ELECTRICAL CONDUCTION IN 3-5 SEMICONDUCTOR STRUCTURES

by

THOMAS P C JUDD

TRINITY COLLEGE
CAMBRIDGE

Thomas Patrick
Charles JUDD

This dissertation is submitted for the degree of
Doctor of Philosophy at the University of Cambridge,
December 1986
This thesis describes experiments performed on three types of 3-5 semiconductor device and the deductions that can be formed from these experiments about the operation of the devices. The experiments concerned the effects on electrical conduction in the devices of low temperatures, magnetic fields and device parameters.

The magnetophoton effect in gallium arsenide field effect transistors has been studied. The results obtained are strongly influenced by the geometry of the device, its orientation with respect to the magnetic field and its doping level. These parameters have been varied over a wide range. The field effect transistor has the advantage that the geometry of the conducting region can be varied by applying a gate voltage, and this has been used to study the effect of reducing one of the dimensions of the conducting region until it is comparable with a critical length, which is either the Debye length or the cyclotron radius, depending on the device.

The triangular barrier switch is a device that was proposed only five years ago and its operation is not yet fully understood. It has two stable states, one of high impedance and one of low. It has been shown that the switching between these states is hysteretic, and new types of behaviour have been seen under the previously uninvestigated conditions of temperatures below 77 kelvin and of high magnetic fields. The best existing model of this device has been extended to account for these effects.

The graded gap diode contains a potential barrier formed by sandwiching a layer of one type of semiconductor between regions of a different type. It is shown that depletion regions are about as important as the barrier in determining the current. The changes with temperature of the mechanism of conduction through this structure can be followed experimentally and have been used to measure properties of the barrier. A novel, unusually large, current instability has been seen at low temperatures. A study has been made of the statistics of this instability to identify its origin.
PREFACE

The research described in this thesis was carried out at the Cavendish Laboratory, University of Cambridge, between October 1983 and December 1986. During this time I have received help and companionship from a large number of people, all of whom I would like to thank. I am particularly grateful to the following:

Dr. Michael Pepper, my supervisor, for his continuous flow of ideas, encouragement and samples and his interest in my welfare;

Dr. Geoff Hill and colleagues at the SERC device fabrication centre at Sheffield University for their invariably high standard of work and innovative solutions when there were problems;

Drs. Mike Kelly, Nigel Couch and Peter Beton from the GEC Hirst Research Centre for allowing me to join their experimental programme and for providing a sounding board for ideas;

Dr. Iain Dorrity and colleagues, also at the GEC Research Centre, for finding time to grow unusual layers for me in the midst of a very busy in-house programme;

Paul Rees, from the University College of Swansea, for continuing to supply me with devices made by himself despite a very high casualty rate as a result of my inexperience;

Professor Karl Berggren, of Linkoping University, Sweden, for technical discussions on the magnetophonon effect in thin layers;

Paul Booth, David Swainston, Peter Flaxman, Ken Gadsby and Mark Cullum, of the LTP workshop, without who very few experiments would be possible.

This thesis is the result of my own work and contains nothing that is the outcome of collaboration.

Tom Judd,

December 1986
# CONTENTS

## Chapter 1 - Introductory concepts

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>3-5 semiconductors</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Calculation of the density of states in low dimensional systems</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>The magnetophonon effect</td>
<td>5</td>
</tr>
</tbody>
</table>

## Chapter 2 - Experimental techniques for the magnetophonon effect

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Magnetic field generation and measurement</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>The sample holder</td>
<td>10</td>
</tr>
<tr>
<td>2.3</td>
<td>Temperature measurement and control</td>
<td>11</td>
</tr>
<tr>
<td>2.4</td>
<td>Extraction of the oscillations</td>
<td>13</td>
</tr>
<tr>
<td>2.5</td>
<td>The data acquisition and processing system</td>
<td>17</td>
</tr>
<tr>
<td>2.6</td>
<td>Origins of noise in magnetophonon measurements</td>
<td>19</td>
</tr>
</tbody>
</table>

## Chapter 3 - Magnetophonon oscillations in GaAs fets

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Device structure and fabrication</td>
<td>24</td>
</tr>
<tr>
<td>3.2</td>
<td>Theory of operation</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Results and discussion</td>
<td>27</td>
</tr>
<tr>
<td>3.4</td>
<td>The magnetophonon effect in thin channels</td>
<td>34</td>
</tr>
<tr>
<td>3.5</td>
<td>Results for thin channels</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Higher doping levels</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Peak shifts in short devices</td>
<td>46</td>
</tr>
</tbody>
</table>

continued
Chapter 4 - The triangular barrier switch

4.1 Device structure and theory of operation 47
4.2 Device fabrication 48
4.3 Experimental results 49
4.4 Experimental techniques 55
4.5 A mathematical model 57
    Changing temperatures 64
    Recombination in the barrier 66
    Low temperature behaviour 68

Chapter 5 - The graded gap diode

5.1 Structure and fabrication 72
5.2 Theory of operation 72
5.3 Results and discussion 76
5.4 Random switching between discrete resistance levels 80
5.5 A model of the switching 82

Appendices

A The magnetophonon program 85
B The electron distribution program 89
C The triangular barrier switch simulation program 92

List of references 96 to 98
1.1 3-5 Semiconductors

These materials are compounds of elements from group 3 of the periodic table, generally Al, In, or Ga, with elements from group 5, generally As or P. They were initially of interest because electrons in the conduction bands of these materials generally have effective masses much lower than in silicon. All other things being equal, this will lead to better high frequency performance, and the transistors currently used at microwave frequencies are smaller versions of the devices used in chapter 3. Another useful property is that the conduction band minimum and valence band maximum both occur at K=0, resulting in a high probability that an electron can move across the bandgap, emitting or absorbing a photon. This makes these semiconductors suitable for optoelectronic devices.

One is not restricted to binary compounds such as GaAs or InP. Ternary compounds, e.g. Al\textsubscript{x}Ga\textsubscript{1-x}As or GaAs\textsubscript{y}P\textsubscript{1-y}, are also possible. For reasonable crystal quality these materials must generally have a lattice parameter equal to that of the substrate that they are grown on, and at present only GaAs and InP substrates are available. Hence ternary materials are usually limited to a single composition at which a lattice match with the substrate is obtained. This composition may still however have more useful properties than GaAs or InP, for instance InGaAs on InP is used for infra-red optoelectronics. The price to be paid for this is an extra source of scattering for conduction electrons; even in an ideal ternary either the group 3 sites or the group 5 sites will be occupied by one type of atom with a certain probability and by the other type of atom from the same group if not. This produces a small non-periodic potential which scatters electrons just as the non-periodic potentials due to defects, impurities and lattice vibrations do. This mechanism is known as alloy scattering.
Quaternary compounds have an extra degree of freedom, which means that even after the lattice matching constraint has been met a range of compositions is still available, so that another parameter of interest, the bandgap for example, can be freely chosen between two limits. However, quaternaries have even worse alloy scattering than ternaries.

A thin layer of electrons with particularly high mobility is formed at the junction of two dissimilar but lattice matched semiconductors. AlGaAs and GaAs are most often used for this. This is utilised by the fastest transistors demonstrated to date.

Devices in which electrons travel through, rather than parallel to, layers of different semiconductors are also possible.

The principal disadvantages of 3–5 semiconductors compared with silicon are the complexity that is a corollary of the versatility, the relatively poor performance of devices based on holes, and the lack of an insulator that will form an interface to these semiconductors with a defect density comparable to the excellent silicon to silicon dioxide interface.

Ideal gas theory is often adequate to describe the conduction electrons in most of these semiconductors because the conduction band is isotropic and almost parabolic at low energies, and the Fermi energy usually lies several times KT below its minimum. The only modification made necessary by the lattice potential is a change in the effective mass, usually a large reduction. This approximation has been made throughout this dissertation.
1.2 Calculation of density of states in low dimensional systems

Electrons that are free to move in the y and z directions, but experience a potential \( V(x) \) that constrains their motion in the x direction, will have energies given by

\[
E = E_n + \frac{\hbar^2 k_y^2}{2m^*} + \frac{\hbar^2 k_z^2}{2m^*} = E_n + \frac{\hbar^2 k_x^2}{2m^*}
\]

where \( k_x = k_y + k_z \), and the \( E_n \) are energies characteristic of the potential in the x direction. These \( E_n \) are the eigenvalues of the one dimensional Schrodinger equation

\[
-\frac{\hbar^2}{2m^*} \frac{d^2 \psi(x)}{dx^2} + V(x)\psi(x) = E_n\psi(x)
\]

This dispersion relation is shown graphically in figure 1.2.1.

The system is strictly two dimensional when \( E_1 \) is sufficiently greater than \( E_0 \) for states on only the lowest parabola ever to be occupied. The faster \( V(x) \) increases the greater is the separation of the parabolaes. The nth parabola contributes a constant to the density of states for energies above \( E_n \) and nothing for lower energies. This can be seen as follows:

For a rectangular piece of semiconductor with dimensions \( a \) and \( b \) in the y and z directions, the permitted values for \( k_y \) and \( k_z \) are those which fit integral numbers of wavelengths of the electron wavefunction into the lengths \( a \) and \( b \). (figure 1.2.2). This gives \( k_y = \frac{n_y}{a} \cdot 2\pi \) and \( k_z = \frac{n_z}{b} \cdot 2\pi \), where \( n_y \) and \( n_z \) are integers. These allowed \( k \) values can be plotted as points on a rectangular grid (figure 1.2.3).

From the dispersion relation, a contour of constant energy \( E \) on this diagram consists of a circle with radius squared equal to \( 2m^*(E-E_n)/\hbar^2 \) for each subband with \( E_n < E \). Because of spin degeneracy the number of electron states with energy less than \( E \) is twice the number of points enclosed by the circle. Hence, if \( a \) and \( b \) are large, the number of states is given by

\[
N(E) = \frac{m^*}{\pi \hbar^2} \frac{(E-E_n)}{ab}
\]
a, b is the area of the sample, and assuming that the result does not depend on the shape of the sample when the dimensions are many electron wavelengths, this result becomes

\[ N(E) = \frac{m^* (E - E_n)}{\pi \hbar^2} \text{ per unit area of the sample.} \]

The density of states is given by \( \frac{dN}{dE} \):

\[ g(E) = \frac{m^*}{\pi \hbar^2} \text{ per unit area for each subband with } E_n \leq E. \]

The density of states for this system is as shown in figure 1.2.4.

---

**FIGURE 1.2.4**
1.3 The Magnetophonon Effect

The magnetophonon effect was first predicted by Gurevich and Firsov in 1961 and detected in the magnetoresistance of n-type InSb at about liquid nitrogen temperatures by Puri and Geballe in 1963. It consists of small oscillations periodic in 1/B superimposed on the usual smooth increase of resistance of a semiconductor with increasing magnetic field. (See figure 1.3.1)

Origin

In a strong magnetic field the allowed energies of electrons in GaAs become

\[ E(n,k_z) = \frac{\hbar^2 k_z^2}{2m^*} + (n+\frac{1}{2})\hbar\omega_c \]

where \( \omega_c \) is the cyclotron frequency, equal to eB/m*, and \( k_z \) is the wavevector along the magnetic field direction.

This corresponds to the motion parallel to the field being unaffected while motion perpendicular to the field is circular. Quantisation of angular momentum leads to the radii of the orbits taking only the values

\[ r_c = \left[ \frac{2\hbar(n+\frac{1}{2})}{eB} \right]^{0.5} \quad n = 0, 1, 2, ... \]

This quantity is called the cyclotron radius. The energy of perpendicular motion is also quantised, as above, into equally spaced levels \( \hbar\omega_c \) apart. Because the density of states is inversely proportional to dE/dk, which is zero at k=0, the density of states has periodic singularities (figure 1.3.2). Most electrons thus have energies corresponding to one of these singularities, and these energies are known as Landau levels. They are equally spaced and their separation is proportional to the magnetic field intensity. (This is treated in more detail in section 3.4).
MAGNETOPHONON OSCILLATIONS

Figure 1.3.1

DISPERSION CURVES AND DENSITY OF STATES

Figure 1.3.2
The $E$ versus $k$ curve for longitudinal optic phonons is very flat. (figure 1.3.3). In fact by the time it intersects the curve for electrons the energy differs from that at the zone centre by less than one part in $10^5$. This means that it is possible to speak of a single energy for the L.O. phonons with the range of wavevectors that can be absorbed or emitted by an electron. This energy corresponds to a temperature of about 400 kelvin in GaAs, so the number of optic phonons present varies very rapidly with temperature, for instance increasing by a factor of about 60 between 50 and 100 kelvin. This increase is generally sufficient for optic phonon scattering to begin to dominate the electrical resistance somewhere in this temperature range.

When the separation between Landau levels is equal to the L.O. phonon energy the probability of absorption or emission of an L.O. phonon increases, since this rate is proportional to the product of the initial and final densities of states, which are both now very large. The same will happen whether the L.O. phonon energy is equal to the separation between one, two or more Landau levels. Hence the condition for this additional scattering is that

$$n\hbar\omega_c = \hbar\omega_p$$

or equivalently

$$\frac{1}{B} = \frac{n.e}{m^*\omega_p}$$

showing that the resistance changes from the additional scattering are periodic in $1/B$.

A density of states exactly as in figure 1.3.2 would produce infinite values of resistance at the resonances. In practice the resistance change is never more than about 10%, and is often very much less. The reason for this is the loss of the singularities in the density of states, the peaks of which are smeared out by approximately the amount of $\hbar$ divided by the finite time an electron spends in a state before being scattered.
FIGURE 1.3.3
The amplitude of the oscillations is found to vary with temperature, sample purity and magnetic field (or equivalently with the index number \( n \) of the peak). The optimum temperature to observe the magnetophonon effect is usually about 150 kelvin; at low temperatures there will be very few L.O phonons with which the electrons can interact, even when the resonance condition is satisfied, and at high temperatures, though most of the scattering is by optic phonons, it is so rapid that the sharp peaks in the density of states are destroyed.

The ratio of the amplitudes of the oscillations in different samples is similar at all temperatures, and the amplitudes are inversely proportional to each sample's concentration of ionised impurities. (Stradling and Wood, 1968). Hence, rather surprisingly, the temperature and impurity variations of the amplitude of a given peak can be separated. Figure 1.3.4 illustrates this.

It is found empirically that the variation with magnetic field of the oscillatory part of the magnetoresistance is well represented by

\[
\frac{\Delta \rho}{\rho_0} = C \times \exp(-\Gamma \omega_p/\omega_c) \times \cos(2\pi \omega_p/\omega_c),
\]

so that the amplitude of the \( n \)th peak is proportional to \( \exp(-\Gamma n) \). \( \Gamma \) is a function of both temperature and impurity concentration. Conditions that make \( C \) small also tend to increase \( \Gamma \), so that oscillations which are weak initially also disappear more rapidly as the magnetic field is reduced. Barker (1970) has a successful theory of the amplitude variations which produces the above expression as the first and largest term of a harmonic series. In his theory \( \Gamma \) is the imaginary part of the electron self energy arising from both impurity and optic phonon scattering.

The above discussion applies when \( B \) is perpendicular to \( j \). The peaks in the density of states occur at \( k_x=0 \), so electrons moving between states in these peaks change their momentum along the direction of \( B \) very little. If this is also the direction of current flow the additional scattering will have little effect on the current. Hence the magnetophonon effect is very
The similarity of the four curves in (a) shows that with j and B at right angles the variation of amplitude can be described by the product of a function of temperature only and a function of impurity level only. This is not the case with j and B parallel, as (b) shows.

(from Stradling and Wood, 1968)
weak for $B$ parallel to $j$, and second order effects involving a second phonon or an impurity are important. These lead to minima rather than maxima in resistance, and these minima may be shifted away from the resonance fields.

**Hot electron conditions**

Magnetophonon oscillations can be observed even at very low lattice temperatures if an electric field is applied that is large enough to heat the electron gas by a few kelvin. The electrons generally exchange energy amongst themselves rapidly enough to establish a nearly Maxwellian energy distribution. At low lattice temperatures the mobility of the electron gas is determined by ionized impurity scattering, which varies inversely with temperature, but most of the energy loss from the electron gas is through the emission of optic phonons by the electrons in the high energy tail of the energy distribution. The emission probability is enhanced at the magnetophonon resonances and so the whole electron gas is cooled. Resistance maxima occur at the resonance fields because of the lowered electron temperature and the ionized impurity scattering. In some 3-5 semiconductors other than GaAs distortions of the energy distribution form a Maxwellian form are important and this is not the case. (see e.g. Yamada, 1973). Because of the way in which these oscillations arise the size and sign of the resistance changes are now almost independent of the direction of $B$ relative to the current.

Harper et al. (1973) have reviewed the experimental and theoretical work on the magnetophonon effect.
EXPERIMENTAL TECHNIQUES FOR THE MAGNETOPHONON EFFECT

2.1 Magnetic field generation and measurement

The size of the magnetic field required for a magnetophonon experiment depends principally on the effective mass of the carriers in the semiconductor. Indium antimonide has an effective mass of only \(0.015m_e\), and its fundamental magnetophonon resonance at about 3 tesla. Its disadvantage is that it is easily available only in bulk samples. Gallium arsenide was used in these experiments as well developed device technology was necessary. It has an effective mass of \(0.067m_e\) and a fundamental resonance field of 22 tesla, so for most experiments it was necessary to use only the resonances with \(n > 3\), with their smaller amplitudes. These occur below 8 tesla, which was the maximum available in house. Magnetic fields up to 25 tesla were available for occasional experiments at the High Field Magnet Laboratory in Nijmegen, Holland.

Both superconducting magnets and electromagnets can be designed to produce 8 tesla. Generally electromagnets are preferable for experiments in which the magnetic field is swept and superconducting magnets are better for steady fields. Only a superconducting magnet was available, and this led to lengthy experiments and hence to some problems with temperature control.

The magnet was an old Oxford Instruments model. This had a large residual field and much hysteresis, and the field was found to vary with current differently in different places. Because of this a field sensor close to the sample was required, which then experienced the variable temperature of the sample. A Hall probe was used as this required only small temperature corrections. The Hall probe characteristic was significantly non-linear, and all the magnetophonon oscillations shown later have as the x-axis the Hall probe voltage. This has been converted to an approximate magnetic field value, denoted by \(B^*\), by assuming linearity. The relationship between \(B^*\) and true magnetic field is shown in figure 2.1.1.
EXPERIMENTAL TECHNIQUES FOR THE MAGNETOPHONON EFFECT

2.1 Magnetic field generation and measurement

The size of the magnetic field required for a magnetophonon experiment depends principally on the effective mass of the carriers in the semiconductor. Indium antimonide has an effective mass of only $0.015m_0$, and its fundamental magnetophonon resonance at about 3 tesla. Its disadvantage is that it is easily available only in bulk samples. Gallium arsenide was used in these experiments as well developed device technology was necessary. It has an effective mass of $0.067m_0$ and a fundamental resonance field of 22 tesla, so for most experiments it was necessary to use only the resonances with $n \geq 3$, with their smaller amplitudes. These occur below 8 tesla, which was the maximum available in house. Magnetic fields up to 25 tesla were available for occasional experiments at the High Field Magnet Laboratory in Nijmegen, Holland.

Both superconducting magnets and electromagnets can be designed to produce 8 tesla. Generally electromagnets are preferable for experiments in which the magnetic field is swept and superconducting magnets are better for steady fields. Only a superconducting magnet was available, and this led to lengthy experiments and hence to some problems with temperature control.

The magnet was an old Oxford Instruments model. This had a large residual field and much hysteresis, and the field was found to vary with current differently in different places. Because of this a field sensor close to the sample was required, which then experienced the variable temperature of the sample. A Hall probe was used as this required only small temperature corrections. The Hall probe characteristic was significantly non-linear, and all the magnetophonon oscillations shown later have as the x-axis the Hall probe voltage. This has been converted to an approximate magnetic field value, denoted by $B^*$, by assuming linearity. The relationship between $B^*$ and true magnetic field is shown in figure 2.1.1.
2.2 The sample holder

The sample had to be maintained at typically 150 Kelvin while surrounded by a superconducting magnet at 4 Kelvin. The heat leak had to be less than one watt to avoid boiling off unreasonable amounts of liquid helium. The sample space was connected to the helium bath of the magnet by a needle valve so that it could be filled with liquid helium and pumped on to reduce the temperature of the sample below 4 kelvin. This needle valve did not seal well enough to allow thermal isolation of the sample merely by evacuating the sample space, so the sample, mounted on an electrically heated copper block and connected by very thin wires, was surrounded by a copper vacuum can. Figure 2.2.1 is a diagram of the assembly used. This was supported by about 1 metre of low conductivity copper-nickel tube, which reached to the top of the cryostat, and served both to carry the connecting wires and as a vacuum line. At the top of the cryostat a side tube was connected for the vacuum pump, while the wires were taken upwards through an araldite seal to a shielded box carrying BNC connectors.

Heat loss was adequately reduced by this assembly from the point of view of helium evaporation, but that which remained, in conjunction with the low thermal mass in contact with the sample due to lack of space, was still sufficient to make difficult the very precise temperature control necessary. The principal problem was heat conduction along the wires, the temperature of which fell almost to that of the helium bath only slightly above the vacuum can, quite close to the sample. In addition the Wood's metal seal was hard to make and required leak testing. There was not room for a conventional "O" ring seal.

Later a new Technology Systems superconducting magnet became available. This would produce slightly higher fields and sweep nearly twice as fast. The insert was designed to produce very low temperatures by pumping on previously condensed He\(^3\), and so was completely sealed from the He\(^4\) bath.
FIGURE 2.2.1

SAMPLE HOLDER (twice actual size)

- Copper-nickel tube
- Wood's metal seal
- Copper vacuum can
- Terminal pins
- Plastic support
- Hall probe
- Copper block on which sample is mounted.
- Heater resistor
The procedure used to attain high temperatures was as follows: dry nitrogen gas was used to flush the sample space while the sample was loaded into the magnet which was at 77K or above. Most of this nitrogen was then removed with a rotary vacuum pump. On cooling the magnet to 4K the remaining nitrogen froze on the walls, leaving an excellent vacuum for the whole distance from the sample to the top of the cryostat. This meant that less heat loss along the wires which stayed far above 4K along their whole length. The extra space freed by being able to do without a vacuum jacket meant the sample could be mounted on a much larger copper block. These factors lead to the thermal time constant of the sample being tens of minutes, which enabled far better temperature control. Except for the absence of the vacuum jacket, the design of the sample holder was along the same lines as that of figure 2.2.1.

2.3 Temperature measurement and control

Temperature control was found to require careful attention. The allowable variation is hard to quantify as only the Fourier components with periods similar to the magnetophonon oscillations (≈1 minute) will pass through the digital filters, but these components must change the device resistance by less than one part in $10^4$ to be comparable to other sources of noise (section 2.5). The device resistance was always very sensitive to temperature variations because the optimum conditions for the magnetophonon effect are where the mobility has (just) started to be determined by scattering by optic phonons, the number of which increases exponentially with temperature. One part in $10^4$ corresponds to perhaps 0.01 kelvin. The controller gain cannot be made large to suppress these variations as the frequencies of interest are not far below the frequency at which large phase shifts appear between variations in heater power and the response of the temperature sensor. The closed loop gain of the controller must be less than one at this frequency or instability will result. The gain should in
fact be kept well below this, or the response to a change of conditions will be damped oscillations, which in this instance may be worse than a gradual recovery without overshoot even if the maximum deviation is increased.

Most of the temperature variations result from the changing magnetic field. Changes in the resistance of the heater are usually small and can be compensated for by placing a resistor of equal value in series with it. The heater power is then independent of changes in heater resistance to first order. The heat losses from the sample holder were found to increase when the magnetic field was swept. This was shown by a fall in the sample temperature after an up and down sweep when using a constant power heater previously carefully adjusted to keep the temperature constant. This may result from the increased evaporation of helium when the magnet is in use cooling other parts of the cryostat.

The temperature sensor was not completely insensitive to the large magnetic field. Both a Lakeshore Cryogenics carbon-glass resistance thermometer and a copper-constantan thermocouple were used. The specification for the resistance thermometer is that $\Delta T/T$ is less than 0.5% at 100 kelvin and 8 tesla. This is considered to be good. The temperature dependence of the thermocouple was found experimentally to be similar in magnitude. In both cases a magnetic field was equivalent to an increase in temperature. Compensation for this effect was not attempted.

The much longer thermal time constant in the second magnet greatly reduced these problems. The greatest remaining problem was considered to be the temperature shift from the effect of the magnetic field on the sensor, so a constant power heater was used while the magnetic field was swept and any slight long-term drift was corrected by a temperature controller when the magnetic field had returned to zero.
2.4 Extraction of the oscillations

Since magnetophonon oscillations are typically much smaller than the monotonic magnetoresistance, a method of separating the two is needed. Two techniques have been widely used: double differentiation, and fitting a smooth curve followed by subtraction. The original analogue methods of achieving these were to use two R-C networks to differentiate twice with respect to time or partially to cancel the monotonic magnetoresistance by a voltage derived from a Hall probe. Stradling and Wood (1970) showed that analogue double differentiation is better than the cancellation method for showing up any fine structure in the oscillations. However, it is much easier to determine the amplitude of the oscillations with a cancellation method.

A double differentiator followed by an inverter has a gain proportional to frequency squared and zero phase shift at all frequencies. The gain must be reduced for frequencies higher than those present in the magnetophonon oscillations or the signal will be lost in much amplified high frequency noise. Unless this gain reduction starts well above the signal frequencies the phase shift at the signal frequencies will no longer be zero, which will appear as a false shift in the peak positions. This can be largely compensated for by averaging the apparent peak positions for experiments with B increasing and B decreasing at the same rate. Typically the time constant for the R-C network is chosen to be 0.1 times the average period of the oscillations.

The cancellation method is generally adequate over only part of the magnetoresistance curve and must be manually adjusted during the experiment. This rapidly becomes harder to do as the oscillations become smaller.

If the magnetoresistance data can be digitised better methods are available. Digital filters can easily be made to combine zero phase shift at all frequencies with the desired gain variation.
The filters used in this work were of the form

\[ s = k \]

\[ Y^s(n) = \sum_{s} a_s Y(n+s) \]

where \( n, s \) and \( k \) are integers, \( Y^s(n) \) is the \( n \)th output value and \( Y(n) \) is the \( n \)th input value. This is known as a non-recursive filter since the expression for \( Y^s(n) \) contains only \( Y \) values and not other \( Y^s \) values. If the interval between data samples is constant, the condition for zero phase shift at all frequencies is that the filter coefficients are symmetrical, i.e. \( a_s = a_{-s} \). The algorithm that is equivalent to double differentiation is \( Y^s(n) = -Y(n+1) + 2Y(n) - Y(n-1) \). This has a response proportional to frequency squared up to the frequency at which the period is comparable with the interval between data points, when it starts to fall at the same rate. The high frequency response could therefore be controlled by modifying this algorithm to operate on points separated by more than one data interval, \( Y^s(n) = -Y(n+2) + 2Y(n) - Y(n-2) \) for example. This provides only crude control, and there are ranges of unwanted higher frequencies where this algorithm again has a fairly large gain. The first version was always used and the high frequency response limited by a separate filter.

The order in which the filters is applied is immaterial. The low pass filter generally used was \( Y^s(n) = Y(n+2) + 3Y(n+1) + 5Y(n) + 3Y(n-1) + Y(n-2) \). The response of algorithms like this is constant at low frequencies and inversely proportional to frequency for higher ones. The transition frequency depends mainly on the separation of the first and last points used in the above average, but the relative sizes of the coefficients is obviously important also. Like the "differentiating" filter this has false responses at high frequencies (known as aliasing) but the larger number of points used makes this less severe and no problems were encountered in practice. The data taking program described later provided paired readings of Hall probe voltage and the corresponding sample voltage or current, denoted by \( X(n) \) and \( Y(n) \). The interval between the Hall probe voltages of
successive readings was not perfectly constant, so it seemed desirable to
correct for this. Other reasons for variable point spacing are possible,
such as conversion of the Hall probe voltages to corresponding B values or
1/B values. The correction used for the low pass filter was to replace X(n)
by the actual average of the X values of the points used in the calculation
of $Y^*(n)$ with the same weighting, i.e. $X^*(n) = X(n+2) + 3.X(n+1) + 5.X(n)$
$+3.X(n-1) + X(n-2)$. The double differentiation algorithm was replaced by
two iterations of $Y^*(n) = [Y(n+1) - Y(n-1)]/[X(n+1) - X(n-1)]$ and $X^*(n) =
[X(n+1) + X(n-1)]/2$, each of which is equivalent to single differentiation.
If $\Delta X$ is constant this is the same as the second double differentiation
algorithm given above.

The approach taken to the curve-fitting method was to calculate the
polynomial of a chosen order that was the best fit to the magnetoresistance
data by the least squares method. Generally a cubic was used. Higher order
polynomials suppress the monotonic resistance more effectively, which can
be important when searching for very small oscillations. The problem with
them is that the inevitable fitting error has more zeros as the order of
the polynomial increases (number of zeros equals order plus one) and
becomes increasingly similar in frequency and hence indistinguishable from
the desired oscillations.

A similar problem arises with the differentiation method. If a system of
filters with a very narrow passband centred on the magnetophonon
frequencies is used it may be very hard to tell whether the result is
mainly oscillations or just filtered noise or the small Fourier component
of the monotonic background at the same frequency. There is no way of
distinguishing without uncertainty between background and oscillation
unless the exact form of one is known beforehand, so some similar problem
will occur with any method.
Another, fundamentally different, technique is to modulate the magnetic field sinusoidally and to extract the second derivative by measuring the component of the sample voltage at twice the modulating frequency with a lock-in amplifier. Measuring the desired quantity directly in this way assists greatly in obtaining a high signal to noise ratio, and this appears to be the optimum technique for magnetophonon experiments in which only moderate fields are required (Blakemore et al., 1975 or Kasai et al., 1978). This is because a 0.1% change in device resistance unrelated to the increase in magnetic field is as large as typical magnetophonon oscillations and will degrade them appreciably if one of the above extraction routines is used. This 0.1% change in device resistance will be accompanied by a change of similar magnitude in $d^2R/dB^2$. Normally the magnetophonon oscillations are a large part of $d^2R/dB^2$, so if this can be obtained directly the effect of the spurious change will now be only 0.1% of the amplitude of the magnetophonon oscillations and probably unnoticeable.

This method has difficulties as well as advantages. The second harmonic signal is generally very small, and is subject not only to noise from preamps etc. but also, more seriously, to coherent interference from second harmonic distortion in the signal generating the modulation. A much larger signal is obtained in a compromise technique in which $dR/dB$ is derived from the component of the sample voltage at the modulation frequency, and then the second differentiation performed numerically. The coherent interference, this time from inductive or capacitative coupling between the sample and the modulating system, may however be greater also. Blakemore et al. (1975) state that the B modulation technique is not the best when large magnetic fields are required, but do not give reasons. The problems with the technique encountered in this work are described in section 2.6.
2.5 The data collection system

This application requires high accuracy (so that the errors will be smaller than the small magnetophonon oscillations) but only a low data rate, as superconducting magnets necessarily sweep slowly. Five and a half digit Thurlby digital voltmeters with RS-232 serial output ports satisfied this. These then had to be interfaced to an available BBC computer, which also has a RS-232 port. Figure 2.5.1 is a diagram of the hardware used. The RS-232 interconnection is designed for only one transmitter, but this was overcome by switching the computer's RS-232 port under computer control between the Thurlby voltmeters. The alternative of converting both the Thurlbys and the BBC to the Hewlett-Packard interface bus (IEEE standard bus for interconnecting instruments) was considered too expensive for this simple application. The RS-232 signal levels are + and −15 Volts, so logic switching cannot be used, but CMOS analogue switches can. Other chips are required to receive and store the desired state of the switch as signalled by the computer.

It was desired to produce graphs of the oscillations for a permanent record and ease of making comparisons. This was achieved by interfacing two 12-bit digital to analogue converters to the BBC computer and using the output to drive an analogue X-Y plotter. The 12 bits gave a resolution of less than 0.1 mm, much less than the line width. The data points were joined by cubic "splines" so that the resulting continuous curve passed through all the points and had no discontinuities in gradient.

The input parameters to the program are the name of the data file to be created, and the lower and upper limits of magnetic field. The point spacing is calculated by assuming 150 equidistant points are to be taken. This is enough even for any quite fine structure in the oscillations to be seen. Taking more points increases processing time later. The output of the RS-232 ports of the Thurlby meters is in standard ASCII and consists of the reading preceded by an "R". A new reading is sent every 0.3 seconds.
**Data Flow**

- LMHz DATA BUS
- 12 bit D to A Converter AD7542
- 12 bit D to A Converter AD7542
- Output Port 74LS373
- X - Y Plotter
- Computer RS232 Port
- THURLBY METERS
- pen up/down

**Address Decoding**

- D to A
- Port
- 74LS02

**Figure 2.5.1**
Because the operations of the computer and the meters are not synchronised, it is possible that the computer will switch to one of the meters part way through an output sequence, resulting in a false reading. To prevent this the program waits for another reading if the first character received is not an "R". Readings are taken continuously and displayed but otherwise ignored until the X (Hall probe) values enter the desired range. Readings are then averaged until the difference between the initial X value and the average X value exceeds half of the desired point spacing. The average X and Y (resistance) values are then stored in memory, the point plotted on a graph on the computer monitor and the variables used as accumulators in the averaging reset to zero.

This procedure will work for increasing or decreasing magnetic fields and gives values that are quite equally spaced. A previous procedure was based on data collection for a point ceasing when the average X value of the current point exceeded the average X value for the previous point plus the desired point spacing (changed appropriately for B decreasing). This developed an instability in which the actual point spacing was alternately shorter and longer than that desired.

The error due to all the Y values being measured just after or just before the corresponding X values was found to be undetectable, as suggested by calculation.

When the X values move out of the desired range the 150 or so averaged X and Y values are stored on disc. The oscillations are then be extracted as in section 2.4. Scaled graphs are drawn of the results at all stages to enable decisions to be made on how to proceed.

The program also provides conversion of resistance data to conductance data and vice versa. Other features that could be included are conversion of Hall probe voltages to B and 1/B values, and Fourier analysis.

An annotated listing of the program appears in appendix A.
2.6 Sources of noise in magnetophonon measurements

Only noise at frequencies similar to those of the magnetophonon oscillations was important as the rest was rejected by the digital filters. The oscillation frequency was set by the sweep rate of the superconducting magnet, giving a period of about a minute.

The three main sources of noise were temperature variations, the instrumentation, and the sample itself. The last two are discussed below.

Noise measurements were made by using the data collection system to record device resistance against time instead of against magnetic field. A voltage ramp generator was used for the X input and set for a sweep time of 15 minutes or so, similar to the time usually taken for a magnetic field sweep. The data was then passed through the digital filters and the noise figures quoted are the amplitude of the output from these divided by the mean value. They are thus directly comparable with the oscillation amplitudes reported later as $\Delta R/R$, which lie in the range $10^{-1}$ to $10^{-5}$. Any contribution to the noise by the X input is proportional to $dY/dX$, which is almost zero in these noise measurements.

The measurements of device resistance were made using either a constant current or a constant voltage. For the measurement on narrow channels the voltage across the device had to be limited to about 10mV to prevent the channel width from being markedly different at the two ends of the channel (section 3.2). Constant voltage measurements had the advantage of needing no adjustment as the channel width was varied, but the small voltage allowable produced large amounts of noise in the current meter used, a Keithley 480 picoammeter. The Thurlby meter measured the analogue output from the rear of this. The noise figure at 10mV was $5\times10^{-4}$. This was independent of device resistance but inversely proportional to the sample voltage.
Constant current measurements were rather better than this. The Thurlby meter now measured the device voltage directly, and the noise figure was found to be $6 \times 10^{-5}$ at 10mV. This is better than the $1\mu V$ resolution of the Thurlby meters, and so is attributed to them. Here too the noise is determined by the device voltage, as the Thurlby meters retain their $1\mu V$ resolution for inputs up to 200mV, equivalent to a noise level of $5 \times 10^{-6}$. Clearly current and voltage sources were required to be stable to this degree. New batteries were found to be excellent, a noise figure of $2 \times 10^{-6}$ being measured by recording the net voltage of two nominally identical PP9s connected back to back. The noise figure was dramatically greater for rather flat batteries, even at currents of a few nanoamps. Resistor noise is quoted as $10^{-7}$ to $10^{-6}$ for good quality types and none could be detected by this apparatus. Variable resistors however were very poor, and occasional trouble was experienced with dirty switch contacts in resistive dividers.

This technique for observing the magnetophonon effect relies on very accurate measurement of the device resistance. The d.c. measurement techniques just described were preferred to the more common a.c. ones because the available oscillators and lock-in amplifiers were much less stable than the above d.c. voltage sources and voltmeters. The principal design objective of the a.c. equipment is the ability to extract signals from a very noisy background rather than extreme accuracy. The method used is sensitive to noise at frequencies of about $10^{-2}$ Hertz. The only source of noise in this frequency range that could have been eliminated by an a.c. measurement is fluctuations in thermal voltages. Other types of noise affect the device resistance itself and so will be present however the resistance is measured. Care was always taken to minimize thermal voltages by changing conductor type at sensible places, so the noise in the thermal voltage caused by temperature fluctuations was much less than that caused by the effect of the temperature fluctuations on the device resistance.
The most severe requirement was the gate voltage supply, because noise in this is amplified by the field effect transistor (FET) being measured. In the experiments of section 3.5 a large gate voltage was used to reduce the conductance of the FET by a factor of typically a hundred from its zero gate voltage value ($\sigma_0$). A change in the gate voltage then caused a fractional change in the device conductance approximately a hundred times as great. A Keithley 230 programmable voltage source was used. This was easily adjustable and at least as quiet as the above batteries when measured in the same way. Many results from section 3.5 suggest the noise figure was about $10^{-6}$. The line regulation specification was only $10^{-4}$ however, so a constant voltage transformer was used in the supply to the instruments to prevent their stability being degraded by fluctuations in the mains supply.

There appeared to be no significant noise arising from variations in the Hall probe voltage or its measurement.

The inherent noise of the devices was large enough to be observable. A noise measurement on one device at 300 kelvin found noise figures of $10^{-4}$ at $\sigma_0/10$ and $10^{-3}$ at $\sigma_0/100$. The noise at $\sigma_0$ was below the level of $6 \times 10^{-5}$ set by the meters. On cooling this device to 77 kelvin the noise figure at $\sigma_0/100$ fell by a factor of 5 to that expected from the gate voltage noise, suggesting that there is device noise present which increases rapidly with temperature and is predominant at 300 kelvin. To check this similar measurements were carried out on a silicon j-fet. This was expected to be quieter because of the larger number of carriers in its channel and the absence of a heterojunction with a high density of traps. The noise figure at both 300 and 77 kelvin was the same as that of the GaAs fet at 77 kelvin, confirming that this noise is from the gate voltage source. At the temperatures of about 130 kelvin used in most of the experiments device noise was generally only a small part of the total. However, with the highest doped layers difficulty was experienced in producing devices with
very low gate leakage currents, and these devices showed noise that increased rapidly with gate voltage and correlated with the measured leakage currents.

Drift in the devices was often a problem. Usually, particularly at high temperatures, drift in the direction expected from relaxation was observed, e.g. an increase in gate voltage would produce an immediate decrease in conductivity, followed by a gradual, partial recovery as traps adjusted to a new equilibrium. This drift was always greatly slowed by cooling to 130 kelvin, then varying between different batches of devices from imperceptible to problematic. In one case, two batches of devices from the same layer showed very different amounts of drift. The processing was similar except that the worse batch had the Schottky gates deposited after the layer had been etched to reduce its thickness. This suggests that the drift is associated with traps at this interface.

Some devices showed a type of drift in which they would gradually turn themselves off completely when set near pinch-off, regardless of whether the gate voltage had just been increased or decreased. This may be related to the large change in pinch-off voltage on cooling also observed by Poole (1982) in similar devices. The causes of both are unknown.

Attempts to use the B modulation technique gave very poor results, but little investigation was performed of the scope for improvement. Coherent interference, apparently from inductive coupling of the modulation coil and the leads to the sample, was the greatest problem. The small permissible d.c. voltage across the sample was a contributing factor since the desired signal is proportional to this.
There were also difficulties in achieving the modulation. The Oxford Instruments magnet had no modulation coil, so one was wound onto the sample holder. This failed because of the vibrations set up by the force of the main field on the modulating coil. The Technology Systems magnet had a modulation coil attached to the cryostat, but there was strong coupling with the main coil. At frequencies below 20 hertz the voltage induced in the main coil audibly varied the switching frequency of the switched mode power supply connected to the main coil. This coupling seemed to greatly reduce the modulation amplitude above about 10 hertz.
3.1 Device structure and fabrication

Two designs of device were available, the metallisation patterns for which are shown in figure 3.1.1. They are a Hall bar with length 750\(\mu\)M, width 115\(\mu\)M and four equally spaced sidearms per side, and a 20\(\mu\)M (gate length 10\(\mu\)M) device 600\(\mu\)M wide. These will be referred to as long and short devices.

Figure 3.1.2 shows a schematic longitudinal section of these devices. Isolation was achieved by etching away the active layer surrounding each device. Bonding pads would normally be deposited on the substrate exposed by this process, and metallisation run up the sides of the device "mesa" to make contact to the correct region. Since many of the layers used had active layers many microns thick, problems were encountered in achieving continuous metallisation over the large step at the edge of the mesa. This was overcome by extending the mesa to include the bonding pads and metallisation regions also.

The very high mobility layers were grown in a hydride VPE machine. This grew excellent material but was not set up with p-type dopants. It was found to be impossible to grow a low-doped n-type layer on a normal p-type substrate with this machine, apparently because of an autodoping effect.

When a back gate was required this problem was overcome by growing a p-type layer on a semi-insulating substrate in a MOCVD reactor, then transferring this to the hydride VPE reactor to grow the active layer after an in-situ etch. Most of the layers used for this work were grown at the G.E.C. Hirst Research Centre and processed at the S.E.R.C. 3-5 semiconductor facility at Sheffield University.

Figure 3.1.3 is a summary of the properties of the main layers used.
3.1 Device structure and fabrication

Two designs of device were available, the metallisation patterns for which are shown in figure 3.1.1. They are a Hall bar with length 750μM, width 115μM and four equally spaced sidearms per side, and a 20μM (gate length 10μM) device 600μM wide. These will be referred to as long and short devices.

Figure 3.1.2 shows a schematic longitudinal section of these devices. Isolation was achieved by etching away the active layer surrounding each device. Bonding pads would normally be deposited on the substrate exposed by this process, and metallisation run up the sides of the device "mesa" to make contact to the correct region. Since many of the layers used had active layers many microns thick, problems were encountered in achieving continuous metallisation over the large step at the edge of the mesa. This was overcome by extending the mesa to include the bonding pads and metallisation regions also.

The very high mobility layers were grown in a hydride VPE machine. This grew excellent material but was not set up with p-type dopants. It was found to be impossible to grow a low-doped n-type layer on a normal p-type substrate with this machine, apparently because of an autodoping effect. When a back gate was required this problem was overcome by growing a p-type layer on a semi-insulating substrate in a MOCVD reactor, then transferring this to the hydride VPE reactor to grow the active layer after an in-situ etch. Most of the layers used for this work were grown at the G.E.C. Hirst Research Centre and processed at the S.E.R.C. 3-5 semiconductor facility at Sheffield University.

Figure 3.1.3 is a summary of the properties of the main layers used.
FIGURE 3.1.2

<table>
<thead>
<tr>
<th>Layer</th>
<th>N doping</th>
<th>Depth</th>
<th>Substrate</th>
<th>77K mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV73</td>
<td>$2 \times 10^{14} \text{cm}^{-3}$</td>
<td>6.0µm</td>
<td>semi-insulating</td>
<td>6</td>
</tr>
<tr>
<td>B257</td>
<td>$2 \times 10^{14} \text{cm}^{-3}$</td>
<td>5.0µm</td>
<td>semi-insulating</td>
<td>14</td>
</tr>
<tr>
<td>B977</td>
<td>$4 \times 10^{14} \text{cm}^{-3}$</td>
<td>5.0µm</td>
<td>P on s.i.</td>
<td>7</td>
</tr>
<tr>
<td>B1967</td>
<td>$2 \times 10^{15} \text{cm}^{-3}$</td>
<td>3.2µm</td>
<td>P on s.i.</td>
<td>-</td>
</tr>
<tr>
<td>B1988</td>
<td>$6 \times 10^{15} \text{cm}^{-3}$</td>
<td>4.0µm</td>
<td>P on s.i.</td>
<td>-</td>
</tr>
<tr>
<td>B1220</td>
<td>$2 \times 10^{16} \text{cm}^{-3}$</td>
<td>1.5µm</td>
<td>P</td>
<td>-</td>
</tr>
<tr>
<td>B1218</td>
<td>$6 \times 10^{16} \text{cm}^{-3}$</td>
<td>0.4µm</td>
<td>P</td>
<td>-</td>
</tr>
</tbody>
</table>

FIGURE 3.1.3
3.2 Mesfet – theory of operation.

Both the metal gate and the p-type layer initially contain unoccupied states with lower energies than those in the n-type active layer. Electrons from the active layer then move into these states, leaving behind regions depleted of free electrons and hence with a net positive charge. This continues until the dipoles created by this charge movement produce a potential barrier large enough to prevent further transfer. By biasing the gates negatively with respect to the source these depletion regions can be extended, reducing the conductivity between the source and drain by narrowing the conducting channel connecting them. If a positive voltage is applied to the drain to produce a source to drain current, the depletion regions will constrict the channel more near the drain as the gate to channel potential is greater there. If the drain voltage is large the depletion regions will meet somewhere near the drain end of the channel. Current will continue to flow as electrons are injected into the depletion region from the channel, but the current is now nearly independent of the drain voltage, since increasing it only moves the point of intersection of the depletion regions slowly nearer to the source, affecting the field in the undepleted region only a little (assuming $L \gg d$). This is the usual operating condition of the device, but for the experiments described later very small drain voltages were used so that the channel width was nearly uniform and the source – drain current obeyed Ohm’s law, with the resistance determined by the gate voltage.

A simple expression for the source – drain conductivity can be obtained by making the following assumptions:

The channel doping level $N_d$ is constant.

The electron mobility is constant.

The depletion regions have zero free electron concentration, but this rises immediately to $N_d$ at the edges. This is called the depletion approximation.
Integration of Poisson's equation shows that the potential change across a depletion region of length \(d\) is \(\frac{e^2 N_d e}{2e}\). Equating this to the gate to source voltage, \(V_g\), the proportion of the channel remaining undepleted is

\[
1 - \left[ \frac{V_g}{V_p} \right]^{0.5},
\]

where \(V_p\) is the "pinch-off" voltage, defined by \(V_p = \frac{d^2 e N_d}{8e}\).

The channel conductivity is proportional to this. The contact potentials between the gates and active layer are of the order of a volt and must be added to the externally applied bias to determine \(V_g\).

The depletion approximation will be poor when the thickness of the channel remaining undepleted is less than the thermal blurring of the edges of the depletion regions. The Debye length, defined as

\[
L_D = \left[ \frac{e^2}{\varepsilon N_d} \right]^{0.5},
\]

is a measure of this blurring. \(L_D\) is the distance into a depletion region at which the potential energy of an electron has risen by 0.5 \(\varepsilon KT\). In the limit of gate voltages much greater than \(V_p\) there will be too few electrons to appreciably affect the potential, which is therefore given by \(x^2 \varepsilon KT / 2L_D^2\), where \(x\) is the distance from the centre of the channel. The electron concentration will vary as \(N_e \exp(-x^2 / 2L_D^2)\) because of this increasing potential. The number of electrons per unit area is then \(N_e L_D (2\pi)^{0.5}\), so the electron distribution has an effective width of \(L_D (2\pi)^{0.5}, \approx 2.5 L_D\).

In addition, because there are too few electrons for screening to occur, the potential in all parts of the active layer will follow changes in the gate potential and the electron concentration will vary throughout as \(\exp(-eV_g/\varepsilon KT)\). This region in which the number of electrons varies exponentially with gate voltage is called the subthreshold region.

In the transition region between thicknesses much larger than \(L_D\) and the subthreshold region the electron distribution can be found by numerical integration, as described in appendix B. The results of this are shown in figure 3.2.1. Note that the width of the electron distribution never falls below about 2.5 \(L_D\).
Three possible measures of the width of the conducting channel, all of which are the same for thicknesses much greater than $L_D$, are:

1. The thickness calculated from the depletion approximation.

2. The number of electrons per unit area divided by $N_0$.

3. The actual width of the electron distribution.

For large values of $V_g$, (1) becomes negative, (2) tends to zero and (3) tends to $L_D$. (2) is most closely related to the sample conductivity and will be denoted by $t$. The numbers by the curves of figure 3.2.1 are $t/L_D$, and so are the number of electrons present per unit area of the channel divided by the number per unit area in an undepleted channel $L_D$ wide.

3.3 Experimental Results and Discussion

Many combinations of doping level, device geometry, temperature and relative orientation of device, current and magnetic field are possible. These have been grouped according to the results obtained. Results for B perpendicular to $j$ will be considered first.

**Long devices, B parallel to layer**

In this configuration the results were the same as for bulk samples:

With all materials resistance maxima occurred at the fields expected for magnetophonon resonance. The relative amplitudes of oscillation were in the same order as the 77K mobilities. The temperature dependence of the oscillation amplitude was as expected for ohmic conditions and is shown for a B257 device in figure 3.3.1.

**Long devices, B perpendicular to layer**

With most materials similar results to those above were obtained, but with B257, the highest mobility material, the oscillations were much smaller and became inverted above a magnetic field of about 5 tesla (figure 3.3.2).

**Short devices**

These showed minima in resistance at resonance for B perpendicular to the layer but maxima for B parallel to the layer (figure 3.3.2).
Figure 3.3.1

$\Delta R/R, \% \ (n = 4 \text{ to } n = 3\frac{1}{2})$

Temperature (kelvin)

$B^*$

235 K
195 K
155 K
95 K
75 K
(1) Long device, B parallel to layer
(2) Short device, B parallel to layer
(3) Long device, B perpendicular to layer
(4) Short device, B perpendicular to layer

Aspect Ratio

375
10
7
0.03
(1) Long device, B parallel to layer
(2) Short device, B parallel to layer
(3) Long device, B perpendicular to layer
(4) Short device, B perpendicular to layer

Aspect Ratio
375
10
7
0.03
Explanation

Figure 3.3.3 defines some symbols used in this section. In a long sample the ratio of voltage to current is proportional to $\rho_{xx}$, but in a short sample it is proportional to $1/\sigma_{xx}$.

This can be seen as follows:

$$j_x = \sigma_{xx} E_x + \sigma_{xy} E_y$$

A short sample will have its Hall voltage shorted out by the low resistance ohmic contacts at each end, so that $E_y = 0$.

Hence

$$\frac{V}{I} = \frac{1}{w t} j_x = \frac{1}{w t} \times \frac{1}{\sigma_{xx}}$$

Since $\sigma$ and $\rho$ are inverse tensors,

$$\frac{1}{\sigma_{xx}} = \rho_{xx} + \left(\frac{\rho_{xy}}{\rho_{xx}}\right)^2$$

$\rho_{xy}$ is zero with no magnetic field, but from the simplest theory of the Hall effect with a magnetic field in the $z$ direction $\rho_{xy}$ is given by

$$\frac{B}{(n.e)}$$

where $n$ is the concentration of electrons. Hence

$$R(B) = \frac{1}{w t} \left( \frac{\rho_{xx}(B)}{(n.e)} + \left(\frac{B}{(n.e)}\right)^2 \times \frac{1}{\rho_{xx}(B)} \right)$$

For small values of $B$, $R$ will be proportional to $\rho_{xx}$ and show maxima at magnetophonon resonance, and for large values of $B$, $R$ varies as $1/\rho_{xx}$ and will show minima. Minima always occur at fields large enough to observe magnetophonon oscillations.

A long device will have end regions in which the Hall voltage is shorted and a central region in which it is not. The magnetophonon oscillations in these regions will be out of phase and may almost cancel. Changes between maxima and minima at resonance can occur if the amplitudes of oscillation in the two regions do not vary identically with $B$.

The length of the region at each end of the device in which the Hall voltage is shorted will be similar to the channel width $w$. This can be seen from figure 3.3.3: the length of the "short circuit" path through the end contact is about $k$. For $k \gg w$ the "short circuit" path will have a higher
region in which $E_y = 0$

$V_t = V_1 - V_5$

$V_B = V_2 - V_3$

$V_h = V_4 - V_3$

FIGURE 3.3.3
resistance than the region between A and B across which the Hall voltage is developed and so will have little effect, and also conversely. The results therefore depend on the length to width ratio of the device seen along the direction of B - if this is large a small proportion only of the device will have its Hall voltage shorted and resistance maxima will occur at resonance, and vice versa.

The results, summarized below, are in accordance with these ideas.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Aspect ratio</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long, Bilayer</td>
<td>375</td>
<td>Maxima, $\Delta R/R = 3%$</td>
</tr>
<tr>
<td>Short, Bilayer</td>
<td>$\geq 10$</td>
<td>Maxima, $\Delta R/R = 1.5%$</td>
</tr>
<tr>
<td>Long, Bilayer</td>
<td>7</td>
<td>Inverts $B \geq 5T$. $\Delta R/R = 0.1%$</td>
</tr>
<tr>
<td>Short, Bilayer</td>
<td>$\geq 0.03$</td>
<td>Minima, $\Delta R/R = -1.5%$</td>
</tr>
</tbody>
</table>

The following simple calculation demonstrates that this effect is of the appropriate size. Consider the model of the device illustrated in figure 3.3.3. The Hall voltage is assumed to be completely shorted within the regions of length $k/2$ and unaffected in the central region.

From this,

$$R = \frac{k}{w.t} \times \frac{1}{\rho_{xx}} + \frac{(1-k)}{w.t} \times \rho_{xx}$$

$$= \frac{1}{w.t} \left[ \rho_{xx} + \frac{k}{1} \left( \frac{B}{n.e} \right) \times \frac{1}{\rho_{xx}} \right]$$

From which

$$\frac{dR}{d\rho_{xx}} = \frac{1}{w.t} \left[ 1 - \frac{k}{1} \left( \frac{B}{n.e} \right)^2 \times \left( \frac{1}{\rho_{xx}} \right)^2 \right]$$

$$= \frac{1}{w.t} \left[ 1 - \frac{k}{1} (\mu B)^2 \right]$$

where $\mu = \frac{1}{n.e.\rho_{xx}}$ and is the drift mobility.

The condition for the oscillations to invert is that this derivative is zero, i.e. $\mu B = \gamma(1/k)$, somewhere in the range of interest. This depends on the aspect ratio and mobility of the device.
\( \mu B \) can be found from the Hall voltage between opposite sidearms of a long device:

\[
\mu B = \frac{E_Y}{E_X} = \frac{1}{w} \frac{V_h}{V_t}
\]

\( 1/k \) can be found from the ratio of the voltage between adjacent sidearms to the voltage across the whole device:

\[
V_S = I \times \frac{l}{w \cdot t} \times \rho_{XX} \quad (l \text{ is the distance between adjacent sidearms})
\]

\[
V_t = I \times \frac{1}{w \cdot t} \times \rho_{XX} \times \left[ 1 + \frac{k}{l} (\mu B)^2 \right]
\]

The last term arises from the extra magnetoresistance of the region in which the Hall voltage is shorted. Hence

\[
\frac{1}{k} = \frac{1}{l \cdot V_t} \frac{V_S (\mu B)^2}{V_S - 1 \cdot V_S}
\]

The calculated value of \( 1/k \) is 19.6 at \( B = 3 \) tesla and 19.7 at 6 tesla. The constancy of this figure suggests that the model works fairly well.

\( 1/k = 19.6 \) means that \( k/w = 0.35 \), which is of order 1 as would be expected. Equating the expressions for \( (\mu B)^2 \) and \( 1/k \) shows that this criterion for oscillations to invert is equivalent to

\[
V_t = 2.1 V_S,
\]

i.e. twice the value expected when \( B = 0 \). For the long devices \( 1/l = 5 \).

Figure 3.3.4 shows \( V_t \) and \( 10 V_S \) on the same axes. They intersect at \( B = 3.8 \), whereas inversion was found at \( B = 5.5 \). This estimate can be improved by taking into account contact resistance, which will cause the channel voltage to be less than \( V_t \). At \( B = 0 \) \( V_t \) was about 20% greater than \( 5xV_S \).

There is also a slight difference in temperature between the two curves of figure 3.3.4.
Prediction of oscillation inversion.

FIGURE 3.3.4
These ideas can be checked experimentally by using the side contacts of a long device. A constant current was passed down a long device and the magnetophonon oscillations appearing between adjacent pairs of contacts are shown in figure 3.3.5, scaled so that the total voltage across the device can be found by adding them as they are drawn. It can be seen that the oscillations in the end regions are out of phase with those in the centre and increase more rapidly with B, eventually becoming large enough to change the phase of the total.

These ideas also explain the data of figure 3.3.6, in which the oscillation amplitude in a short device with B parallel to the layer at first increases as the conducting channel is narrowed by application of a negative gate voltage. In this case the aspect ratio is sufficient to prevent inversion, but the amplitude is noticeably reduced. Narrowing the conduction channel will increase the aspect ratio and prevent this.

B parallel to j

Minima in resistance near the resonant fields occurred. With long devices these minima were shifted to fields about 3% lower than the resonant fields, consistent with the results reported in Harper, Hodby and Stradling (1973), but no such shift was apparent with the short devices. Other differences are that the oscillation amplitude was much greater in the short devices, and the ratio of the resistance at 6 Tesla to that at 3 Tesla was 2.8, compared to only 1.2 under similar conditions for the long devices. One explanation for this is that in the short devices the gate occupies only 1/2 of the distance between the source and drain, and the depletion region below it is a large proportion of the layer depth. This means that there is a significant component of the current perpendicular to the layer. The interaction of this with the magnetic field will produce a Hall voltage across the 600μM. dimension of the device, which will be shorted by the contacts, as with Blj, Bilayer, producing large magnetoresistance and large, inverted oscillations at the resonant fields.
Initial increase in oscillation amplitude in B257 short device. Traces are for 100, 50 and 25 per cent of zero gate voltage conductance. Vertical scale: 2cm corresponds to $\Delta R/R$ is 0.8%.

FIGURE 3.3.6
Hot electron oscillations

By applying large electric fields (above $5 \times 10^3$ Vm$^{-1}$) it was possible to observe magnetophonon oscillations in MV73 devices even at low temperatures. The field required was appreciably greater for $B$ perpendicular to $j$ compared with $B$ parallel to $j$, presumably because of the extra magnetoresistance. At low temperatures the "impurity series" oscillations described in section 1.3 were observed. These were used to check that the Hall probe used for magnetic field measurements had been correctly calibrated. This calibration was carried out by measuring the positions of normal, ohmic magnetophonon oscillations in a wide variety of devices. These agreed well with each other, but some uncertainty remained because of the difficulty of deciding which experimentally observed peak corresponds to which resonance and hence what value of $B$ to assign to its position. The calibration selected was the one which brought the peaks of the impurity series closest to their expected positions. A correction was made for the change in gain with temperature of the Hall probe of a few percent. This was measured by cooling the Hall probe in the constant field of a superconducting magnet in persistent mode.

Between 65 and 75 kelvin there was a rapid change from the impurity series to normal magnetophonon oscillations, similar to that seen by Nicholas and Stradling (1975). At intermediate temperatures a single set of peaks appear at intermediate positions rather than both sets of peaks being resolved. This is expected if the peaks in the density of states are greatly broadened by scattering.

With smaller electric fields at low temperatures jagged, non-periodic structure was seen in the magnetoresistance instead of oscillations. The amplitude of this increased as the temperature was reduced. This may be associated with changes in the degree of impact ionisation of the donors, but it could not be demonstrated to be the same from run to run as would be expected if the experimental conditions were perfectly controlled.
$X\%$ means that 20mm represents $\Delta R/R = X\%$.

FIGURE 3.3.7
B257 devices, which had a similar carrier concentration, also showed this jagged structure at low fields and temperatures. At high fields the magnetoresistance curve became very smooth, showing neither oscillations nor any jagged structure. The reason for this difference is not known but it may be related to the higher degree of compensation in the MV73 devices.
3.4 The magnetophoton effect in narrow conducting channels.

Introduction

In a strong magnetic field the density of states function becomes sharply peaked at energies given by $\hbar\omega_C$, as described in section 1.3. As the magnetic field is increased these peaks move further apart and hence pass successively through the Fermi energy. In crossed electric and magnetic fields a free electron moves at right angles to both. A component of motion parallel to $E$ is only introduced if scattering is present. The only states into which scattering can occur are those within a few $kT$ of the Fermi energy. At low temperatures, where $kT$ is much less than $\hbar\omega_C$, the number of these varies significantly as a peak passes through the Fermi energy and the resulting change in current can be measured. This is known as the Shubnikov – de Haas effect. The condition that $\hbar\omega_C = E_F$ can be rearranged to show that the oscillations are periodic in $1/B$.

An electron in one of the peaks in the density of states has a circular orbit perpendicular to the magnetic field with a radius (the cyclotron radius) that increases with $n$, the number of quanta of angular momentum. Figure 3.4.1 shows a cross-section of the circular probability distribution of an electron orbiting in a magnetic field and a potential well narrow enough to affect the tails of the wavefunction. This will raise the energy of the electron, increasingly so as the well is made narrower. Hence peaks in the density of states with values of $r_C$ small compared with the well width $t$ will be little affected, but peaks with larger values will be shifted to higher energy.

A potential well like this occurs in the channel of a field effect transistor which is nearly pinched off, only a thickness of about $r_C$ remaining undepleted. Poole et al. (1982) were able to demonstrate this shift experimentally by showing that Shubnikov – de Haas oscillations in a fet were periodic in $1/B$ for large values of $B$, at which $r_C$ was smaller than $t$, but deviated from this at lower values of $B$ in a way that
corresponded to a calculation based on the above ideas.

The magnetic fields at which magnetophonon resonances occur also depend on the positions of the peaks in the density of states; the remainder of this chapter is an investigation of whether this shift can also be observed using the magnetophonon effect.

**Theory**

Magnetophonon resonances occur when the energy difference between two peaks in the density of states is equal to the optic phonon energy. When these peaks are sharp they may be considered to form discrete degenerate energy levels in which almost of the electrons reside. In the conventional magnetophonon effect these levels are equally spaced in energy, so that the resonant fields for transitions between many pairs of levels are the same. In narrow devices this is not generally true. However, transitions to and from the lowest level are the most important as other levels contain fewer electrons because of their higher energies. For this reason calculations have been performed only for transitions to the lowest level, though the fields for the other resonances could easily be calculated if required.

The sample geometry that will be considered is shown in figure 3.4.2. The magnetic field is in the z direction. $\mathbf{B} = \text{Curl } \mathbf{A}$, so a possible and convenient choice for the magnetic vector potential $\mathbf{A}$ is $(0, Bx, 0)$.

The appropriate Schrödinger equation is

$$\frac{1}{2m^*}(\mathbf{p} - e\mathbf{A})^2\psi + V_E(x)\psi = E\psi$$

where $\mathbf{p} = -i\hbar \nabla$. $V_E(x)$ is the potential from the depletion regions on each side of the channel. The depletion approximation is assumed here, so that $V_E(x)$ is zero in the undepleted region of thickness $t$, and

$$V_E(x) = \frac{e^2N_A}{2\varepsilon}(|x| - t/2)^2 \quad \text{for } |x| > t/2.$$

Berggren and Widlund (1985) have shown numerically that this is a good representation of the true potential at zero temperature.
With the above choice of \( A \), and trying a solution of the form
\[ \psi = \psi(x)e^{i(k_y y + k_z z)} \]
this equation for \( \psi(x) \) and \( E \) results:
\[
-\frac{h^2}{2m^*} \frac{d^2}{dx^2} \psi + \left[ V_F(x) + \frac{B^2 e^2}{2m^*} \left( x + \frac{h}{e B} k_y \right)^2 \right] \psi = E' \psi
\]
where \( E' \) is the total energy minus the kinetic energy of the free motion in the \( z \) direction. From this it can be seen that the effect of the magnetic field is to add to \( V_F(x) \) a quadratic potential centred at a point displaced from the centre of the channel by an amount dependent on the electron momentum in the \( y \) direction. This is illustrated in figure 3.4.3. The position of the electron in the channel now depends on \( k_y \). This equation can be solved exactly in the two limits \( t=0 \) and \( t \) large:

If \( t=0 \) (Maa, 1984) then \( V_F(x) \) is parabolic. When added to the magnetic potential the total, \( V(x) \), will also be parabolic. The electrons oscillate in \( V(x) \) at a frequency \( \omega \), where
\[
\omega^2 = \omega_C^2 + \omega_o^2 \quad \text{and} \quad \omega_o^2 = \frac{e^2 N_a}{m^* \varepsilon} .
\]
\( \omega_o \) is the electron's frequency of oscillation in \( V_F(x) \) alone.

The minimum value of \( V(x) \) occurs at
\[
x_{\text{min}} = \frac{\hbar k_y \omega_C}{m^* \omega^2} ,
\]
and the minimum energy is
\[
V_{\text{min}} = \frac{\hbar^2 k_y^2 \omega_o^2}{2 m^* \omega^2} .
\]
Hence the dispersion relation is
\[
E = E_n + \frac{\hbar^2 k_y^2}{2m^*} \times \frac{\omega_o^2}{\omega^2} + \frac{\hbar^2 k_x^2}{2m^*} .
\]
The constant energy contour corresponding to this (see section 1.2) is an ellipse enclosing a number of states
\[
N(E) = \pi \frac{2m^*}{\hbar^2} \left( E - E_n \right) \cdot \frac{\omega}{\omega_o} \cdot \frac{2}{\pi^2}
\]

36
By differentiation the density of states is

\[ g(E) = \frac{4 \, m^*}{\pi \, \hbar^2} \cdot \frac{\omega}{\omega_0} \]

per subband for energies above \( E_n \), which is shown in figure 3.4.4 (a). As this has no peaks there will be no magnetophophon oscillations.

If \( t \) is as large as the active layer depth \( d \) (Landau and Lifshitz, 1958) then \( V_E(x) \) is zero everywhere and the energy will be independent of \( k_y \).

Values of \( k_y \) outside the range \( \pm eBd/2\hbar \) are not allowed as the electron would then be outside the active layer.

The substitution

\[ x' = x + \frac{\hbar k_y}{eB} \cdot dx' = dx \]

produces the Schrödinger equation for a one dimensional harmonic oscillator, giving subband energies

\[ E_n = (n + \frac{1}{2}) \hbar \omega \]

and a dispersion relation

\[ E = E_n + \frac{\hbar^2 k_y^2}{2m^*}. \]

The constant energy contour is now a rectangle with length \( eBd/\hbar \) and width \((8m^*(E-E_n))^0.5\), from which the density of states is

\[ g(E) = \frac{2eBd}{\pi^2 \hbar^2} \cdot \left( \frac{2m^*}{E-E_n} \right)^{0.5}. \]

This is shown in figure 3.4.4 (d) and will produce strong magnetophonon oscillations.

Calculation for intermediate thicknesses requires an approximation. The WKB approximation (see e.g. Park, 1964) is attractive here because it gives answers that are exactly correct for the two limits of very thick and very thin channels. It leads to this integral equation for the eigenenergies of a one dimensional potential:

\[ \frac{(2m^*)^{0.5}}{\hbar} \int_{x_1}^{x_2} (E_n-V(x))^{0.5} \, dx = (n+\frac{1}{2}) \pi. \]

\( x_1 \) and \( x_2 \) are the classical turning points of a particle in the potential.
In this case $V(x)$ is $V_E(x)$ plus the magnetic potential. For this $V(x)$ an analytic expression can be obtained for the integral, which with $k_y^2=0$ reduces to this implicit equation for $E$: (Poole, 1982)

$$2m^* \omega_C \left[ \frac{t}{4} \left( \frac{2E}{m^* \omega_C} - \frac{t^2}{4} \right)^{0.5} + \frac{E}{m^* \omega_C} \sin^{-1} \left( \frac{t}{2} \left( \frac{m^*}{2E} \right)^{0.5} \right) \right] + (2m^* A)^{0.5} \left[ a^2 \left( \frac{\pi}{2} - \sin^{-1} \left( \frac{t}{2} - B/2A \right) \right) \right] - (t/2 - B/2A) \right)^{0.5} = (n+\frac{1}{2}) \hbar$$

$$n = 0, 1, 2, \ldots$$

where $A = \frac{1}{2} m^* \omega_C \left[ 1 + \frac{e^2 N_A}{\epsilon m^* \omega_C} \right]$

$$B = \frac{e^2 N_A t}{2 \epsilon}$$

$$a^2 = \frac{1}{A} \left[ E - \frac{eN_A t}{8 \epsilon} \right] + \frac{B^2}{4A^2}$$

Since the peaks in the density of states always occur at $k_y^2=0$, this expression is adequate to predict the magnetic fields at which magnetophonon resonance will occur. A calculation of the density of states requires the energies of states with $k_y^2=0$. Berggren and Newson (1986) have extended the WKB calculation to find the dispersion relation. The resulting implicit equation for $E$ has several different forms, with ranges of validity depending on whether the approximate wavefunction, which is finite in extent, lies wholly in the rising part of $V_E(x)$ or across one or both of the boundaries of the region in which it is zero. The exact density of states requires numerical integration and has not been performed, but the major features can be determined without this. For $t \gg 2r_C$ the density of states has infinitely high peaks, rising less fast as $t$ decreases. This is because for small values of $k_y$ the wavefunction remains in the region with $V_E(x) = 0$, so these states contribute a smaller version of the density of states when $V_E(x)$ is zero throughout the device. For $2r_C \geq t \geq 0$ the peaks now have a finite height, again becoming less sharp as $t$ decreases. This is shown in figure 3.4.4 (b) and (c). The approximation also predicts that the energy of an electron with $k_y=0$ is not changed until $t \leq 2r_C$, so that the onset of the shift in the magnetophonon resonances is coincident with the
rapid reduction of the peaks in the density of states on which the effect depends.

The solution divides into discrete regions in this approximation because the approximate wave function is finite in extent, being zero outside the classical turning points. The correct wavefunction has exponentially decaying tails out to an indefinite distance, so there will be smooth transitions from one type of behaviour to the next. The predictions should still be qualitatively correct though, suggesting that the optimum condition to observe this effect will be with \( t \) similar to a typical value of \( r_0 \), say 30 nanometres. Larger values of \( t \) will produce no shift and smaller ones will make the oscillations too small to observe.

The shift in field expected for thicknesses in this range can be calculated from the previous expression for electron energies with \( k_y = 0 \), by choosing a value of \( B \), solving iteratively for the energies, and repeating this for suitably altered values of \( B \) until the difference in energy between a chosen pair of subbands is equal to the optic phonon energy. This \( B \) value is then compared with the value obtained when \( t \) is large to obtain the shift. These \( B \) values for large \( t \) are not exactly those found experimentally - for various reasons the observed peaks are not perfectly periodic in \( 1/B \), especially for small \( n \) - but most of this small error should disappear when the two \( B \) values are subtracted. The results of such a calculation, for a doping level of \( 1.8 \times 10^{16} \text{cm}^{-3} \), are shown in figure 3.4.5. The lower graph shows the number of mm. by which the resonance is expected to shift to lower fields on the scale of 1 tesla = 50mm used for the experimental recordings.

Only resonances for \( n \leq 8 \) are large enough to be seen even under ideal circumstances. The minimum shift that is observable depends greatly on the signal to noise ratio, never being less than several mm. and often much more. It is clear that the shifts with this doping level become quite large
enough to observe when $t$ is made small, provided that the oscillations are still detectable.

So far the depletion approximation has been assumed, i.e., the thermal blurring of the edges of the depletion regions has been ignored. It seems reasonable to use the depletion approximation in these calculations when the Debye length is less than the cyclotron radius, as for thicknesses much less than the cyclotron radius the magnetophonon oscillations will be quenched anyway. This corresponds to doping levels greater than about $10^{16}\text{cm}^{-3}$. Figure 3.4.8 summarises the expected results for various thicknesses for devices with doping levels higher than this. Measurements have also been made on devices with doping levels much lower than this, so the limit of $L_D \gg r_C$ must also be considered.

For $t \gg L_D$, and hence $t \gg r_C$ also, the blurring is unimportant and the oscillations will be unaffected. For values of $t$ less than $L_D$ the electron distribution becomes little narrower and instead the electron concentration is reduced fairly uniformly over the whole channel. The limit of this is the subthreshold region in which the electron concentration varies as $\exp (-eV_g/kt)$ throughout and is largest in the centre of the channel, falling off as $\exp (-x^2/2L_D^2)$ to each side. The width of the electron distribution does not fall below the subthreshold value of about 2.5 $L_D$ anywhere in the gradual transition between the linear region and the exponential subthreshold region. Some representative electron concentration profiles are shown in section 3.2. The variation of $N_e$, the electron concentration in the centre of the channel, with $t$ is shown in figure 3.4.6.

This distribution of electrons affects $V_E(x)$, but only slightly since $N_e$ is now much less than $N_d$. Approximately, for distances less than $L_d$ from the centre of the channel the electrons cause $V_E(x)$ to increase as $x^2e^2(N_d-N_e)/2\varepsilon$ rather than as $x^2e^2N_d/2\varepsilon$ expected from the ionized donors alone. $\omega_0$ is reduced by $(1-N_0/N_d)^{-0.5}$, so when $\omega_C \gg \omega_0$ the density of states is increased by the same factor. This enhancement persists until
\( x \geq L_D \), when \( E-E_n = kT \left(1-N_0/N_d\right) \). The resulting approximate densities of states at various values of \( t/L_D \) are illustrated in figure 3.4.7. These suggest that magnetophonon oscillations should disappear rapidly as \( t \) is reduced below \( L_D \), and hence will not be observable when \( t \) has been reduced enough to shift what remains of the peaks in the density of states. Figure 3.4.9 contrasts this behaviour with that expected for \( N_d \geq 10^{16} \text{cm}^{-3} \).

Scattering blurs features in the density of states over an energy range \( \Delta E \) of about \( \hbar/\tau \), where \( \tau \) is the momentum relaxation time. In these layers with low doping levels this is less than 1 meV. The ratio of the peak separation to \( \Delta E \) is \( \omega_c\tau \), which is greater than 10 for these devices, so the densities of states predicted above and the conclusions drawn from them are not greatly affected by scattering.
3.5 Experimental results for thin layers

To compare the experimental results with the theory of the previous section a means of finding the channel thickness is required. A technique based on $V_g$ will be very inaccurate because both $N_d$ and $d$ are known only roughly and because of the unknown and variable occupancy of traps under the gate. The thickness must therefore be estimated from the experimental $I$ versus $V_g$ curve in some way. The method used was to estimate the thickness with $V_g=0$ ($t_o$) using a C-V profile of the layer taken before processing and allowing for the depletion depth from contact potentials. The thickness is related to the conductance by $t/t_o = \mu_o/\mu \times \sigma/\sigma_o$, where $\mu$ is an average drift velocity. It was assumed that $\mu_o/\mu$ remained close to one. The C-V profile also yielded a value of $N_d$ from which to calculate the Debye length.

Figure 3.5.1 shows the behaviour of the magnetophonon oscillations in a B977 device as $V_g$ was varied. The results were typical of those obtained with low doped layers. The numbers underneath each curve are the values of $t/L_D$ derived as above, and those at the right are the vertical scale in units of $10^{-3}$. The drain voltage was kept constant at only 5mV, so the noise level is quite high at about $10^{-3}$. It can be seen that the amplitude does fall as $t/L_D$ decreases, but has halved only when this parameter has decreased to 0.46, rather less than expected. Oscillations are still present for very much smaller values of $t/L_D$, which is quite contrary to predictions unless $\mu$ is falling so rapidly that the thickness is much greater than estimated. Even this is open to the objection that a large decrease in $\mu$ should itself produce a large decrease in oscillation amplitude.

To investigate possible changes in $\mu$ the Hall voltage appearing between opposite sidearms of a long device was measured as a function of gate voltage. With a constant voltage between source and drain this is proportional to $\mu$. Figure 3.5.2 shows both the device conductance and the Hall voltage as a function of gate voltage. At first $\mu$ decreases very
\[ V_b - V_c \text{ (mV)} \]

-30

-2

-1

\[ V_g \]

\[ B = 0.5 \text{ tesla} \]

\[ B = -0.5 \text{ tesla} \]

\[ \text{FIGURE 3.5.2} \]
gradually, but after $V_g = -2$, where the conductance has been reduced by a factor of about seven, the curves for $B=\pm 0.5$ tesla are no longer mirror images of each other. The conductance curve has only a barely perceptible change of gradient at this point. Figure 3.5.3 shows the variation with $V_g$ of individual sidearm voltages with respect to the source. There is now no magnetic field. $V_D$ rises rapidly towards the drain voltage below $V_g = -2$ but $V_C$ does not, suggesting that the bottom right hand corner of the illustrated device has ceased to conduct, and also that there is now little conduction between points $b$ and $c$. Non-uniform conduction explains the persistence of oscillations when the total conductivity is very small; the parts of the device which are still conducting may still be several Debye lengths thick. It also means that there is no point in using more sophisticated models to determine the channel depth.

Variations in either $N_g$ or $d$ could be responsible for the differing pinch-off voltages. C-V profiles from different parts of the layer indicate that sufficiently large variations occur across the layer, but the expected variation within the dimensions of a device is much less. It seems more likely that conduction along the edges of the mesa is being observed. The mesas are formed by wet etching and so the edges are shaped as in figure 3.5.4. The bottom edge of the mesa can be seen clearly under an optical microscope as a line 10 microns or so outside the edge of the gate metallization. The edges will require a larger gate voltage to deplete them than the rest of the device because they are further from the gate metallization. The width of these regions is of the order of the thickness of the active layer, and so the proportion of current that flows in them when $V_g = 0$ will be roughly $d/w$, the layer thickness divided by the device width. This is about 4% in these devices, so if the edges are depleted much more slowly than the rest this effect is a suitable size. The observed asymmetry requires the gate metallization to be imperfectly aligned with the mesa, but this is expected on the scale considered here. Edge
Figure 3.5.3

Figure 3.5.4
conduction also explains the apparent decoupling of the voltages at b and c.

Higher doping levels

Section 3.4 predicts that magnetophonon oscillations will shift before they disappear only in devices with \( N_d > 10^{16}\text{cm}^{-3} \) or so. The amplitude of magnetophonon oscillations decreases sufficiently rapidly as the concentration of ionized impurities increases to cause experimental difficulties in this range. Figure 3.5.5 is a plot of oscillation amplitude against \( N_d - N_a \) for all the devices studied. The results for devices grown by hydride VPE at GEC are shown by * and lie close to the smooth curve. The results of all other devices studied lie below this curve, probably because these devices were more highly compensated. This, and a comparison of mobilities with published values, suggests that significantly better material is not available.

The noise of the instrumentation used was reduced to about \( 5 \times 10^{-5} \) by the methods described in chapter 2. The highest doped devices which showed oscillations usefully larger than this (\( 5 \times 10^{-4} \) maximum) were from B1220. Only the \( n=3 \) and \( n=4 \) peaks were large enough to be detectable. \( N_a - N_d \) for B1220 was about \( 2 \times 10^{16}\text{cm}^{-3} \), only just meeting the above approximate criterion.

The problem of edge conduction is expected rapidly to become less severe as the thickness of the active layer is reduced because both the height and width of the edge region are roughly equal to this. The trends in the results of sidearm voltage versus conductance experiments suggest that edge conduction should have been adequately reduced in devices with active layers of thickness 1.5 microns, that of layer B1220. A different etch which did not attack the metallization was used to form the mesas of the devices made from this layer, allowing the top of the mesa to be etched right back to the edge of the gate. This should further reduce the edge conduction. Figure 3.5.6 shows the sidearm voltage and conductance of a
Magnetoophonon amplitude versus doping level.

FIGURE 3.5.5
FIGURE 3.5.6

$V_D = 4.7\, \text{mV}$
Bl220 device. The behaviour is now quite different; the sidearm voltages are almost constant then plunge abruptly to near zero. This is probably because the sidearms are now becoming depleted before the main part of the device. The extra etching of the mesa will cause the resist over the sidearms to become undercut, so that the sidearms are less thick than the main part of the device and hence more easily depleted.

In all cases the noise level in these devices increased with gate voltage. The rate of increase correlated with the gate leakage current, which varied from around 0.1 nanoamp upwards. (To keep the drain voltage sufficiently small when the device was near pinch-off, the measuring current was only about 100 nanoamps. Hence the magnetophonon oscillations had an amplitude of about 0.1 nanoamps). In the best device it was possible to decrease the conductance by a factor of 15 before the signal to noise ratio became hopeless. This corresponds very roughly to a channel thickness of 100 nanometres. Within the large uncertainty set by a signal to noise ratio of about 2, no shift or reduction of amplitude had occurred.

Conclusion
The magnetophonon effect has been found to be much less satisfactory than the Shubnikov - de Haas effect for observing the energy levels produced by electric and magnetic fields in combination. This is because the magnetophonon effect is inherently much weaker, so lower doping levels are required to give adequate peaks in the density of states. This cannot be taken as far as is necessary for the same signal to noise ratio because the magnetophonon effect requires reasonably high temperatures at which the Debye length for low doping levels becomes too great for any shift to be observed. High temperatures also lead to problems with gate leakage. Lower doping levels also require correspondingly thicker active layers, leading to possible problems in fabricating devices in which conduction is still uniform when the channel has been narrowed to about 1% of its initial depth.
Peak shifts in short devices

In some cases, particularly in short devices, the magnetophonon peaks did shift on application of a gate voltage, though not in the manner expected from Landau levels being raised in energy by a potential well. This effect was seen in devices from several layers, but most thoroughly investigated in a short 1988 device with B parallel to the layer and perpendicular to the current. The aspect ratio under these conditions is sufficient for the oscillations not to be inverted. The doping level of layer 1988 is $6 \times 10^{15}$ cm$^{-3}$, so the oscillation amplitude was about $10^{-3}$. The signal to noise ratio was thus about 15, and in this device it deteriorated only by a factor of three or so near pinch off. This ratio was sufficient for a peak position to be repeatable to within $\pm 2\text{mm}$ (.04 tesla) when successive measurements were made at the same gate voltage. When a gate voltage sufficient to reduce the conductance by a factor of ten was applied ("at $\sigma_0/10$") the n = 3, 4 and 5 peaks moved to higher fields by 6 to 7 mm. This shift was apparent, though only about half as large, at $\sigma_0/3$ and persisted unchanged to about $\sigma_0/200$, after which it decreased again. The current distribution would be highly non-uniform by this stage. All these results were repeatable within a few mm., and the repeats for a given gate voltage were interspersed by repeats for others, including $V_g=0$, in case some systematic error was being observed. The adjacent long device (figure 3.1.1) showed peaks with no perceptible shift with gate voltage. These occurred in positions between the extreme positions of the short device peaks, though rather nearer to the $\sigma_0$ positions.

This effect may be due to ambiguities in separating the oscillations from the monotonic magnetoresistance. As the shape of the magnetoresistance curve changes with gate voltage smaller or larger amounts of the magnetoresistance will pass through the digital filters and shift the apparent peak positions slightly. Short devices have much greater monotonic magnetoresistance because of their smaller aspect ratios.
THE TRIANGULAR BARRIER SWITCH

4.1 Structure and theory of operation

The triangular barrier switch was first suggested by Board et al. (1981). It consists of an n-type planar doped barrier (Malik et al., 1980) extended by forming a p-n junction with one of the n-type regions so that holes can be injected into the barrier. The structure is shown in figure 4.1.1. The planar doped barrier consists of a nominally undoped region sandwiched between two n-type layers. A very thin (typically 20nm) plane of acceptors is placed in the undoped region to provide an inbuilt potential barrier. The areal doping density in the thin acceptor plane is sufficiently low for almost all of the acceptors to capture electrons. Most of these electrons come from the outer n-type regions, so the electric field between these regions and the acceptor plane is fairly constant, producing a triangular potential barrier. This is modified by the potential across the depletion regions in the n-type layers, the finite width of the acceptor distribution and any charge in the nominally undoped region (figure 4.1.2). The I-V characteristics can be made quite asymmetrical by placing the acceptor plane unsymmetrically within the undoped region. Kazarinov and Luryi (1981) have made calculations of these I-V characteristics, including the effects of the depletion regions and the space charge of the electrons crossing the barrier. Generally the undoped region is quite wide, so that current flow is predominantly by thermionic emission except at very low temperatures. Mechanisms with lower activation energies, such as thermally assisted tunnelling, then become important. Gossard et al. (1982) found a reduction in activation energy at low temperature which they attribute to this change of conduction mechanism.

The height of the potential barrier can be reduced by injecting holes into the undoped region. These holes will be attracted towards the acceptor plane and partially compensate the negative charge there. In the triangular barrier switch this is accomplished by forward biasing the p-n junction and
FIGURE 4.1.1

FIGURE 4.1.2
making the n-type region between the p⁺ layer and the undoped region sufficiently thin for some of the holes injected from the p⁺ layer to cross it without recombining, as in the base of a transistor.

The ratio of electron current to hole current for a given total is determined by the properties of the p-n junction. As the total current is increased the device voltage at first increases to thermonically emit the increasing electron current over the barrier. The concentration of holes held under the barrier increases with the number injected and hence with the total current. At some point the barrier lowering by these holes becomes more than sufficient to accommodate the increased electron current without extra forward bias, and the device voltage will now decrease with increasing current. At high currents the main impediment to current flow becomes the p-n junction, the conductivity of the undoped regions being greatly increased by carrier injection. The device thus shows three regions of behaviour: a high impedance (off) state, a negative resistance region, and a low impedance (on) state. (A detailed analysis is presented later.)

In this, and in the fact that it is current controlled, the device resembles a thyristor. Other similarities are that another contact can be made to the n region of the device and used as a gate (a small current on this contact will switch a much larger current on the cathode) and that the device can be switched by optically generated electron-hole pairs.

4.2 Device fabrication

The material for these devices was grown at the GEC Hirst Laboratory by molecular beam epitaxy, using a Vacuum Generators V80H machine. The layers were grown at 640°C and 25nm per minute using silicon and beryllium as the n and p type dopants. Circular, stepped mesas were produced by wet etching. AuGe/Ni/Au was used as the n-type metallisation and Ti/Au for the p-type. Gold wires were bonded directly to the metallisation from TO18 or ceramic headers.
**FIGURE 4.2.1**

**SEM-I-INSULATING SUBSTRATE**

<table>
<thead>
<tr>
<th>Layer</th>
<th>Thickness</th>
<th>MBE 111</th>
<th>MBE 219</th>
<th>MBE 628</th>
</tr>
</thead>
<tbody>
<tr>
<td>p+</td>
<td>300nm</td>
<td>$5 \times 10^{18} \text{cm}^{-3}$</td>
<td>400nm</td>
<td>$2 \times 10^{18} \text{cm}^{-3}$</td>
</tr>
<tr>
<td>N2</td>
<td>300nm</td>
<td>$5 \times 10^{17} \text{cm}^{-3}$</td>
<td>50nm</td>
<td>$1 \times 10^{16} \text{cm}^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>350nm</td>
<td>$2 \times 10^{18} \text{cm}^{-3}$</td>
</tr>
<tr>
<td>U2</td>
<td>350nm</td>
<td>undoped</td>
<td>200nm</td>
<td>undoped</td>
</tr>
<tr>
<td>p-p1</td>
<td>20nm</td>
<td>$1 \times 10^{18} \text{cm}^{-3}$</td>
<td>20nm</td>
<td>$5 \times 10^{17} \text{cm}^{-3}$</td>
</tr>
<tr>
<td>U1</td>
<td>50nm</td>
<td>undoped</td>
<td>200nm</td>
<td>undoped</td>
</tr>
<tr>
<td>N1</td>
<td>300nm</td>
<td>$5 \times 10^{17} \text{cm}^{-3}$</td>
<td>400nm</td>
<td>$2 \times 10^{18} \text{cm}^{-3}$</td>
</tr>
<tr>
<td></td>
<td>50nm</td>
<td>$5 \times 10^{18} \text{cm}^{-3}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A cross-section of the three types of device used is shown in figure 4.2.1 and figure 4.2.2 is a diagram of the metallisation. The author is grateful to Mr. Paul Rees of University College, Swansea for supplying these devices.

4.3 Experimental results

The I-V characteristics of these devices were displayed on a curve tracer, the essentials of which are shown in figure 4.3.1. The theory of operation of the TBS given in section 4.1 suggests that these devices will have an I-V characteristic similar to figure 4.3.2. The following results would be expected from this combination:

If the resistance of the external circuit (R in figure 4.3.1) is small, then the device will switch discontinuously between high and low resistance states as shown by the dashed lines in figure 4.3.3. These "load lines" represent the equation \( V_T = IR + V \), and hence have a slope of \(-1/R\) and cross the voltage axis at the instantaneous value of \( V_T \). The possible states of the circuit for a particular value of \( V_T \) are given by the intersection of the appropriate load line with the device I-V characteristic.

As \( R \) is increased, the load lines become more horizontal and the difference between the currents at which the device enters and leaves the low impedance state (\( I_2 \) and \( I_3 \)) decreases. It is assumed that \( V_T \) is increased as necessary to keep the maximum current large enough to take the device through the switching points. Eventually the load lines become flatter than the flattest part of the device characteristic and no discontinuous switching will occur. While the load line intersects the device characteristic only in the negative resistance region, as in figure 4.3.4, there will probably be oscillations. The origin of these may be seen from the circuit of figure 4.3.5, in which the capacitor represents the device capacitance and that of the external circuit. The voltage across the device cannot now change infinitely fast and so any jumps between points on the
I–V characteristic must be vertical. $I_0$ must be between $I_g$ and $I_h$ for oscillations to occur. If the capacitor is initially uncharged the operating point will gradually move along the characteristic towards point A, where the device jumps to a high current state. The device current is now greater than $I_0$, so the capacitor voltage will fall, bringing the operating point down the characteristic towards point B, where the device turns off. The capacitor now starts to charge again and the oscillation repeats. The principal factor determining the frequency is the slow charging of the capacitor by the current source between points C and A. The capacitor is then discharged much more rapidly by the TBS.

The behaviour actually observed was in accordance with this only for small values of $R$. Figure 4.3.7 shows the large $R$ characteristic for a device from the layer MBEIII. The behaviour of all three types of TBS was qualitatively similar in most respects, so MBEIII data will be presented as typical and differences discussed later.

Oscillations are present but have a much smaller amplitude than predicted and die away before the end of the negative resistance region. This is attributed to the required frequency of operation exceeding the upper frequency limit of the TBS. Using an oscilloscope and current source, the waveform was checked to have the expected amplitude and sawtooth form at low frequencies. The rising slope was noticeably curved because of the finite impedance of the current source and the increasing current drawn by the TBS as point D was approached. The frequency could be increased over a large range by increasing the current, but the amplitude decreased dramatically above 500 kilohertz and the device would not oscillate above 1 megahertz. In particular the voltage at which the device switched to the high current state was lessened at high frequencies, suggesting that it is the turn-off time that limits the frequency response. Rees (1985) has shown that, when driven from a low impedance source, at turn-on the current through these devices increases approximately exponentially with a time
constant of about 100 picoseconds. The limiting factor is the time required for holes to diffuse through the N₂ region between the p-n junction and the barrier. The p-plane charge in these devices is about 3×10⁻¹¹ coulombs. Allowing for some recombination losses, a current pulse containing about 10⁻¹⁰ coulombs should be enough to change the device to its low impedance state. Circuit capacitances greater than about 10 picofarads will supply this amount of charge as the device voltage falls, so in this and most other circuits the turn-on time is determined by the above time constant and is less than a nanosecond. An external current of about 1mA is required to replace the holes continuously being lost from the barrier in the low impedance state (I₉). If the loss rate remained constant during turn-off the hole charge would disappear in about 10⁻⁷ seconds, which is comparable to the times seen in the above experiment.

The frequency at a given current could be decreased by adding an external capacitor. The oscillations disappeared at almost the same frequency and hence at a higher current. It was possible to produce oscillations over the full length of the negative resistance region with a suitable capacitor. Using too large a capacitor caused destruction of the device because most of the energy stored in the capacitor is dissipated in the TBS in a fraction of a microsecond as it switches.

Despite the large value of R the discontinuous jumps and hysteresis are still present. I₁ is now larger than I₃, which cannot happen with a characteristic like figure 4.3.1. The observations are consistent with a characteristic that curves back on itself, like figure 4.3.8. The section between T and H has positive differential resistance, so these states would not be obscured by oscillations, but they are not reached in normal operation. Active circuitry could in theory provide a load line that intersected the device characteristic only in this region, but the model described in the next section suggests that these states are only unstable equilibria, and so will never be observed.
The I-V characteristic of these devices varies greatly with temperature. This is illustrated in figure 4.3.9. The hysteresis loop at first increases in area as the device is cooled, and then disappears completely. At low temperatures the device voltage is much above the voltage of a p-n junction even at high currents, showing that the barrier retains some influence. These features were common to all devices examined and reproducible from run to run. The behaviour of devices from the same layer was always much closer than that of devices from different layers.

The device behaviour around the temperature at which the hysteresis disappears is always very complex, and the I-V characteristic can show a sequence of features which are quite repeatable but exist over a temperature range of less than 1 kelvin. Devices from different layers behave differently at the transition and minor details are unique to each device.

In all cases both $I_T$ and $I_H$ increased with decreasing temperature. For MBE111 and MBE628 devices (figure 4.3.10) $I_H$ increased more rapidly and hysteresis disappeared when $I_H$ became equal to $I_T$. After this point the characteristics showed a small negative value of resistance at moderate currents, changing gradually to a small positive value at large currents.

Representative MBE219 characteristics are shown in figure 4.3.11. These devices showed little increase in $I_H$ and observation of the hysteresis was limited by $I_T$ becoming greater than the maximum allowable device current.

It is believed that the hysteresis does disappear because the characteristic at 41 kelvin shows a transition to a low impedance state that is continuous though abrupt. Below 30 kelvin hysteresis reappears, apparently very similar to that at high temperatures.
FIGURE 4.3.11
MBE 219
Figure 4.3.12 shows the effect of a large magnetic field on an MBELL1 device at two temperatures, one "high" and one "low". The effect depends on whether \( B \) is parallel or perpendicular to the current. The zero field characteristics are shown for comparison. At high temperatures a parallel magnetic field reduced \( I_H \) dramatically and \( I_T \) slightly, whereas a perpendicular field reduced \( I_H \) by less and significantly increased \( I_T \). These effects will be discussed in terms of the model described in the next section.

Switching voltage \( (V_S) \) followed a similar pattern for all devices, showing a broad maximum near the temperature at which hysteresis disappeared. \( I_S \) most often increased slowly below room temperature and then fell very rapidly near the temperature at which hysteresis disappeared, but there were wide variations in its behaviour and size, probably for the reason given below.

Experiments were also performed on the variation of off-state current with voltage and temperature. The potential barrier is very high in these devices at zero bias, so appreciable current flow is expected only for large voltages and high temperatures. A calculation almost identical to that of section 5.2 for a graded gap diode suggests that current flow will be by thermionic emission above a temperature

\[
T_{\text{therm}} = \frac{\hbar}{K d} \left[ \frac{e}{m^*} \left( V_0 - rV \right) \right]^{0.5},
\]

where \( r = d_1/d_1 + d_2, d = d_1 + d_2(V_0 - rV)/(V_0 - rV + V) \) and \( V_0 \) is the zero bias barrier height, about 1.3 volts. This reduces to the expression of section 5.2 when \( d_2 = 0 \).

\( T_{\text{therm}} \) is less than 100 kelvin for all device voltages \( V \) from zero to \( V_0/r \), often much less. The current from thermionic emission can be estimated from the simple formula

\[
J_{\text{therm}} = \frac{4\pi e m^* K^2}{\hbar^3} T^2 \exp[(E_B - E_F)/K T]
\]

for forward voltages greater than \( K T/e \). \( E_F \approx K T \ln(N_d/N_C) \), which is close
FIGURE 4.3.12
MBE 111
to zero in these devices. Below the switching point $E_b$ is approximately $e(V_0-rV)$. Taking the device area as $8 \times 10^{-9} \text{m}^2$ the current can be calculated to be about $10^{-14}$ amps at 300 kelvin and 3 volts.

A typical value for the current actually observed under these conditions was $1 \mu\text{A}$, and the voltage and temperature dependences were not as predicted. In addition there was wide variation between nominally identical devices, suggesting that the off-state current is dominated by leakage, perhaps over the unprotected surface of the mesa. It proved impossible to obtain behaviour approximating to the thermionic emission predictions even with large voltages and temperatures up to 80°C.

An interesting off-state I-V characteristic for an MBE628 device is shown in figure 4.3.13. Different results were obtained over part of this curve each time it was measured. Figure 4.3.14 is the temperature variation of the leakage current of another MBE628 device which gave repeatable results. Plotting $\log I$ against $T$ to the power of $1/3$ or $1/4$ gives quite a good straight line over a wide range. The reasons for these observations are unknown.

Summary

The I-V characteristics of these devices have been found to be unexpectedly complex near the switching point. The switching behaviour is strongly temperature dependent. A variety of behaviour is seen at low temperature though all the devices behaved similarly at room temperature. The off-state characteristics are very variable and seem to be dominated by leakage.
4.4 Experimental details

The switching, threshold and holding voltages and currents were measured on a Tektronix type 575 curve tracer. This feeds a full-wave rectified 50 hertz sine wave with an amplitude up to 200 volts to the device through a variable resistor, and displays a graph of device current against voltage on a cathode ray tube. The device will switch slightly before the threshold and holding points on the characteristic if the impedance of the measuring circuit is less than infinite. These devices are limited by the bond wires to a maximum current of about 10mA, so the circuit impedance was always at least 20kΩ. The effect of reducing this to 10kΩ was not discernible. The sensitivity and accuracy were not very great; changes greater than 1% or so could be detected and absolute values measured to within about 5%.

The brightness of the low impedance part of the characteristic increases above the point to which the TBS jumps from the transition point because this section is traced with current both increasing and decreasing. From this it can be verified that the jump takes place along the load line as expected. It is generally harder to see what happens after the holding point. Increasing the brightness of the display shows that the spot initially moves along the load line but then either disappears or moves vertically downwards and joins the high impedance line. Sometimes oscillations occur with a different envelope from those with current increasing, giving the hump in the oscillation envelopes of figure 4.3.2 for instance.

To check that the hysteresis was a steady state phenomenon, an oscilloscope, resistance box and function generator were assembled to mimic the curve tracer, but now with variable frequency. The voltage available was much less than the 200 volts of the curve tracer but this did not matter as only changes in the characteristic were being sought. No reduction in hysteresis was detected even for periods as long as 30 seconds. For periods shorter than about 1 millisecond the whole
characteristic distorted greatly, no part being the same for increasing and decreasing current. This was attributed to device and circuit capacitance. In addition the device was left for many minutes at a point on the characteristic just before the threshold and holding points, but it did not switch.

Most measurements were made in an Oxford Instruments temperature controlled continuous flow helium cryostat. The sample temperature was sensed by a four wire rhodium-iron resistance thermometer which was sensitive over a very wide temperature range. High magnetic field measurements were made with the same equipment as for magnetophonon measurements (sections 2.1 to 2.3).

For the I-V and I-T characteristics of the off-state of the device, voltages were supplied by a switch-settable source (Uren, 1981) and current measured with a Keithley 480 picoammeter.
4.5 A mathematical model

Several approximate models of the TBS have been published, e.g. Board et al. (1981) and Najjar et al. (1982). This section is based on the model of Habib and Board (1983), which is the most recent and comprehensive of those known to the author.

The notation used in this section is defined in figure 4.5.1 and below:

- $\beta$ reciprocal thermal voltage, $e/KT$
- $A^*$ effective Richardson constant for electrons
- $J_n$ electron current per unit area
- $J_p$ hole current per unit area
- $J_{rec}$ current per unit area from recombination in the p-n junction depletion region
- $N_a$ concentration of acceptors in the p region
- $N_d$ concentration of donors in the n1 and n2 regions
- $N_i$ intrinsic carrier concentration
- $N_C$ effective density of states in the conduction band
- $N_t$ number of acceptors per unit area in the p-plane
- $P_o$ equilibrium hole concentration in the n regions, $N_i^2/N_d$
- $D_p$ hole diffusion coefficient in the n regions
- $\tau$ recombination time
- $L_p$ hole diffusion length, $\sqrt{\tau D_p}$

At the start of the calculation a value is selected for $V_j$, the external voltage appearing across the junction. This determines the device current. There are two unknowns: $V_1$ and $V_2$, the heights of each side of the barrier. The continuity of both the electron and hole currents in the device provide the two conditions necessary to determine these. The assumptions of this model and the sequence of the calculations performed to derive J-V characteristics from them will now be presented in detail. The flow of the calculation is shown diagramatically in figure 4.5.2.
Figure 4.5.1
A value is selected for $V_j$. The width $W$ of the junction depletion region is then given by

$$W^2 = 2\varepsilon (V_{bi} - V_j)/(eN_d),$$

(1)

where

$$N_d = N_aN_d/(N_a + N_d) = N_d$$

(2)

and $V_{bi}$, the "built-in" junction voltage, is given by

$$V_{bi} = \ln(N_aN_d/N_i^2)/\beta.$$  

(3)

The expression for the recombination current is

$$J_{rec} = \frac{eN_i W}{2\tau} [\exp(\beta V_j/2) - 1]$$

(4)

This is an electron current on the barrier side of the p-n junction. The electron diffusion current and the electron current from holes that recombine after the depletion region are both much smaller than $J_{rec}$, so $J_n = J_{rec}$. $J_n$ flows over the barrier by thermionic emission:

$$J_n = A^* T^2 N_a N_d [\exp(-\beta V_1) - \exp(-\beta V_2)]$$

(5)

The problem is to find the two unknowns $V_1$ and $V_2$. At this point a value for $V_2$ is guessed and $V_1$ calculated from the above equation.

It is assumed that the holes in the region of the barrier peak act only to reduce the area concentration of charge from the acceptors in the p-plane. The electric fields $E_1$ and $E_2$ in the undoped regions are then constant. The holes have a maximum concentration $P_m$ at the peak of the barrier and are assumed to be in quasi-equilibrium with this throughout the undoped region, so that the concentration falls exponentially with distance from the barrier peak at a rate set by the electric fields. $E_1$ is calculated from

$$V_1 = E_1 x d_1 + \frac{\varepsilon E_1^2}{2eN_d}$$

(6)

and similarly for $E_2$. The last term in these expressions is the voltage across the n1 and n2 depletion regions. By integration the number of holes is $P_m \times s$, where $s$, the effective width of the hole distribution, is given by

$$s = [1/E_1 + 1/E_2]/\beta.$$  

(7)
The hole concentration at the end of the n1 depletion region is \( P_m \exp(-\beta V_1) \). Holes are lost by diffusion to the cathode contact (where the density of holes is taken to be \( P_0 \)) and by recombination on the way, giving

\[
J_p(X_1) = \frac{e D_p}{L_p} \coth \left[ \frac{L_n}{L_p} \right] [P_m \exp(-\beta V_1) - P_0].
\] (8)

Similarly the hole concentration at the end of the n2 depletion region is \( P_m \exp(-\beta V_2) \), and solving the diffusion equation taking recombination in the n2 region into account gives

\[
J_p(X_2) = \frac{e D_p}{L_p \sinh(L_2/L_p)} \left[ P_0 \exp(\beta V_j) - 1 \right] - \cosh \left[ \frac{L_2}{L_p} \right] \times [P_m \exp(-\beta V_2) - P_0].
\] (9)

The condition \( J_p(X_1) = J_p(X_2) \) allows \( P_m \) to be calculated.

Applying Gauss's theorem to the p-plane gives

\[
e (N_t - P_m X_t) = e (E_1 + E_2).
\] (10)

This equation will probably not be satisfied by the values just calculated.

This is determined by calculating \( V_2^* \), the value of \( V_2 \) that does satisfy it given the values of \( P_m, s \) and \( E_1 \) just found. \( V_2^* \), the corresponding voltage is calculated from

\[
V_2^* = E_2^* \times d_2 + \frac{s E_{22}^*}{2 e N_d}.
\] (11)

This is then compared with \( V_2 \). If it is the same, then the required values of \( V_1 \) and \( V_2 \) have been found. If not, the guess must be suitably modified and the calculation repeated until the guessed value and the corresponding \( V_2^* \) are sufficiently close.

The current is then calculated to the required accuracy. \( J_{rec} \) is the largest contributor until well into the on state. \( J_{po} \) and \( J_{no} \), the hole and electron diffusion currents each side of the p-n junction depletion region, can be added if desired.

\[
J_{po} = \frac{e D_p}{L_p \sinh(L_2/L_p)} \left[ P_0 \left[ \exp(\beta V_j) - 1 \right] \exp(L_2/L_p) - [P_m \exp(-\beta V_2) - P_0] \right].
\] (12)

For a p\textsuperscript{+}-n junction \( J_{no} \ll J_{po} \) and so has been ignored.

The total device voltage is given by

\[
V = V_j + V_2 - V_1.
\] (13)
Some calculated J-V characteristics from Habib and Board (1983) are reproduced in figure 4.5.3. No hysteresis is shown or referred to. The device current depends almost entirely on $V_j$ ($J_{PD}$ has a slight dependence on $P_m$ and $V_2$ but is small), so hysteresis implies more than one solution for $V_1$ and $V_2$ for a single value of $V_j$. Because hysteresis had been found experimentally, a method of calculation was adopted for the simulations that follow which would find more than one solution for a given value of $V_j$ should more than one exist.

Using the parameter values for curve "c" of figure 4.5.3 it was found that there was generally only one solution, but that there were three for each value of $V_j$ in part of the negative resistance region. The corresponding I-V characteristic is shown in figure 4.5.4. The intermediate voltage states in the region of hysteresis (the dashed part of the characteristic) are not observed in practice as described in the previous section. When the device current is increased and then decreased again the device voltage will follow the path indicated, showing hysteresis and two discontinuous jumps as in the experiments. The ratio of the two currents $I_T$ and $I_H$ at which the jumps occur is about 2.

The remainder of the characteristic is similar to curve "c" suggesting that the model has been correctly implemented. The small discrepancies are believed to be due to different estimates of minor parameters for which Habib and Board do not quote values.
Fig. 4  Calculated J/V characteristics of a GaAs TB switch

The switch has the following parameters: \( N_{d1} = N_{d2} = N_d = 4.7 \times 10^{17} \text{ cm}^{-2}, N_s \)
(doping of the \( n^+\)-region) = \( 5 \times 10^{16} \text{ cm}^{-2}, d_2 = 8d_1 = 2000 \text{ Å}, L_{nn} = 1 \mu\text{m}, L = 0.2 \mu\text{m}, T = 300 \text{ K}, \text{ and } t_{s1} = t_{s2} = 10^{-9} \text{ s}\)

\( a \)  \( N_s = 10^{12} \text{ cm}^{-2}, \phi_{bn} = 0.46 \text{ eV} \)
\( b \)  \( N_s = 2 \times 10^{12} \text{ cm}^{-2}, \phi_{bn} = 1.10 \text{ eV} \)
\( c \)  \( N_s = 3 \times 10^{12} \text{ cm}^{-2}, \phi_{bn} = 1.386 \text{ eV} \)
\( d \)  \( N_s = 4 \times 10^{12} \text{ cm}^{-2}, \phi_{bn} = 1.411 \text{ eV} \)

FIGURE 4.5.3
The parameter values were then altered to simulate an MBE111 device. The minority carrier lifetime is by far the least well known quantity. Using \( \tau = 10^{-9} \) seconds as before gave values of \( I_S \), \( I_H \) and \( I_T \) which were all far smaller than observed experimentally. These three currents are very sensitive to the value of \( \tau \) because a decrease in this will not only reduce the proportion of holes that cross the n2 region without recombining, but also increase the loss of holes by diffusion to the cathode. This occurs because decreasing \( \tau \) causes \( J_{\text{rec}} \) to increase, requiring a smaller value of \( V_j \) and hence increasing the concentration of holes at \( X_1 \). It is assumed here and in later verbal explanations that switching occurs at an almost constant value of hole charge \( P_{\text{m}} \times 8 \), and that all three currents \( (I_S, I_H \) and \( I_T) \) behave similarly. Hence an increase in \( V_j \), which increases \( J_{\text{po}} \) faster than \( J_{\text{rec}} \), is required before switching will occur. The larger value of both \( V_j \) and \( 1/\tau \) leads to greatly increased switching currents.

In this case reducing \( \tau \) by a factor of 10 to \( 10^{-10} \) seconds increased all three currents by a factor of about 150. (figure 4.5.5). However, the shape of the characteristic and in particular the ratio of \( I_T \) to \( I_H \) are very little changed. The calculated values of \( I_S \), \( I_H \) and \( I_T \) are now 20\( \mu \)A, 0.2mA and 0.4mA, compared with 50\( \mu \)A, 0.5mA and 2.3mA found experimentally. The experimental and calculated magnitudes are now comparable, but significantly more hysteresis is seen experimentally than predicted by the model.

An idea of the origin of the hysteresis can be gained from the comparison on page 62 of the derived parameters of the device in the three states corresponding to a single value of \( V_j \). \( J_P(X_2) \) is the sum of two independent, oppositely directed, hole diffusion currents. That due to the excess hole concentration at the pn junction has been called \( J_P(X_2) \) (Fwd). The small variations in this arise from changes in the length of the depletion region at \( X_2 \). \( J_P(X_2) \) (rev) is the current that flows back towards the anode when an excess hole density exists at \( X_2 \).
MBE 111
300 kelvin
$\tau=10^{-10}$ seconds

FIGURE 4.5.5
Parameters of the three equilibrium states of an MBE111 device at one particular current

\[ V_j = 1.100 \text{ volts} \quad \tau = 10^{-10} \text{sec.} \quad T = 300 \text{ kelvin.} \]

\[ P_0 = 6.4 \times 10^{-6} \text{ cm}^{-3}. \quad N_i = 1.8 \times 10^6 \text{ cm}^{-3}. \]

| State | \( V_1 \) | \( V_2 \) | \( V \) | \( J_{\text{rec}} \) | \( J_{\text{po}} \) | \( J_{p}(X_1) \) | \( J_{p}(X_2) \) (\( \text{fwd} \)) | \( J_{p}(X_2) \) (\( \text{rev} \)) | \( P_{m}^{x8}/N_L (%) \) | \( P_m \) (\( \text{cm}^{-3} \)) | \( \delta \) (\( \text{nm} \)) | \( P_0 \exp(\beta V_j) \) | \( P_m \exp(-\beta V_2) \) | \( P_m \exp(-\beta V_1) \) |
|-------|--------|--------|--------|----------------|----------------|----------------|------------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| State 1 | .285 | .308 | 1.123 | 74,131 | 14,125 | 3,115 | 4,437 | 1,322 | 80 | 4.6\times10^{17} | 35 | 1.9\times10^{13} | 3.1\times10^{12} | 7.5\times10^{12} |
| State 2 | .299 | 1.350 | 2.151 | 74,131 | 14,791 | 4,539 | 4,539 | small | 70 | 1.1\times10^{18} | 12 | 1.9\times10^{13} | 2.5\times10^{-5} | 1.1\times10^{13} |
| State 3 | .299 | 3.408 | 4.209 | 74,131 | 14,791 | 4,749 | 4,749 | small | 48 | 1.2\times10^{18} | 8 | 1.9\times10^{13} | small | 1.2\times10^{13} |
State 1 has a lower maximum hole concentration than state 3, but the low barrier leads to a wide hole distribution, allowing sufficient holes almost to neutralize $N_t$. The low value of $P_m$ means that fewer holes are lost by diffusion to the cathode, but this is compensated by appreciable diffusion back to the anode from $X_2$, where the hole concentration has been greatly increased by the decrease in $V_2$.

State 2 is similar to state 1 at currents near to $I_H$ and to state 3 at currents near to $I_T$. Hysteresis occurs for values of $V_1$ between 1.071 and 1.104 volts, so state 2 is here more similar to state 3. The principal difference is the value of $\delta$. In state 2 a larger value of $\delta$ produces a smaller value of $E_2$ and $V_2$ by reducing the net flux from the $p$-plane. $V_2$ does not fall sufficiently to alter the current flows which are almost the same as in state 3. The lower value of $E_2$ allows the larger value of $\delta$ to persist.

These characterisations do not refer to the motion of the depletion regions at $X_1$ and $X_2$, suggesting that this is not relevant to the origin of the hysteresis. This is confirmed by the results of an early, simplified simulation which ignored the existence of these depletion regions altogether and used a fixed value of $W$, but which produced a very similar value of $I_T/I_H$. These effects do significantly alter the absolute values of the switching voltages and currents though.

Further information can be obtained by considering the stability of these states against fluctuations in $\Omega$, the number of holes under the barrier. In the standard calculation $V_1$ and $V_2$ are chosen to satisfy $J_{\text{rec}} = J_{\text{therm}}$ and $J_p(X_2) = J_p(X_1)$. The value of $\Omega$ is a product of the calculation. Here the value of $\Omega$ is decided in advance and $V_1$ and $V_2$ are instead chosen to satisfy $J_{\text{rec}} = J_{\text{therm}}$ and $\Omega = \text{the predetermined value}$. $\delta$ is calculated from $V_1$ and $V_2$ as before and $P_m$ obtained from $P_m = \Omega/\delta$.

(14) $J_p(X_2)$ and $J_p(X_1)$ can then be calculated from $P_m$, $V_1$ and $V_2$. If $J_p(X_2)$ minus $J_p(X_1)$ is positive then $\Omega$ will increase and vice versa.
Figure 4.5.6 is the flow diagram of this calculation and has been drawn for easy comparison with figure 4.5.2.

The results are presented in figure 4.5.7. They show that small fluctuations from states 1 and 3 are self-correcting, but are self-reinforcing from state 2, i.e. states 1 and 3 are stable equilibria but state 2 is an unstable equilibrium and will spontaneously convert to either 1 or 2. Hence it is not observed.

Since state 3 is stable against small fluctuations in charge, as the current increases through the range in which both state 3 and state 1 are allowed, the device will not jump from state 3 to state 1 until state 3 disappears. This of course is what is observed experimentally. There is a close analogy here with a particle in the potential of figure 4.5.8 in that the force curve for this particle has the same form as figure 4.5.7. The particle remains in the local minimum until it is converted to a point of inflection, as in figure 4.5.9, which is when states 2 and 3 become identical and disappear, as can be seen from the corresponding force curve.

**Changing temperatures**

The values of all the input parameters to the program are likely to vary to some extent with temperature, but $N_i$ clearly varies much faster than any of the others. It was taken to vary as $\exp(-E_g/2KT)$, which spans such a range over the temperatures considered that it was hard to keep calculations within the permitted numerical range. (See appendix C). The next effect to become important as the temperature is reduced is expected to be the failure of donors and acceptors to ionize fully. This has not been included in any calculations. Tunnelling may be important at lower temperatures still. The values of other parameters are expected to remain comparatively constant unless derived from $N_i$ or $N_d$. 

64
The effects of allowing only $N_i$ to vary with temperature are shown in figure 4.5.10. $I_T$ and $I_H$ increase rapidly as temperature decreases, and the other voltages and currents increase slowly, as seen in the experiments. Since $J_{\text{rec}}$ is proportional to $N_i \exp(\beta V_1/2)$ and $J_{\text{po}}$ is proportional to $N_i^2 \exp(\beta V_1)$ the effect of a smaller value of $N_i$ can be compensated for both currents simultaneously by a larger value of $\beta V_1$. Thus little change in the currents is expected with temperature, and the changes observed certainly are small compared with the factor of $10^{48}$ or so by which $N_i$ varies between 100 and 300 kelvin. The changes that do occur arise from more slowly varying quantities, such as the width of the hole distribution, which is proportional to $T$, and the $T^2$ prefactor in the expression for thermionic emission over the barrier. A narrower hole distribution will require a larger value of $P_m$ for switching, so hole losses will be larger and the necessary current increased. Similarly, at a given current a decrease in the prefactor will require an increase in $\exp(-\beta V_1)$, allowing more holes to diffuse to the cathode and delaying switching.

Neither of these effects reverses as temperature falls further and the trend in the simulation results shows no sign of reversing either, so this model only describes the effects seen above about 80 kelvin. In addition the quantitative agreement is poor: between 370 and 120 kelvin, $V_S$, $I_H$ and $I_T$ increase by factors of 1.5, 1.6 and 2.0 respectively, whereas the corresponding calculated ratios are 1.4, 4.5 and 9.7.
FIGURE 4.5.10

Graphs showing temperature vs. current for two different currents, $I_T$ and $I_H$, and temperature vs. current density for $I_s$. The temperature range is from 100 to 420 kelvin.

- $I_T$: Current in mA decreases with increasing temperature.
- $I_H$: Similar trend as $I_T$.
- $I_s$: Current density in μA remains relatively constant with increasing temperature.

Temperature range: 100 to 420 kelvin.
Recombination in the barrier

A typical minority carrier lifetime in high purity silicon is $10^{-3}$ seconds, but this can be brought down enormously to the $10^{-10}$ second range by adding $10^{17}$ cm$^{-3}$ of impurities such as gold with energy levels near the mid-gap. (Bullis, 1966). In general minority carriers in gallium arsenide and other direct gap semiconductors have much shorter lifetimes. Sze (p. 851) suggests an approximate value of $10^{-8}$ seconds without further comment. Alfano (1984, p. 93) reports lifetimes of 20 to 100 nanoseconds from photoluminescence experiments at low intensities, though a lifetime as short as 0.2 nanoseconds was seen at extremely high intensities. He also states that little is known about the dominant recombination mechanisms in gallium arsenide, so it is hard to decide by how much $\tau$ might plausibly be reduced by impurities, but the value of $10^{-10}$ seconds used here seems uncomfortably low. Because of this, effects were sought that had been omitted from the model but which might increase the switching currents and hence allow a larger value of $\tau$.

One such effect is recombination in the barrier. The expression used in this model for the recombination rate is

$$R = \frac{p_n - N_1^2}{\tau (n+p)},$$

which arises from a simple model of recombination via states in the bandgap (Sze, p. 37). Because the holes in the barrier are assumed to be in quasi-equilibrium, the pn product is constant on each slope of the barrier, though far greater than $N_1^2$. The recombination rate thus has its maximum value of $1/2\tau \times [P_m N_d \exp(-\beta V_1)]^{0.5}$ at the point on the cathode slope at which $n = p$. The recombination rate varies as $\text{sech}(\beta E_1 l)$, where $l$ denotes distance from this point, and integration shows that a current density of

$$J_{pr} = \frac{\pi}{2\tau} \beta E_1 [P_m N_d \exp(-\beta V_1)]^{0.5}$$

is required to replace the holes being lost. A similar expression for recombination on the other slope is included, but this is much smaller.
except when the current is well above $I_T$. To take recombination into account, the condition on the hole currents is altered to

$$J_p(X_2) = J_{pr} + J_p(X_1).$$

This modification increased the ratio of $J_{po}$ to $J_{rec}$ at switching nearly to unity. This ratio increases as temperature is reduced, so the approximation $J_n \approx J_{rec}$ is no longer valid. Ideally, the current from all the holes that recombine before the barrier would be added to $J_{rec}$, but the expressions for recombination contain parameters such as $P_m$ which are not known at the beginning of the calculation when the value of $J_n$ is required to obtain $V_1$.

To get round this the expression

$$J_n = J_{rec} + \frac{eD_0 P_0}{I_p} \left( \exp(\beta V_j) - 1 \right)$$

was used, in which the second term is an approximation to $J_{po}$ which contains no unknown quantities. At worst, this term may be too large by a factor of perhaps two.

With $\tau$ equal to $10^{-10}$ seconds these changes increased the calculated value of $I_H$ by more than an order of magnitude. $\tau$ was therefore increased to $2 \times 10^{-10}$ seconds, which restored $I_H$ at room temperature to close to the experimental value.
Low temperature behaviour

These modifications produced significant quantitative changes but no qualitative ones; in particular there was no disappearance of switching at low temperatures. If switching is not to occur the product of $P_m$ and $\delta$ must remain smaller than $N_t$ for all values of $V_j$. There is an upper limit to $P_m$ which decreases as temperature falls. This is because for $J_{po} \gg J_{rec}$ the ratio of $J_{po}$ to $J_n$ no longer increases with $V_j$. Hole losses by diffusion to the cathode are proportional to $\exp(-\beta V_1)$ and hence to $J_n$. This means that, when this loss mechanism dominates, hole losses increase as fast as hole injection, which is proportional to $J_{po}$, and so an upper limit is placed on $P_m$. This decreases as temperature falls because of the $T^2$ factor in the thermionic emission formula. Even so, this model will always predict a complete collapse of the barrier because, by the thermionic emission formula, as current increases $V_1$ must decrease, allowing $\delta$ to increase without limit and hence the hole charge to become equal to $N_t$ however low the maximum value of $P_m$. With this final version of the program, at 100 kelvin $I_T$ was about 10mA, more holes were lost by diffusion to the cathode than by recombination in the barrier and $V_1$ had decreased so far that $1/\beta E_1$ was 10 nanometers, 20% of $d_1$. The approximations used to describe the barrier are clearly poor when $1/\beta E_1$ becomes a significant fraction of $d_1$, so let us take a closer look at the width of the hole distribution.

Figure 4.5.11 shows the p-plane as a delta function, one side of the hole distribution and the corresponding electric field. The hole concentration initially falls rapidly at a rate set by the full field from the p-plane ($E_B$), and is well below $P_m$ before the rate of fall decreases to that corresponding to $E$, the field far from the barrier. If the hole charge is small, $E$ will be little less than $E_B$ and the effective width will be close to $1/\beta E_B$. If the hole charge is almost as great as $N_t$, then $E$ will be small, the distribution will have a long tail and the effective width will be between $1/\beta E_B$ and $1/\beta E$. Since the hole concentration is highest where it
Electric field against distance from the P-plane for the case $E/E_B = 0.3$.

Corresponding hole concentration (solid line): $P_m \exp(-\beta E_B x)$ (dashed line).

Figure 4.5.11
is falling fastest a value nearer to $1/\beta E_D$ is expected.

Very minor changes to the program of appendix B allow it to be used to calculate the shape of the hole distribution. This shows that the effective width never exceeds $2/\beta E_D$ even when $E$ tends to zero. This suggests an alternative treatment of the barrier in which $\delta$ is taken as $1/\beta E_D$. $E$ and $E_D$ are then related by

$$E_D - E = \frac{P_m}{\beta E_D} \times e/\epsilon.$$

$V_1$, $E_2$, and $P_m$ are obtained as before, then this equation is used to find $E_D$. Gauss's theorem yields $E_{D2}$, from which the above equation gives $E_2$ and $V_2$, which can be compared with the guessed value.

The results of these assumptions are shown in figure 4.5.12. They are quite similar to the behaviour of MBE 628 devices (figure 4.3.10) in that the minimum voltage at low temperatures is little short of that which would flatten the barrier without any hole injection. There is no hysteresis even at high temperatures, which is associated with the fact that in this model $\delta$ is independent of $V_2$. Previously a small value of $V_2$ led to a large value of $\delta$ and hence enough holes to cancel the p-plane charge, consistent with the small value of $V_2$ and also conversely. The ratio of $\delta$ in state 1 and state 3 of the earlier table is greater than 4, and $1/\beta E_2$ changes by more than a factor of 10, but even so the hysteresis is less than seen experimentally. Even a more exact treatment of this model, which recognises the fact that the width of the hole distribution on the $U_2$ side of the barrier can increase by a factor of up to 2 if $E_2$ is reduced nearly to zero, seems unlikely to produce adequate hysteresis.

$E_D$ will be of the order of $N_0/\epsilon$, so $2/\beta E_D$ can be estimated as about 1 nanometre at 100 kelvin. Since the width of the p-plane itself is 20 nanometres in these devices, it is clear that this will greatly affect the value and perhaps the behaviour of $\delta$. Limiting $\delta$ to 20 nanometres when $1/\beta E_1 + 1/\beta E_2$ is greater than this gives qualitatively correct results but has not been justified. An adequate approximation has not been found for
Figure 4.5.12
this case. An exact solution could be found by appendix B techniques as follows:

$V_1$, $V_2$ and $P_m$ are known at the point in the calculation at which $\delta$ is required. $\delta$ is used with $P_m$ to calculate $V_2^*$ form comparison with $V_2$. It is also known that the electric field is zero at the peak of the barrier. The only thing unknown is the position of the peak within the p-plane. This is therefore guessed, and the potential change between the peak and the bottom of the N$_1$ side of the barrier found by numerical integration, taking into account the distribution of negatively charged acceptors forming the p-plane and the holes in equilibrium with $P_m$. This should be equal to $V_1$. If it is not, the guessed position of the barrier peak is altered until it is. $V_2^*$ can then be found by integration down the N$_2$ slope of the barrier.

In attaining this improved accuracy the evaluation of equations 7, 10 and 11 has been replaced by a numerical integration that must be repeated many times. This will take a great deal longer, and the evaluation is a critical step as it is part of the iteration loop of the main program. For this reason this improvement has not been attempted.

The behaviour of these devices is very complex. The three types examined may exhibit three different types of behaviour at low temperatures because, although 111 and 628 devices show some similarities, there is a marked difference in the extent of the negative resistance region at low temperatures. The 628 characteristic is similar to that expected merely from the applied voltage flattening the barrier, whereas in 111 devices appreciable reduction of the barrier height by hole injection must still be occurring. 219 devices even show hysteresis. It may be that many more effects will have to be included to simulate this variety of behaviour.
Summary

It has been shown that the model of Habib and Board gives a qualitative account of many of the features observed experimentally, including the hysteresis and the variation of the switching points at moderate temperatures. Quantitative agreement is poor. Hole recombination in the barrier has been shown to be quantitatively significant but more detailed modelling of the hole distribution is likely to have a greater quantitative effect. A different treatment of the hole distribution led to a qualitative explanation of the low temperature behaviour. No simple model has been found that predicts both the high temperature hysteresis and its absence at low temperatures. An algorithm for a more accurate simulation has been proposed.
5.1 Device structure

A cross-section of this device is shown in figure 5.1.1. It was grown by molecular beam epitaxy on a (100) n+ GaAs substrate. A buffer layer of GaAs 0.5μm thick and n doped to 5x10^{18} cm^{-3} was grown first, followed in sequence by 0.5μm of 10^{16} cm^{-3} n doped GaAs, a compositionally graded region of AlGaAs, 0.5μm of GaAs n doped to 10^{16} cm^{-3} and finally a contact layer of 0.5μm of n^+ GaAs doped to 10^{18} cm^{-3}. The n dopant was silicon. The graded region was nominally undoped and started at an aluminium fraction of 0.3, decreasing linearly to zero over 50nm. Transmission electron microscopy of the material confirmed an abrupt interface followed by a 50nm region in which the aluminium concentration fell smoothly to zero. Aluminium concentrations appeared uniform across the layer. Simple rectangular mesas in a range of sizes up to 180μm square were fabricated by photolithography and etching, and ohmic contacts made to the top of the mesa and the back surface of the substrate.

5.2 Theory of operation

Junctions between two semiconductors with different bandgaps have properties that depend on the way the conduction and valence bands align. The GaAs/AlGaAs system has the alignment shown in figure 5.2.1, so that this device has the band diagram of figure 5.2.2. If it is biassed so that electrons are accelerated up the slope in the graded region, on reaching the junction these electrons will gain an amount of energy which depends on the maximum Al fraction, and can be hundreds of meV. At most temperatures this is much greater than KT, so that the electrons will have only a narrow spread of energy. The structure has been used as an effective emitter of hot electrons in hot electron transistors and other devices (e.g. Long et al., 1986).
It is simple to determine the bandgaps of GaAs and AlGaAs, but rather
d harder to discover how much of the difference appears as the conduction
band offset and how much as the valence band offset. Calculations and an
early experiment suggest that the conduction band offset should be about
88% of the total, but more recent experiments have found values nearer to
65%. (see Adachi, 1985 and refs. 70 to 73 therein). This leads to some
uncertainty in the expected barrier height even if the maximum Al
concentration is known accurately.

The number of electrons with a particular energy traversing the barrier is
the product of the number of such electrons approaching the barrier and the
transmission probability at that energy. Both of these quantities vary
rapidly with energy, but in opposite directions, so that their product has
a sharp maximum and appreciable current flows only at energies close to
this. At low temperatures the number of electrons falls off very rapidly as
energy increases, so states at the bottom of the conduction band will
contribute most to the current. As temperature increases, the electron
distribution falls off less rapidly and so the energy of the maximum
increases towards the limit of the top of the barrier. This is always true
(Rhoderick, 1978), but the temperature range in which this behavior is seen
depends on the shape and size of the barrier. To calculate this an
expression is required for the variation with energy of the tunnelling
probability for this barrier. Because of the simple form of this potential
an analytic solution of Schrodinger's equation does exist, but it involves
Bessel functions of fractional order, which are not everyday objects. The
WKB approximation can be used for wide barriers with any shape, giving
tunnelling probability = $e^{-Z}$, where

$$Z = \frac{2}{\hbar} \left( 2em^* \right)^{0.5} \int_{x_1}^{x_2} (V_B - V_e)^{0.5} dx.$$  

$x_1$ and $x_2$ are the classical turning points of the electron and $V_B$ and $V_e$
the potential energy of the barrier and the electron energy expressed as
For this particular barrier (figure 5.2.3)

\[ V_b = \left( V_0 - V \right) \frac{X}{d} \quad \quad x_1 = \frac{v_e d}{V_0 - V} \quad \quad x_2 = d \]

and

\[ Z = \frac{4d}{3h} \left( \frac{V_0}{V_0 - V} \right) (2\omega m^*)^{0.5} \left( V_0 - V - V_e \right)^{1.5} \]

The energy at which most current flows is approximately where the product of the Boltzmann factor exp \(-E/RT\) and the tunnelling probability exp \(-Z\) is greatest, since the other factors such as the density of states vary much more slowly. This maximum occurs at

\[ V_e(\text{max}) = V_0 - V - \frac{e}{8m^*} \left[ \frac{\hbar}{d} \frac{(V_0 - V)}{KT} \right]^2 \]

This allows a prediction of the temperature at which \(V_e(\text{max})\) begins to fall below the top of the barrier. \(V_e(\text{max})\) has fallen to 7/8 of \(V_0 - V\), the barrier height with bias applied, when

\[ T = \frac{\hbar}{d} \frac{1}{K} \left[ \frac{e}{m^*} \left( V_0 - V \right) \right]^{1/2} \]

and \(V_e(\text{max}) = 0\) when \(T\) has fallen by a further factor of \(\sqrt{8}\). The approximation becomes less good as \(V_e(\text{max})\) approaches zero, because the density of states starts to vary rapidly, falling to zero for \(V_e(\text{max})\) less than zero. This means that \(V_e(\text{max})\) will tend to zero as the temperature decreases, not to minus infinity as suggested by this formula, and hence \(V_e(\text{max})\) is expected to be small but still above zero at this temperature. These values can be compared with experiment, since when \(V_e(\text{max})\) does not change rapidly with temperature the slope of a graph of \(\ln(I)\) against \(e/RT\) can be shown to be \(-V_e(\text{max})\).

When an external voltage is applied to the device, negative charge is built up in front of the barrier and depleted from behind it. Treating the barrier as the dielectric of a parallel plate capacitor leads to a calculation of the amount of charge in these regions which predicts that the length of the depletion region and the potential drop across it will be given by

\[ L_d = \frac{V_e}{N_d e d} \quad \text{and} \quad V_d = \frac{V^2 e}{2 N_d e d^2} \]

where \(V\) is the barrier voltage.
The amount of charge in the accumulation region is equal to that in the depletion region, but it is distributed differently, usually being accommodated in a shorter length so that the potential drop across this region is much less. This can be seen from the Thomas-Fermi theory of the screening of a positive potential by a free electron gas (e.g. Ashcroft, 1976), which predicts that the excess charge density will decrease as \( \exp\left(-\frac{\text{distance/screening length}}{\text{distance}}\right) \) on moving away from the barrier. The screening length is given by

\[
L_s = \left[ \frac{\pi^2 \hbar^2 e}{m^* e^2 K_F} \right]^{1/2}, \text{ where } K_F = 3\pi^2 N_d.
\]

Integration shows that the correct amount of charge is present for an accumulation voltage of

\[
V_a = V \cdot \frac{L_s}{d}
\]

This is a constant, in this case small, proportion of \( V \). \( V_d \) on the other hand increases quadratically with \( V \), and can become very significant. Figure 5.2.4 is an energy band diagram showing these effects. It can be seen that the barrier height for thermionic emission is changed to \( V_0-(V_L-V_d) \), and that \( V_0(\text{max}) \) will now be equal to \( V_a \) rather than zero for tunnelling at the lowest temperatures.

The total voltage across the device is given by

\[
V_t = V + V_a + V_d.
\]

This can be combined with the above expressions to give the barrier voltage as a function of the total voltage across the device:

\[
V = A \left[ \left( 1 + \frac{2dV_t}{A(d+L_s)} \right)^{0.5} - 1 \right],
\]

where

\[
A = \frac{e d (d+L_s) N_d}{e}.\]

Figure 5.2.5 shows the calculated variation of \( V, V_d \) and \( V_a \) with \( V_t \).
FIGURE 5.2.5
5.3 Results and discussion

The static characteristics of this device were taken with a Keithley 480 picoammeter in conjunction with a voltage ramp generator and an X-Y plotter when the current range to be covered was narrow enough for range changes to be unnecessary, and otherwise with the same ammeter in conjunction with a switched voltage source and manual data recording. Constant currents were produced from a 10 volt stabilised voltage source with a suitable series resistor. The non-commercial pieces of equipment have been described previously. (Davies, 1979)

An Oxford Instruments continuous flow helium cryostat was available which permitted rapid and easily controlled temperature changes between about 4K and 350K. A four terminal rhodium-iron resistance thermometer was placed close to the sample and this gave good results over this wide temperature range. The magnetic field experiments were performed in the magnet used for the magnetophonon experiments.

Figure 5.3.1 shows the variation with temperature of the current through the device for various forward voltages. The above theory suggests that, although the barrier voltage is not equal to the external voltage, it is constant, so that figure 5.3.1 should accurately reflect the variation of $V_0(\text{max})$ with temperature. The curve for $V_T = 10\text{mV}$ has a constant slope over 5 decades of current, suggesting thermionic emission is dominant. This slope corresponds to $V_0(\text{max}) = 160 \pm 5\text{mV}$, so the corresponding barrier height is about 170meV. The bandgap of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ exceeds that of GaAs by 1.26x meV (Adachi, 1985), so with $x = 0.3$ a barrier of at least 225 meV is expected ($\Delta E_C/\Delta E_V \geq 60/40$). The maximum value of $x$ for a 170 meV barrier is less than 0.23, well below that expected for the growth conditions.
FIGURE 5.3.1
Unintentional doping can lead to charge in the barrier that will alter the barrier profile from that shown in figure 5.1.2. The potentials at the edges of the barrier are not affected by this doping though, so it cannot reduce the barrier height if the aluminium concentration remains as grown.

Diffusion of aluminium across the abrupt interface during contact alloying would reduce the barrier height. 15 nm is a typical diffusion distance that would be required. Another explanation in terms of irregularities in the barrier produced by impurities will be discussed later.

Above 90K the curve for $V_t = 100$ mV has a constant slope corresponding to a $V_e(\text{max})$ about 44mV less than that for $V_t = 10$ mV. This is in excellent agreement with the change in V predicted from figure 5.2.5.

The higher voltage curves can be followed to temperatures at which tunneling becomes important. Figure 5.3.2 shows the variation of the current with temperatures down to a few kelvin. The predicted temperatures for the start and finish of the curved region corresponding to activated tunnelling are 83K and 29K, using the formulae of the previous section. These are in reasonable agreement with the experimental results.

The current still decreases very slowly as the temperature falls, even well below 29K. The apparent activation energy is about 0.3 meV, but as this is less than KT the relationship between slope and activation energy will no longer hold. This is attributed to electrons in the $10^{16}\text{cm}^{-3}$ doped regions "freezing out" onto the donors. This happens much more slowly than would be expected from the 5.8meV ionisation energy for isolated silicon donors in GaAs. This is because the donors are now close enough for the wavefunctions of the unbonded electrons to overlap, causing the formation of a band of donor levels. Though this is centred on $-5.8$ meV, the highest levels are much nearer the conduction band. Above an n-type doping of about $3\times10^{16}$ cm$^{-3}$ in GaAs this band has become wide enough to overlap the conduction
FIGURE 5.3.2

thermionic emission

activated tunneling

freezeout

$\log I$

$-6$

$-7$

$-8$

$-9$

$-10$

0 20 40 60 80 100 120 140

$\frac{1000}{T}$
band, and electrons can then be moved throughout the sample however low the temperature. This is known as the metal-insulator transition. The effect of a magnetic field is to reduce the size of the donor wavefunctions and to raise the bottom of the conduction band by about $\hbar\omega_c/2$ relative to the donors. This once again separates the donors from the conduction band and the current drops by a large amount at low temperatures, as in figure 5.3.3. There is little difference between B parallel and B perpendicular to the current, suggesting that it is the above effects that are important, rather than the magnetic field preventing electrons from gaining enough energy from the electric field to impact ionize donors.

Figure 5.3.4 shows the I-V characteristics for this device at several temperatures. At 130K the current flow is mainly by thermionic emission and at 4K by tunnelling at the Fermi energy.

The barrier heights for forward and reverse thermionic emission are $V_o-(V_t-V_d)$ and $V_o+V_d$. Hence $I = I_o \exp(-eV_d/KT) \left[ \exp(eV_t/KT) - 1 \right]$. $I_o$ could be calculated but has instead been used as a fitting parameter, found by plotting $I$ against $\exp(eV_t/KT)$ for small values of $V_t$ where $V_d$ is expected to be negligible. The value that results is 28nA. Two I-V characteristics have been calculated from this, and are shown dotted on fig.5.3.4. The first ignores the formation of a depletion region, and the second assumes the values of $V_d$ from figure 5.2.5. The second is a great improvement, but the values of $V_d$ appear too small. A rough calculation suggests that much of the remaining discrepancy would be removed if allowance were made for the accumulation region extending into the barrier because of its gradual onset. The capacitance of the structure would then be increased, leading to a larger depletion region voltage.

A similar calculation has been performed for the low temperature tunnelling. From section 5.2 the slope of the I-V characteristic is given by $-dZ/dV_t$, with $V_e(\text{max})$ equal to $V_a$. Making this substitution in the
FIGURE 5.3.3

V = 120 mV
T = 30 K
FIGURE 5.3.4

$T = 130$ kelvin

$T = 4$ kelvin
previous expression for $Z$ gives

$$Z = \frac{4d}{3h} (V_0 - V)^{0.5} (V_0 - V_a - V)^{1.5}.$$  

Since $Z$ is a function of both $V$ and $V_a$

$$\frac{\partial Z}{\partial V_t} = \frac{\partial Z}{\partial V} \times \frac{\partial V}{\partial V_t} = \left[ \frac{\partial Z}{\partial V} + \frac{\partial Z}{\partial V_a} \times \frac{\partial V_a}{\partial V_t} \right] \frac{\partial V}{\partial V_t}.$$  

$$= \frac{2d}{3h} \left[ \frac{2em^2}{V_0 - V} \right]^{0.5} \left[ 1 + 3 \frac{\partial V_a}{\partial V} \right] \frac{\partial V}{\partial V_t}.$$  

Voltage dropped across the accumulation region is three times as effective at increasing the current as voltage across the barrier itself, so the proportion of $V_t$ estimated to appear across the accumulation region will have a large effect on the calculated value of $dZ/dV_t$. $dV_a/dV \approx 0.22$ from section 5.2 and figure 5.2.5 provides values of $V$ and $dV/dV_t$. The calculation for three values of $V_t$ spanning the experimental range is tabulated below. The last two columns are the calculated and experimentally found slopes.

<table>
<thead>
<tr>
<th>$V_t$</th>
<th>$V$</th>
<th>$dV/dV_t$</th>
<th>$-dZ/dV$</th>
<th>$-dZ/dV_t$</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>47</td>
<td>.357</td>
<td>x</td>
<td>229</td>
<td>82</td>
</tr>
<tr>
<td>150</td>
<td>68</td>
<td>.289</td>
<td>x</td>
<td>252</td>
<td>73</td>
</tr>
<tr>
<td>210</td>
<td>85</td>
<td>.250</td>
<td>x</td>
<td>276</td>
<td>69</td>
</tr>
</tbody>
</table>

Again agreement is reasonable, but would be improved by $V_a$ increasing more rapidly with $V_t$.

The conclusion of this section is that the important influences on the behaviour of this device have been identified and are reasonably well described by the simple model of section 5.2.
5.4 Random switching between discrete resistance levels.

At low temperatures and currents above 0.1µA the resistance of these devices was found to jump between discrete values at apparently random times. Figure 5.4.1 shows the changes of current with time for constant voltages of 197mV and 208mV. \((I_{\text{max}}-I_{\text{min}})/I_{\text{max}}\) is very large at about 25% and a typical time in either state is of the order of 1 second. Similar changes were found in all of these devices that were not either open or short circuits. All of the devices were from the same layer. There was considerable variation in the amplitude of the changes between devices. Occasionally there were differences between experimental runs on the same device. The changes of figure 5.4.1 were the largest seen repeatedly, but single events with values of \(\Delta I/I\) of nearly 100% were seen at very low currents and temperatures where the switching rate is very low. Some devices exhibited more than two levels, at least 10 being present in one device (figure 5.4.2). The switching rate increased with current, as in figure 5.4.1, and also increased as temperature increased, but \(\Delta I/I\) and the mark to space ratio decreased with temperature until above 90K the changes were below the noise level.

A study was carried out of the variation of the mean times spent in the high and low resistance states \(t_{h}\) and \(t_{l}\) for the device of figure 5.4.1, since it showed large changes and only two levels. The random switching of the device was generally too fast for a chart recorder to follow, so the measurements of the mean time in each state were carried out as follows: a constant current was passed through the device, and the resulting voltage connected to the input of a comparator. The threshold voltage of this comparator was set between the voltages of the high and low resistance states so that it switched in sympathy with the device. The mean output voltage of the comparator was found by an integrating R-C network. If this
mean output is denoted by \( V \), and the high and low output voltages of the comparator by \( V_h \) and \( V_l \), then \( t_h/(t_h+t_l) = (V-V_l)/(V_h-V_l) \). \( t_h + t_l \) was found by counting the number of times the device switched in a fixed time with a frequency counter connected to the output of the comparator. Both \( V \) and the number of counts varied because of the random nature of the switching. The variation in \( V \) was more significant. The procedure adopted was to use an integration time of 10 seconds, to allow at least 30 seconds for the output to come to equilibrium and then to observe the output for about 10 seconds noting the highest and lowest values and averaging them. This was repeated for different currents and then the temperature was altered and the whole procedure repeated.

The results are presented in figures 5.4.3 and 5.4.4. The logarithms of the mean times have been used as the variation in the times themselves is so large. \( t_l \) shows little variation with temperature, but \( t_h \) decreases rapidly above about 30K. This method gives nearly the same absolute errors in \( t_h \) and \( t_l \), but \( t_h \) was often much less than \( t_l \), so the fractional errors in its estimate were sometimes large. \( t_l \) varies rather closely as \( I^{-3} \). In the range where \( t_h \) is independent of temperature it varies roughly as \( I^{-2} \).

Similar waveforms from other devices have been analysed to confirm the exponential probability distribution of \( t_h \) and \( t_l \) expected if there is a constant probability per unit time that the device will change to the other state. This requires more sophisticated data acquisition equipment than was available.
FIGURE 5.4.3
FIGURE 5.4.4

- 25 kelvin
- 41 kelvin

line has gradient 3

lines have gradient 2
5.5 A model for the switching

Similar, though generally smaller, variations of current with time have been seen in many systems, including both forward and reversed biased p-n junctions (reviewed by Buckingham, 1976) and small mosfets (Uren et al., 1986). Theories that have been advanced for the origin of these changes include the spontaneous formation and quenching of microplasmas, traps modulating surface channels, traps modulating the current through defects in the p-n junction (all in Buckingham, 1976) and traps in the dielectric modulating the number of electrons in the channel (Uren et al., 1986). In these devices the current flow is expected to controlled by the barrier height, perhaps at small currents by local low points in the barrier. At the temperatures at which switching is observed, conduction is by tunnelling somewhere below the barrier peak rather than by thermionic emission. One explanation for the different resistance levels is that they are the result of the changing occupancy of one or more traps near the peak of the barrier. This would alter the probability of tunnelling through the adjacent region. If this is correct, then the high resistance state must be the more negative state of the trap. A donor becoming positive could cause an exponentially large local increase in current, but an acceptor becoming negative could only shut off what current already exists. Thus it seems more likely that a donor could cause changes in current as large as those observed. This does not apply if the current distribution in the device is highly non-uniform even when the trap is neutral, owing to some defect or local low point in the barrier. An acceptor could then shut off the large local current density around this.

The Fermi level is near to the conduction band in the n-type GaAs each side of the barrier, so in equilibrium it will be about 200meV below the conduction band near the peak of the barrier. If the trap is a donor, i.e. above this Fermi level, then at low currents (i.e. near equilibrium) the trap will generally be empty and the device will be in the low resistance
state, which is what is observed. This prediction is reversed if the trap is an acceptor, and in fact an acceptor based model gives wrong predictions generally.

The change from high to low resistance will result from the excitation of an electron from the trap into the conduction band. This interpretation suggests that the ionization energy of the donor will be given by the slope of the emission rate curves for the lowest currents, about 30meV. This is consistent with silicon in Al$_{27}$Ga$_{73}$As (Chand et al., 1984). It may also be consistent with other types of donor in AlGaAs of this composition, since in GaAs at least the donor ionization energy is almost independent of the donor atom. The apparent loss of the activated nature of the emission process below about 30K is attributed to local Joule heating of the region around the trap, which must carry a large proportion of the current for $\Delta I/I$ to be as large as 25%. This explains why the emission time decreases as the current is increased at low temperatures, but is less current dependent at higher temperatures. There is some evidence that the temperature above which activation is apparent increases with current, which also supports this theory. Another possibility is that at low temperatures the electron leaves the trap by tunnelling rather than by thermal excitation to the conduction band.

The change from low to high resistance results from electron capture by the trap. Since the trap is more than 100meV above the quasi-Fermi energy anywhere in the barrier, at first sight it is rather surprising that at constant current the capture time is almost independent of temperature. (The slope of the log $t$ versus $1/T$ curves corresponds to an energy of zero, ± a few meV). However, such a situation is possible: when thermionic emission predominates the current depends on the number of electrons near the peak of the barrier times the thermal velocity, which varies only slowly. Hence at constant current the voltage drop across the barrier will adjust itself so that the number of electrons near the peak of the barrier
is independent of temperature. A similar result may apply in this case, where tunnelling predominates. The reason for the capture rate varying as $R^3$ is not understood. A simple argument suggests it should be proportional to the attempt rate, i.e. to $R^1$.

If there is only one silicon atom in the last 10nm of the barrier then the concentration of silicon is about $10^9$ atoms per cm$^3$. This is lower than the expected unintentional silicon doping level by a factor of about $10^4$. Thus traps producing this effect must have to satisfy some additional requirement. The most likely is that at low currents only a small proportion of the device conducts, and that only donors near this region can affect the current. The conducting region may be one in which the aluminium fraction is unusually low, or the inhomogeneity may be in the depletion region, which is only a few donor separations (at $10^{16}$cm$^{-3}$) in length and sustains a large proportion of the total voltage. Devices that were nominally identical except for a doping level of $10^{18}$cm$^{-3}$ in the regions each side of the barrier did not show this effect.

The effective barrier height for thermionic emission has been measured as 170±5meV (section 5.2) at temperatures above 100K, at which the device is expected to be in the low resistance state almost all the time. A similar measurement cannot be made for the high resistance state, since at the low temperatures at which the device will remain in this state for an appreciable period the predominant mechanism of current flow is not thermionic emission. However, the minimum barrier height expected is about 200meV, using a maximum aluminium fraction of 0.27 and 60% conduction band offset. The difference between these figures may be a result of the barrier lowering by the donor. The amount of reduction is similar to the donor ionization energy.

This model could be verified if the activation energy is seen to track the rapidly varying donor ionization energy in other devices where the maximum aluminium concentration is varied between 20 and 40 per cent.
Appendix A: Magnetophonon program

The functions performed by this program have been described in section 2. The program grew in an unplanned fashion from its simple original form and so is anything but model software. Hence only the sections which are unique to this application are described, the graphics routines and overall structure being omitted. The first section collects data from the Thurlby meters.

1130 INPUT "ENTER DATA FILE NUMBER "AS
1140 AS="DATA."+AS e.g. "DATA.50".
1150 INPUT "BMAX (7) "BMAX limits within which data to be collected.
1160 INPUT "BMIN (3.5)"BMIN
1170 DX=(BMAX-BMIN)/200 spacing of data points.
1180 N%=20207 set print format.
1190 J%=0 data point counter.
1200 *FX2,1 input from RS232 instead of keyboard.
1210 MODE 4 clears screen.
1220 CLOSE#0 closes any open files.
1230 *FX3,6 stops RS232 data appearing on screen.
1240 ?FD04=253 select Thurlby 1.
1250 INPUT XS take reading.
1260 IF NOT (LEFT$(XS,1)="R") THEN GOTO 1250 check it.
1270 ?FD04=255 select Thurlby 2.
1280 INPUT YS
1290 IF NOT (LEFT$(YS,1)="R") THEN GOTO 1280
1300 X=.18*VAL(RIGHT$(XS,9)) separate numerical part of reading.
1310 Y=VAL(RIGHT$(YS,9))
1320 *FX3,0 screen back to normal.
1330 MOVE X*160,Y*100 prepare to draw graph.
1340 PRINT TAB(0,0);""
1350 PRINT TAB(0,0);X,Y
1360 IF X<BMIN OR X>BMAX THEN GOTO 1230 wait until data in desired range

1370 LAST=X points now in range.
1380 J%=J%+1 increment point counter.
1390 B=0 used to average data.
1400 A=0 ditto.
1410 N%=0 number of points averaged.
1420 *FX2,1 this section as above.
1430 *FX3,6
1440 ?FD04=253
1450 INPUT XS
1460 IF NOT (LEFT$(XS,1)="R") THEN GOTO 1450
1470 ?FD04=255
1480 INPUT YS
1490 IF NOT (LEFT$(YS,1)="R") THEN GOTO 1480
1500 N%=N%+1
1510 A=A+.18*VAL(RIGHT$(XS,9)) average readings...
1520 B=B+VAL(RIGHT$(YS,9))
1530 S=A/N%
1540 IF ABS(LAST-S)<=DX THEN GOTO 1440 until time for next point.
1550 LAST=.18*VAL(RIGHT$(XS,9))
1560 *FX3,0
1570 *FX2,0
1580 F(J%,0)=A/N%  
1590 F(J%,2)=B/N%  
1600 DRAW F(J%,0)*160,F(J%,2)*100  
1610 PRINT TAB(0,0);"  
1620 PRINT TAB(0,0);F(J%,0),F(J%,2),N%,J%  
1630 IF F(J%,0)>BMIN AND F(J%,0)<BMAX THEN GOTO 1380 next point, if any.  
1640 PRINT F(J%,0)  
1650 X%=OPENOUT AS  
1660 FOR P%=1 TO J%  
1670 PRINT#X%, F(P%,0), F(P%,2)  
1680 NEXT P%  
1690 CLOSE#X%  

The data is now read back into the array F and graphed. There are W% points. Double differentiation is carried out as follows:  
2360 K%=3 differences taken between points 2K% intervals apart.  
2370 FOR N%=1 TO 2 single differentiation performed twice.  
2380 FOR R%=1+W% TO W%-K% i.e most of the points  
2390 F(R%,3)=(F(R%+K%,2)-F(R%-K%,2))/(F(R%+K%,0)-F(R%-K%,0))  
2400 F(R%,1)=(F(R%+K%,0)+F(R%-K%,0))/2  
2410 NEXT R%  
2420 FOR R%=1 TO K% special treatment for end points  
2430 F(R%,3)=(F(R%+K%,2)-F(1,2))/(F(R%+K%,0)-F(1,0))  
2440 F(R%,1)=(F(R%+K%,0)+F(R%-K%,0))/2  
2450 NEXT R%  
2460 FOR R%=W%-K%+1 TO W%  
2470 F(R%,3)=(F(W%,0)-F(R%-K%,2))/(F(W%,0)-F(R%-K%,0))  
2480 F(R%,1)=(F(W%,0)+F(R%-K%,0))/2  
2490 NEXT R%  
2500 FOR R%=1 TO W%  
2510 F(R%,2)=F(R%,3) when calculation complete original  
2520 F(R%,0)=F(R%,1) data can be written over.  
2530 NEXT R%  
2540 NEXT N%  

Alternatively the oscillations can be extracted by finding the deviation of the data from a least squares fit polynomial. This gives a more easily visualised measure of the amplitude.  
3440 N%=4 polynomial has 4 coefficients, i.e. cubic.  
3450 I%=1 decide whether data has + or - slope, i.e resistance  
3460 IF (F(W%,0)-F(1,0))*(F(W%,2)-F(1,2))<0 THEN I%=-1 or conductance.  
3470 FOR R%=1 TO N%  
3480 FOR S%=0 TO N%  
3490 M(R%,S%)=0 set matrix M to zero  
3500 NEXT S%  
3510 NEXT R%  

continued
3520 FOR R%=1 TO W%
3530 FOR S%=1 TO N%  M is filled with the coefficients of the
3540 FOR T%=1 TO N%  simultaneous equations to be solved.
3550 M(S%,T%)=M(S%,T%)+F(R%,0)!=S%-1)
3560 NEXT T%
3570 NEXT S%
3580 FOR S%=1 TO N%
3590 M(S%,0)=M(S%,0)+F(R%,2)*F(R%,0)!(S%-1)
3600 NEXT S%
3610 NEXT R%

3620 FOR P%=1 TO N%  the polynomial has N% coefficients.
3630 FOR R%=0 TO N%
3640 FOR S%=1 TO N%
3650 Q(S%,R%)=M(S%,R%)
3660 NEXT S%
3670 NEXT R%

3680 FOR M%=N% TO 2 STEP -1  find one of these coefficients
3690 FOR S%=1 TO M%-1  by pivotal condensation.
3700 FOR T%=0 TO M%-1  start by making a copy of M called Q.
3710 R(S%,T%)=Q(S%+1,M%)*Q(S%,T%0-Q(S%,M%)*Q(S%+1,T%)
3720 NEXT T%
3730 NEXT S%

3740 FOR R%=0 TO M%-1  R is a matrix with one less coefficient than Q.
3750 FOR S%=1 TO M%-1
3760 Q(S%,R%)=R(S%,R%)
3770 NEXT S%
3780 NEXT R%

3790 NEXT M%  copy R to form new Q.

3800 A(P%)=R(1,0)/R(1,1)
3810 FOR S%=1 TO N%  and repeat the elimination.
3820 M(0,S%)=M(S%,1)
3830 NEXT S%
3840 FOR R%=1 TO N%-1
3850 FOR S%=1 TO N%
3860 M(S%,R%)=M(S%,R%+1)
3870 NEXT S%
3880 NEXT R%
3890 FOR S%=1 TO N%
3900 M(S%,N%)=M(0,S%)
3910 NEXT S%
3920 NEXT P%

3930 FOR R%=1 TO W%  A(1), A(2).... are the coefficients
3940 A(0)=0  of the polynomial.
3950 FOR S=1 TO N%  this section rearranges M so that repeating
3960 A(0)=A(0)+A(S%)*F(R%,0)!=S%-1)  the preceding sections will find another
3970 NEXT S%  coefficient.
3980 F(R%,2)=I%*5000*(F(R%,2)-A(0))/A(0)  calculate y values from the polynomial
3990 NEXT R%  for each existing x value.

4000 A(0)=0
4010 FOR S=1 TO N%
4020 A(0)=A(0)+A(S%)*F(R%,0)!=S%-1)  A(0) is the calculated y coordinate.
4030 NEXT S%
4040 F(R%,2)=I%*5000*(F(R%,2)-A(0))/A(0)  ΔR/R.
4050 NEXT R%
This is the routine for driving the X-Y plotter:

2620 INPUT"ENTER PLOT POSITION "P%  set vertical position on graph paper.
2630 X%=INT(F(1,0)*500)  calculate coordinates
2640 Y%=INT(3000-400*P%+40*F(1,2)) of first point.
2650 F(W%+1,0)=2*F(W%,0)-F(W%-1,0) extrapolate extra point to
2660 F(W%+1,2)=2*F(W%,2)-F(W%-1,2) avoid end effects.
2670 GOSUB 2840  pen to (X%,Y%).
2680 GOSUB 3010  2 second delay.
2690 ?&FD04=251  pen down.
2700 GOSUB 3010  2 second delay.

2710 FOR R%=1 TO W%
2720 G1=(F(R%+1,0)-F(R%,0))/(F(R%+1,0)-F(R%-1,0))
2730 G2=(F(R%+1,0)-F(R%,0))/(F(R%+2,0)-F(R%,0))
2740 C=3*(F(R%+1,2)-F(R%,2)-2*G1-G2) calculate cubic splines to
2750 D=-2*(F(R%,2)-2*F(R%+1,2)+G1+G2) give intermediate points.

2760 FOR F=0 TO 1 STEP.1
2770 X%=INT(500*(F(R%+1,0)-F(R%,0)) plot intermediate points
2780 Y%=INT(3000-400*F%+40*F(R%,2)+F%*C+F%*F%*D))
2790 GOSUB 2840  pen to new point.
2800 NEXT F

2810 NEXT R%  next data point.
2820 ?&FD04=255  pen up when finished.

Routine to move pen to (X%,Y%):

2840 IF Y%>4095 THEN Y%=4095  12 bits is 0 to 4095.
2850 IF Y%<0 THEN Y%=0

2860 H%=INT(X%/256)  digital to analogue converter must be
2870 ?&FC02=H%  given the 12 bits as 3 lots of 4 bits
2880 M%=INT((X%)/16)-16*H%  each, written to separate addresses.
2890 ?&FC01=M%
2900 L%=X%+16*M%-256*H%  now send the y value to the other
2910 ?&FC00=L%  d to a converter similarly.

2920 H%=INT(Y%/256)  signal the d to a converters to
2930 ?&FD02=H%  act on the new information.
2940 M%=INT((Y%)/16)-16*H%  
2950 ?&FD01=M%
2960 L%=Y%+16*M%-256*H%
2970 ?&FD00=L%

2980 ?&FC03=0  
2990 ?&FD03=0

3000 RETURN

The delay routine is very simple:

3010 FOR B%=1 TO 1000
3020 C=50
3030 NEXT B%
3040 RETURN
Appendix B: Electron distribution in a fet

This calculation relates to a fet with a uniformly doped channel being depleted from both sides (figure B.1). The centre is at l=0. The equation to be solved is Poisson's equation, \( \frac{d^2V}{dl^2} = \frac{e(N_d-N_e)}{\varepsilon} \).

For almost all combinations of doping and temperature \( N_e = N_d \times \exp(eV/KT) \) is a good approximation. The zero of potential is the Fermi energy far from the gate.

The equation is simplified by using the dimensionless variables \( x = l/l_D \), \( K = N_e(0)/N_d \) and \( C = e(V-V(0))/KT \), where \( N_e(0) \) and \( V(0) \) are the values at \( l=0 \). The equation then becomes \( \frac{d^2C}{dx^2} = 1 - K \times \exp(-C) \).

The boundary conditions are \( C(0) = 0 \) and \( dC/dx(0) = 0 \). The value of \( K \) may be chosen anywhere between 0 and 1, and this produces the different solutions. The numerical solution proceeds by calculating values of \( C \) at values of \( x \) separated by a small amount \( s \), starting from the known values at \( x = 0 \):

\[
X_n = s \times n
\]

\[
\frac{d^2C}{dx^2}\big|_{X_n+1} = 1 - k \times \exp(-C_n)
\]

\[
\frac{dC}{dx}\big|_{X_n+1} = \frac{dC}{dx}\big|_n + s \times \frac{d^2C}{dx^2}\big|_{X_n+1}
\]

\[
C_{n+1} = C_n + s \times \frac{dC}{dx}\big|_{n+1}
\]

The resulting \( K \times \exp(-C) \) (electron density) and \( X \) values are stored in memory until the end of the calculation, when they are written to disc as a data file compatible with the magnetophonon program. They can then be plotted.
As each value of C is calculated the approximate number of electrons in that interval, which is \( s \times K \times \exp(-C) \), is added to a total. This always overestimates the number of electrons (figure B.2). This is largely corrected for by subtracting \( s/2 \) from the total, which is equivalent to approximating the actual distribution by the trapezia instead of the rectangles.

This integration is stopped when the electron concentration becomes very low, or at the gate electrodes if it is desired to find the number of electrons present as a function of gate voltage. C gives the gate voltage relative to the centre and \( C - \ln(K) \) gives it relative to the distant Fermi level. The result of this for a device with a total width of 30 Debye lengths (typical for the devices and temperatures in section 3.5) is shown in figure B.3. The dimensionless pinch-off voltage is given by 
\[
\left( \frac{l_{\text{max}}}{L_D} \right)^2/2,
\]
here 112.5, at which the calculated dimensionless number of electrons is still 0.5. Some of this is calculation error however; putting \( K \) equal to zero should give \( C = 112.5 \) at \( X = 15 \), but the value obtained with a step length \( s \) of 0.1 is 115, where the dimensionless number of electrons is 0.2. Because of this, the appreciable uncertainty in both the Debye length and the device width, and the problem of variable trap occupancy under the gate, it was decided not to use the ratio of this curve to an actual device I-V characteristic to extract the electron mobility. On the other hand the program should be quite accurate when calculating the total number and overall distribution of electrons for a given \( K \) value.
FIGURE B.3
Electrons in the channel as a function of gate voltage.
Electron distribution program

Lines starting with REM (e.g. 330) are sometimes useful but not required most of the time. When desired they can be activated by removing the REM statement.

100 DATA .09, .17, .37, .605, .85, .979, .9996
110 DIM X(200)
120 DIM Y(100)
130 READ K
140 N%=0
150 T=0
160 C=0
170 B=0
180 S=.1
190 X=S*N%
200 A=1-K*EXP(-C)
210 B=B+S*A
220 C=C+S*B
230 REM PRINT X, INT(100*EXP(-C)), .01*INT(100*T)
240 X(100+N%)=3.8+.5*S*N%
250 X(100-N%)=3.8-.5*S*N%
260 Y(N%)=K*EXP(-C)
270 IF EXP(-C)≥.5 THEN D=X
280 N%=N%+1
290 T=T+S*K*EXP(-C)
300 IF EXP(-C)≥1E-3 THEN GOTO 190
310 REM IF X<7.6 THEN GOTO 190
320 PRINT "K=";K,"WIDTH=";D,"ELECTRONS=";T-.5*S*K,"POTENTIAL=";C-LN(K)
330 REM PRINT"K=";K,"WIDTH=";D,"ELECTRONS=";T-.5*S*K,"DELTA V=";C
340 PRINT
350 AS="DATA."
STR$(INT(.5*10*(T-.5*S*K)))
360 PRINT "WRITING ";AS
370 U%=OPENOUT AS
380 FOR P%=-N%+1 TO N%-1
390 PRINT#U%, X(100+P%), Y(ABS(P%))
400 NEXT P%
410 CLOSE#U%
420 GOTO 130
Appendix C: The Triangular Barrier Switch Simulation Program

This program is written in BBC Basic, as are all the others. The next three pages contain a listing of the program split into sections with different functions. The notes preceding each section describe the function of that whole section. The notes in italics by a particular line refer to that line only.

The greatest difficulty in implementing the equations in section 4.5 is the range of the quantities involved. Some equations have been put into a more complex form to avoid successive multiplications by small or large numbers, leading to under or overflow of the running products. Many quantities have a very large dynamic range. In some cases this has been dealt with by using alternative expressions in different ranges of the quantity, as in lines 970 to 990.

$P_0$ and $N_i$ varied enormously with temperature and were not amenable to this treatment. Calculations at low temperatures where $P_0$ and $N_i$ are small and $\exp(\beta V_j)$ correspondingly large were performed by noting that the expressions for $J_{rec}$, $J_{po}$ and $J_p(X_2)$ can be written in terms of the product $N_i \exp(\beta V_j/2)$. When temperature halves, $N_i$ decreases by a factor of $\exp(\beta V_j/2)$. If $N_i$ is instead kept constant the correct currents will be produced by a value of $V_j$ that is too low by $V_g/2$. The lowest temperature at which there is no overflow or underflow can thus be reduced by a factor of two by using the value of $N_i$ appropriate to twice the temperature and adjusting the values of $V_j$ obtained by $V_g/2$ where necessary (e.g. in the expressions for the total device voltage and for $W$). This can be extended if necessary, and is far more effective than multiplying quantities by constants to keep them within range and subsequently correcting for this.
This section contains and calculates the numerical constants, prints them as a table and writes the headings for columns of results.

100 REM TRIANGULAR BARRIER SWITCH SIMULATION
110 T=300
120 B=11594/T B = 1/thermal voltage
130 REM DEVICE PARAMETERS
140 d1=60E-9
150 d2=350E-9
160 Nt=5E24
170 Nd=5E23
180 Nt=2E16
190 L10=300E-9
200 L20=300E-9
210 REM MATERIAL PARAMETERS
220 t=2E-10 different lifetimes in barrier
230 tl=2E-10 and in P-N junction if desired
240 Dp=5E-4
250 Vg=1.424
260 Nc=4.7E23
270 Ni=1.60E24*EXP(-B*Vg/2)
280 m=.067
290 e=1.16E-10 absolute permittivity
300 REM OTHERS
310 Po=Nt*Nt/Nd
320 q=1.602E-19 electronic charge
330 Lp=SQR(Dp*t)
340 @%=&1030C
370 @%=10
380 PRINT
390 PRINT " Vj VL V2 V J Jrec Jpo Jp(X2)"
400 PRINT

This section looks for a solution for \( V_2 \) between \( VL \) and \( VH \).

410 FOR Vj=1.142 TO 1.200 STEP .001 start of main calculation
420 N=0
430 Vbi=(LN(Nd/Ni)+LN(Na/Ni))/B
440 W=SQR(2*e*(Vbi-Vj)/(q*Nd))
450 C=50E-3
460 VL=.05
470 VH=.1
480 PROC(VL) PROC(V) calculates \( S = V_2^* - V_2 \)
490 H=S
500 PROC(VH)
510 IF H=S<0 THEN GOTO 590 i.e. if a solution exists between VL and VH
If no solution is found then try the next interval, unless it is likely that all solutions have been found.

520 VL=VH
530 VH=C+VL
540 IF N=1 AND Vj<.05 THEN GOTO 820
550 IF N=0 AND VL≥1.4 AND Vj≥.05 THEN VH=10
560 IF N=1 AND (VH≤.15 OR VL≥5) THEN GOTO 820
570 IF N=2 THEN VH=10
580 GOTO 500

If a solution exists in the interval VL to VH then this section locates it accurately by successively halving the interval.

590 N=N+1
600 V2MAX=VH
610 V2MIN=VL
620 V2=(V2MAX+V2MIN)/2
630 PROC(S(V2)
640 IF ABS((V2MAX-V2MIN)<1E-3 AND S≥0 THEN GOTO 700
650 IF H*S≥0 THEN GOTO 680
660 V2MAX=V2
670 GOTO 620
680 V2MIN=V2
690 GOTO 620

When a solution has been found, this section calculates the currents and displays the results.

700 Jxrc=q*Ni*xW*(EXP(B*Vj/2)-1)/(2*t)
710 Jp1=q*DP/Lp*FNCOTH(L1/Lp)*(EXP(-B*V1)*Q/D-Po)
720 Jp2=q*DP*EXP(B*Vj/2)/(Lp*FNSINH(L2/Lp))*(Po*EXP(B*Vj/2)-Po*EXP(-B*Vj/2)
+FNCOSH(L2/Lp)*EXP(-B*Vj/2)*(Po-Q/D*EXP(-B*V2/2)*EXP(-B*V2/2)))
730 Jp0=q*DP*Po*EXP(B*Vj/2)*EXP(B*Vj/2)*EXP(L2/Lp)-Q*EXP(-B*V2)/D
/(Lp*FNSINH(L2/Lp))
740 Jp=PI/2*q*SQR(Nd)*SQR(Q/D)*(EXP(-B*V2/2)/(B*E2*t1)
+EXP(-B*V1/2)/(B*E1*t1)+1E-30
750 IF B*Vj>80 THEN GOTO 770
760 Jpo=q*DP*Po*((B*Vj-1)*EXP(L2/Lp)+1-Q*EXP(-B*V2)/(Po*D))
/(Lp*FNSINH(L2/Lp))
770 J=Jxrc+Jp0
780 PRINT INT(.5+1000*Vj),INT(.5+1000*V1),INT(.5+1000*V2),
INT(.5+1000*(Vj+V2-V1)),INT(.5+1000*LOG(Jrec)),
INT(.5+1000*LOG(Jp0)),INT(.5+1000*LOG(Jp2)),INT(.5+1000*LOG(Jp1)),
INT(.5+1000*LOG(Jp1))
790 PRINT "Q-";Q
800 H=H
810 IF N<3 AND VH<10 THEN GOTO 520
820 NEXT Vj
830 END

Look for next solution unless probably all found
This section performs the calculation described in section 4.5 to derive a value of $V_2^*$ from a guessed value of $V_2$. If they are the same, i.e. $S = V_2^* - V_2 = 0$, then the solution has been found.

840 DEFPROC(V2)
850 V1=V2
860 IF B*Vj<1E-5 THEN GOTO 890
870 A=1.333E-25*Nc/(m*T*Nd)*(N1*Nw/(2*t)*(EXP(B*Vj/2)-1)
     +Dp*Po/Lp*EXP(B*Vj/2)*EXP(B*Vj/2))
880 V1=-LN(A+EXP(-B*V2))/B 
     this ensures that $J_{rec} = J_{therm}$
890 E1=d1*q*Nd*(SQRT(1+2*e*V1/(q*Nd*d1*d1))-1)/e
900 E2=d2*q*Nd*(SQRT(1+2*e*V2/(q*Nd*d2*d2))-1)/e
910 D=(1/E1+1/E2)/B
920 L1=L10-e*E1/(q*Nd)
930 L2=L20-e*E2/(q*Nd)-W
940 Q=Po*EXP(B*V1/2)*D*EXP(B*V1/2) 
     these two lines apply if
950 IF B*V1>80 THEN GOTO 1020  
     Vj and hence $J \approx 0$
960 E=PI/2*q*SQRT(Nd)*((EXP(-B*V2/2))/(B*E2*t1)+EXP(-B*V1/2)/(B*E1*t1)) \approx J_n
970 F=q*Dp*EXP(B*Vj/2)*Po*EXP(B*Vj/2)/(Lp*FNSINH(L2/Lp)) \approx J_p(X_2)
980 IF B*Vj>80 THEN GOTO 1000  
     990 more accurate but would overflow
990 F=q*Dp*Po*(EXP(B*Vj)-1+FNCOSH(L1/Lp)+FNCOSH(L2/Lp))/(Lp*FNSINH(L2/Lp))
1000 G=q*Dp*(FNCOTH(L1/Lp)*EXP(-B*V1)+FNCOTH(L2/Lp)*EXP(-B*V2))/Lp \approx J_p(X_1)
1010 Q=D*(E/E+G/E+F*SQR(E/G/E+G4*F/G))/(2*G) gives $J_p(X_2) = J_p(X_1) + J_r$
1020 E2=q*W/(Nt-Q)+E1
1030 S=e*E2*(E2/(2*q*Nd)) + E2*d2 - V2
1040 IF E2<0 THEN S=-1 
     rejects false solutions
1050 ENDPROC

These functions are needed but not provided in BBC basic.

1060 DEF FNSINH(X)=((EXP(X)-EXP(-X))/2
1070 DEF FNCOSH(X)=((EXP(X)+EXP(-X))/2
1080 DEF FNCOTH(X)=FNCOSH(X)/FNSINH(X)
REFERENCES


BERGGREN K. and NEWSON D.J. (1986): to be published.


Attention is drawn to the fact that the copyright of this dissertation rests with its author. This copy of the dissertation has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author. In accordance with the Law of Copyright no information derived from the dissertation or quotation from it may be published without full acknowledgement of the source being made nor any substantial extract from the dissertation published without the author's written consent.