complex models. For the model to be valid the disagreement must arise from a source other than the model itself. One possible source is the values for the parameters used in the model.

The value of Fermi energy used in the current expression corresponded to the expected doping level. Any dopant level fluctuation would require a different Fermi level which could improve the agreement.

The electron temperature used assumed an electron distribution in thermal equilibrium with the lattice. Since the current is controlled by the electrons at the contact/AlGaAs interface the important electron population may not be in equilibrium with the lattice. A possible mechanism which could drive the electrons away from thermal equilibrium is the energy gained by the electrons crossing from emitter to collector. This energy could, for high enough electron densities, be shared amongst the electrons until a distribution was established corresponding to a Fermi distribution with an electron temperature greater than the lattice temperature.

The other major parameters are the barrier characteristics, the barrier width, separation and mole fraction of Al.

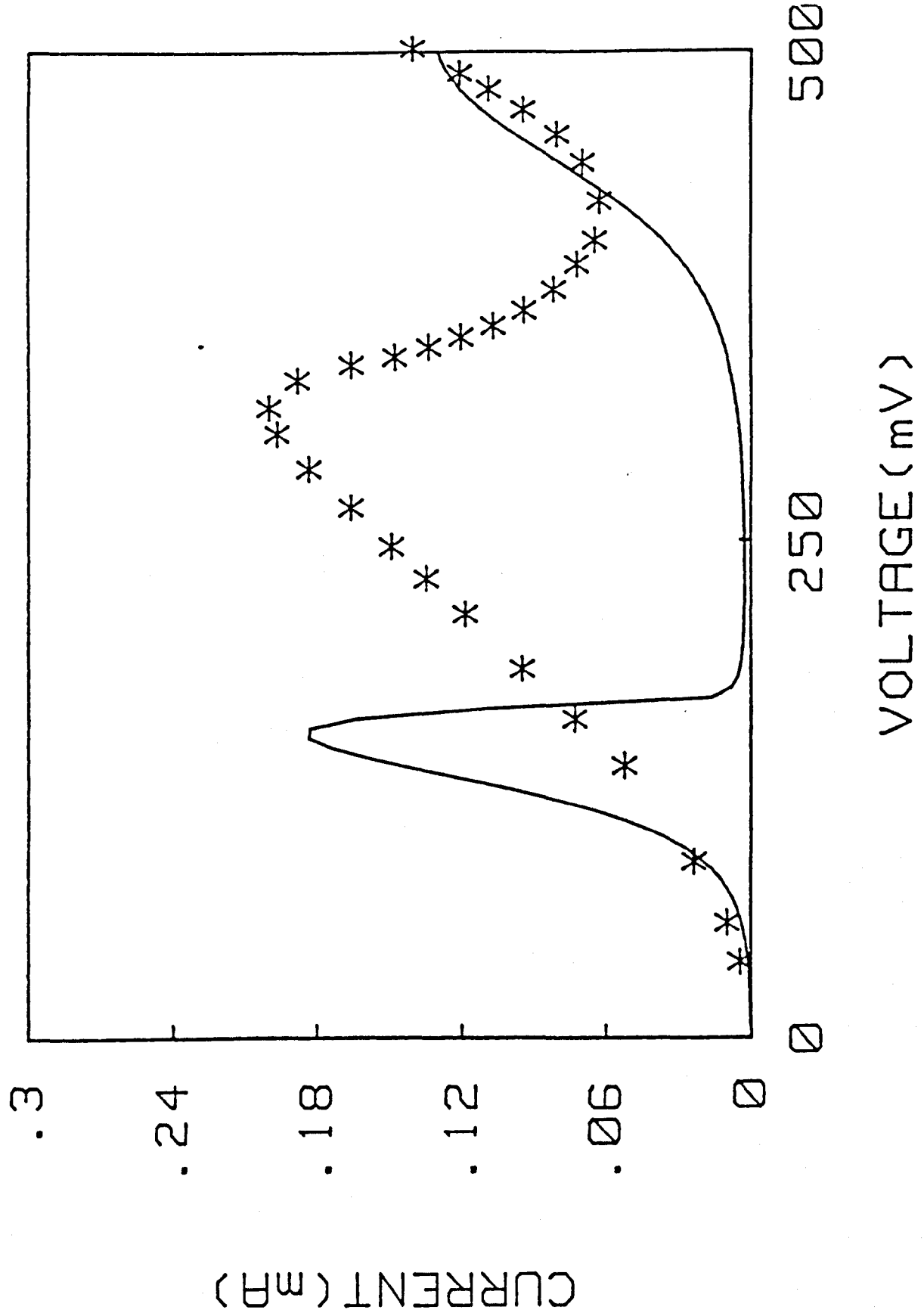

Figure 5.4 A comparison of the observed (*) and predicted (-) current voltage characteristics of Sollner's double barrier structure

TABLE 5.2

System No	Potential Distribution	$\Delta E_c/\Delta E_g$	I_{res} mA	V_{res} mV
1	Potential across AlGaAs	0.60	0.247	150 ± 5
2	Potential across AlGaAs	0.65	0.190	160 ± 5
3	Potential across both AlGaAs and GaAs	0.60	0.183	150 ± 5
4	Potential across both AlGaAs and GaAs	0.65	0.137	160 ± 5
	Observed		0.199	320

To study the affect of varying all parameters simultaneously over all possible values would take a prohibitively large amount of time and resources. It was therefore decided to study the affect of varying each parameter individually.

A variation of the Fermi energy showed the resonant voltage to be inversely proportional to the Fermi energy. However, the calculated value only changed by a few tens of millielectron volts. This insensitivity and the associated decrease in predicted resonant current means that a variation in doping cannot be the major cause of the disagreement.

Since the current expression was obtained assuming a Fermi distribution of electrons any deviation of the electron population from thermal equilibrium can only be examined using a Fermi distribution characterised by an elevated temperature. The temperature in the current expression was therefore varied, the contacts being characterised by the same temperature. This assumption can be justified because the contacts are only separated by 75Å of electron free material, so that there is very little screening of one distribution from the other. With a fixed Fermi level the position of the resonant voltage was found to only vary by approximately 10 mV as the temperature varied from 4 K to 450 K. It was therefore concluded that a different electron temperature alone was insufficient to explain the lack of agreement.

The use of full current calculations to study the dependence of the resonant voltage upon the barrier parameters can be avoided. For all the systems studied it is found that the relation

$$V_r \approx 2E_r(0)/e$$

where V_r is the resonant voltage and $E_r(0)$ is the resonant energy level with no applied voltage, was correct to approximately 10 mV. The dependence of the resonant voltage upon the system parameters can therefore be obtained from the dependence of $E_r(0)$. Again it was found that any variation in the parameters, within acceptable limits, was insufficient to explain the lack of agreement between the calculated and observed resonant voltage.

This examination of the affect of varying the parameters in the model leads to the conclusion that the lack of agreement between the calculated and observed resonant voltages cannot be attributed to a variation in any one parameter. The lack of sensitivity to any parameter suggests that a variation in all parameters would also fail to obtain agreement.

Although none of the four model systems agree with the resonant voltage, two systems do agree with the resonant current to within ten percent. This apparently large error is acceptable because of the simplicity of the model and the sensitivity of the resonant current to the parameters used in the current expression. The parameters used could be altered, within acceptable bounds, to fit the resonant current without affecting the predicted resonant voltage by more than a few millivolts. Such a fitting exercise would be lengthy and, with five variable parameters, unlikely to produce a unique set of parameter values. It is therefore more instructive to accept the values given by assuming that the device was manufactured as designed, whilst remembering that a best possible fit has not been obtained.

Of the two best model systems only system 3 will be considered further.

5.6 Affect of external devices

To find the cause of the disagreement in resonant voltages consider the experimental technique. This involves measuring the voltage across the external contacts of the sample and not the voltage across the GaAlAs layers alone. Assuming the difference can be modelled as a series device the experimental measurement V_E and the theoretical prediction V_T are linked by the expression

$$V_E = V_T + V_D$$

where V_D is the voltage across the series device. Consider two simple models for the series device.

First, assume that the series device can be modelled as a diode, with a current voltage relationship

$$J = J_s \, (\exp(eV_D/kT)-1)$$

giving the saturated current density J_s as a fitting parameter. The study of the dependence of the resonant voltage on the parameters used in the current expression has shown it to be insensitive to parameter variations. This property makes the resonant voltage an ideal fitting criterion for J_s , since any errors in the values used in the current expression only cause a small error in V_r . The expression for J_s is then

$$J_s = \frac{J_r}{(\exp(e\Delta V_r/kT)-1)}$$

where ΔV_r is the difference between the observed and predicted resonant voltages, and J_r is the current density at resonance. The voltage added to the theoretical value to give the predicted voltage for a current

density J is then

$$V_D = \frac{kT}{e} \ln \frac{J+J_s}{J_s}$$

The logarithmic dependence of V_D on J means that points of similar current density are shifted by very similar voltages. This explains results such as those shown in figure (5.5a) where the peak position is moved 160 mV with only a slight distortion in the peak shape.

The assumption of a series diode can therefore lead to a fit to the observed resonant voltage and current, but, the overall fit to the peak shape is not good enough to claim a fit between the observed and predicted characteristics.

If the series device is assumed to be a resistance, then

$$V_D = IR$$

Again the parameter, R, is fitted using the resonant voltage, so that

$$R = \Delta V_r / I_r$$

where I_r is the current at resonance. Using this formula to obtain the resistance required to explain Sollner's data gives R=930 Ω and with the predicted voltage V, obtained from the theoretical voltage and current, V_T and I_T

$$V = I_T R + V_T$$

a comparison between the predictions and observations shows a much improved agreement. Not only are the peak voltage and current predicted approximately but so is the shape of almost the entire curve, figure (5.5b).

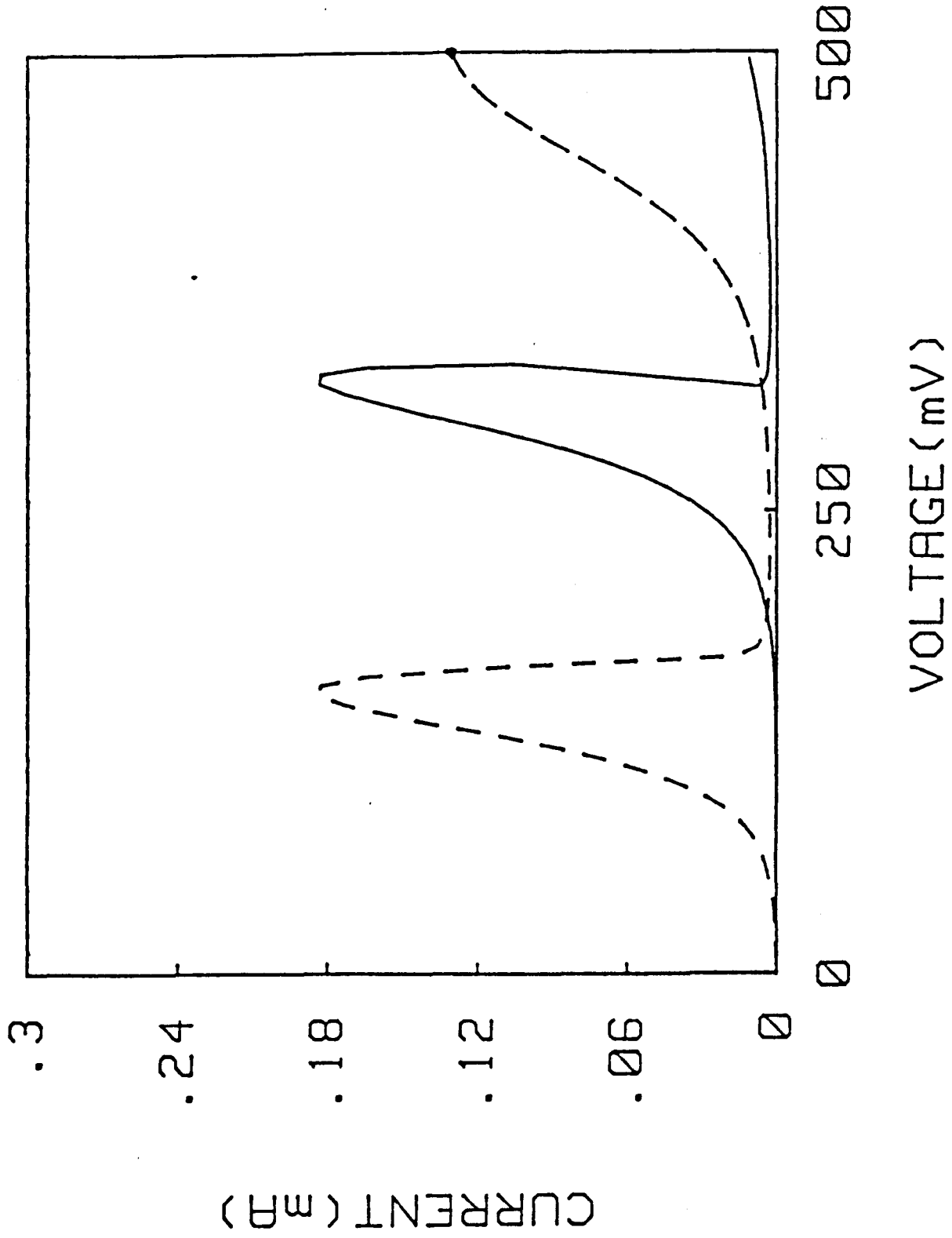

Figure 5.5a The affect of a diode in series with the predicted results
(--) predicted results
(—) predicted results and diode

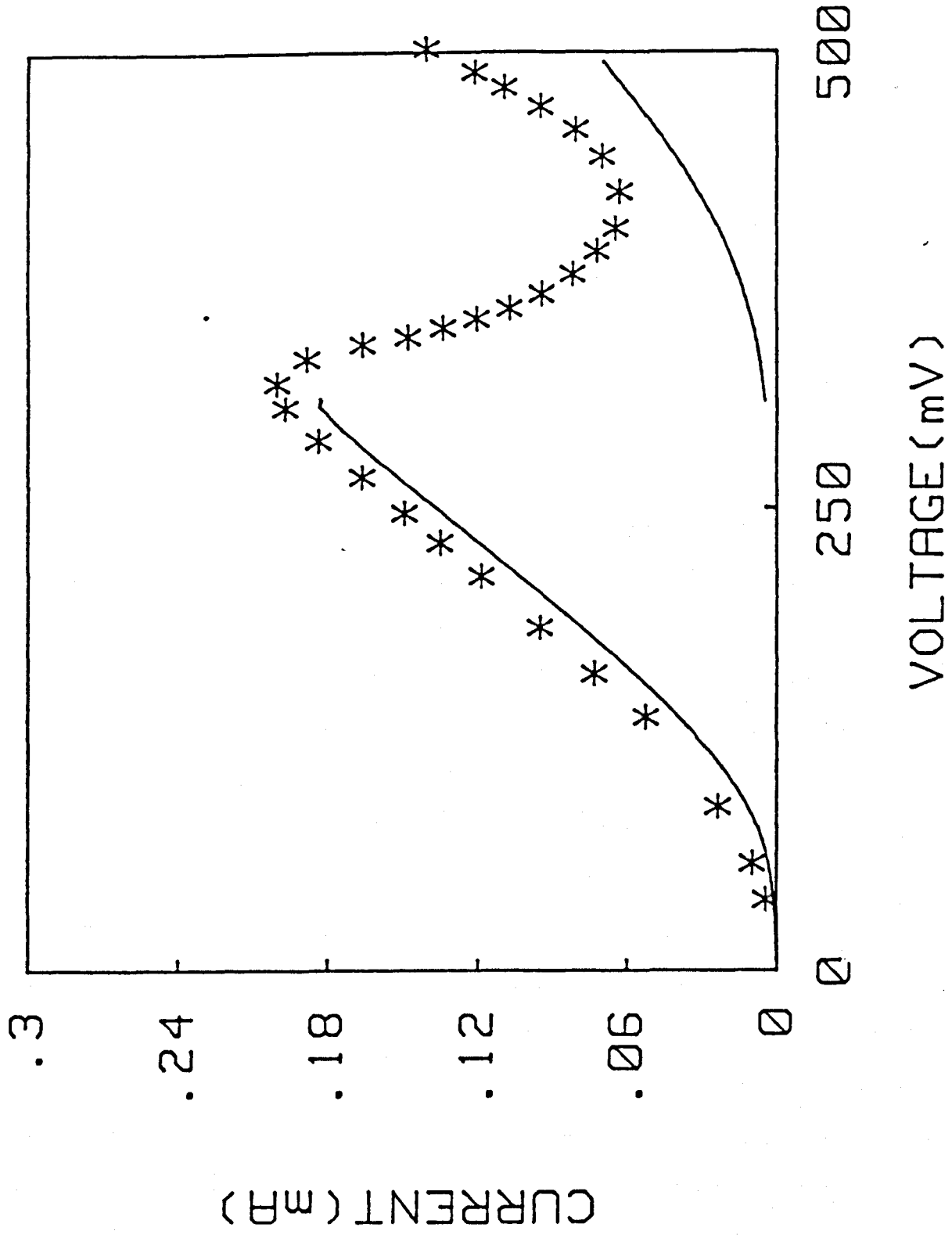

Figure 5.5b Comparison between the observed (*) and predicted characteristics. The predicted characteristics include a resistor in series with the structure. (The predicted characteristics are plotted to exclude the multi-valued part of the curve.)

To understand the cause of the remaining disagreement, the region in which the current decreases, the affect of a series resistance must be considered. The resistance causes a voltage shift which is proportional to the current, this shifts the resonant peak to much higher voltage than the region of small current which immediately follows it. This differential shift leads to a predicted current voltage characteristic which is multivalued under the peak.

The exact current voltage characteristics which are predicted now depends upon the experimental technique used. Consider figure (5.5c), if a constant voltage technique is used with increasing voltage, a discontinuous jump is predicted from point A to point B as the voltage sweeps through A, a type of behaviour observed by Shewchuk et al (5).

In contrast to the constant voltage technique a constant current technique would show a discontinuous change from A to C with increasing current. A measurement using decreasing current would not necessarily give a discontinuity from C to A. As the current increases beyond A, C is the only possible next voltage to produce an increased current. However, as the current decreases beyond C there are three possible voltages which would generate the required decrease in current. The system is most likely to follow a small change in current with a small change in voltage. This would mean that discontinuity in voltage will not be observed until a current value below C, then a transition would occur to a point other than A, leading to a hysteresis loop. A similar behaviour could be expected in a constant voltage measurement since as the voltage increases the transition A to B must occur but as it decreases there is no reason to expect the transition B to A.

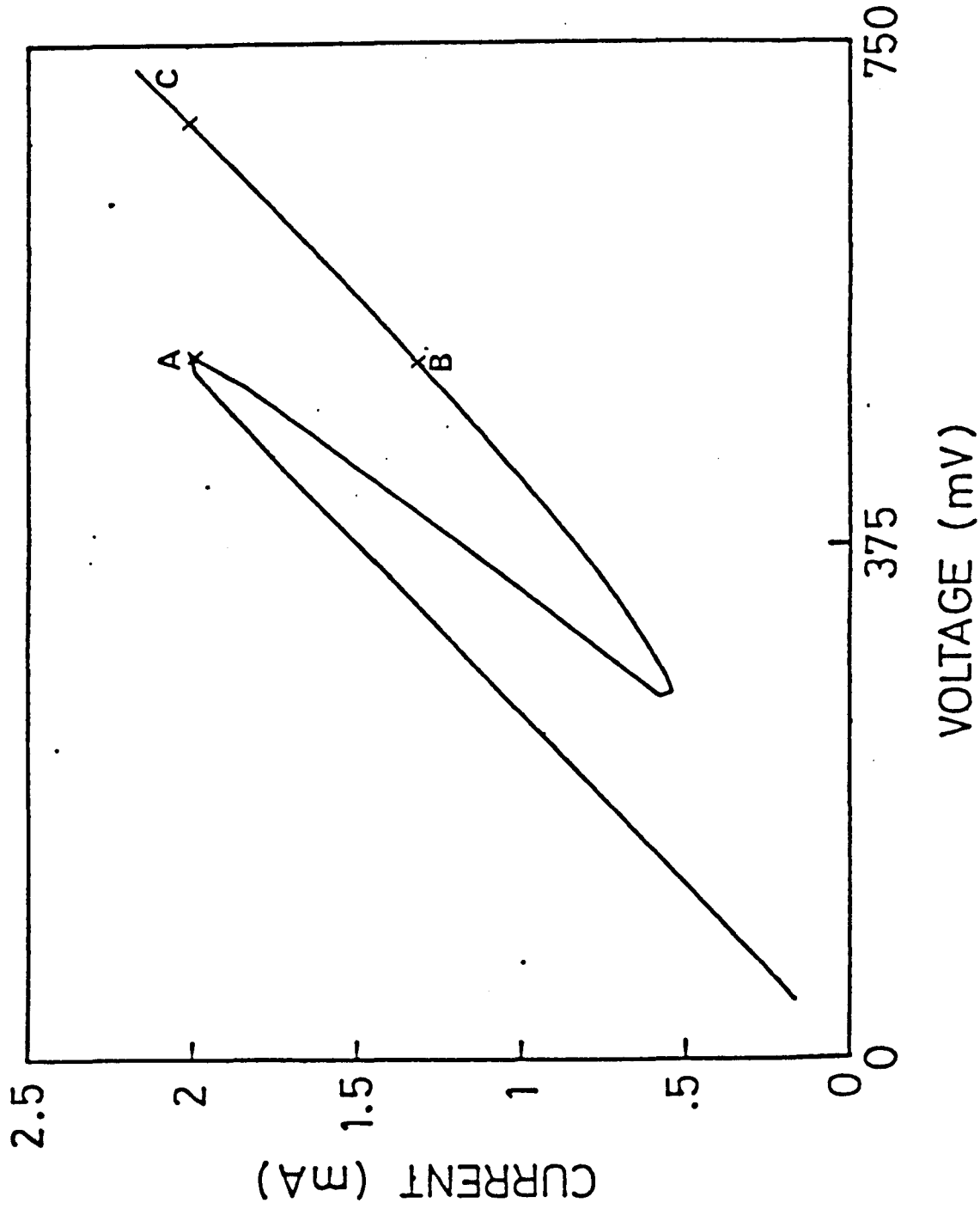

Figure 5.5c A resonant peak with a series device showing the various possible discontinuities
A-B constant voltage mode of measurement
A-C constant current mode of measurement

The mechanism to cause these discontinous changes is not included in the model and definite statements that the transitions C to A and B to A will not occur are impossible. However, the model predicts hysteresis loops in the current voltage characteristics, the exact nature of the loops depending on the technique used to obtain the characteristics. Recently the predicted hysteresis has been observed in a double barrier system at 77 K (57). The nature of the observed loop indicating a constant current mode of measurement.

There is a third possible technique for obtaining a current voltage characteristic which uses a small AC signal imposed on a DC background. The small AC signal allows the electronic differentiation of the current voltage characteristics.

The shape of the Sollner curves, which is characteristic of all the data sets they obtained, may be due to the use of this technique. Although the exact experimental technique used is not discussed the presence of differentiated characteristics supports the assumption that a small AC signal was used. The remaining disagreement between predicted and observed behaviour may be due to this technique, the smooth decrease in current being due to the ammeter averaging the current arising from voltages on each side of the discontinuity, as the voltage oscillates.

Another important consequence of the series resistance is the decrease in the peak to valley ratio. This is important in determining the power output possible from the device when it is oscillating (4). The shift in the peak position means that the true minimum in the current voltage characteristics of the layer system becomes inaccessible. For the system described by Shewchuk (5) the peak to valley ratio decreases from 50 to

2 if a resistance of 300Ω is placed in series with the layers.

The conclusion of this section is that the observed characteristics of double layer systems depend upon the experimental technique used. The series resistance causes hysteresis and a decrease in the observable peak to valley ratio. The fit obtained by assuming a series resistance could be improved by assuming that the manufactured device did not quite reach the design specification. This fitting process would be lengthy and the asymmetry always oserved in these structures (4, 5) would make it fruitless until the source of the asymmetry is known.

5.7 Asymmetry in double barrier systems

Double layer structures are usually designed to be reflection symmetric about the centre of the central GaAs layer. If this reflection symmetry is achieved during fabrication the current magnitude is independent of the direction of the applied voltage, a behaviour that has not been observed in any published data.

If external sources are excluded there are still several possible sources of asymmetry. A study of all of these would be lengthy. This discussion will therefore be restricted to two sources, a difference in contact doping and a difference in the barrier parameters.

A difference in contact doping density could arise during fabrication, by a process similar to that which produces a transient in the Aluminium flux (58). Since the transient would be correlated with the initialisation of the growth of the second GaAs contact it would occur in the position to give the maximum impact on the current voltage characteristics, hence,

only a small transient may be necessary to give rise to an observable asymmetry.

As an example of the affects of different doping densities consider the results shown in figure (5.6) where a doping difference of two has been assumed in a system based on that fabricated by Sollner (56).

The predicted resonant voltages are equal, to a resolution of 10 mV, however, the bias dependence of the current would mean that a series resistance of 930Ω, needed to explain the previous results, would produce a difference in observed resonant voltages of 46.5 mV.

Another affect of the doping difference arises from the experimental definition of the voltage origin, as the voltage which produces zero current. Consideration of the current density expression shows that zero current flows in a system with a Fermi level difference if

$$eV = \mu_e - \mu_c$$

This means that the experimental and theoretical voltage scales are shifted with respect to each other, with the experimental origin being at $\mu_e - \mu_c$.

Since the contact with the higher doping produces the higher current, the shifts produced by the change of origin and difference in shifts produced by the resistance are different in sign. This means that the contact with the higher doping can be identified as the one which produces the higher peak current, but, it will not necessarily be furthest from the origin.

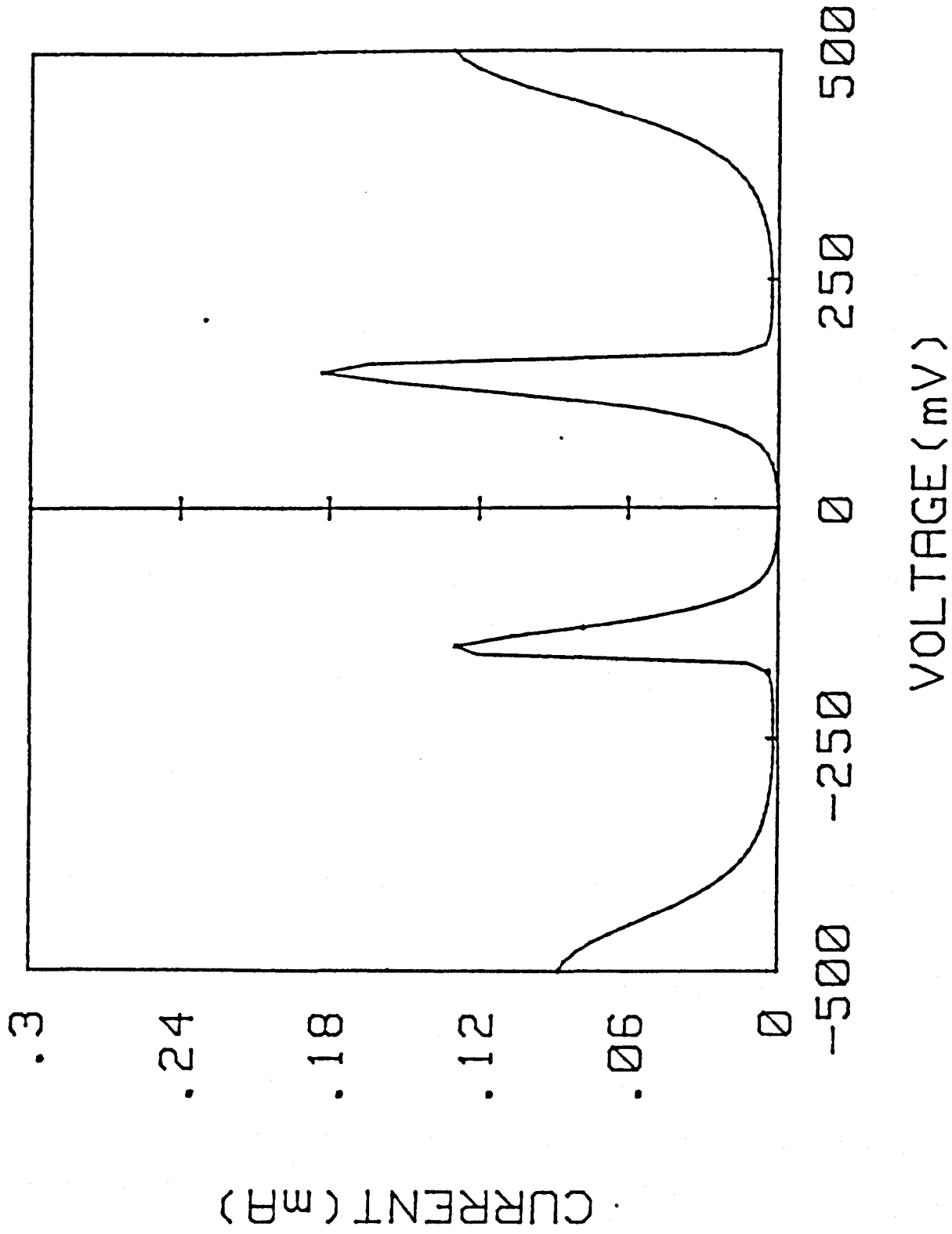

Figure 5.6 Affect of different doping. The temperature is 100 K and the two barriers have an Aluminium fraction of 35%. The two contacts have doping densities of $10^{18}cm^{-3}$ and $5 \times 10^{17}cm^{-3}$

An asymmetry in barrier width or composition is much less probable to occur acidentally, but intensional asymmetry may be useful.

In section 4.3 the WKB approximation was used to indicate that the transmission resonance attains its maximum value only if the transmission coefficients of the individual barriers are equal. The resonant current occurs when a finite bias is applied to a double barrier system. The asymmetry introduced reduces the amplitude of the transmission resonance. It should therefore be possible to fabricate a system with a more reflecting barrier next to the collector so that at a finite bias the transmission coefficients of the two different barriers become equal at the resonant energy, producing an increased peak current.

Without an analytical expression for finding the optimum combination of barriers the optimum can only be found by trial and error, a lengthy process. The choice of barriers can be reduced by imposing restrictions.

As an example consider the case in which an increased peak current is required at approximately the same voltage as a known system. Since the resonant voltage is set by both the energy of the transmission resonance at zero bias and the potential of the central GaAs layer at a finite voltage, the requirement for a constant resonant voltage means that both the barriers must have the same width as in the known system, whilst the barrier near the emitter must also have the same Al molefraction. This leaves one variable against which to optimise the peak current, the Al mole fraction of the barrier next to the collector.

A study of a system based on that fabricated by Sollner (56) showed that, under the conditions imposed above, the optimum solution gave an increase

in peak current of only 0.01 mA, from 0.18 mA to 0.19 mA.

This example indicates that if an increase in peak current is desired it is advisable to increase the doping in the emitter slightly rather than altering the layer parameters.

5.8 Other double barrier systems

An attempt in the previous section to maximise the peak current only gave limited success. Ricco has proposed an effective method of maintaining the barrier symmetry (59). Rather than apply the voltage across the contacts, the proposal is to apply it between the contacts and the central GaAs layer. As in other systems the resonance remains a fixed distance above the centre of this layer, the applied voltage bringing the resonance into the energy range occupied by electrons. A current is obtained by doping the contacts to different levels, and applying the small bias necessary to maintain the current flow.

This method can be shown to give rise to sharp resonances with large peak to valley ratios. A system, of two 41% Al, 50Å layers separated by 50Å of GaAs, with contact Fermi levels at 5 mV and 0 mV gives a current decrease by a factor of 314 over a range of 15 mV,with an actual peak to valley ratio of 750 over 40 mV. The current could be increased by increasing the doping difference, however, the peak to valley ratio, which is also important, decreases. If the same system had contacts with 10 mV and 0 mV Fermi levels, operated at 4 K the peak to valley ratio would be 10^4 and the current voltage characteristic would have a very sharp peak only 20 mV in width.

The problem with this design is how to contact the central GaAs layer which is only 50Å thick. A possible mechanism has already been demonstrated by Sollner (60). Rather than directly contacting the central region its potential is depressed by ionising donors in the AlGaAs and GaAs layers. One handicap of this method has already been observed by Sollner et al (60). They found that removing donors from the layers could enhance the current by two orders of magnitude, therefore reintroduction must be expected to cause a reduction in current.

Sollner et al modelled the observed characteristic. Although the calculated current was plotted on a scale 200 times that used for the observed current good agreement was claimed. The details of the model were not published, however, it is known that a temperature of 60 K was used. This elevated temperature being justified by the observation that the current was temperature independent below 100 K. This result, which is in in agreement with the results from section 3, was used to argue the presence of electron heating and hence the use of an elevated temperature. Their model failed to give the increase in current which occurs at the higher voltages. The conclusion was therefore reached that this increase is due to phonon assisted electrons.

Figure (5.7) shows that the current can be modelled to a factor of approximately 10. The high voltage tail is present and is caused by electrons with energy close to the barrier height. The increased ionisation lowers the barriers, moving the tail to lower voltage as observed.

Some of the overestimate in current may be caused by an overestimate of the dopant levels. Above, the interface doping was assumed to be $10^{18}cm^{-3}$

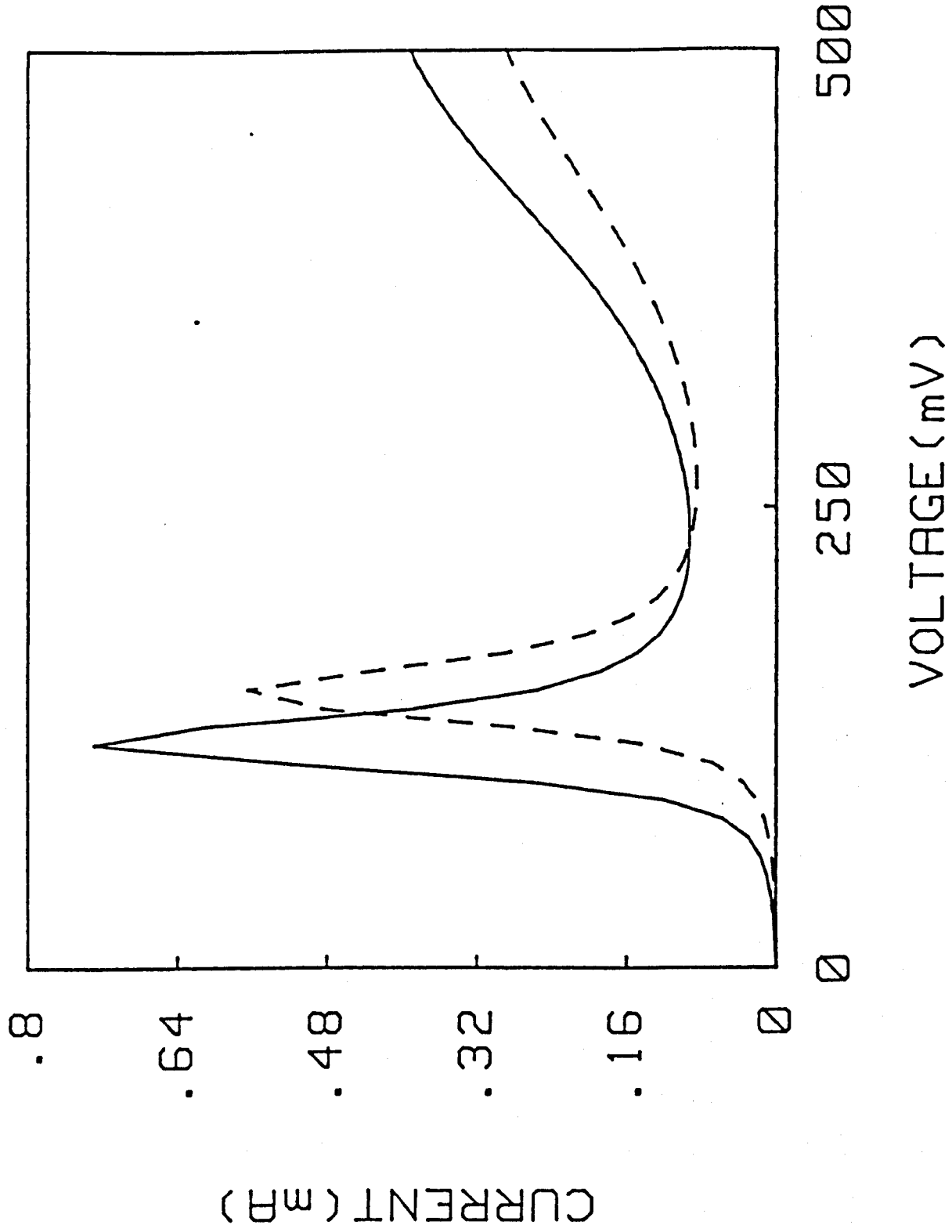

Figure 5.7 The affect of a depressed central layer on the current voltage characteristics
(--) system as fabricated
(—) system with central layer depressed 10 mV below the contacts
(The system is that fabricated by Sollner)

as designed. However, Sollner et al estimate a dopant density in the central layer of $4 \times 10^{17} cm^{-3}$. This is only 40Å from the interface and if it is assumed to be equal to density at the interface the peak current decreases to 0.41 mA. The remaining difference in current magnitude may be due to scattering by the ionised donors. Except for the difference in magnitude the model predicts the general behaviour of the system, in contrast to Sollner's model.

5.9 Systems with more than two layers

There have been several proposals to use systems with graded layer thicknesses to act as rectifying elements (61). The larger systems have become known as CHIRP superlattices. The explanation of how the devices should operate is always based on calculating the individual properties of each GaAs layer before combining these properties to predict the behaviour of the system.

To demonstrate the failure of this approach consider a system of three 50Å, 24.9% Al layers separated by 50Å and 45Å of GaAs. The simplistic approach would use the individual transmission coefficient resonances at 82.2 meV and 73.6 meV to predict a resonance with an applied field of 26.6 meV.

The transmission coefficient for the system was calculated. In figures (5.8a) and (5.8b) the positions and heights of the two resonant peaks are plotted as a function of voltage. It can be seen that there are two voltages corresponding to unit transmission coefficient. One of these values does approximately correspond to the predicted value. However, the resonant energy is too large for this resonance to have a significant

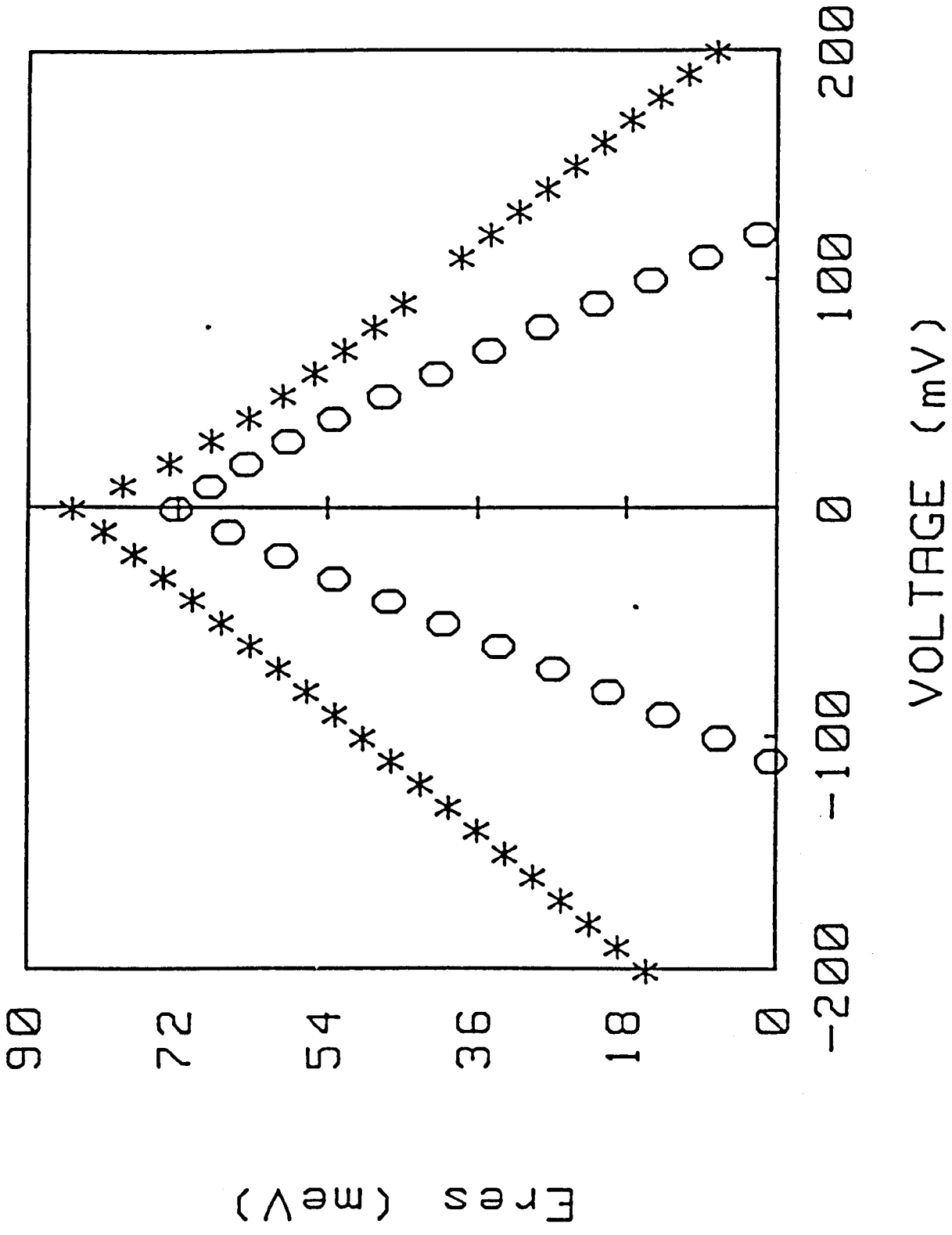

Figure 5.8a The position of the resonances in the three barrier system discussed in the text plotted as a function of the potential difference between the contacts

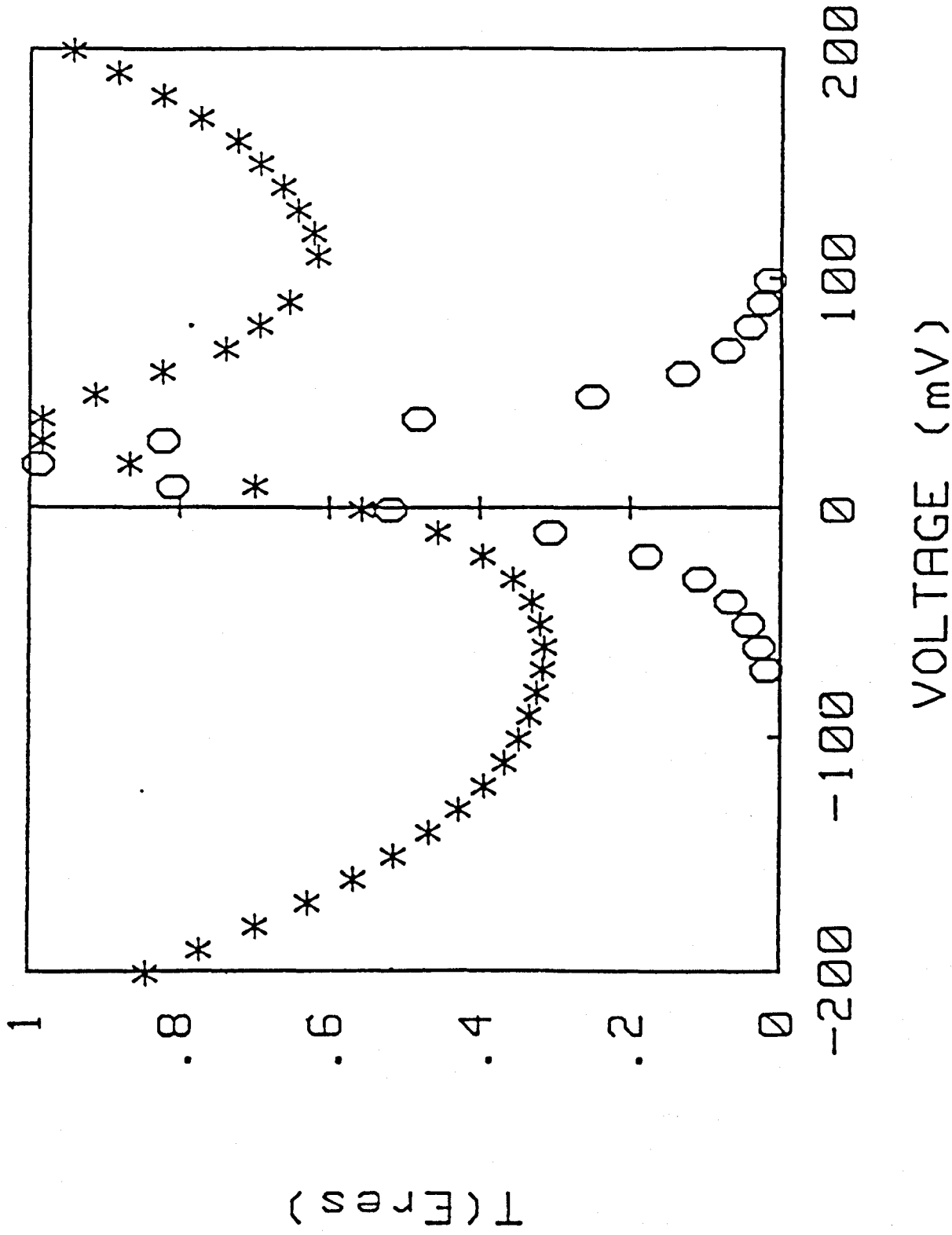

Figure 5.8b Amplitude of transmission resonances as a function of the applied voltage showing the rapid extinction of the lowest resonance. (The symbols are the same as those used in figure 5.8a)

impact on the current. The rapid decrease in its magnitude means that the current will be dominated by the other peak which maintains a high transmission probability until it moves below the conduction band edge.

The complex behaviour of the transmission coefficient gives rise to a complex current voltage characteristic as shown in figure (5.9). The characteristics show none of the behaviour of an element which could be used as a rectifier. All the peaks are due to the persistent resonance, the predicted resonance having no significant impact.

It may be possible for the simplistic approach to be valid for systems with a sufficiently slow variation in layer thicknesses. Nakagawa et al (76) have manufactured a system with the very slow variation of layer thickness. The observed current voltage characteristics showed a monotonically increasing current for forward bias with a small peak for reverse bias.

If the affect of a series resistance on the characteristics in figure (5.9) is considered, the large minimum following the reverse bias peak would lead to it being distinguishable for values of series resistance which would make the forward peaks indistinguishable. The observation of a small peak in reverse bias by Nakagawa et al may therefore be due to a large resistance in series with a characteristic similar to that in figure (5.9).

The overall conclusion must be that the graded layer devices cannot be described in a simplistic way and are not a good design for rectifying devices.

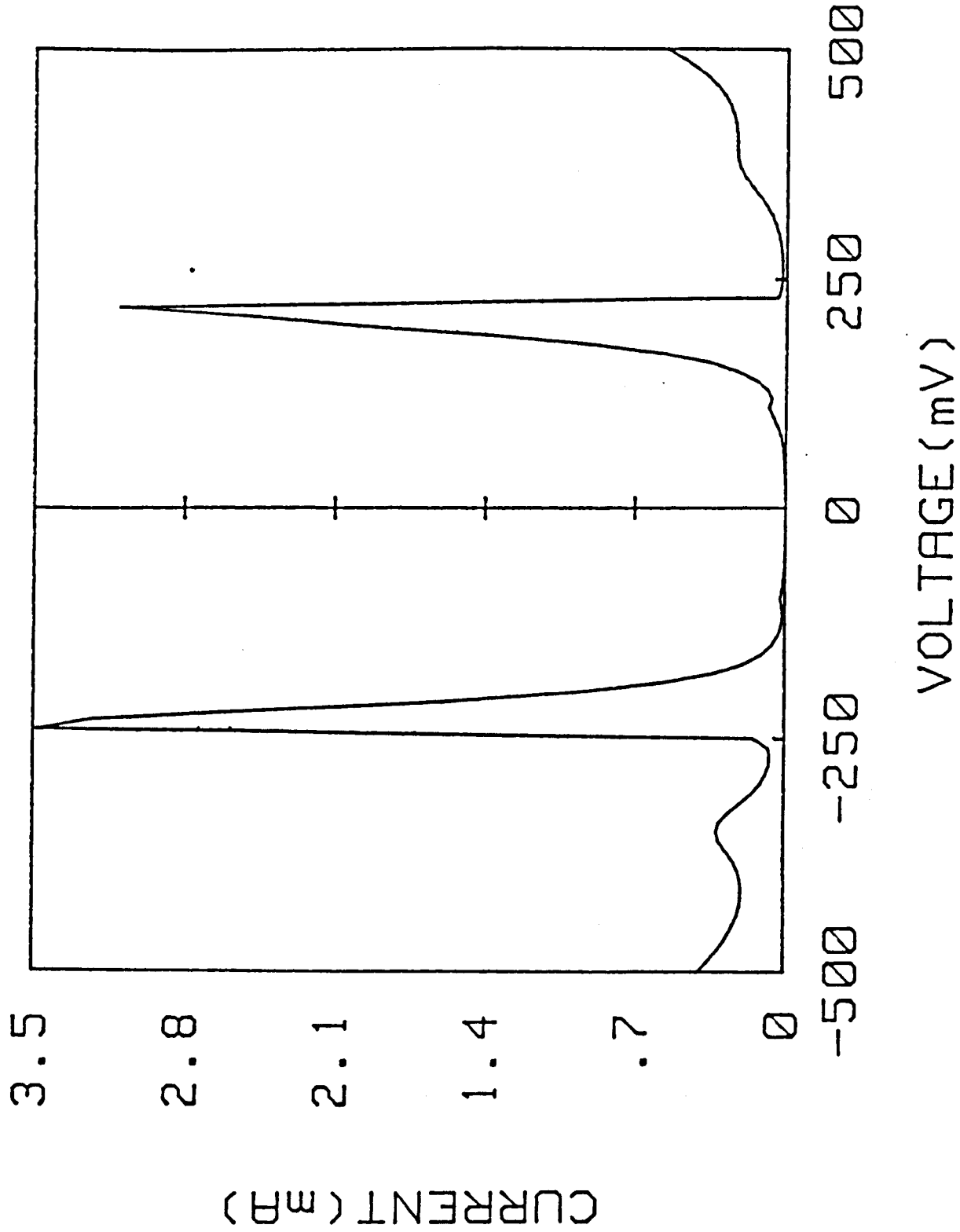

Figure 5.9 Current voltage characteristic for the three barrier potential discussed in the text. The calculation was performed for a temperature of 100 K with emitter (collector) doping levels of 10^{18} (5×10^{17})cm^{-3} for a 1.5 μm circular mesa

5.10 Triangular and parabolic well potentials

Unlike the graded layer structures Allyn et al (62) has shown that triangular potentials possess good rectifying characteristics. Of the three systems studied by Allyn two contained doped barrier layers. The reduction in current due to ionised donor scattering means that these two systems are not ideal for modelling. This section will therefore concentrate on the remaining system. This was fabricated by linearly increasing the Al flux during the depositin of a 500Å layer before abrupt termination of the flux.

The system is modelled as a 500Å layer over which the Aluminium concentration increases followed by a 10Å region over which it decreases. For simplicity the applied field is distributed equally over the 510Å. The predicted and observed characteristics are compared in figure (5.10). Since both predicted and observed characteristics show no current under reverse bias only forward bias is shown.

Figure (5.10) shows that good agreement is achieved between the observed and predicted behaviour if a diode is assumed to be in series with the barrier. The values of the saturated current density were 90 mAm^{-2} at 300 K and 1.5×10^{-23} mAm^{-2} at 77 K.

Hayes et al (63) have proposed using the triangular barrier as a hot electron spectrometer. To show how this would operate consider the dominate term from the current expression

$$J = \frac{2mkT}{(2\pi)^2\hbar^3} \int_0^\infty dE\ T(E)\ D(E) \qquad (5.10.1)$$

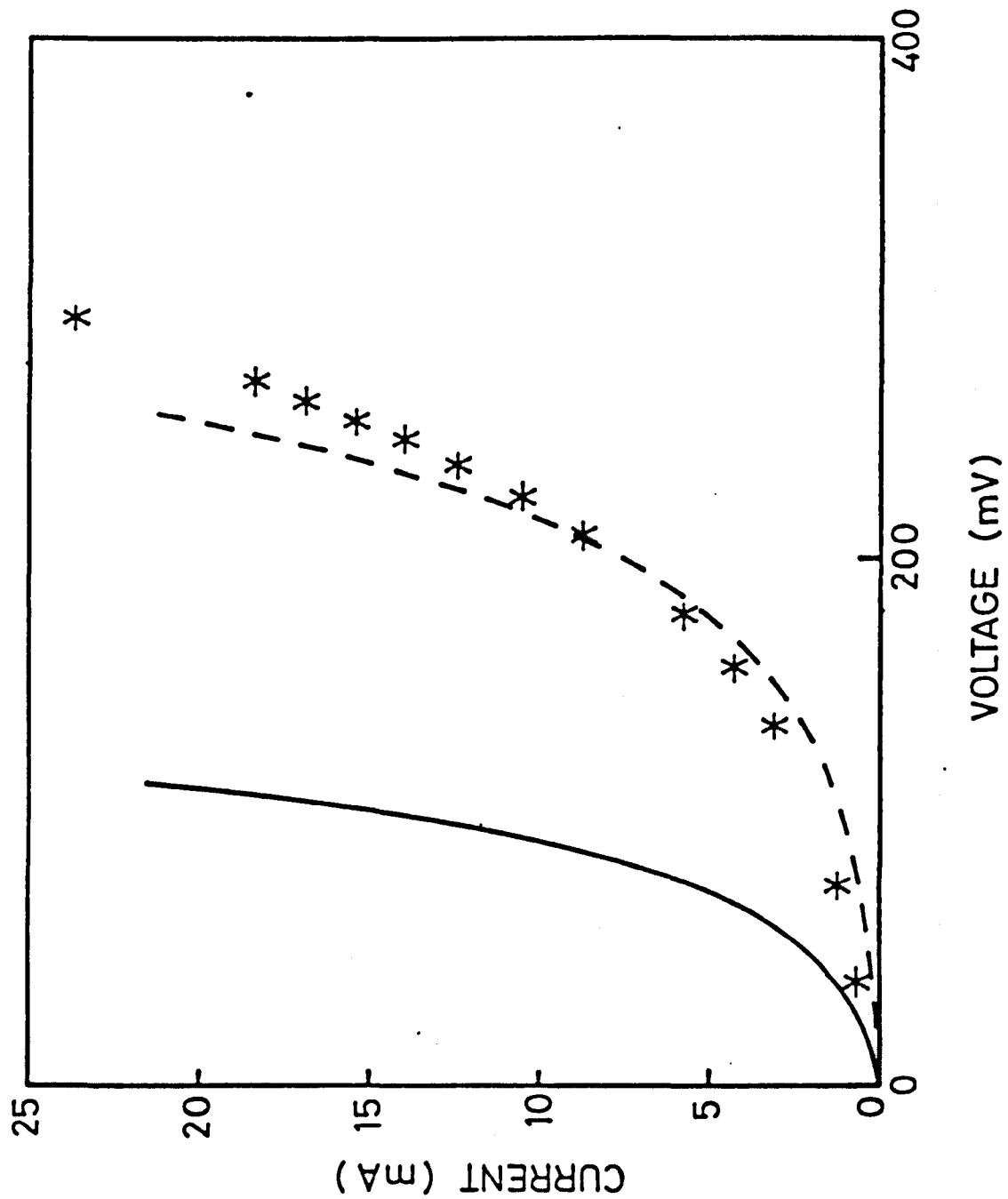

Figure 5.10a Results of modelling Allyn's triangular barrier at 300 K
* observed
— predicted
-- predicted + diode (J_s = 90 mAm^{-2})

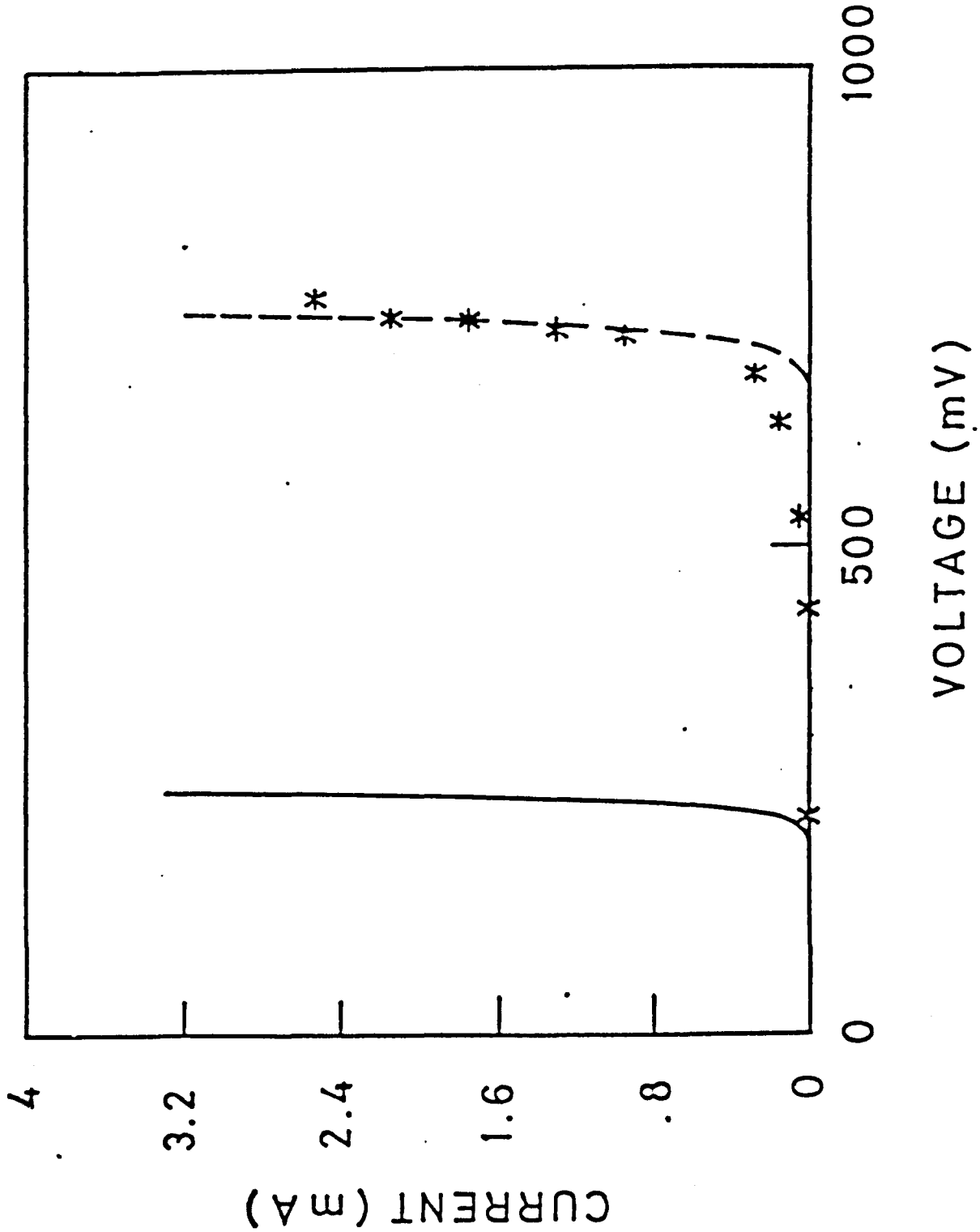

Figure 5.10b Results of modelling Allyn's triangular barrier at 77 K

* observed
— predicted
-- predicted + diode (J_s = 1.5 x 10^{-23} mAm^{-2})

In (5.10.1) D(E) is the result of the transverse integral and is not restricted to the analytic result obtained by assuming a Fermi distribution of electrons. If the transmission coefficient can be approximated as a step function at the maximum potential in the system, $\phi_m(V)$ then .

$$J = \frac{2mkT}{(2\pi)^2\hbar^3} \int_{\phi_m}^{\infty} dE \; D(E)$$

Differentiation with respect to the applied voltage V then gives

$$\left(\frac{\partial J}{\partial V}\right)_V = \frac{2mkT}{(2\pi)^2\hbar^3} \; D\,(\phi_m(V)) \; \left(\frac{d\phi_m}{dV}\right)_V$$

This shows that if the second term of (4.2.4) is neglected and the approximation for the transmission coefficient is valid, the derivative of the current can be used to study the supply function, not the electron distribution function as stated by Hayes et al. The neglect of the term from the current expression may be justified because the interesting results are obtained at high voltage. The validity of the approximation used for the transmission coefficient can only be demonstrated by a comparison of results.

A point which should be made is that the approximation for the transmission coefficient can only be expected to be valid if $\phi_m > 0$, a point neglected by Hayes et al.

The test system chosen was a triangular barrier which increased to a maximum Al mole fraction of 29% over 100Å and decreased back to zero over 20Å. The applied voltage V was distributed equally over the 120Å to give a maximum potential ϕ_m

$$\phi_m = \phi_o - 5eV/6$$

where ϕ_o is the potential corresponding to 29% Al (217 meV). The accuracy of the technique can be tested by comparing the derivative of the current with the supply function used in its calculation. The comparison in figure (5.11) shows that the derivative overestimates the electrons energy dispersion.

The transmission coefficient in figure (5.12) shows that it is the smoothness of the transmission coefficient which causes the observed differences between the supply function and the derivative. This smooth transmission coefficient is contrary to the results obtained by Kelly (64), where the oscillations observed are probably due to inaccuracy accumulated from the use of library subroutines to evaluate Airy functions. The smooth transmission coefficient also means that the resonance seen in the current, figure (5.13), is not due to a bound state, but, to the decrease in low energy transmission probability as the field is increased.

The results of this test system indicate that the use of a triangular barrier as a hot electron spectrometer is at best difficult and gives rise to misleading results.

Capasso and Keihl (65) have suggested using the evenly spaced energy

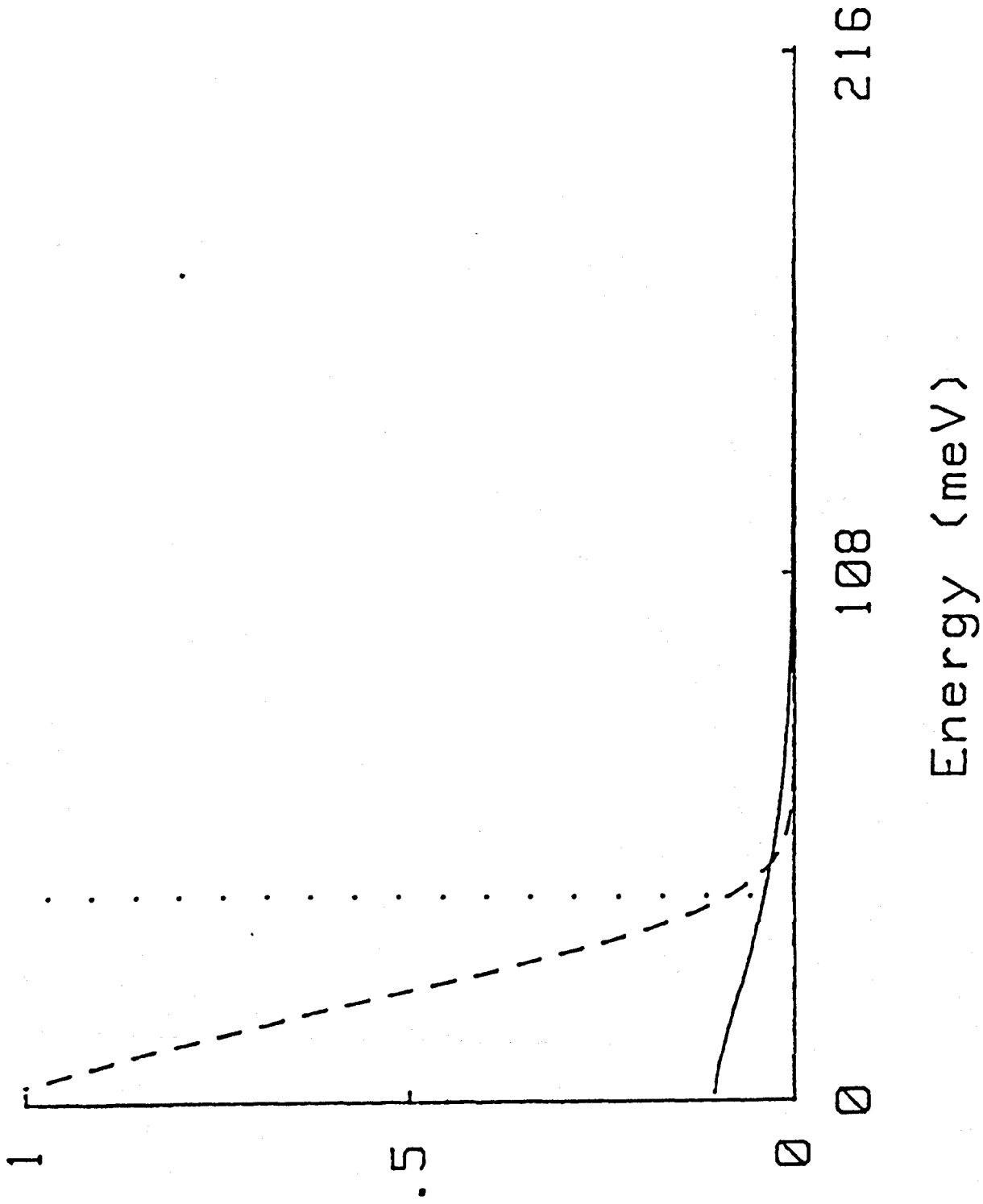

Figure 5.11 A comparison between the derivative of the current density (—) and the supply function (--) for a triangular barrier showing that results of a hot electron spectrometer are misleading.
[(...) is the Fermi level]

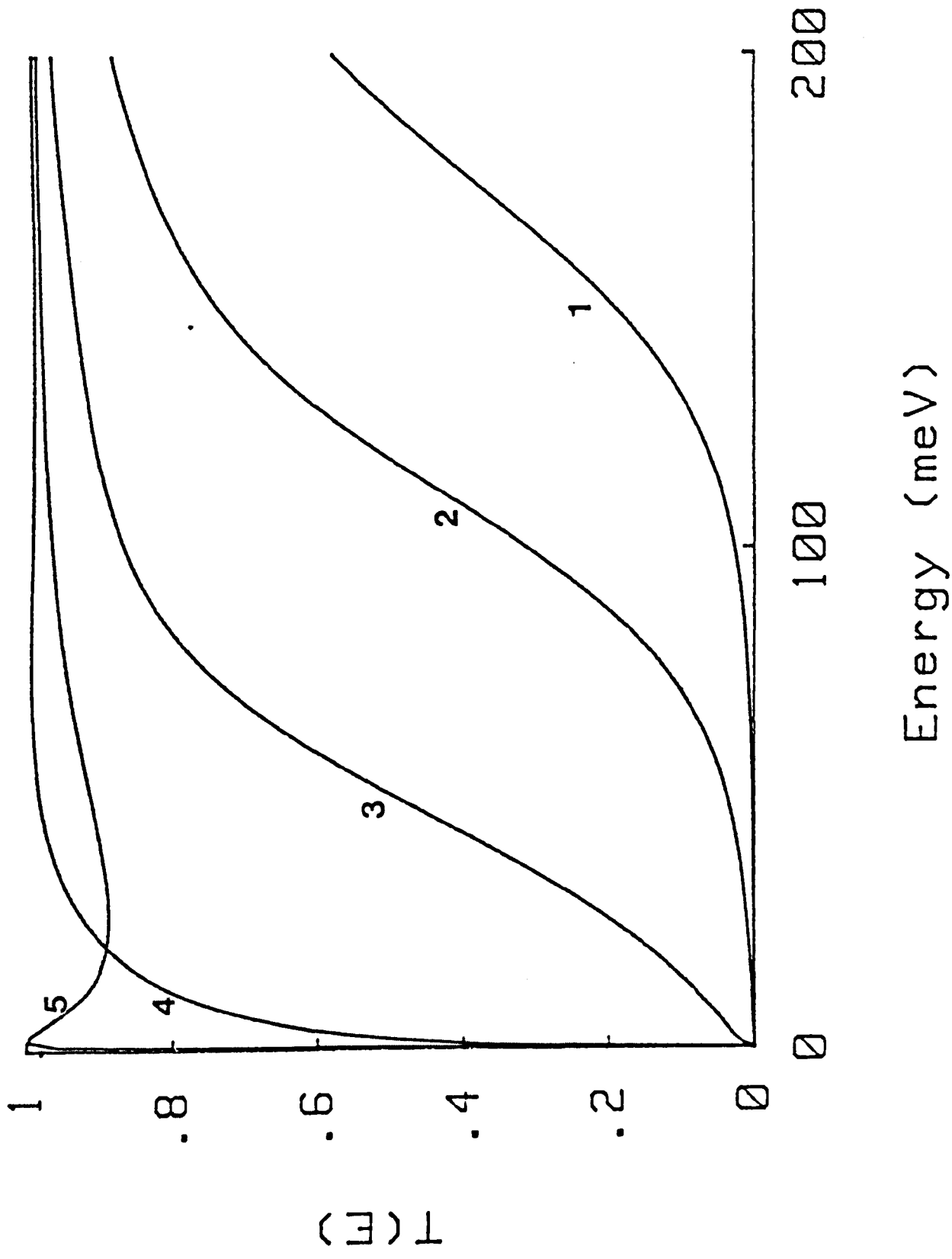

Figure 5.12 The transmission coefficient of a triangular barrier at various voltages. As the voltage increases (curves 1 to 5), the edge of the transmission coefficient decreases in energy. There are never any oscillations, as shown by Kelly.

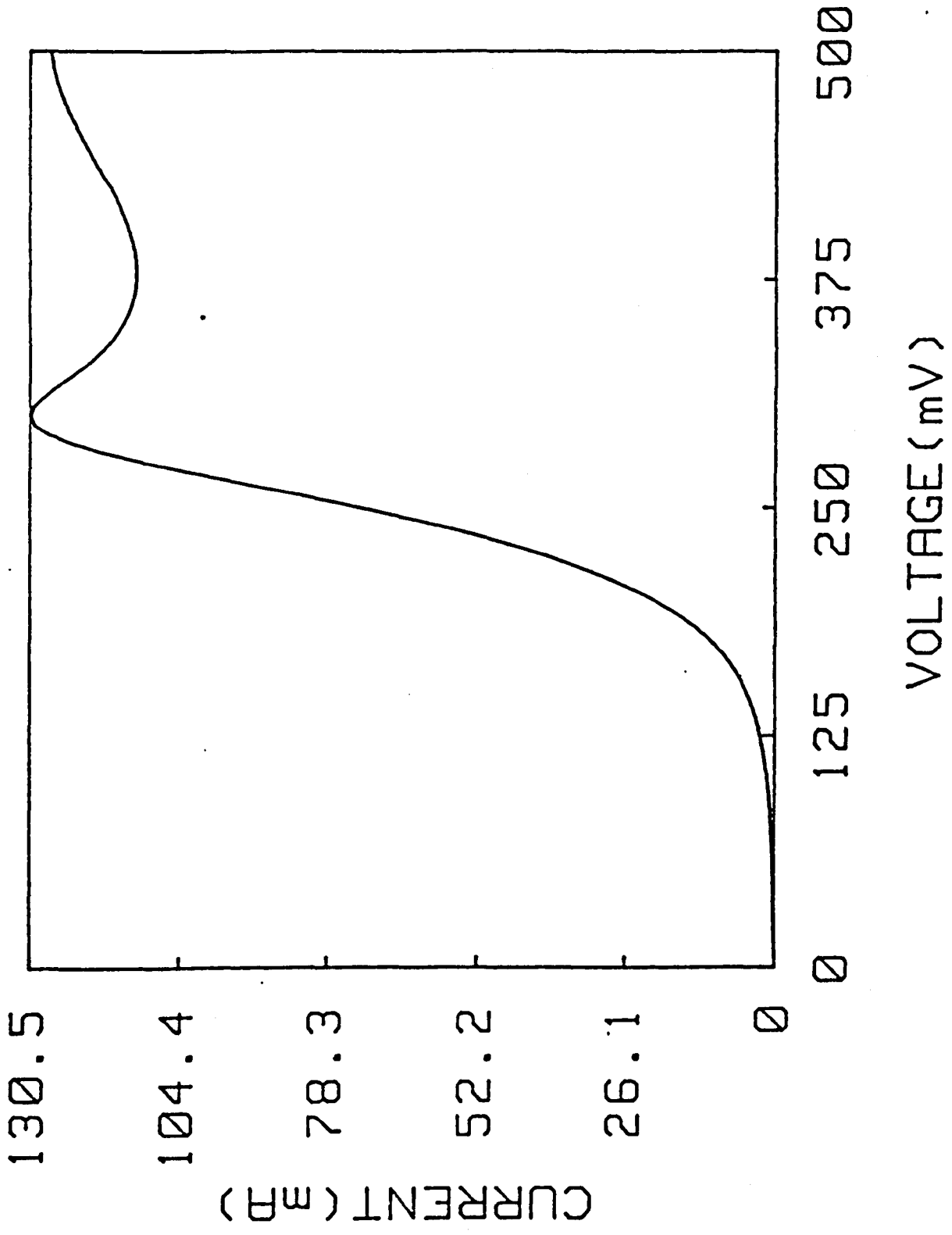

Figure 5.13 The current voltage characteristics for the triangular barrier, showing the peak caused by the behaviour of the transmission coefficient at low energy. It is the region of the curve before the peak which is useful in the "spectrometer".

levels of a parabolic well to design a device which could be used in multiple level logic. The well would be produced by varying the mole fraction of Aluminium. Since the lattice restricts the length scale over which variations can occur, a model system with a 200Å parabolic well was chosen, figure (5.14). The even spacing of the energy levels is a property of infinite system, however, examination of the transmission coefficient showed that 25Å of GaAlAs each side of the well was sufficient to produce four almost evenly spaced levels (42 meV, 125 meV, 208 meV, 293 meV).

The calculated current voltage characteristics are shown in figure (5.15), with an assumed contact doping density to 5 x $10^{18}cm^{-3}$, and a contact area of 100 μm^2. At 77 K the device is seen to have three distinct peaks in the first 500 mV corresponding to the first three resonances. The increased width of the higher energy resonances counters the decrease in magnitude produced by the potential induced asymmetry, to give resonances of approximately the same magnitude. An increase in temperature produces the expected general increase in current, however, the temperature sensitivity of the minima means that by 300 K the first peak can only be distinguished as a shoulder.

These preliminary results show that the suggested device is promising for application to multiple logic, however, it may prove difficult to engineer a structure which will operate at normal temperatures with small voltages.

5.11 Shewchuk's data

So far any discrepancy between the predicted and observed current voltage characteristics have been attributed to a device in series with the

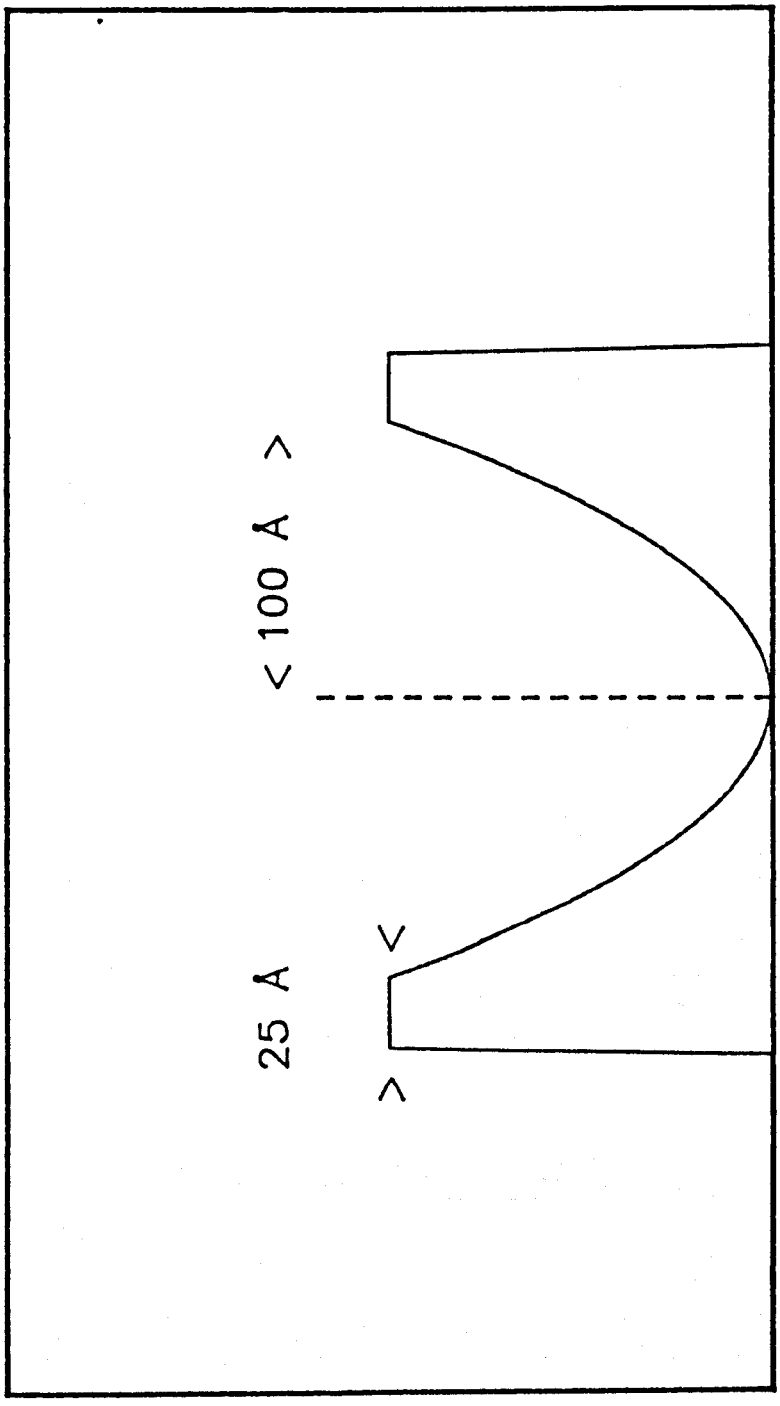

Figure 5.14 A schematic diagram of the parabolic well potential used in calculating the current voltage characteristics.

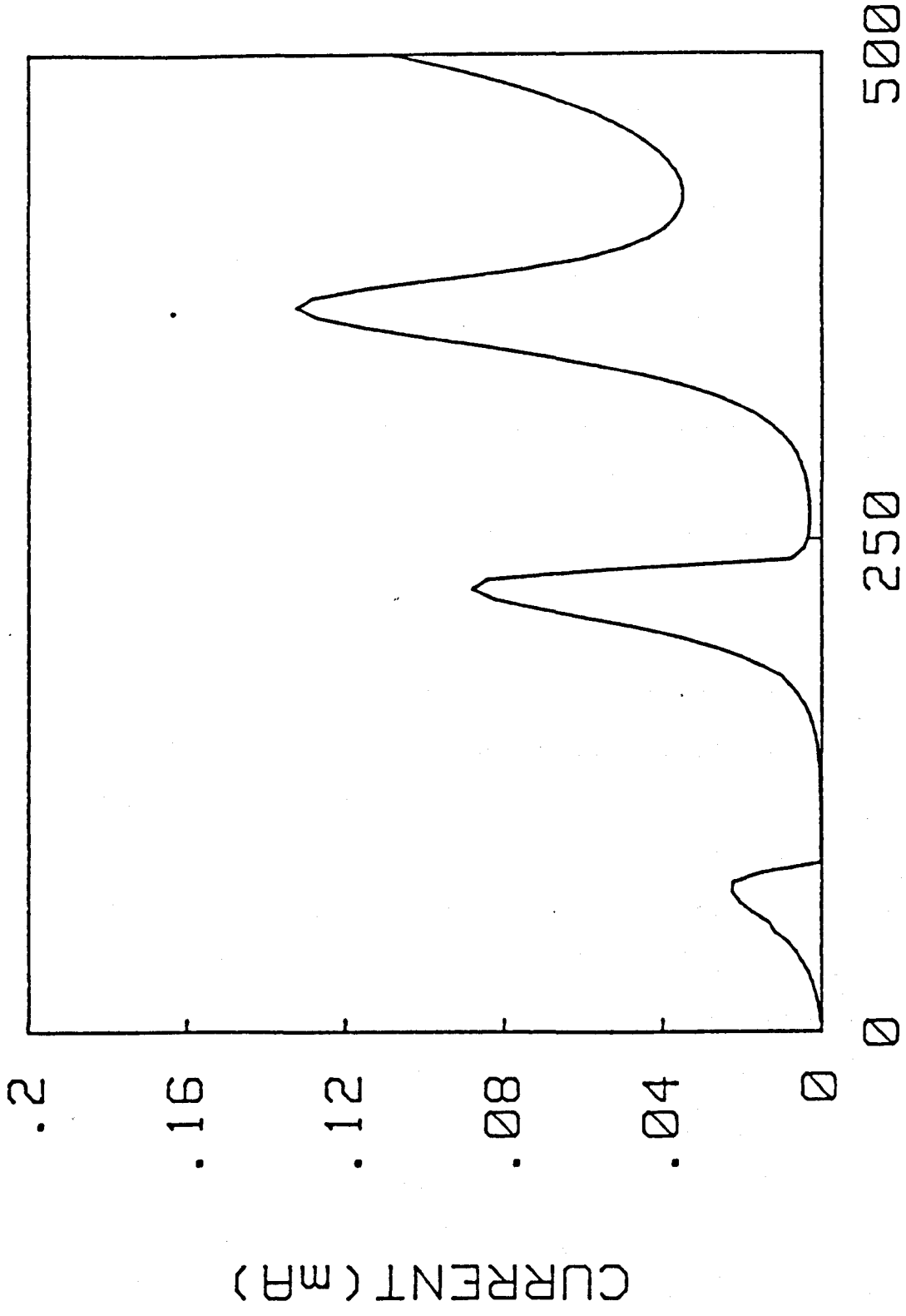

Figure 5.15 The current voltage characteristics of the potential shown in figure 5.14, with a maximum potential of 307 mV. The system was at 77 K with emitter (collector) Fermi levels at 6.69 (0) meV and a contact area of 100 μm^2.

structure. In this section a system will be considered for which the discrepancy cannot be explained in this way.

Shewchuk et al (5) fabricated a system of two 50Å, 25% Al GaAlAs layers separated by a 50Å GaAs layer, with a contact doping density of $5x10^{16}cm^{-3}$. Although the system should have a symmetric current voltage characteristic the observed characteristics were asymmetric. An attempt to explain the asymmetry in terms of a series resistance and doping level difference failed, the Fermi level difference required in the explanation having the wrong temperature dependence.

Despite the difficulties apparent in the asymmetry, the fit obtained for the 300 K data assuming a series resistance of 200Ω is satisfactory, as shown in figure (5.16).

The difficulty arises in modelling the data obtained at 77 K. If all the given parameter values are used, a peak current of 1.08 mA is predicted rather than the observed values of 2.24 mA or 1.54 mA. The difference could be due to asymmetric layers which would account for the asymmetric characteristics. An attempt to fit the data using asymmetric barriers would be lengthy and was not attempted.

The fit was not attempted because consideration of both sets of data shows that the current at 77 K is greater than at 300 K, contrary to the temperature dependence obtained from the current expression in section 5.2. This leads to the conclusion that, at 77 K, the current expression (4.2.4) is not valid for this system. The agreement at 300 K would indicate that it is a temperature dependent phenomena which invalidates the expression. If this conclusion is correct the major cause of the

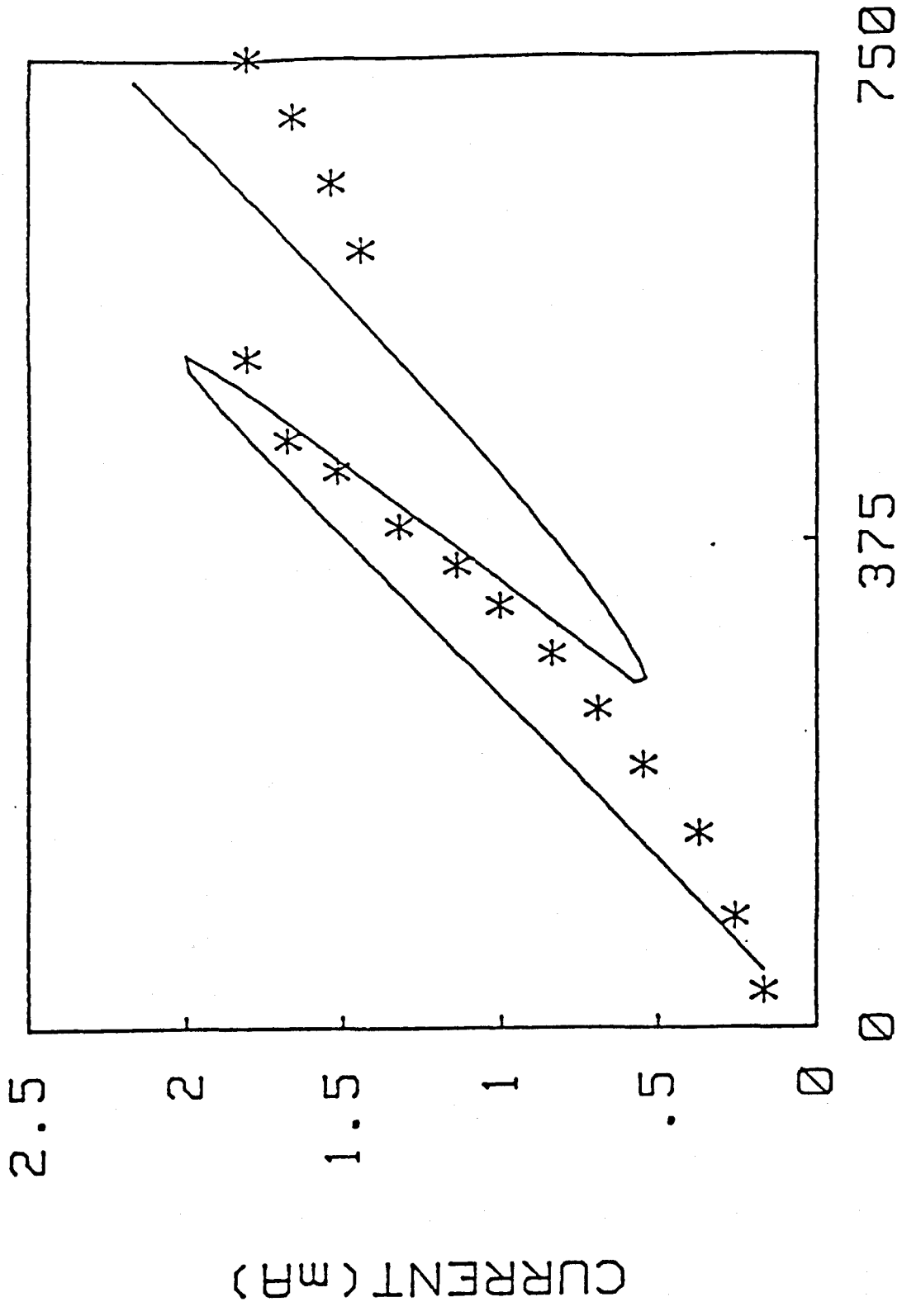

Figure 5.16 A fit to Shewchuk's data taken at 300 K (*). A 200 Ω resistor is assumed to be in series with the system leading to the multi-valued section of the curve.

difficulty would be attributable to the representation of the electron distribution. Because of the low doping density and temperature the electron population may be insufficient to maintain an electron temperature close to that of the lattice.

If the electrons cannot be represented by a Fermi distribution the current expression (4.2.4) can no longer be used. Modelling the system would be complex, requiring a self-consistent solution for the current, electron distribution and energy loss to the lattice. The assumption of a hot Fermi distribution would considerably simplify the problem requiring the solution of simplified current and energy balance equations.

The possible success of this approach can be estimated using an elevated electron temperature in (4.2.4). In the full model the temperature would become a function of the voltage, however, using a uniform temperature of twice the lattice temperature, 154 K, a peak current of 1.49 mA was obtained in comparison to an observed reverse bias resonance of 1.54 mA.

This preliminary study and the results of section 5.7 indicate that the asymmetric 77 K data may be explained using a small asymmetry in the system parameters and a Fermi distribution with an elevated temperature.

5.12 Summary

By comparing the predicts of the theoretical model with the observed behaviour of systems several results have been obtained.

A comparison between the supply function and the derivative of the current density for a triangular barrier showed that the approximation used for

the transmission coefficient is not accurate for the results of the "spectrometer" to be useful.

The results obtained for the three barrier system suggested as a rectifier demonstrated thaf systems must be treated as systems. System behaviour cannot be predicted or explained from considerations of the behaviour of isolated sections.

The model demonstrated that a key factor in determining the observed current voltage characteristics is the device which is in series with the GaAlAs structure. It has been shown that these devices; produce shifts in the resonant voltage, accentuate any asymmetry in the current voltage characteristics, reduce the peak to valley ratio and produce discontinuities and hysteresis in the current voltage characteristics.

Although each of these results is important in its own context the outstanding result is the one implied by all the others. A simple model, based on a simple representation of a GaAlAs system, a one particle picture of tunnelling and the neglect of all scattering processes, has been used for the first time to explain the observed behaviour of GaAlAs systems. The accuracy of the model is approximately ten percent which is good enough for it to be used as an aid for device design, where simple models to give approximate results are sufficient because of the difficulties in translating exact specifications from model calculations into fabricated devices. If the accuracy is not sufficient the model can be further improved without abandoning the basic simplifying step, the simple representation of the structure, by including non-parabolicity.

Other attempts which have been made to model resonant tunnelling devices

have failed to achieve the same level of accuracy as the model presented in this thesis (81). Because the same basic approach seems to have been adopted in these models this failure is difficult to understand. Experience has shown that there are two major reasons which could be the source of this failure. These are the methods of calculating the transmission coefficient and the contact Fermi levels.

The calculation of the transmission coefficient could give rise to errors in two ways. The use of the WKB method has been shown to be inaccurate in section 4.6, this could be the source of the failure. However, the errors in the WKB technique are known (49) and its use is unlikely. The more likely source of errors is in the method of defining the transmission coefficient. The Tsu-Esaki formula does not make clear whether the transmission coefficient is defined as a ratio of currents or amplitudes. In section 4.2 it was shown that the definition in terms of currents must be used. The use of a definition in terms of amplitude would lead to a large error, since it is equivalent to multiplying T(E) in (4.2.4) by $(1+V/E)^{\frac{1}{2}}$, which, with V> E, can be a significant factor. It is therefore possible that other models fail simply because they use the wrong definition of the transmission coefficient.

The calculation of the Fermi level in the contacts can also give rise to errors. In section 5.4 the Fermi level was calculated using a simple balance equation. In device modelling it is common to assume that all the donors are ionised. This assumption can give rise to electron densities which are factors of two or three greater than those obtained from the balance equation. This means that if a model assumes fully ionised donors it will overestimate the current, simply by overestimating the number of electrons available to tunnel.

SUMMARY AND FURTHER WORK

Summary of the thesis

The focal point of this thesis was the heuristic current density expression (4.2.4). The first part of the thesis was a justification for the assumptions and approach used to obtain the expression, whilst the remainder of the thesis was a discussion of the method of evaluating the expression and the results obtained.

The first chapter showed how the complex many body system which arises in a heterostructure can be modelled using a simple one dimensional equation. The simplification is achieved using assumptions which can be justified but which reduce the range of applicability of the model. A critical analysis of some of the parameters needed in modelling GaAlAs systems indicated that it is reasonable to expect further characterisation of GaAlAs to show that the presently accepted parameter values contain large errors. These errors are important in explaining the lack of agreement between the results of superlattice emission/absorption experiments and the results of the model. These errors are also a possible explanation for recent experimental results which showed that incorrect continuity conditions gave the best fit to the data (67).

The simple model could be adopted on the grounds of its simplicity alone. However, published criticism would indicate that its use generates large errors. Consideration of this criticism showed that due to misunderstandings both Osbourn (35) and Marsh and Inkson (9) overestimated the errors involved in using the model. The overall conclusion of this section was that an error of approximately ten percent is expected when

modelling GaAs-AlAs heterojunctions, with less error in the GaAs-GaAlAs heterojunctions used in practical devices.

The tunnelling of an electron through a potential is a time dependent process. A picture of the time dependence of tunnelling was therefore obtained from both an analytic and numerical solution of the S.E. for a Gaussian wavepacket incident upon a square potential.

The picture that emerged is of a wavepacket which evolves as it would in the absence of the potential until impact. On impact part of the wavepacket is reflected, giving rise to interference between the incident and reflected components, and a little later a transmitted wavepacket emerges on the opposite side of the barrier. Numerically it was confirmed that the total transmission probability is given by

$$P = \int dk \; |f(k)|^2 |t(k)|^2$$

where $f(k)$ is the fourier component of the initial condition and $t(k)$ is the transmission probability calculated from the time independent equation. This result is important in two respects. Firstly it shows that the calculation of the transmission probability requires the solution of the time independent equation. Secondly, since this expression can be derived by a consideration of plane waves incident upon the potential, it indicates that a heuristic current density expression can be obtained from a consideration of plane waves.

Another important result to emerge from this picture is the time taken for a transmitted wavepacket to emerge from the potential. The analytic result showed that if the wavepacket had a narrow momentum distribution

then the time to tunnel across a potential was related to the derivative of the phase difference between the transmitted and incident waves. This expression showed that the wavepacket traverses the potential in less than 10 fs, this compares to a mean free time in GaAs of 0.1 ps to indicate that a majority of the electrons can be expected to traverse the potential without scattering.

The study of the tunnelling of a wavepacket through a resonant level showed that resonant tunnelling and elastic hopping via an energy level are different descriptions of the same process occurring under different initial conditions. If the wavepacket has a large momentum spread compared to the spread of the resonant level then a trapped wavepacket is formed. Transmission can then be described as occurring via an intermediate state and a hopping description emerges. In the other extreme, in which the momentum spread of the wavepacket is narrow, a transmitted wavepacket emerges before the incident wavepacket has completed its interaction. Since most of the wavepacket is transmitted this picture can be termed true resonant tunnelling. The study also showed that the presence of a quasi-bound state is not a necessary condition for resonant tunnelling to occur. This is contrary to the picture proposed by Ricco and Azbel from consideration of the time independent equation. Since this picture was used by Luryi to argue that the frequencies at which resonant currents are observed excludes resonant tunnelling as the mechanism underlying resonant currents, the study contradicts Luryi's argument and re-establishes resonant tunnelling as a candidate for the underlying mechanism.

A background was therefore established which justified a heuristic derivation of a current voltage expression by considering plane waves and

neglecting scattering. The resulting expression was reduced to a one dimensional integral by assuming parabolic conduction bands, position independent mass and electrons characterised by a Fermi distribution. To use this expression to calculate the current voltage characteristics of a general one dimensional structure requires a method of calculating the transmission coefficient of a general potential. A comparison between the results of two possible general methods and the results of an exact expression showed the numerical method to be more accurate than the WKB method. Consequently, the numerical method was adopted for calculating current voltage characteristics.

The current density expression itself was then studied using model transmission coefficients to show that the observed current voltage characteristics of a single layer can be qualitatively understood using the model. This approach was also used to study the temperature dependence of the current at low voltages. It was found that for degenerate systems the current is approximately independent of the temperature whilst for non-degenerate systems the current is temperature dependent. The temperature dependence of the non-degenerate systems is itself a function of the position of the Fermi level and may change as the temperature varies.

For double barrier systems containing one resonance the behaviour of the resonant current can be predicted using a model transmission coefficient which contains the main features of a resonant transmission coefficient. It was found that with degenerate contacts at low temperatures, the maximum current is approximately independent of the temperature, whilst non-degenerate contacts and high temperatures gave rise to a temperature sensitive maximum current. Another important result to emerge was that

the current resonance occurs at the voltage which brings the transmission resonance to the bottom of the conduction band in the emitter, the transmission resonance being a fixed energy above the bottom of the conduction band in the middle of the central layer.

The current density expression was then integrated numerically to fit the observed current voltage characteristics of several experimental systems. Although the major parameters of the systems were given, there remained some areas of doubt concerning the distribution of the applied voltage and the value of the conduction band discontinuity. A system of one GaAs layer separating two GaAlAs layers was chosen for detailed modelling to examine these areas of doubt. There was no attempt to solve Poisson's equation in the model, so there were many possible field configurations which could have been used. Two were chosen as representing extremes. In the first the field is equal in all three layers whilst in the other the field is zero in the GaAs layer and equal in the GaAlAs layers. In both configurations the field in the contacts was assumed to be zero. Using these two configurations and two values of the conduction band discontinuity, four possible current voltage characteristics were calculated which predicted currents of the correct order of magnitude but gave no agreement when comparing the voltage at which the current resonated.

A detailed consideration of the effect of errors in the system parameters which could have arisen during fabrication showed that any difference between the design and fabrication of the system could not explain the differences between the predicted and observed characteristics. An explanation for these differences had to be proposed if the model is to be used with any confidence.

The required explanation was found by considering the experimental situation. The voltage measurements are taken across a sample which includes components other than the double layer structure considered in the theoretical analysis. The effect of these other components was modelled by assuming that they have the same current voltage characteristics as a diode or a resistor in series with the double layer structure. The resonant voltage which was found to be insensitive to the system parameters provided an ideal criterion for fitting the parameter associated with the series device. A resistance of 930 ohms was found to be sufficient to obtain agreement with the resonant voltage and to give good overall agreement with the shape of the characteristics, except for a small range just following resonance.

The series device was found to have an important effect on the observed current voltage characteristics, discontinuities appear because of the multivalued nature of the predicted characteristics. The exact nature of the discontinuity depends upon the experimental procedure used to observe the characteristics, and hysteresis is possible. The remaining disagreement between the predicted and observed characteristics of the test system appear at the position where the discontinuity is expected. It therefore appears justifiable to assume that the remaining small area of disagreement is attributable to experimental technique. The series device was also found to cause a sharp decrease in the observed peak to valley ratio.

A detailed fit of the predicted to the observed characteristics was not attempted. The remaining disagreement in the magnitude of the current, is sensitive to the exact contact doping. An improved fit could be obtained by assuming a slightly different doping concentration but this would not

give any more insight into the operation of these devices. It is therefore preferable to leave the comparison unaltered until a very accurate set of measurements is availale to check the accuracy of the model. The presently available data is not sufficiently accurate for this purpose, a fact which is most clearly demonstrated by the asymmetry in the data which should not be present, and whose presence is attributed to errors in fabrication.

The asymmetry which gives rise to the asymmetric current voltage characteristics must be introduced during fabrication process. It is possible that intentionally introduced asymmetries in barrier parameters could be used to enhance the magnitude of the transmission resonance at the resonant voltage and hence increase the resonant current. A preliminary investigation of the use of an increased layer thickness showed that a disappointingly small increase of about ten percent in the resonant current can be expected.

A more effective way of increasing the resonant current was found to be to use only a small bias or Fermi level difference between the two contacts and to change the current by varying the potential between each contact and the central GaAs layer. The predicted characteristics show a very sharp resonance making the fabrication of such systems desirable. The difficulty is in contacting the central GaAs layer which is only of the order of a few tens of Angstroms thick. One possible method has been demonstrated by Sollner et al using photo-ionised donors to produce a potential difference between the GaAs layer and the contacts. The model was more accurate than the unpublished model developed by Sollner et al, reproducing the shape of the current voltage characteristic but overestimated the magnitude of the current by a factor of four. This

overestimate may be attributed to the scattering due to the donors showing that although scattering by donors is not included in the model the overall shape of the curve can still be predicted.

A system of five layers suggested as a rectifying device was shown to demonstrate no rectifying behaviour. This example demonstrates the errors which arise from an analysis which uses the behaviour of isolated components in attempting to predict the behaviour of whole systems. The behaviour of the system can only be predicted by considering the system as a whole unless it is reasonable to expect the electrons to be scattered between two components of the system. The scattering destroys the coherence of the wave function allowing the components to be considered in isolation.

Unlike the layered systems triangular barriers have been shown to demonstrate good rectifying characteristics. Agreement between the observed and the predicted behaviour could only be obtained if a diode were assumed to be in series with the structure. The suggested use of a triangular barrier as a hot electron spectrometer was shown to be based on an approximate transmission coefficient which, at least for the system investigated, was not sufficiently accurate for the results of the spectrometer to represent the electron distribution. Although this source of error could be minimised by system design, the energy scale would still be affected by the series device, making interpretation more difficult. Another point which should be emphasised is that Hayes et al stated that the electron distribution could be obtained, however the integrand contains the supply function and not the electron distribution. This means that the analysis proposed will give the supply function from which an electron distribution can only be inferred.

A short study of the parabolic well system proposed by Capasso and Keihl showed that the evenly spaced levels gave rise to evenly spaced resonances in the current if a series device were not included. The magnitude of the resonances was found to be only a weak function of voltage, making the device a good candidate for use in multi-level logic.

The final section of the thesis discussed the only data for which no fit was possible. The failure to obtain agreement is probably due to a failure of the assumption that there are sufficient electrons to maintain a thermal distribution. A failure which would occur in any model which assumed a thermalised electron distribution.

Despite the results of the last section, Chapter 5 shows that expression (4.2.4), in conjunction with the methods of calculating the transmission coefficient and Fermi levels, can be used to predict and explain the behaviour of GaAs/GaAlAs tunnel devices.

Further work

The model presented in this thesis has one major area in which improvements could be made, there is no attempt to solve Poisson's equation to obtain a realistic distribution of the applied potential. Poisson's equation in one dimension is

$$\frac{d^2V}{dx^2} = \rho(x)$$

where $\rho(x)$ is the net charge density. This has two important consequences.

The first is that finite change in the electrical field over zero distance is unphysical, requiring infinite charge density. This means that the field distribution assumed in the model, zero in the contacts but finite in the barriers, is unphysical.

The second is that a full treatment of the problem requires the self-consistent solution of both Poisson's equation and the equation governing electron motion.

Although a full treatment requires the self-consistent solution of these two equations a useful approximation can be obtained by decoupling the equations. Assuming a constant field in the barrier region and that the electrons in each contact are in equilibrium, Poisson's equation can be solved separately in each contact using the barrier electric field as one boundary condition and zero field at infinity as the other. The situation is then that shown in figure (6.1) and the total voltage drop, V_T, is given by

$$V_T = V_{acc} + V_{dep} + V$$

The current can then be calculated using the potential shown in figure (6.1), the region between the two regions of zero electric field acting as the barrier.

The device which was placed in series with the barrier region can now be seen to be a simple model of the voltage dropped in the accumulation and depletion layers.

Solving Poisson's equation as described shows that the sum of the two voltages V_{acc} and V_{dep} is of the same order as the voltage dropped across

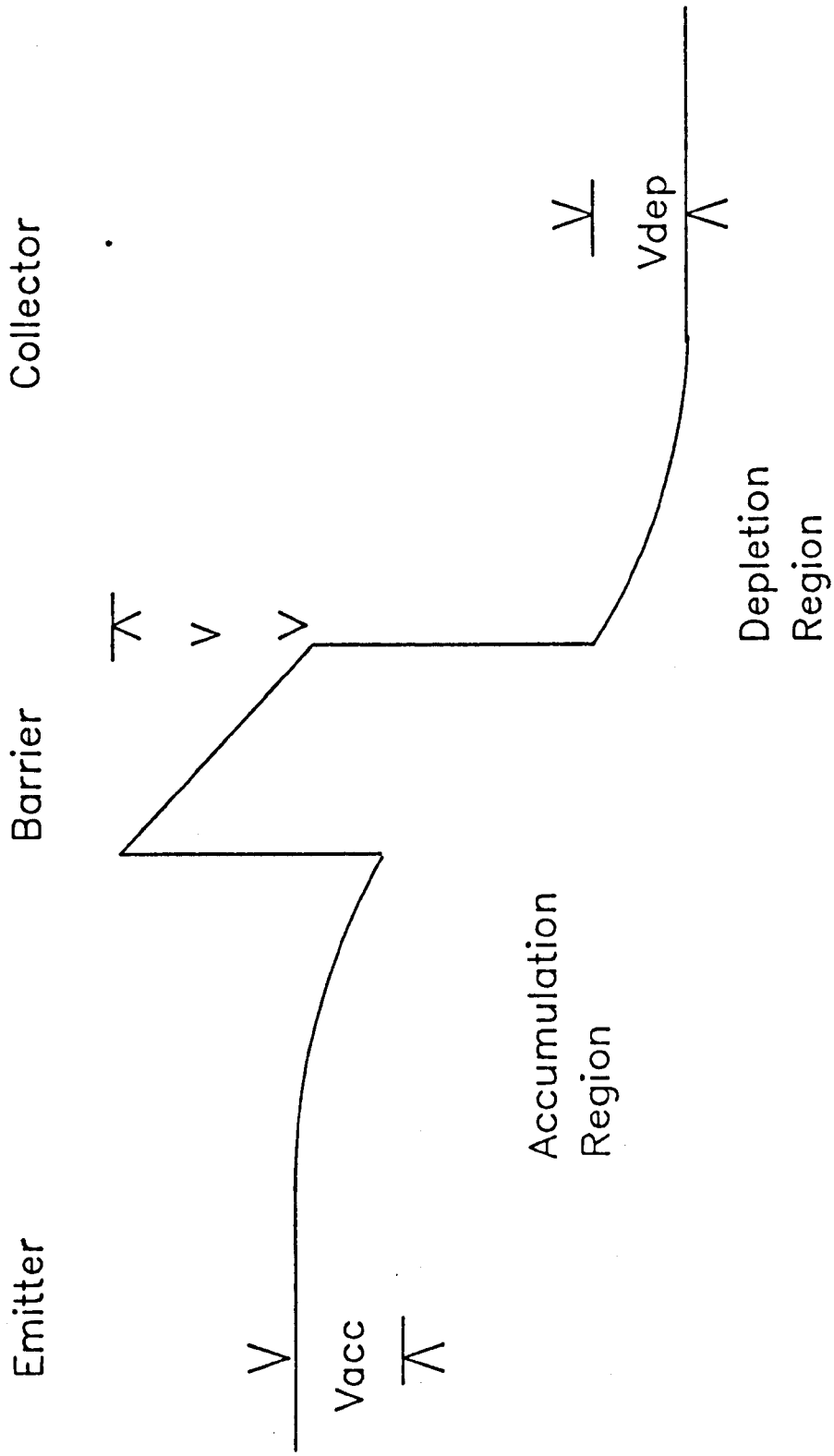

Figure 6.1 The accumulation and depletion voltages which arise from the solution of Poisson's equation with a constant field boundary condition.

the "device" (82). These calculations also show that the sum of the two voltages is approximately linearly dependent upon the voltage dropped across the barrier.

This can be used to explain the fit obtained to Sollner's data using a series device. The best fit was obtained in the region before resonance. This is approximately Ohmic, so that the assumption of a series resistance adds on a voltage which is linearly dependent upon the voltage across the barrier. The resistance is therefore a good representation of the accumulation and depletion voltages. The only affect of these voltages not accounted for in the model is the net lowering of the barrier by V_{acc}, however, this may be accounted for in the uncertainty in the initial barrier height.

These simple considerations indicate that adding Poisson's equation into the model in this simple way will be useful in eliminating the need for the assumption of a series device and making the basis for the model more physical.

The only requirement for using (4.2.4) is that the electrons traverse the accumulation and depletion regions and the barrier without scattering. Since the barrier traversal time is only an order of magnitude less than the mean free time the neglect of scattering can only be justified using a calculation which includes scattering. Although this involves a non-trivial extension of the present work it would allow inelastic tunnelling to be studied.

REFERENCES

(1) SANO N, KATO H, NAKAYAMA M, CHIKA S and TERAUCHI H
Jap J Appl Phys 23 L640 (1984)

(2) TSU R and ESAKI L Appl Phys Lett 22 562 (1973)

(3) BOHM D Quantum Mechanics
Prentice Hill (1951)

(4) SOLLNER T C L G, GOODHUE W D, TANNENWALD P E, PARKER C D and PECK D D Appl Phys Lett 43 588 (1983)

(5) SHEWCHUK T J, CHAPIN P C, COLEMAN P D, KOPP W, FISCHER R and MORKOC H Appl Phys Lett 46 508 (1981)

(6) BASTARD G Phys Rev B 24 5693 (1981)

(7) COLLINS R T, LAMBE J, McGILL T C and BURNHAM R D Appl Phys Lett 44 532 (1984)

(8) PICKETT W E, LOUIE S G and COHEN M L Phys Rev B 17 815 (1978)

(9) MARSH A C and INKSON J C J Phys C 17 6561 (1984)

(10) SCHULMAN J N and CHANG Y-C Phys Rev B 31 2056 (1985)

(11)MAILHOT C, McGILL T C and SCHULMAN J N J Vac Sci Technol B1 439 (1983)

(12)KANE E O J Phys Chem Sol 1 249 (1957)

(13)MERIAN M and BHATTACHARJEE A K Sol State Comm 55 1071 (1985)

(14)BLAKEMORE J S J Appl Phys 53 R123 (1982)

(15)CHEN A-B and SHER A Phys Rev B 23 5360 (1981)

(16)MONEMAR B Phys Rev B 8 5711 (1973)

(17)ONTON A Proc Tenth Int Conf Phys Semicond 107 (1972)

(18)YIM W M J Appl Phys 42 2884 (1971)

(19)STUKEL D J and EUWEMA R N Phys Rev 188 1193 (1969)

(20)BRAUNSTEIN R and KANE E O J Phys Chem Sol 23 1423 (1962)

(21)BEROLO O and WOOLLEY J C Can J Phys 49 1335 (1971)

(22)CASEY Jr H C J Appl Phys 49 1335 (1971)

(23)JAROS M Rep Prog Phys 48 1091 (1985)

(24)WALDROP J R and GRANT R W Phys Rev Lett 43 1686 (1979)

(25)DINGLE R Festkorprobleme XV 21 (Ed H J Queisser)

(26)MILLER R C, KLEINMAN D A and GOSSARD A C Phys Rev B 29 7085 (1984)

(27)DAWSON P, DUGGAN G, RALPH H I, WOODBRIGHT K and 't HOOFT G W Superlattices and Microstructures 1 231 (1985)

(28)KROEMER H Appl Phys Lett 46 504 (1985)

(29)WATANABE M O, YOSHIDA J, MASHITA M, NAKANISI T and HOJO A J Appl Phys 57 5340 (1985)

(30)HEIBLUM M, NATHAN M I and EIZENBERG M Appl Phys Lett 47 503 (1985)

(31)WANG W I,MENDEZ E E and STERN F Appl Phys Lett 45 639 (1984)

(32)TERSOFF J Phys Rev B 30 4874 (1984)

(33)MARGARITONDO G Phys Rev B 31 2526 (1985)

(34)MONEMAR B, SHIK K K and PETTIT P D J Appl Phys 47 2604 (1976)

(35)OSBOURN G C J Vac Sci Technol 17 1104 (1980)

(36)MARSH AC and INKSON J C Solid State Comm 52 1037 (1984)

(37)COLLINS S, LOWE D and BARKER J R J Phys C 18 L637 (1985)

(38)MARSH A C and INKSON J C preprint

(39)MESSIAH A Quantum Mechanics North Holland (1961)

(40)GOLDBERG A, SCHEY H M and SCHWARTZ J L Am J Phys 35 177 (1967)

(41)MACCOLL L A Phys Rev 40 621 (1932)

(42)WIGNER E P Phys Rev 98 145 (1955)

(43)HARTMAN T E J Appl Phys 33 3427 (1962)

(44)STEVENS K W H J Phys C 16 3649 (1983)

(45)BUTTIKER M Phys Rev B 27 6178 (1983)

(46)SCHNUPP P Thin Solid Films 2 177 (1968)

(47)PSHENICHOV E A Sov Sol State 4 819 (1962)

(48)STEVENS K W H J Phys C 17 5735 (1984)

(49)VIGNERON J P and LAMBIN Ph J Phys A 13 1135 (1980)

(50)HEADING J J Atmos Terr Phys 25 519 (1963)

(51)KANE E O Tunneling Phenomena in the Solid State
Ed E Burstein and S Lundquist
Plenum Press (1969)

(52)VIGNERON J P and LAMBIN Ph J Phys A 12 1961 (1980)

(53)DAVIES R A, KELLY M J and KERR T M Phys Rev Lett 55 1114 (1985)

(54)CHANG L L, ESAKI L and TSU R Appl Phys Lett 24 593 (1974)

(55)VOJAK B A, KIRCHOEFER S W, HOLONYAK Jr N, CHIN N, DUPUIS R D and DAPHUS P D J Appl Phys 50 5830 (1979)

(56)SOLLNER T C L G, TANNENWALD P E, PECK D D and GOODHUE W D Appl Phys Lett 45 1319 (1985)

(57)TSUCHIJA M, SAKAKI H and YOSHINO J Jap J Appl Phys 24 466 (1985)

(58)THRUSH E J,WALE-EVANS G, WHITEWAY J E A, LAMB B L, WIGHT D R, CHEW N G, CULLIS A G and GRIFFITHS R J H J El Mater 13 969 (1984)

(59)RICCO B and SOLOMON P M IBM Tech Dis Bull 27 3053 (1984)

(60)SOLLNER T C L G, LE H Q, CORREA C A and GOODHUE W D Appl Phys Lett 47 36 (1985)

(61)KIRCHOEFER S W, NEWMAN H S and COMAS J Appl Phys Lett 47 36 (1985)

(62)ALLYN C L, GOSSARD A C anh WIEGMANN W Appl Phys Lett 36 373 (1980)

(63)HAYES J R, LEVI F J and WIEGMANN W El Lett 20 851 (1985)

(64)KELLY M J El Lett 20 771 (1984)

(65)CAPASSO F and KIEHL R A J Appl Phys 58 1366 (1985)

(66)RICCO B and AZBEL M Ya Phys Rev B 29 1970 (1984)

(67)KAWAI H, KANEKO J and WATANABE N J Appl Phys 58 1263 (1985)

(68)WALL H Analytic Theory of Continued Fractions (1967)

(69)MORROW R A and BROWNSTEIN K R Phys Rev B 30 678 (1984)

(70)ALTARELLI M Physica 117B+118B 749 (1983)

(71)GRADSTEYN I S and RYZHIK I W Tables of Integrals, Series and Products Academic Press (1965)

(72)ABRAMOWITZ M and STEGUN I A Handbook of Mathematical Functions with Formulae, Graphs and Mathematical Tables Dover (1981)

(73)VASSELL M O, LEE J and LOCKWOOD H F J Appl Phys 54 5206 (1983)

(74)HRACH R Czech J Phys 20B 32 (1970)

(75)ESAKI L and CHANG L L Phys Rev Lett 33 495 (1974)

(76)NAKAGAWA T, IMAMOTO H, SAKAMOTO T, KOJIMA T, OHTA K and KAWAI N J El Lett 21 882 (1985)

(77)LURYI S Appl Phys Lett 47 490 (1985)

(78)BLAKEMORE J S Semiconductor Statistics
Pergamon (1962)

(79)LAWAETZ P Phys Rev B 4 3460 (1971)

(80)BARKER J R Private communication

(81)ARCHIBALD I E Private communication