Some parts of this thesis may have been removed for copyright restrictions.

A STUDY OF FIELD EMISSION

FROM CADMIUM SULPHIDE

A THESIS SUBMITTED FOR THE DEGREE

OF PHILOSOPHIAE DOCTOR IN THE

UNIVERSITY OF ASTON IN BIRMINGHAM

-4.JUL72 152307

Leslie Thomas James Salmon, B.Sc.

Department of Phycics

May 1972

ABSTRACT

It was decided to investigate field emission from cadmium sulphide because many workers have found that the agreement between theory and experiment for this material, and other semiconductors, is poor. An electron energy analyser, similar to those used in most of the previously reported experiments, was, therefore, built. The performance of the analyser was thoroughly investigated both theoretically and practically and the results of these investigations were used in conjunction with a tungsten emitter. Excellent agreement was obtained between the usually accepted total energy distribution for tungsten and the corresponding distribution measured with the present analyser.

A method of obtaining reliable cadmium sulphide emitter was developed. These emitters were then used in the analyser and it was found that the agreement between theory and experiment was poor.

Previous explanations of the lack of agreement are considered and are found to be doubtful.

The theory of field emission from semiconductors is reviewed and possible reasons for the discrepancy between theory and experiment are proposed. Finally, further experiments are described which should prove or disprove the conclusions arrived at in this work.

o o 0 o o

INDEX

o o O o o

1.1 Introduction

"Field emission" is the name given to the emission of electrons from the surface of a metal or a semiconductor caused by the action of a high (typically 10^7 volt cm^{-1}) electric field. In addition to the importance of the field emission effect in other areas, for example in the study of electrical breakdown in vacuum, the field emission cathode is itself of great interest since in its usual form it can act as a point source capable of delivering a current density as high as 10^6 amp cm^{-2}. The topic has, therefore, attracted a great deal of experimental and theoretical research and, although it is only recently that a field emission cathode has been used successfully [1], the theory of field emission from metals has been well understood for many years. The same, however, cannot be said of the theory of field emission from semiconductors which, in spite of a great deal of work and interest, has not yet been proved satisfactorily.

1.2 Historical Background

1.2.1 Field Emission From Metals

Field emission undoubtedly occurred in many early investigations [2] where high voltages were used, but the first experimental study of the effect was not carried out until 1922. In that year Lilienfeld [3] was able to show that a field emission cathode could be activated by caesium but he was not able to achieve a stable emission current. His work did, however, lead to the first attempt at a theory of field emission, which was made by Schottky [4] in 1923.

Schottky assumed that the potential barrier at the surface of the metal was reduced by the applied field, thus allowing electrons to escape as shown in Fig. 1.1 . He showed that the lowering of the barrier, $\triangle\phi$, is given by the expression:

$$\triangle\phi = \epsilon\sqrt{eF} \qquad\qquad 1.1$$

Fig. 1.1

(a) BASIC POTENTIAL BARRIER
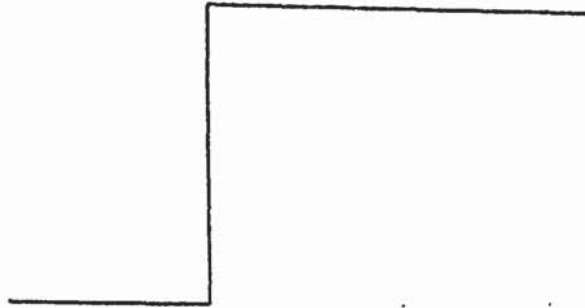
(b) POTENTIAL BARRIER
    WITH IMAGE FORCE
    CORRECTION
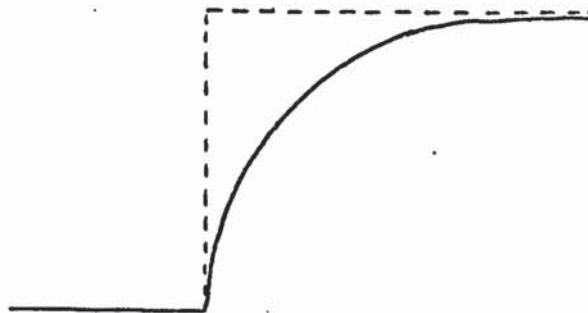
(c) REDUCTION OF BARRIER
    HEIGHT BY APPLIED FIELD

$\Delta\phi$

F
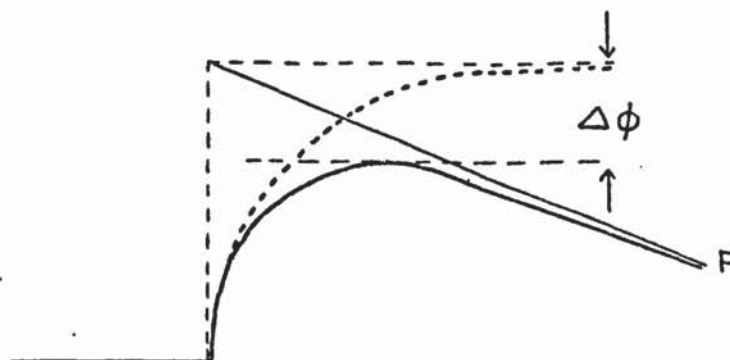
where e is the electronic charge and F is the field strength. The theory predicts a linear relationship between log i and V , i being the emission current density and V the applied voltage. Further, for tungsten, the surface barrier should be lowered considerably at a field of 1.4×10^8 volt cm^{-1} when an appreciable temperature effect should exist.

Lilienfeld [3] had already shown that large currents could be drawn from tungsten at a field as low as 10^7 volt cm^{-1}. This Schottky explained by assuming that the emission originated from "sub microscopic protruberances" which enhanced the local field by a factor of ten to a hundred.

In the years that followed, many investigators attempted to verify Schottky's theory. Thus, Gossling [5], by using the high field produced at an etched point, found no linear dependence of log i on V and no appreciable temperature effect from 300 degrees Kelvin (^{o}K) to $1000^{o}K$. Millikan and Eyring [6] using cylindrical wires of small diameter also found no temperature dependence up to $1000^{o}K$ and Piersol [7] found that the emitted current did not change when a field emission tube was immersed in liquid air.

An important discovery was made by Millikan and Lauritsen [8] who found that log i was linearly dependent on 1/F. They, therefore, suggested that the law governing field emission was of the form:

$$\log i = A \exp (-B/F) \qquad\qquad 1.2$$

where A and B are constants. They further suggested, in company with Gossling, that the discrepancies between theory and experiment could perhaps be explained on the grounds that the theory had been formulated in terms of classical mechanics rather than quantum mechanics. Fowler and Nordheim [9] produced the first theory of field emission based on quantum mechanics in 1923. They assumed that an electron could escape from a metal by tunnelling through the potential barrier rather than by going over it, as shown in Fig. 1.2., and by solving the

Fig. 1.2

(a) ESCAPE OVER THE POTENTIAL BARRIER
   BY THERMIONIC EMISSION

FERMI LEVEL
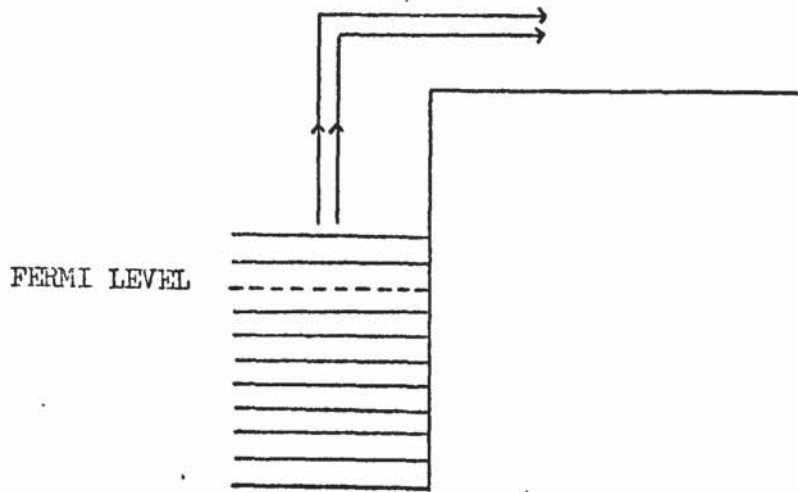
(b) ESCAPE THROUGH THE POTENTIAL BARRIER
   BY TUNNELING

FERMI LEVEL
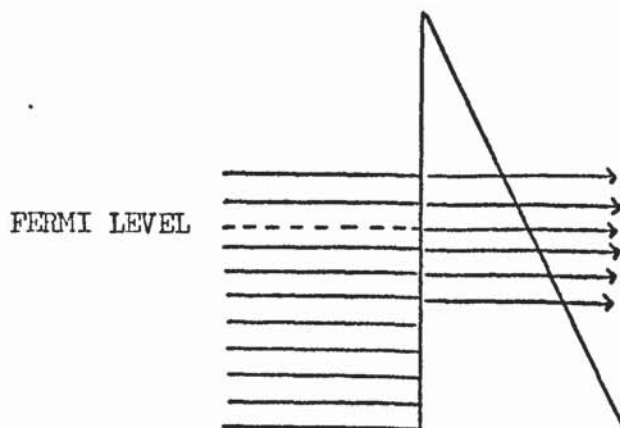
Schrödinger equation for a triangular barrier, they were able to find
the possibility of escape of an electron with any given energy. From
the Sommerfeld theory of metals they deduced the number of electrons
within a given energy range arriving at the barrier. Integration of
the product of these two quantities over all possible electron energies
led to an expression for the field emitted current density of the form:

$$J = \frac{4\sqrt{\mu\phi}}{\mu+\phi} \frac{e^3}{8\pi h\phi} F^2 \exp\left(-\frac{8\pi\sqrt{2m}}{3he} \frac{\phi^{3/2}}{F}\right) \qquad 1.3$$

where m is the mass of the electron, h is Planck's constant and μ and
ϕ are the Fermi energy and work function of the metal respectively.
The theory was further refined by Nordheim [10] to include the
deformation of the potential barrier by the image force.
The Fowler Nordheim formula predicts appreciable field emitted currents
at field strengths which are approximately five times lower than those
needed by the Schottky theory and, except for the experimentally
difficult to determine F^2 term, has the same form as the equation
suggested by Millikan and Lauritsen. The theory also shows that field
emitted electrons originate from levels at or near the Fermi level,
which explains why no temperature dependence had been observed.
Although many investigators attempted to verify the Fowler Nordheim
equation during the thirties, very little progress was made because it
proved impossible to produce a well defined cathode. It should also
be remembered that the field emitted current density depends on the
work function of the emitter which, in turn, is a function of the
cleanliness of the emitting surface. Ultra high vacuum conditions are
therefore needed for successful field emission work and in the thirties
this branch of physics was still in its infancy.
Beams [11], Quarles [12] and Moore [13] used the smooth surface of
liquid mercury as an emitter but found that the field necessary to
produce an arc discharge was well below that predicted by the Fowler

Nordheim equation, even after repeated distillation and the use of micro-second pulses to prevent surface distortion.

Other workers attempted to verify the relationship between work function and field emitted current. Ahearn [14], for example, used a thoriated tungsten cathode and measured its work function by thermionic means as he varied its activation. When he used the same cathode as a field emitter he found no dependence of the current on the work function and came to the conclusion that the current originated from the small protrusions which did not experience the same change in activation as that of the bulk material. A similar experiment was performed by Müller [15], who varied the work function of his etched tungsten emitter by bombarding it with electropositive substances. He found that the current was dependent on the third power of the work function instead of the expected 3/2 power and came to the same conclusion as Ahearn in explaining his results.

The existence of small protusions on the cathode was convincingly demonstrated by Wehnelt and Schilling [16] who showed, by means of an electron microscope, that the emitted current came from small, randomly appearing and disappearing spots along the edge of their emitter. The first well defined field emitter was obtained by Müller [17], who used his field emission microscope to show that, at high temperatures, surface migration of atoms transformed an etched tungsten point into a hemispherical emitter which was smooth down to atomic dimensions. He further showed, by measuring the radii of the emitters with an oil immersion microscope, that the fields so obtained were in good agreement with those predicted by the Fowler Nordheim equation. This work was extended by Haefer [18] who used an electron microscope to find the emitter geometries and hence to calculate the electric fields more accurately. He was able to prove the validity of the Fowler Nordheim equation to an accuracy of 15% and by coating a tungsten emitter with barium, caesium and potassium in a field emission microscope he

demonstrated the dependence of the field emitted current density on the 3/2 power of the work function.

1.2.2 The Energy Distribution Of Electrons Field Emitted From Metals

A further experimental proof of the Fowler Nordheim theory was attempted by Henderson and his co-workers [19][20] who tried to measure the energy distribution of field emitted electrons. In accordance with the theory they found that the electrons came from the Fermi level but the narrow distribution of energies above and below this level was beyond the resolution of their analyser.

It should be mentioned at this point that Henderson et. al. were trying to measure the normal energy distribution. The probability of escape of an electron depends only on the energy associated with its component of momentum normal to the potential barrier. It is these energy valves that make up the normal energy distribution. Richter [21] pointed out that energy could be transferred from the direction normal to the barrier to other directions by protrusions on the emitter and also by aberrations in the lenses used in the analyser, thus leading to errors in the measured distribution. Müller [22] tried to overcome these problems by using a very smooth emitter in conjunction with a carefully designed analyser in which electron lens aberrations were reduced to a minimum. In this way he obtained good agreement between the normal energy distribution and his experimental results.

It was not until many years later that Müller's results were challenged when Young and Müller [23][24], using a similar analyser, found energy distributions which were a third as wide as the normal energy distributions. To explain this anomaly, they suggested that their analyser was, in fact, measuring the total energy of the field emitted electrons instead of just the normal "component" of the energy. Young[23] reformulated the Fowler-Nordheim equation in order to include the energy associated with directions other than the direction normal to the potential barrier and obtained an expression for the total energy

distribution which was in good agreement with their experimental results.

1.2.3 Field Emission From Semiconductors

The study of field emission from semiconductors was a natural extension of the work on metals because the carrier statistics of a semiconductor, being Maxwellian, are a limiting case of the Fermi Dirac statistics found in a metal and also because the potential barriers at the surfaces of both materials are similar. There are, however, special effects associated with a semiconductor caused, in the first place, by the state of its surface and, secondly, to the fact that an applied electric field can have an appreciable penetration into its interior.

Until the mid fifties, very little experimental work on semiconductors had been reported, probably because of the difficulties experienced in obtaining materials and in preparing samples. Apker and Taft [25], for example, were able to produce emitters from needle-like crystals of cadmium sulphide by electron bombardment which, although very sharp, were too irregular in shape to yield useful field emission micrographs. They did, however, note the dependence of the field emitted current on illumination and that an appreciable fraction of the applied voltage could be lost along the emitter caused, they suggested, by its large ohmic resistance.

A new era of productivity started in 1955 when Stratton [26] published the first extensive paper on the theory of field emission from semi-conductors. He assumed a Fowler Nordheim type of barrier at the surface to which the image force correction had been added and modified the expression for the transmission probability in order to include the dielectric constant of the material. This barrier was then used in conjunction with two models of a semiconductor, one in which there were no surface states and another in which surface states were present. From the case with no surface states, Stratton showed that field penetration causes the conduction band to become degenerate at the

surface and derived an expression for the value of electric field at
which degeneracy occurs. This model was then extended in order to
produce equations which gave the field emitted current density as a
function of applied field with and without field penetration.
The second model was used to derive an expression for the field
emitted current density in the presence of surface states and showed
that the presence of these states causes an internal potential drop
at the semiconductor surface. At low fields, therefore, there is
little emitted current because of the low electron population in this
region. As the field is increased the internal potential barrier
decreases until at some critical field the conduction band becomes
flat, a situation similar to the case of no surface states and no field
penetration. On further increasing the electric field penetration
occurs and the conduction band becomes more and more degenerate until
at high enough fields the electron concentration at the surface of the
semiconductor is similar to that of a metal.
A plot of log i vs 1/F is not a straight line for a semiconductor with
surface states, but rather consists of two straight lines connected by
a curve. The Fowler Nordheim plot of such a material is, therefore, very
useful since not only does it show the presence of surface states, but
also it can be used to give an estimate of their density.
A great deal of experimental work followed from Stratton's investigation
into the theory of field emission from semiconductors. Most of the early
experiments were directed towards obtaining Fowler Nordheim plots which
deviated from the straight line behaviour of metals due to the presence
of surface states. Perry [27], and later Elison and Vasiliev [23],
produced Fowler Nordheim plots from germanium which showed departures
from linearity, but this characteristic could not be attributed
exclusively to the germanium because field emission patterns taken at
the same time suggested that their emitters were contaminated.

Perry [29] was able to obtain emission from clean silicon samples by using cleaning techniques similar to those employed by Allen [30] and D'Asaro [31]. These samples gave two kinds of results, one in which the Fowler Nordheim plot was linear and another in which the Fowler Nordheim plot consisted of two straight lines with approximately the same slope connected by a third straight line of slope different to the first two. Perry initially assumed that the different types of results were due to n type and p type material but this hypothesis was disproved when he found that both kinds of Fowler Nordheim plot could be obtained from n type and p type silicon.

Allen [30] found that very strong p type layers were present on the surface of silicon field emitters that had just been removed from a field emission microscope. This fact led Perry to suggest that linear Fowler Nordheim plots were due to these layers whereas the non linear Fowler Nordheim plots were due to surfaces which, although still p type, were not degenerate enough to exhibit metal-like behaviour. Further support was given to this argument when Perry found that linear plots could be obtained from samples which had been cleaned by heating, the strong p type condition then being due to diffusion of boron into the surface, and that non linear plots were obtained from emitters which had been cleaned by D.C. field desorption which caused a sudden removal of silicon atoms, thereby exposing a relatively uncontaminated surface. He was not, however, able to produce quantitative proof of Stratton's theory because he used the voltage applied to the emitter instead of the electric field as one of the parameters of his Fowler Nordheim plots. It should also be noted that the general shape of Perry's graphs did not agree with the theoretical shape predicted by Stratton.

A less qualitative proof of Stratton's theory was provided by Klimin [32] who investigated the temperature dependence of the field emission effect in germanium. He found that a plot of log i versus the reciprocal of

temperature consisted of two straight line segments with different slopes, the slope of the high temperature part being greater than the slope of the low temperature part. Further, the slope of the high temperature part was strongly dependent on the anode voltage and decreased as the latter increased.

In explaining these results Klimin used Stratton's [26] model of a semiconductor with no surface states. He supposed that at low temperatures, when impurity conduction prevailed, the Fermi level was near the bottom of the conduction band. The electron distribution at the surface, therefore, became degenerate even at low anode voltages, leading to an emission current which was virtually independent of temperature. At high temperatures it was supposed that intrinsic conduction caused the Fermi level to fall to approximately mid way between the conduction band and the valence band. Under these conditions the surface was far from degenerate and the emission current was, therefore, temperature dependent. This kind of emission became independent of temperature only at very high anode voltages when field penetration caused surface degeneracy again.

Klimin was also able to obtain numerical values of the band bending and of the corresponding electron densities by using Stratton's [26] equations. The experiments described above showed that Stratton's 1955 theory is at least complete enough to provide a qualitative picture of the processes that occur when electrons are field emitted from a semiconductor. It is, however, rather general and does not consider the more fundamental properties of a semiconductor, for example, its band structure variation. Stratton overcame some these limitations in a paper published in 1962[33]. He again used a Fowler Nordheim type of barrier and was able to produce expressions that not only gave the emitted current density from both the conduction and the valence band but also contained correction terms which could be applied to different band shapes.

The equations are rather difficult to solve analytically because they

involve integrals which contain the expression for the transmission probability. This difficulty was removed by expanding the transmission probability about the Fermi level, for positive values of the Fermi energy, and about the bottom of the conduction band for negative values of the Fermi energy. Both expansions lead to the same result for small Fermi energy values.

The first problem considered was that of emission from a conduction band containing free electrons. It was found that the shape of the Fowler Nordheim plot depends only on the value of the Fermi energy at the surface of the emitter. The equations were then modified to include the electron effective mass corresponding to a parabolic conduction band and by similar reasoning an expression for emission from a parabolic valence band containing holes was obtained.

Stratton also considered the effect of surface states in more detail. He showed that the value of the Fermi energy at the surface depends on the field due to the surface states which, in turn, is a function of both the applied field and the surface states density. The exact relationship between the Fermi level at the surface and the field due to the surface states had already been obtained by Siewatz and Green for germanium [34].

Two surface state configurations were then considered, one consisted of a set of acceptor levels of equal energy and the other of a set of acceptor levels with a continuous spread of energies. Previous work [35] had shown that these configurations could exist on a clean germanium surface. Stratton then proceeded to compute theoretical Fowler Nordheim plots for conduction and valence band emission from germanium and found that there was a change in slope at a field corresponding to the flat band condition. This change was more abrupt in the case of single acceptor energy levels then in the case of acceptor levels with a spread of energies. The results also showed that field emission from the valence band predominates over that from the conduction band in

germanium except when the density of surface states is sufficiently low (or the applied field is sufficiently high) to allow field penetration. The possibility of field emission of electrons from the actual surface state levels was not considered.

Stratton used Allen's [36] experimental results in an attempt to justify his theory because they were taken from an emitter which field emission microscopy had shown to be clean. Allen's Fowler Nordheim plots were in the form of log I vs 1/V, where I is the total current and V is the applied voltage. It was found that the experimental curve and the theoretical curve could be expressed in the same form, further, by making reasonable assumptions concerning the emitter geometry, Stratton was able to show that emission from a conduction band with a continuous set of acceptor energy levels was slightly more plausible than valence band emission, indicating that field penetration must have occurred. Allen's results were definitely inconsistent with emission from a conduction band containing a set of acceptor states at the same energy level.

A similar comparison with Klimin's experimental results again showed that emission from the conduction band was more plausible than emission from the valence band. Further proof of this argument was supplied by the fact that the total current increased with temperature. Unfortunately the values of the parameters used in the comparison proved to be smaller that those predicted by theory, but Stratton showed that this could be due to the presence of an oxide layer which would reduce the tunnelling probability. Application of the above arguments to the work of Elison and Vasiliev [28] led to the same conclusions and it was found that the comparison parameters had similar values to those in Klimin's work. In both cases the experimental conditions and the state of the emitter were similar.

The results described in the predeeding paragraphs again only provide a qualitative proof of Stratton's theory. A more rigorous proof

necessitates a knowledge of the field at the emitter and the current density. An investigation of the theory on these lines was carried out by Ernst [37] who used field evaporated germanium samples to ensure surface cleanliness.

Ernst proceeded in the following way: before taking results from any particular emitter he determined the voltage necessary for evaporation of the germanium, then, if the field necessary for evaporation was known, the constant relating the two could be determined. This constant was then used to convert subsequent voltage readings into field values. The field necessary for evaporation had already been determined by Arthur [38] who assumed that the field necessary for an optimum hydrogen ion image of germanium was the same as that required for an optimum hydrogen ion image of tungsten. Since the latter was already known, by measuring the voltage necessary for the optimum germanium hydrogen ion image, he was able to determine the constant relating field and voltage. This constant was then used to determine the field necessary for evaporation from the corresponding voltage. The current density was obtained from the total current by using the area from which the emission occurred. This area was determined from field emission micrographs of the emitting samples.

Ernst compared his results with a theoretical Fowler Nordheim plot which he had obtained from Stratton's work by assuming that the emission came from a degenerate conduction band. A single surface acceptor level was also included since at that time it was thought to be the most probable surface state configuration.

He found that the measured current density for a given applied field was much smaller than the corresponding calculated value and that the work function, which had been determined from the slope of the experimental curves, was larger than the usually accepted value for germanium. Ernst suggested that these discrepancies were caused either by limitations in Stratton's theory or by anisotropy in the emission, but he was not able

to differentiate between these two possible reasons.

The validity of Stratton's theory, therefore, seems to be somewhat doubtful. It especially fails to predict the emission characteristics of semiconductors which have relatively few electrons in their conduction band, a condition which occurs in p type material and high resistivity n type material of large band gap. These emitters can give rise to extremely non linear emission characteristics and usually exhibit a saturation region where the emitted current no longer increases with the applied voltage [39][40]. Various attempts [41][42] have been made to explain this phenomenon but it was not until fairly recently that a successful theory was produced by Baskin [43] et. al. They, unlike Stratton [33], did not restrict their theory to a Fermi level which was constant up to the surface of the emitter and were thus able to show that a depletion layer can form on a p-type semiconductor which limits the rate of carrier arrival at the surface, therefore leading to a saturation region in the emission characteristic. A similar effect can occur in the high resistivity n type materials when the bulk of the emitter is not able to supply a sufficient number of carriers to the surface region. Numerical analysis of their equations showed that the theory at least partly explained the phenomenon and they were able to obtain estimates of the width of the depletion layer for various semiconductors.

1.2.4 The Energy Distribution Of Electrons Field Emitted From Semiconductors

The probability that an electron will escape from the emitter depends only on the value of the "component" of energy in the direction normal to the emitter surface. Each of these components has a tangential component associated with it and together they form the total energy of the electron. It therefore follows that for any given value of this energy there must be a range of escape probabilities. In deriving his expression for the emitted current density, Stratton [33] used a double integral, the first integral was performed over the range of normal

- 13 -

components corresponding to a given energy and the second was performed over all possible values of the electron energy. By considering the first integral only, Stratton [44] was able to obtain an expression for the number of electrons escaping with one particular value of total energy i.e. the total energy distribution. He also included a correction term because, in a semiconductor, the values of the normal components depend on the shape of the energy bands.

The integration was carried out by expanding the tunnelling probability about an arbitrary energy following the work of Good and Murphy [45]. Since the tunnelling probability is a very sensitive function of energy, the total electron energy was used in the final expression because most of the emission comes from electrons with normal components near this value. The final equation, however, still contains the correction term. Fischer [46][47] also obtained an expression for the total energy distribution from a degenerate and a non-degenerate conduction band by expanding the tunnelling probability about the Fermi level and the bottom of the conduction band. Stratton showed that his equations reduced to the same form as Fischer's, but he used Fischer's in the remainder of his paper because they were more convenient. He was able to find an analytical expression for the energy at which the peak of the distribution occurs but was only able to find an expression for the half width corresponding to low temperatures. It proved impossible to obtain an expression for the half width in the case of non-degenerate emission. The problem of emission from the valence band was then considered. Fischer [46][47] had also derived an expression for this case, but it was in error since it had been based on an incorrect equation originally published by Stratton [33]. By proceeding in a way analogous to that of conduction band emission, Stratton was able to obtain expressions for emission from a valence band with degenerate and non-degenerate hole distributions. An equation for the energy corresponding to the peak of the distribution was also found.

Numerical analysis of the above equations predicted conduction band half widths of 2 KT to 7 KT and the valence band half widths of 3 KT to 12 KT for all degeneracies and for the typical range of fields used in practice, where K is Boltzmann's constant and T is the ambient temperature of the emitter in degrees Absolute. It will be seen that experimental results from a range of semiconductors have so far produced half widths in excess of these values.

Russell [48] used an analyser based on the design of Young and Müller [24] to investigate the total energy distribution of electrons field emitted from silicon. He first calibrated the analyser by using a tungsten emitter and was then able to show that the highest energy electrons originated from energy levels which were less than half the band gap below the Fermi level of his sample. This result was consistent with earlier work [49] which showed that silicon emitters acquire a p type surface when used in pyrex vacuum systems due to boron contamination. The valence band, therefore, bends upwards and its "distance" from the Fermi level at the surface is less than in the bulk material.

As the anode voltage was increased further, a larger value of retarding potential was needed before collection of the high energy electrons occurred. Russell suggested that this effect was due to the ohmic potential drop along the emitter.

It was found that the energy distributions were similar to those of a metal except that the onset of collected current occurred at a larger value of retarding potential. At very high applied voltages the sample fractured and the energy distribution then contained two peaks. Russell [50] assumed that this was due to emission from the valence band and the conduction band since the energy separation of the peaks had a value similar to that of the band gap of silicon. Although this hypothesis was based on an n type surface whereas the results of the previous paper were based on a p type surface the two are not incompatible since the breaking of the emitter could have led to the removal of the boron contamination. Russell's interpretation of his results does, however,

seem doubtful when the resolution of the analyser is considered. The calibration experiment with tungsten gave a half width of 0.6 eV whereas the "correct" value [24] is approximately 0.23 eV. It therefore seems improbable that Russell's analyser could resolve the band gap with such precision. The half widths of both peaks were far greater than the values predicted by Stratton's theory [44].

Shcherbakov and Sokol'skaya [51] used a similar kind of analyser in their work on cadmium sulphide. The calibration was again performed with a tungsten emitter and, although the shapes of the distributions were somewhat different, the half width of the distribution they obtained agreed fairly well with that reported for tungsten by Young and Müller [24]. The total energy distribution of electrons field emitted from cadmium sulphide was found to have a half width of approximately 0.6 eV, which is again much larger than the value predicted by Stratton [44]. The half width was also found to increase with the anode voltage. This phenomenon was explained in terms of the ohmic potential drop along the emitter. Previous work [40][52] had suggested that electrons could gain energy and, therefore, become hot under the action of the electric field caused by this potential drop, leading to emission from a wider range of energy levels and a consequent broadening of the energy distribution.

The existence of hot electrons was inferred from Fowler Nordheim plots in which evidence of carrier generation had been observed. This conclusion must, however, be treated with caution since the electron emission occurred from several "micro tips" instead of one well-defined emitter. Although this condition may not have affected the energy distribution, which was taken from one "micro tip", it may have affected the Fowler Nordheim plots in which current from the whole emitter, and presumably several "micro tips", was measured.

The total energy distribution was also found to broaden with temperature. Although the emitter was heated by the passage of emission current,

Shcherbakov and Sokol'skaya claimed that this effect was not the cause of the increase in half width with anode voltage because the total current never exceeded 10^{-7} Amp., and previous work had shown that with currents of this magnitude appreciable emitter heating did not occur.

The first direct comparison between theory and experiment was performed by Arthur [53] shortly after the publication of Stratton's paper. He used the results of that work in order to produce theoretical total energy distributions for germanium corresponding to degenerate and non-degenerate conduction and valence band emission. The equations showed that the current densities from the conduction and valence bands were of comparable magnitude only when the surface was practically non-degenerate and that when the surface became degenerate, only one peak could be expected because emission from the conduction band then predominated. Since previous work [54] had shown that field penetration could occur even in p type germanium emitters, the distribution corresponding to a degenerate conduction band was used for comparison with the experimental results.

The total energy distributions were again measured with a Young and Müller analyser [24] which was calibrated with a tungsten emitter and thereby shown to have good resolution. The distributions from field evaporated samples of all dopings proved to be approximately Gaussian in shape with half width values approximately twice that of tungsten and in all cases became narrower after the emitters were annealed. The higher resistivity p type samples gave rise to an extra peak which disappeared with annealing. Although the half widths of the measured distributions were again wider than those predicted by theory, being approximately 0.4 eV instead of 0.2 eV, Arthur concluded that emission from a degenerate conduction band had, in fact, taken place since the emission had originated from energy levels just below the Fermi level and no temperature dependence had been observed. He noted, however,

that this conclusion may not be strictly true since there was evidence
of electron emission from energy levels well below the Fermi level and,
even for the case of no surface states when field penetration should
be at its maximum, Stratton's theory [44] showed that the conduction
band should not dip more than approximately 0.3 eV below this level
with the values of electric field used in the experiment.

The half widths of the distributions and the presence of the second
peak were very dependent on the state of the surface. It would there-
fore seem that the second peak and the emission from levels well below
the Fermi level could be due to surface states associated with surface
imperfections. Although this hypothesis is very attractive, Arthur
points out that it is not tenable because in the first place, the
existence of degeneracy places a limit on the number of possible surface
states which is too low to support sufficient state emission and secondly,
the adsorption of approximately one layer of oxygen, which would
presumably have affected the surface state distribution, had very little
effect on the energy distributions.

Similar effects were observed by Hughes and White [55] in their work on
gallium arsenide emitters which had been cleaned by d.c. field desorption
in the presence of hydrogen. [56] In order to obtain good agreement
between the theoretical total energy distribution for tungsten and the
distribution measured by their Young and Müller [24] analyser, Hughes
and White used the latter's "empirical correction formula" because their
uncorrected distributions were far narrower than expected, a result
which indicates the presence of appreciable secondary electron [24]
effects in their analyser.

Critical bias measurements showed that the emission originated from
energy levels below the Fermi level and therefore led Hughes and White
to believe that they were observing valence band emission. They also
found that, although the total energy distributions usually consisted
of one peak, a second peak would sometimes appear when the emitter was

carefully positioned.

In explaining their results, Hughes and White used Stratton's [44] theory containing the relevant gallium arsenide parameters to produce a theoretical total energy distribution corresponding to valence band emission. They found that the half widths of the measured distributions were again larger than the theoretical value, being 0.7 eV and 0.3 eV respectively, and suggested that this discrepancy could be caused in part by the oversimplified band structure configuration used in Stratton's theory [44]. The appearance of the second peak was assumed to be caused by emission from surface states with energy levels within the forbidden gap. It was thought that the surface state density in gallium arsenide was large enough for this to occur since there had been no evidence of degeneracy.

The work reviewed above shows that, although the theory of field emission from semiconductors has been quite successful in providing a qualitative explanation of the experimental results obtained so far, a quantitive explanation is still outstanding. Three main reasons have been suggested to account for this, namely, limitations in the theory itself, hot electron effects and emission from surface states. There is also the further possibility of experimental errors. Most investigations have used equipment based on the design of Young and Müller [24] and nearly all the early work was performed with apparatus which was of low resolution. In fact, it is only fairly recently that good agreement has been obtained with their results for tungsten in an analyser which was subsequently used to investigate field emission from a semiconductor. In view of this it was decided not only to design and build a Young and Müller analyser, but also to carry out a fairly extensive investigation of its characteristics. Cadmium sulphide was chosen for the energy distribution experiments because it exhibits at least one of the reasons suggested for the broadened energy distributions, namely that of hot electron emission, and also because experience in the use of this

material for field emission had previously been gained [57].

o o 0 o o

CHAPTER 2

The Energy Analyser

2.1 The Collection Mechanism

Electron energies will be measured with respect to the "vacuum level" which is the energy level of an electron at rest just outside the emitter. [58] Consider the case of field emission from a metal. Fig. 2.1 shows the equilibrium condition when the emitter and the collector have a common Fermi level. An electron leaving the emitter with an energy of E eV below the vacuum level will not be collected because its energy is insufficient to surmount the potential barrier at the collector.

Suppose the collector is made V volts positive with respect to the emitter as shown in Fig. 2.2 . The electron can now be collected provided:

$$|E| = \left(\phi_e + V - \phi_c \right) e \qquad\qquad 2.1$$

where ϕ_c is the work function of the collector and ϕ_e is the work function of the emitter. By varying the value of V, therefore, electrons with different energies may be collected and in particular, when $V = \phi_c$, the collector will register electrons from the Fermi level of the emitter. Let $N(E)dE$ be the number of electrons leaving the emitter per unit area per unit time with total energies in the range E to E + dE. Then the collected current, I_c , is given by:

$$I_c = e \int_{-E}^{\infty} N(E)\, \partial E \qquad\qquad 2.2$$

where e is the electronic charge. The total energy distribution, N(E), is therefore equal to $\frac{dI_c}{dE}$, and if the collector is V volts positive with respect to the emitter, it follows that:

Fig. 2.1

EMITTER AND COLLECTOR IN EQUILIBRIUM
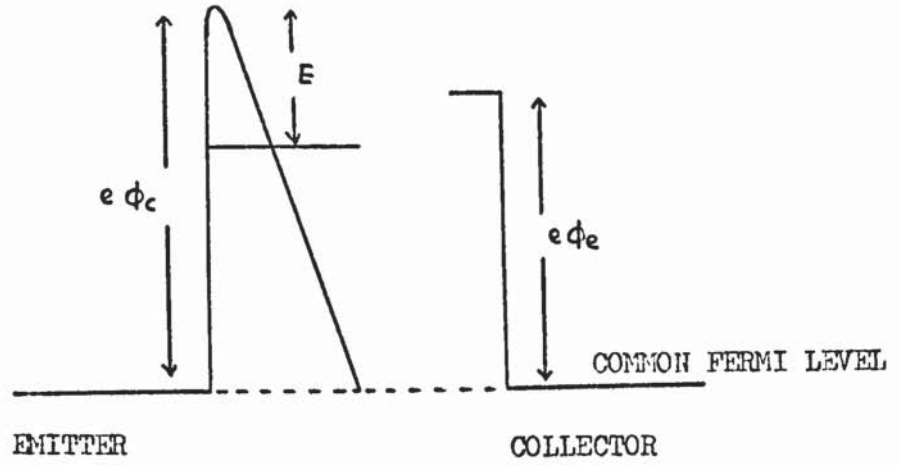
Fig. 2.2

COLLECTOR V VOLTS POSITIVE WITH
RESPECT TO THE EMITTER.
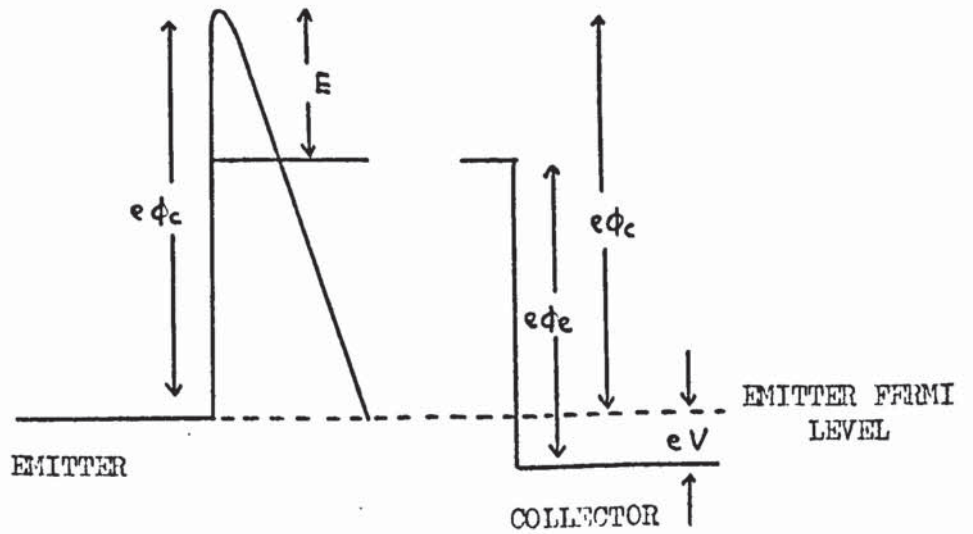
$$N(E) = \frac{d i_c}{d(\phi_e + V - \phi_c)} \qquad 2.3$$

Thus, provided the work function of the collector and of the emitter remain constant:

$$N(E) = \frac{d i_c}{d V_c} \qquad 2.4$$

and the slope of the collector current vs collector-emitter potential difference curve is a measure of the total energy distribution.

2.2 The Importance Of The Electron Trajectories

Any practical electron energy analyser, in fact, measures the momentum associated with the particle and from this value of momentum the energy is inferred. It is therefore important that the electrons should arrive at the collector normally, otherwise a smaller component of their momentum is available to overcome the potential barrier which exists there. If normal incidence does not occur, an electron will be collected at a different value of emitter-collector potential difference than that corresponding to its total energy and it follows from equation 2.4 that the measured total energy distribution will then be in error. The resolution of any electron analyser is therefore very dependent on the trajectories followed by the electrons whose energies it is attempting to measure.

2.3 The Design Of The Analyser

Fig. 2.3 shows a schematic diagram of the analyser. T is the field emitter, C the collector, A the anode and X1, X2 are apertures through which the field emitted current reaches the collector.

Electrons leaving T are accelerated up to A and then are de-accelerated between A and C. The energy they gain due to acceleration in the emission field is therefore lost in the retarding field so that they arrive at the collector with the same spread of energies with which

Fig. 2.3

T >   X₁ A X₂   C

ANODE   COLLECTOR

Fig. 2.4

B

E

A

EMITTER   COLLECTOR

——— ANODE
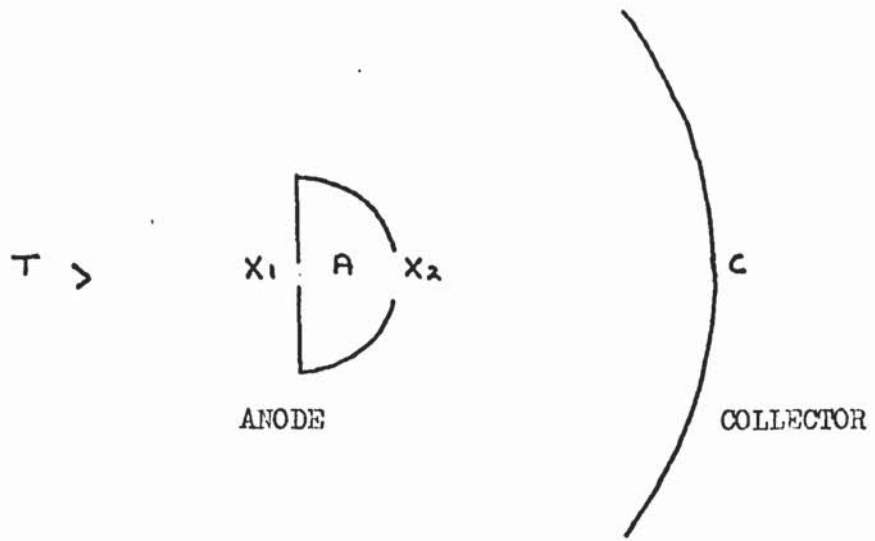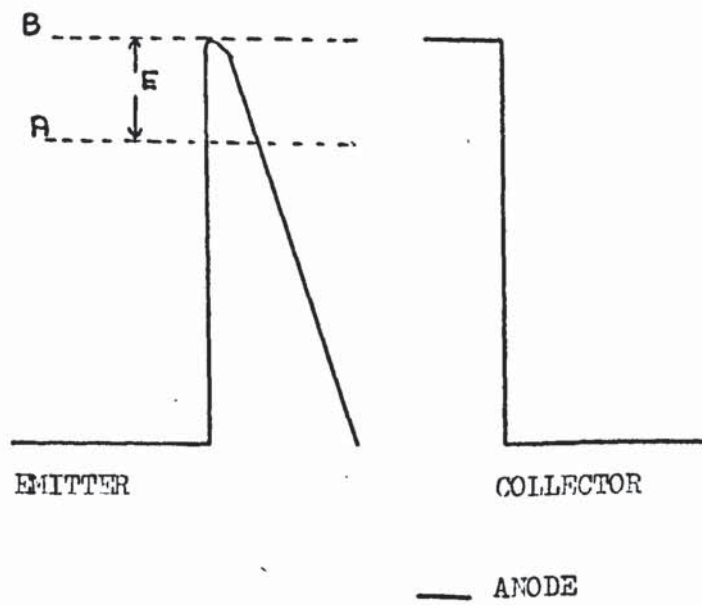
they left the emitter. This distribution of energies is then measured as explained previously. It is usual to keep the collector at earth potential and to make the emitter negative with respect to it. In this way the static fields and equipotentials in the retarding region remain constant.

The electrons must travel along field lines to impinge on the collector normally and must therefore come, or appear to come, from the centre of the spherical collector. In general they do not follow field lines in the retarding region but instead have elliptical paths. [59] The field emitter can, however, be imaged at the centre of the collector if use is made of the lens effect associated with aperture X2.
It can be shown [60] that an aperture of diameter d in an electrode held at a potential VO has a focal length f given by:

$$f = \frac{4\ V_C}{E_1 - E_2} \qquad 2.5$$

where E_1, E_2 are the electric field strengths on either side of the aperture and VC, the potential at the centre of the aperture is given by [61]:

$$V_C = VO - \frac{d}{2\pi}(E_1 - E_2) \qquad 2.6$$

These formulae are strictly applicable to plane parallel electrodes only and further, VO/d must be greater than E_1 and E_2.
The value of E_2 can be set equal to zero since it is the field strength to the left of X_2 in Fig. 2.3 and this region is virtually surrounded by the equipotential surface of the anode. E_1 can be estimated from simple electrostatic theory by using the relationship:

$$E(r) = \frac{V}{r^2}\ \frac{ab}{b-a} \qquad 2.7$$

where $E(r)$ is the field strength at a radius r between two concentric spheres of radii a and b (b greater than a) with a potential difference

of V between them.

A 5-litre glass flask was chosen for the collector because of its accurate sphericity and because its volume was compatible with the requirements of the vacuum system. This leads to a value of 10.6 cm for b. A convenient size for a proved to be 4 cm and by taking d equal to 1 cm the condition that VO/d should be greater than E_1 and E_2 was fulfilled. By using these values in equations 2.5, 2.6 and 2.7, the focal length of aperture X_2 was found to be 9.3 cm.

In order that the emitter be imaged at the centre of the collector it is necessary to place it a distance u from the aperture X_2, where u is given by:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \qquad\qquad 2.8$$

The image distance, v, and the focal length, f, are both negative because of the divergent nature of the lens. By using the results of the previous paragraph, u was found to equal 7 cm, therefore, the emitter was placed 3 cm to the left of aperture X_1.

2.3.1 Lens Effect Of Aperture X_1

The field strength to the right of aperture X_1 can be set equal to zero for reasons previously mentioned and the field strength to the left of the aperture can be estimated in the following manner:

Imagine the emitter to be a small sphere of radius b placed a distance a from the plane electrode in which aperture X_1 is situated. It can be shown by the method of electrostatic images [62] that the capacitance of this sphere is given by:

$$C = 4\pi\epsilon_0 b \left(1 + m + m^2 + m^3 + \ldots \right) \qquad\qquad 2.9$$

where Eo is a constant and m = b/2a. For a field emitter, b is typically 1000 Angstroms or less, therefore, m approximately equals 2×10^{-6}. and:

$$C \doteq \frac{4\pi \cdot 2 \times 10^{-6}}{36\pi \times 10^9} \quad \text{Farad.} \qquad\qquad 2.10$$

C also equals Q/V, where Q is the charge on the sphere and V is the

potential difference applied between it and the electrode.

The field strength $F(r)$, a distance r from the centre of the sphere is given by the expression:

$$F(r) = \frac{Q^2}{4\pi\epsilon_0 r^2} \qquad 2.11$$

By substituting the values relevant to the problem into equations 2.9, 2.10 and 2.11, a value of 10^{-14} volt cm^{-1} is obtained for the field strength to the left of aperture X_1. It follows, therefore, from equation 2.5 that this aperture has negligible lens effect and acts purely as a stop.

The above results were used in building the analyser. A method of varying the distance from the emitter to the first aperture was included to allow for errors in the calculated value of this distance caused by the approximate nature of the equations used.

2.4 Factors Affecting The Resolution Of The Analyser

There are four main causes of poor resolution, namely, patch effect on the collector, work function variation over the surface of the emitter, stray electric and magnetic fields and aberrations in the electrostatic lens. At this point the lens aberration effect alone will be mentioned since the other effects have either already been dealt with or are external to the analyser.

The two main aberrations which limit the resolution of the analyser are spherical aberration and chromatic aberration. They are inherrent to the system whereas others may be eliminated by careful positioning of the emitter and accurate fabrication of the analyser components.

2.4.1 Spherical Aberration

This form of aberration can theoretically be reduced to zero by making aperture X_1 as small as possible. Unfortunately, the total current available for measurement also falls as the diameter of the aperture is reduced so that a compromise must be reached.

2.4.2 Chromatic Aberration

Consider Fig. 2.4 which again shows the case of emission from a metal. In order to clarify matters the contact potential difference between the emitter and the collector is shown equal to zero i.e. $\phi e = \phi c$. Electron B, whose energy corresponds to that of the vacuum level, will fall through a potential of VA volts after leaving the emitter thus acquiring a speed at the anode corresponding to this voltage. Electron A, which is energetically lower than electron B, will fall through a potential of VA-VE volts, where eVE = E. This electron will therefore arrive at the anode with a different speed to electron B and, because of the chromatic aberration in the lens, will follow a different trajectory to that of electron B in the region between the anode and the collector.

When a potential difference is applied between the collector and the emitter in order to collect electrons A they are, in effect, being "pre-energised" so that they will have sufficient energy to overcome the potential barrier at the collector when they arrive there. The number of electrons corresponding to energy E is measured by the differential increase in current when the potential difference is applied. The original collected current was due to electrons B at the vacuum level. The energy of these electrons is also increased by the applied potential difference so that, by the arguments of the previous paragraph, although they may have been arriving normally at the collector when at their original energy, they may not be doing so now. It follows that although they may now possess more than enough energy for collection, their trajectories may be so modified that this does not in fact happen. The increase in current due to electrons A may, therefore, be offset by a decrease in current due to non collection of the higher energy electrons leading to errors in the measured value of N(E). This mechanism may explain the decrease in collected current at high values of collector-

emitter potential difference reported by Müller [24] which he attributes to the production of secondary electrons.

Suppose that a contact potential difference now exists between the collector and the emitter. It follows that all electrons must be brought to an energy of $|\phi_e - \phi_c|$ units below the vacuum level in order to be collected whereas, in the case of zero contact potential difference, it was necessary to bring them to the vacuum level. If the analyser is adjusted so that, for example, the trajectory of minimum aberration corresponds to the zero contact potential case, it is not necessarily correctly adjusted for the finite contact potential difference case. The optimum collection condition must, therefore, be different for emitters of different work functions.

Because the resolution of the analyser is very dependent on the electron trajectories it was decided to determine them as accurately as possible. This necessitated the solving of Laplace's equation for the system. The solution was obtained numerically since an analytical solution, based for example on the method of Spherical Harmonics [63], would have contained truncated infinite series and, therefore, would have been no more accurate.

2.5 The Numerical Solution of Laplace's Equation

The numerical equations used in obtaining the solution are a slightly modified version of those published by Gallaway et. al.[64] Their main advantage is that they permit the use of unequally spaced mesh points thus allowing curved boundaries to be accurately specified. The conversion from the mesh notations used by Gallaway et. al. to that used in the present work is shown in Fig. 2.5 and the parameter describing a variation in permittivity can be ignored because the system to be solved is under vacuum. For each mesh point there is an equation of the form:

$$\phi_1 A_1 + \phi_2 A_2 + \phi_3 A_3 + \phi_4 A_4 - \phi_0 A_0 = 0 \qquad 2.12$$

Fig. 2.5
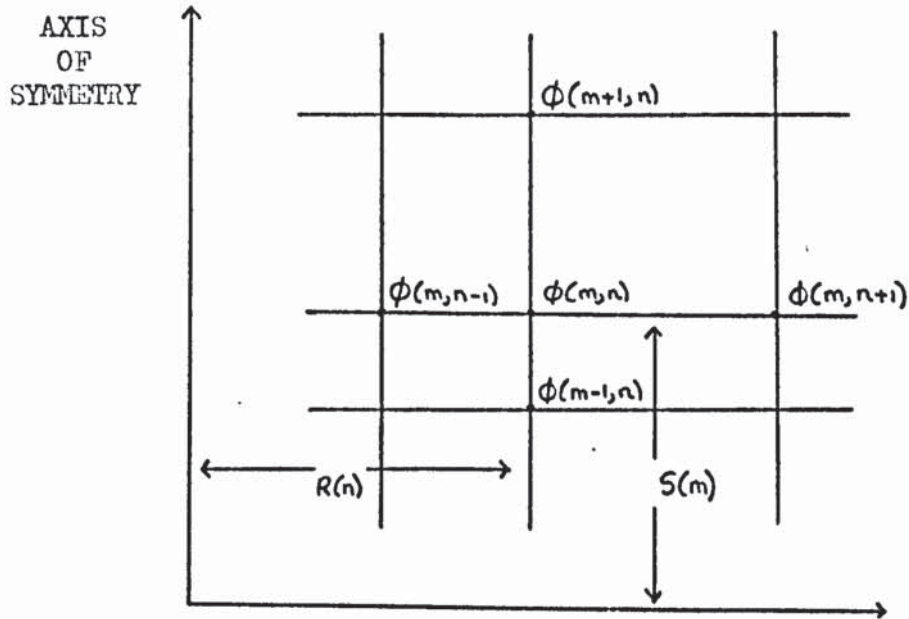
MESH NOTATION

$$h_1 = S(M) - S(M-1)$$

$$h_2 = R(N+1) - R(N)$$

$$h_3 = S(M+1) - S(M)$$

$$h_4 = R(N) - R(N-1)$$

WHERE THE h_n ARE THE CONSTANTS
USED BY GALLOWAY

Fig. 2.6

POINTS ALONG AXIS OF SYMMETRY
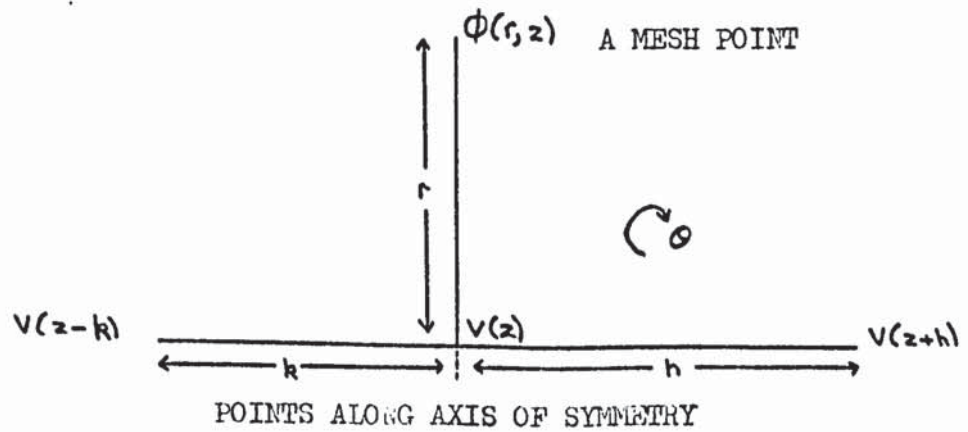
where Φ_1 ..... are the values of potential at different points in the mesh and A_1 ..... are constants which are given by the equations:

$$A_1 = \frac{1}{[S(m)-S(m-1)]} \left[[R(n)-R(n-1)] \left(1 - \frac{[R(n)-R(n-1)]}{2R(n)} \right) \right.$$

$$\left. + [R(n+1)-R(n)] \left(1 + \frac{[R(n+1)-R(n)]}{2R(n)} \right) \right] \qquad 2.13(a)$$

$$A_2 = \frac{1}{[R(n+1)-R(n)]} \left[[S(m+1)-S(m-1)] \left[1 + \frac{[R(n+1)-R(n)]}{2R(n)} \right] \right] \quad 2.13(b)$$

$$A_3 = \frac{1}{[S(m+1)-S(m)]} \left[[R(n)-R(n-1)] \left(1 - \frac{[R(n)-R(n-1)]}{2R(n)} \right) \right.$$

$$\left. + [R(n+1)-R(n)] \left(1 + \frac{[R(n+1)-R(n)]}{2R(n)} \right) \right] \qquad 2.13(c)$$

$$A_4 = \frac{1}{[R(n)-R(n-1)]} \left[[S(m+1)-S(m-1)] \left[1 - \frac{[R(n)-R(n-1)]}{2R(n)} \right] \right] \quad 2.13(d)$$

$$A_0 = \left[1 - \frac{[R(n)-R(n-1)]}{2R(n)} \right] \left[[R(n)-R(n-1)] \left[\frac{1}{[S(m)-S(m-1)]} + \right. \right.$$

$$\left. \frac{1}{[S(m+1)-S(m)]} \right] + \frac{[S(m+1)-S(m-1)]}{[R(n)-R(n-1)]} + \left[1 + \frac{[R(n+1)-R(n)]}{2R(n)} \right]$$

$$\left[[R(n+1)-R(n)] \left[\frac{1}{[S(m)-S(m-1)]} + \frac{1}{[S(m+1)-S(m)]} \right] + \right.$$

$$\left. \left. \frac{\left[\, S(m+1) - S(m-1)\,\right]}{\left[\, R(n+1) - R(n)\,\right]} \right] \right] \cdot \qquad \qquad 2.13(e)$$

The set of equations which represents the points of the mesh can only be solved if the boundary of the mesh is completely specified. It therefore follows that in an axially symmetric system, the values of potential along the axis of symmetry must be known. Since these are the values needed in order to calculate the electron trajectories and are, therefore, unknown, another set of equations is needed which will relate the axis of symmetry values to those of the mesh.

2.5.1 Equations For Points Along The Axis Of Symmetry

Consider Fig. 2.6. $V(z - k)$, $V(z)$ and $V(z + h)$ are points along the axis of symmetry. $\phi\,(r, z)$ is a mesh point and therefore has an equation which contains $V(z)$.

By Taylor's theorem [65]:

$$V(z+h) = V(z) + h\dot{V}(z) + \frac{h^2}{2}\ddot{V}(z) + . \qquad \qquad 2.14$$

and

$$V(z-k) = V(z) - k\dot{V}(z) + \frac{k^2}{2}\ddot{V}(z) + \qquad \qquad 2.15$$

where $\dot{V}(z) = \partial V(z)/\partial z$ and $\ddot{V}(z) = \partial V(z)/\partial z^2$ etc.

Since the values of potential at the mesh points must also satisfy Laplace's equation it can be shown [66] that:

$$\phi(r,z) = V(z) - \frac{r^2}{4}\ddot{V}(z) + \qquad \qquad 2.16$$

Ignoring all derivatives of order higher than two leads to the result:

$$\ddot{V}(z) = \frac{2}{hk(h+k)}\left(kV(z+h) + hV(z-k) - (h+k)V(z)\right) + E_{in} \qquad 2.17$$

where:

$$E_{1n} = \frac{2}{hk(h+k)} \left(\frac{(kh^3 - hk^3)}{6} \overset{'''}{V}(z) + \frac{(kh^4 + hk^4)}{24} \overset{''''}{V}(z) + \dots \right) \qquad 2.18$$

It follows that:

$$\overset{''}{V}(z)_{accurate} = \overset{''}{V}(z)_{numerical} - E_{1n} \qquad 2.19$$

and that:

$$V(z)_{accurate} = \phi(r,z) + \frac{r^2}{4} \overset{''}{V}(z)_{accurate} + \frac{r^4}{64} \overset{''''}{V}(z)_{accurate} + \qquad 2.20$$

Therefore:

$$V(z)_{accurate} = V(r,z) + \frac{r^2}{2} \left(\frac{kV(z+h) + hV(z-k) - (h+k)V(z)}{h\,(h+k)} \right) - \frac{r^2}{4} E_{1n} - \dots \qquad 2.21$$

The remaining terms in the expansion of $V(r, z)$ may be written:

$$E_{2n} = -\frac{r^4}{64} \overset{''''}{V}(z) + \frac{r^6}{64.36} \overset{'''''}{V}(z) + \qquad 2.22$$

By using the above equations the following expression for $V(z)$ is obtained:

$$V(z)_{acc} = \underbrace{\frac{2hk}{(2hk+r^2)} V(r,z) + \frac{2hk}{(2hk+r^2)} \left(\frac{kr^2 V(z+h) + hr^2 V(z-k)}{2\,hk\,(h+k)} \right)}_{A} + \underbrace{\frac{2hk}{(2hk+r^2)} \left(E_{2n} - \frac{r^2}{4} E_{1n} \right)}_{B} \qquad 2.23$$

Where expression A is the value of $V(z)$ obtained when derivatives of order higher than two are ignored and expression B is the error involved in doing this.

2.5.2 The Error Terms

(1) $\quad \dfrac{2hk\,r^2}{(2hk+r^2)} \quad$ which can be re-written as $\quad \left(\dfrac{1}{r^2} + \dfrac{r^2}{2hk} \right)^{-1}$

This term diminishes for a constant value of 2hk as r is made smaller and, for a constant value of r, as 2hk is made smaller.

(2) $-\dfrac{r^2}{4} E_{1n}$ which can be written as $-\dfrac{r^2}{4} \displaystyle\sum_{n=3}^{\infty} \dfrac{2}{n!} \dfrac{(h^{n-1} + (-1)^n k^{n-1})}{(h+k)}$

Although small values of h and k make the denominator small, the effect is offset because the first term of the expansion has the numerator $(h^2 - k^2)$. It follows that small h and k values make this error term diminish and it completely disappears when $h = k$ for odd values of n.

(3) E_{2n} The terms of this expansion decrease rapidly if r is less than unity. For accurate numerical values, therefore, the mesh should be drawn so that:

(i)   h, k and r are as small as possible, preferably less than unity.

(ii)   h and k should be as nearly equal as possible.

(iii) The error is strongly dependent on r so that, if a choice is to be made, r should be small even at the expense of h and k.

Another set of equations has therefore been found for the points along the axis of symmetry. They are of the form:

$$A \, \phi(r,z) + B \, V(z+h) + C \, V(z-k) - V(z) = 0 \qquad 2.24$$

where:

$$A = \frac{2hk}{(2hk+r^2)}, \quad B = \frac{kr^2}{(h+k)(2hk+r^2)} \quad \text{and} \quad C = \frac{hr^2}{(h+k)(2hk+r^2)}$$

2.5.3 The Solution Of The Equations

The linear simultaneous equations were solved by using the standard algorithm 'BANDSOLVE'[67]. In order to use this algorithm, the sets of coefficients corresponding to each mesh point must be written in the form of a band matrix. This was achieved by numbering the mesh points as shown in Fig. 2.7 and then by inserting the sets of coefficients into the matrix diagonally. There is a further advantage in this method because it reduces the amount of storage space needed in the computer, only the matrix elements in the band are stored.

The coefficients were calculated by means of Programme One which produced them in the order in which they were needed for the matrix to an accuracy of six decimal places. The version of Programme One in the Appendix

Fig. 2.7

A SIMPLE MESH

THE CORRESPONDING BAND MATRIX
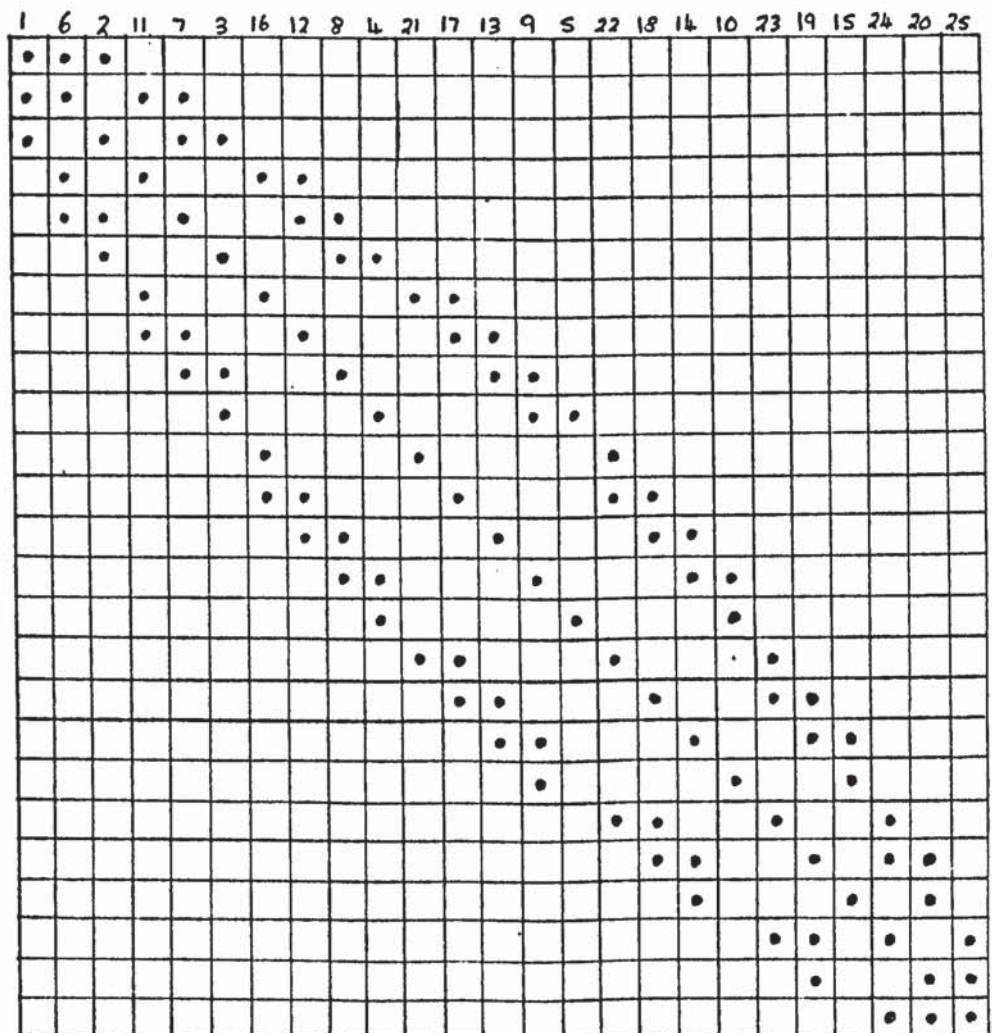
refers to the mesh which was used for the analyser. All programmes were written in I.C.L. 1900 Algol.

2.6.0 Application Of The Numerical Equations To The Analyser

2.6.1 Test Of Accuracy

Most of the loss of accuracy in numerical calculations of the kind used in this work occurs at sharp edges and re-entrant corners because of the large number of points needed to represent them accurately. The electrode system first solved by Loh [68] was used to test the accuracy of the numerical methods because it has similar boundary conditions and re-entrant corners to those found in the analyser.

Unfortunately, Loh's equations are expressed in terms of toroidal co-ordinates which makes direct comparison with the mesh used in the numerical methods rather difficult. They were therefore converted to cylindrical polar co-ordinates. An example similar to that quoted by Loh was then divided into a mesh using the conditions for accuracy previously derived and numerical values of potential at certain points in the mesh were obtained. Comparison between these values and the analytical values derived from Loh's converted equations then gave a measure of the accuracy of the numerical equations.

2.6.2 Conversion Of Toroidal Co-ordinates To Cylindrical Polar Co-ordinates

Let u, v be the toroidal co-ordinates of a point in the mesh and ρ, z be the corresponding cylindrical co-ordinates.

It can be shown that: [68]

$$\rho = \frac{a.\sinh u}{\cosh u - \cos v} \quad \text{and} \quad z = \frac{a.\sin v}{\cosh u - \cos v} \quad \text{where a is a scale factor.}$$

Thus:
$$\frac{\rho}{z} = \frac{\sinh u}{\sin v} \quad \text{and} \quad \sinh u = \frac{\rho}{z} \sin v \qquad 2.25$$

also:
$$\rho^2 + z^2 = a^2 \frac{(\sinh^2 u + \sin^2 v)}{(\cosh u - \cos v)^2} \qquad 2.26$$

and since $\sinh^2 u = \cosh^2 u - 1$ and $\sin^2 v = 1 - \cos^2 v$

- 32 -

$$\rho^2 + z^2 = a^2 \frac{(\cosh u + \cos v)}{(\cosh u - \cos v)} \qquad 2.27$$

Re-arrange this equation and:

$$\frac{2 \cosh u}{2 \cos v} = \frac{\rho^2 + z^2 + a^2}{\rho^2 + z^2 - a^2} \qquad 2.28$$

and finally: $\cosh u = \left(\dfrac{\rho^2 + z^2 + a^2}{\rho^2 + z^2 - a^2}\right) \cos v \qquad 2.29$

Square and subtract equations 2.25 and 2.29 and re-arrange which leads to the expression:

$$\cos^2 v = \left(\frac{\rho^2 + z^2}{z^2}\right)\left(\left(\frac{\rho^2 + z^2 + a^2}{\rho^2 + z^2 - c^2}\right)^2 + \frac{\rho^2}{z^2}\right)^{-1} \qquad 2.30$$

Thus v can be determined and consequently Cos v, Cosh v, Sin v and Sinh v. In a similar manner Cosh u etc. can be found.

The co-ordinate system is shown in Fig. 2.8. The values of u and v also depend on which part of the mesh the point is to be found. Programme Two was used to calculate the values of potential.

Fig. 2.9 shows the mesh used to represent the example and the following table compares the analytical and numerical results.

Analytical Value	Numerical Value
997	997
999	998.7
998	998.2
996	996
989	991.1
971	976.5

Fig.2.8

THE TOROIDAL COORDINATE SYSTEM

V CONSTANT

U = CONSTANT

AO = OB = a
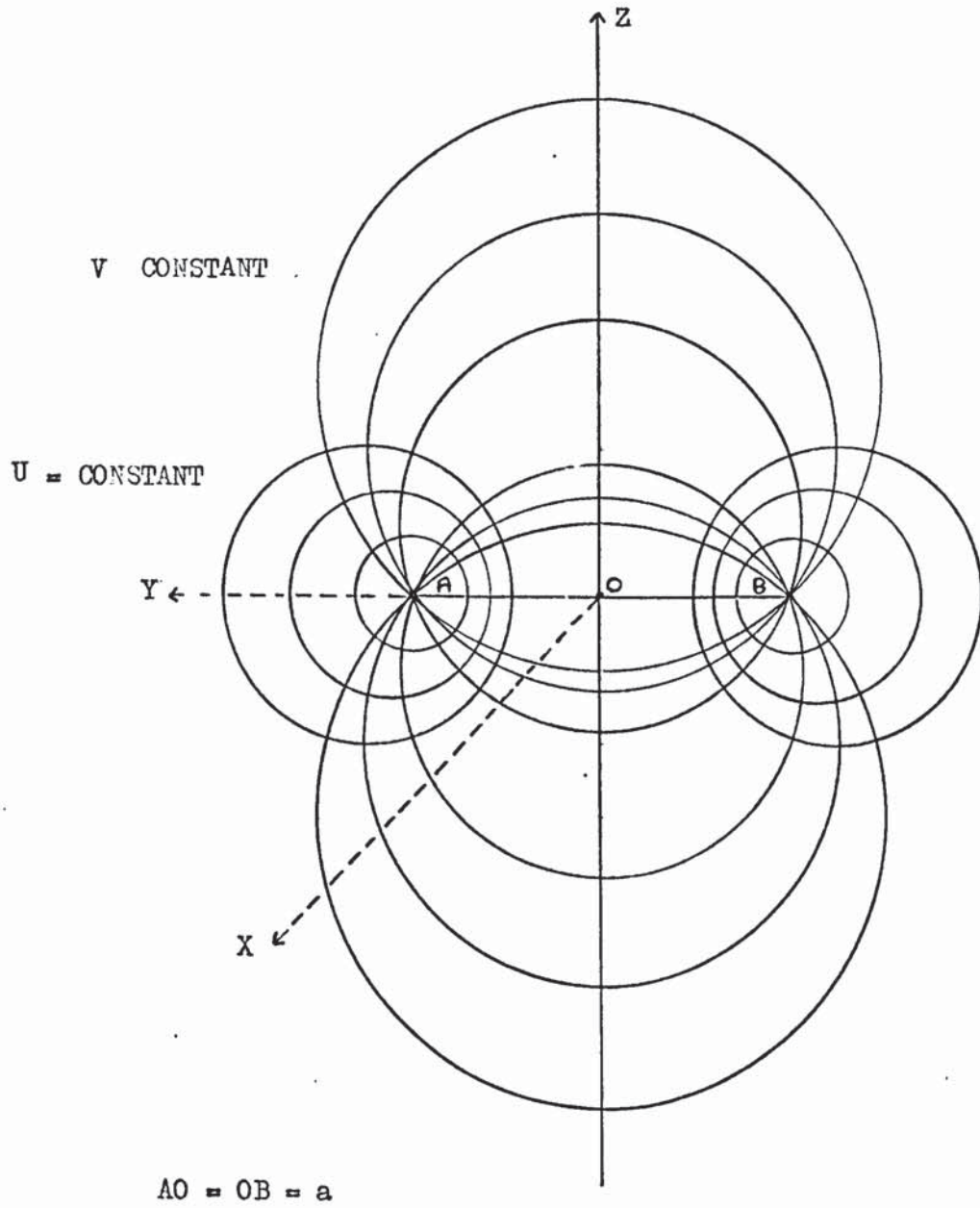
MESH USED FOR
ACCURACY CHECK

Fig. 2.9
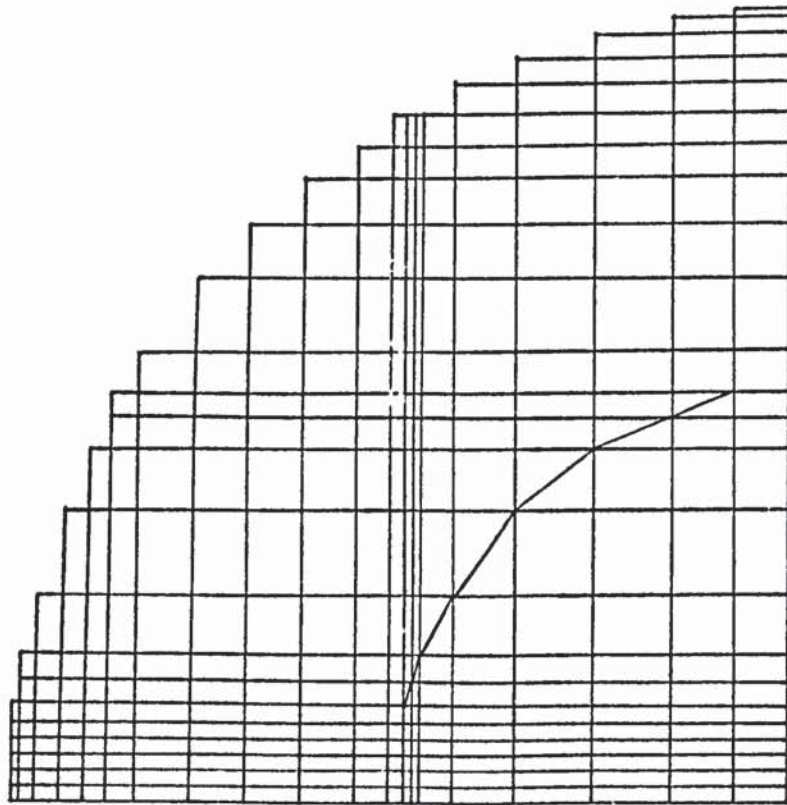
Analytical Value	Numerical Value
950	945.8
943	950.2
935	946.2
918	925
872	878
791	795
714	718.2
651	654.4
594	596.58
568	566.1
549	550.3
532	532.4
513	512.55
502	501.5
497	497

It can be seen that the agreement is very good, the maximum percentage error being approximately 0.7%.

2.6.3 Application Of The Method To The Analyser

The mesh chosen for the analyser is shown in Fig. 2.10. The maximum error conditions were again adhered to and the number of mesh points per unit area was made approximately the same as in the test problem. It therefore follows that the percentage error should be no worse than that obtained above. Programme Three was used to calculate the values of potential and the results for points along the axis of symmetry are given below:

Distance Along Axis Of Symmetry (cm.)	Potential
10.5	1000.00
10.275	999.92

Fig. 2.10

MESH USED FOR ANALYSER
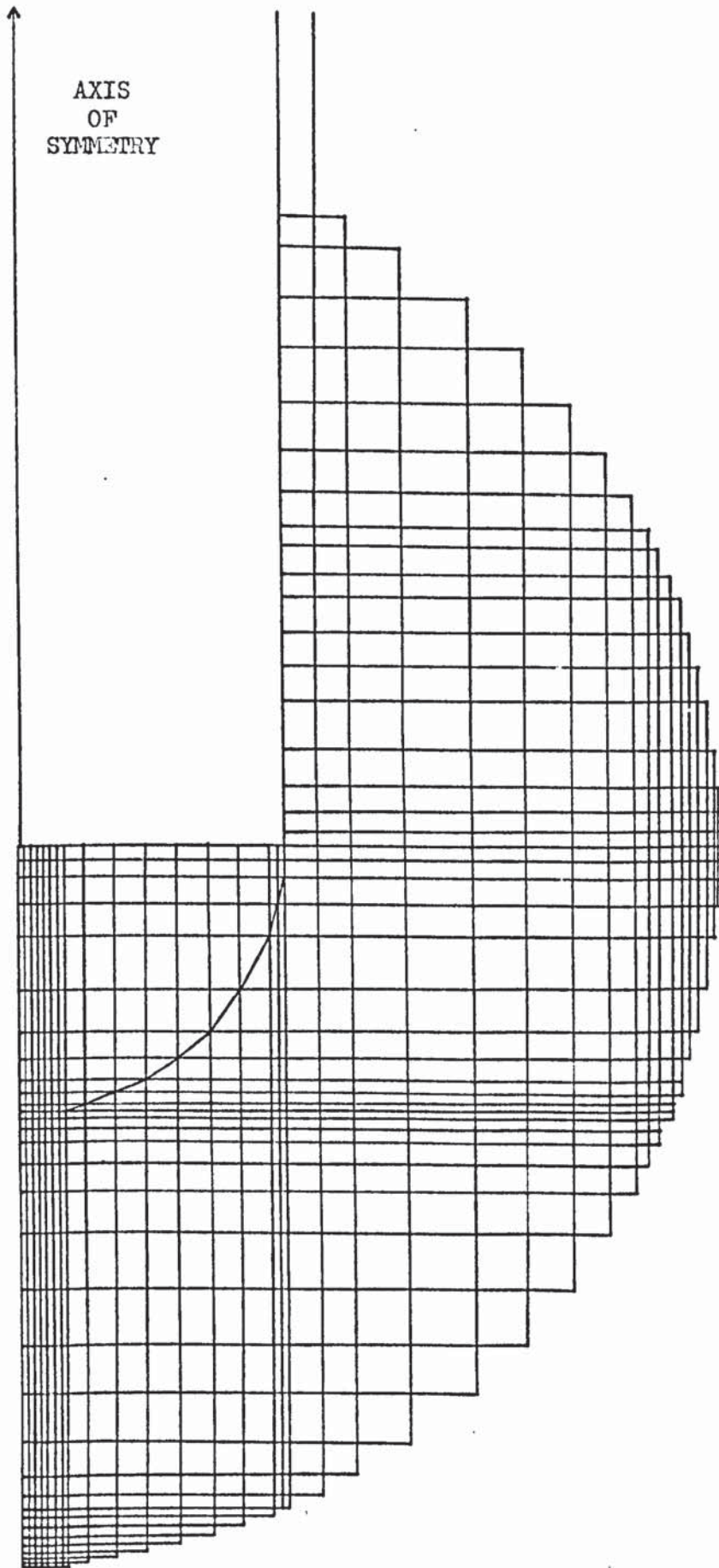
AXIS
OF
SYMMETRY

Distance Along Axis Of Symmetry (cm.)	Potential
10.025	999.80
9.61	999.62
9.11	999.29
8.35	998.39
7.75	996.39
7.325	992.15
7.02	983.94
6.80	970.44
6.65	953.91
6.55	938.40
6.45	919.14
6.30	884.14
6.10	830.84
5.80	748.12
5.42	650.68
4.825	520.73
4.025	382.37
3.23	274.13
2.55	198.23
1.85	132.53
1.35	91.65
1.025	67.36
0.83	53.57
0.725	46.37
0.565	35.68
0.400	24.99
0.275	17.07
0.175	10.82
0.10	6.18

Distance Along Axis Of Symmetry (cm.)	Potential
0.05	3.09
0	0

The distance along the axis of symmetry is measured with respect to the collector. In order to obtain equally spaced values of potential along the axis of symmetry, the above results were plotted on graph paper and readings were taken every half millimetre. Although the values could be read accurately, there was some scatter due to the estimation of decimal places and because such a large scale was used that the graph covered several sheets of graph paper. To obtain a smooth set of values, therefore, the results were processed by using the standard algorithm 'SMOOTH' [69] until they converged to their final values.

2.6.4 The Calculation Of The Electron Trajectories

The values of the radial separation of an electron from the axis of symmetry as it passes through the system were obtained by using the following equations:

The electron trajectory equation is of the form: [71]

$$\ddot{r} = \frac{(1+\dot{r}^2)}{2\phi}\left[\frac{\partial\phi}{\partial r} - \dot{r}\frac{\partial\phi}{\partial z}\right]$$ 2.31

where $\phi = \phi(r,z)$ and $\dot{r} = dr/dz$ etc.

From Taylor's Theorem: [65]

$$r(z+h) = r(z) + h\dot{r}(z) + \frac{h^2}{2}\ddot{r}(z) + \frac{h^3}{6}\dddot{r}(z) + \ldots\ldots\ldots$$ 2.32

and:

$$\dot{r}(z+h) = \dot{r}(z) + h\ddot{r}(z) + \frac{h^2}{2}\dddot{r}(z) + \ldots\ldots\ldots\ldots$$ 2.33

From equation 2.31 it follows that:

$$\overset{\shortparallel}{r} = \frac{1}{2\phi}\frac{\partial\phi}{\partial r} + \frac{r'^2}{2\phi}\frac{\partial\phi}{\partial r} - \frac{r'}{2\phi}\frac{\partial\phi}{\partial z} - \frac{r'^3}{2\phi}\frac{\partial\phi}{\partial z} \tag{2.34}$$

Also $\overset{\shortmid\shortmid\shortmid}{r} = \dfrac{d}{dz}\overset{\shortparallel}{r}$ therefore:

$$\overset{\shortmid\shortmid\shortmid}{r} = \frac{1}{2\phi}\frac{d}{dz}\left(\frac{\partial\phi}{\partial r}\right) + \frac{1}{2}\frac{\partial\phi}{\partial r}\frac{d}{dz}\left(\frac{1}{\phi}\right) + \frac{2r'r''}{2\phi}\frac{\partial\phi}{\partial r} + \frac{r'^2}{2\phi}\frac{d}{dz}\left(\frac{\partial\phi}{\partial r}\right)$$

$$+\frac{r'^2}{2}\frac{\partial\phi}{\partial r}\frac{d}{dz}\left(\frac{1}{\phi}\right) - \frac{r''}{2\phi}\frac{\partial\phi}{\partial z} - \frac{r'}{2\phi}\frac{d}{dz}\left(\frac{\partial\phi}{\partial z}\right) - \frac{r'}{2}\frac{\partial\phi}{\partial z}\frac{d}{dz}\left(\frac{1}{\phi}\right)$$

$$-\frac{3r'^2 r''}{2\phi}\frac{\partial\phi}{\partial z} - \frac{r'^3}{2\phi}\frac{d}{dz}\left(\frac{\partial\phi}{\partial z}\right) - \frac{r'^3}{2}\frac{\partial\phi}{\partial z}\frac{d}{dz}\left(\frac{1}{\phi}\right) \tag{2.35}$$

Further:

$$\frac{d}{dz}\left(\frac{\partial\phi}{\partial r}\right) = \frac{\partial^2\phi}{\partial z\,\partial r} + \frac{\partial^2\phi}{\partial r^2}r' \tag{2.36}$$

and

$$\frac{d}{dz}\left(\frac{\partial\phi}{\partial z}\right) = \frac{\partial^2\phi}{\partial z^2} + \frac{\partial^2\phi}{\partial r^2}r' \tag{2.37}$$

Substitution of equations 2.36 and 2.37 into equation 2.35 leads to the result:

$$\overset{\shortmid\shortmid\shortmid}{r} = \frac{1}{2\phi}\left(\frac{\partial^2\phi}{\partial z\,\partial r} + \frac{\partial^2\phi}{\partial r^2}r'\right) + \frac{1}{2}\frac{\partial\phi}{\partial r}\frac{d}{dz}\left(\frac{1}{\phi}\right) + \frac{2r'r''}{2\phi}\frac{\partial\phi}{\partial r}$$

$$+\frac{r'^2}{2\phi}\left(\frac{\partial^2\phi}{\partial z\,\partial r} + \frac{\partial^2\phi}{\partial r^2}r'\right) + \frac{r'^2}{2}\frac{\partial\phi}{\partial r}\frac{d}{dz}\left(\frac{1}{\phi}\right) - \frac{r''}{2\phi}\frac{\partial\phi}{\partial z}$$

$$-\frac{r'}{2\phi}\left(\frac{\partial^2\phi}{\partial z^2} + \frac{\partial^2\phi}{\partial r^2}r'\right) - \frac{r'}{2}\frac{\partial\phi}{\partial z}\frac{d}{dz}\left(\frac{1}{\phi}\right) - \frac{3r'^2 r''}{2\phi}\frac{\partial\phi}{\partial z}$$

$$-\frac{r'^3}{2\phi}\left(\frac{\partial^2\phi}{\partial z^2} + \frac{\partial^2\phi}{\partial r^2}r'\right) - \frac{r'^3}{2}\frac{\partial\phi}{\partial z}\frac{d}{dz}\left(\frac{1}{\phi}\right) \tag{2.38}$$

Hence:

$$\overset{\shortmid\shortmid\shortmid}{r} = \frac{1}{2\phi}\frac{\partial^2\phi}{\partial z\,\partial r} + \frac{r'}{2\phi}\frac{\partial^2\phi}{\partial r^2} + \frac{1}{2}\frac{\partial\phi}{\partial r}\frac{d}{dz}\left(\frac{1}{\phi}\right) + \frac{2r'r''}{2\phi}\frac{\partial\phi}{\partial r}$$

$$+ \frac{r'^2}{2\phi} \frac{\partial^2 \phi}{\partial z \partial r} + \frac{r'^3}{2\phi} \frac{\partial^2 \phi}{\partial r^2} + \frac{r'^2}{2} \frac{\partial \phi}{\partial r} \frac{d}{dz}\left(\frac{1}{\phi}\right) - \frac{r''}{2\phi} \frac{\partial \phi}{\partial z}$$

$$- \frac{r'}{2\phi} \frac{\partial^2 \phi}{\partial z^2} - \frac{r'^2}{2\phi} \frac{\partial^2 \phi}{\partial r^2} - \frac{r'}{2} \frac{\partial \phi}{\partial z} \frac{d}{dz}\left(\frac{1}{\phi}\right) - \frac{3r'^2 r''}{2\phi} \frac{\partial \phi}{\partial z}$$

$$- \frac{r'^3}{2\phi} \frac{\partial^2 \phi}{\partial z^2} - \frac{r'^4}{2\phi} \frac{\partial^2 \phi}{\partial r^2} - \frac{r'^3}{2} \frac{\partial \phi}{\partial z} \frac{d}{dz}\left(\frac{1}{\phi}\right) \qquad\qquad 2.39$$

Rearrange and collect like terms:

$$r''' = \frac{1}{2\phi} \frac{\partial^2 \phi}{\partial r^2} (r' + r'^3 - r'^2 - r'^4) + \frac{1}{2} \frac{d}{dz}\left(\frac{1}{\phi}\right)\left(\frac{\partial \phi}{\partial r} + r'^2 \frac{\partial \phi}{\partial r} - r' \frac{\partial \phi}{\partial z} - r'^3 \frac{\partial \phi}{\partial z}\right)$$

$$- \frac{1}{2\phi} \frac{\partial^2 \phi}{\partial z^2} (r' + r'^3) + \frac{1}{2\phi} \frac{\partial^2 \phi}{\partial z \partial r} (1 + r'^2) + \frac{r''}{2\phi}\left(2r' \frac{\partial \phi}{\partial r} - \frac{\partial \phi}{\partial z} - 3r'^2 \frac{\partial \phi}{\partial z}\right) \quad 2.40$$

The coefficient of $\frac{d}{dz}\left(\frac{1}{\phi}\right)$ can be written as:

$$\frac{1}{2} (1 + r'^2) \left(\frac{\partial \phi}{\partial r} - r' \frac{\partial \phi}{\partial z}\right)$$

which from equation 2.31 equals $r'' \phi$ therefore:

$$r''' = \frac{1}{2\phi} \frac{\partial^2 \phi}{\partial r^2} (r' + r'^3 - r'^2 - r'^4) + r'' \phi \frac{d}{dz}\left(\frac{1}{\phi}\right) - \frac{r'}{2\phi} (1 + r'^2) \frac{\partial^2 \phi}{\partial z^2}$$

$$+ \frac{1}{2\phi} \frac{\partial^2 \phi}{\partial z \partial r} (1 + r'^2) + \frac{r''}{2\phi}\left(2r' \frac{\partial \phi}{\partial r} - \frac{\partial \phi}{\partial z} - 3r'^2 \frac{\partial \phi}{\partial z}\right) \qquad 2.41$$

Also:

$$\frac{d}{dz}\left(\frac{1}{\phi}\right) = \frac{\partial}{\partial z}\left(\frac{1}{\phi}\right) + \frac{\partial}{\partial r}\left(\frac{1}{\phi}\right) r'$$

$$\frac{\partial}{\partial z}\left(\frac{1}{\phi}\right) = - \frac{1}{\phi^2} \frac{\partial \phi}{\partial z}$$

$$\frac{\partial}{\partial r}\left(\frac{1}{\phi}\right) = - \frac{1}{\phi^2} \frac{\partial \phi}{\partial r}$$

thus:

$$\frac{d}{dz}\left(\frac{1}{\phi}\right) = -\frac{1}{\phi^2}\left(\frac{\partial\phi}{\partial z} + \dot{r}\,\frac{\partial\phi}{\partial r}\right) \qquad\qquad 2.42$$

Substitution of equation 2.42 into equation 2.41 leads to the result:

$$\dddot{r} = \frac{1}{2\phi}\frac{\partial^2\phi}{\partial r^2}(\dot{r}-\dot{r}^2+\dot{r}^3-\dot{r}^4) - \frac{\ddot{r}\phi}{\phi^2}\left(\frac{\partial\phi}{\partial z}+\dot{r}\,\frac{\partial\phi}{\partial r}\right) - \frac{\dot{r}}{2\phi}(1+\dot{r}^2)\frac{\partial^2\phi}{\partial z^2}$$

$$+\frac{(1+\dot{r}^2)}{2\phi}\frac{\partial^2\phi}{\partial z\partial r} + \frac{\ddot{r}}{2\phi}\left(2\dot{r}\,\frac{\partial\phi}{\partial r} - \frac{\partial\phi}{\partial z} - 3\dot{r}^2\frac{\partial\phi}{\partial z}\right) \qquad\qquad 2.43$$

which on rearranging becomes:

$$\dddot{r} = \frac{1}{2\phi}\frac{\partial^2\phi}{\partial r^2}(\dot{r}+\dot{r}^3-\dot{r}^2-\dot{r}^4) - \frac{\dot{r}}{2\phi}(1+\dot{r}^2)\frac{\partial^2\phi}{\partial z^2} + \frac{(1+\dot{r}^2)}{2\phi}\frac{\partial^2\phi}{\partial z\partial r}$$

$$+\frac{\ddot{r}}{2\phi}\left(2\dot{r}\,\frac{\partial\phi}{\partial r} - \frac{\partial\phi}{\partial z} - 3\dot{r}^2\frac{\partial\phi}{\partial z} - 2\dot{r}\,\frac{\partial\phi}{\partial r} - 2\frac{\partial\phi}{\partial z}\right) \qquad\qquad 2.44$$

Finally, collect like terms and:

$$\dddot{r} = \frac{1}{2\phi}\frac{\partial^2\phi}{\partial r^2}(\dot{r}+\dot{r}^3-\dot{r}^2-\dot{r}^4) - \frac{\dot{r}}{2\phi}(1+\dot{r}^2)\frac{\partial^2\phi}{\partial z^2} + \frac{(1+\dot{r}^2)}{2\phi}\frac{\partial^2\phi}{\partial z\partial r}$$

$$-\frac{3\ddot{r}}{2\phi}(1+\dot{r}^2)\frac{\partial\phi}{\partial z} \qquad\qquad 2.45$$

It can be shown that:[60]

$$\phi(r,z) \equiv \phi = \phi(z) - \frac{r^2}{4}\ddot{\phi}(z) + \frac{r^4}{64}\overset{\prime\prime\prime\prime}{\phi}(z) - \frac{r^6}{2304}\overset{\prime\prime\prime\prime\prime\prime}{\phi}(z) + \ldots \qquad 2.46$$

therefore:

$$\frac{\partial\phi(r,z)}{\partial r} \equiv \frac{\partial\phi}{\partial r} = -\frac{r}{2}\ddot{\phi}(z) + \frac{r^3}{16}\overset{\prime\prime\prime\prime}{\phi}(z) - \frac{r^5}{384}\overset{\prime\prime\prime\prime\prime\prime}{\phi}(z) + \ldots \qquad 2.47$$

and:

$$\frac{\partial\phi(r,z)}{\partial z} \equiv \frac{\partial\phi}{\partial z} = \dot{\phi}(z) - \frac{r^2}{4}\overset{\prime\prime\prime}{\phi}(z) + \frac{r^4}{64}\overset{\prime\prime\prime\prime\prime}{\phi}(z) + \ldots \qquad 2.48$$

It follows that the values of $r(z+h)$ and $\dot{r}(z+h)$, and hence the electron trajectory, can be found from equations 2.34, 2.45, 2.46 and 2.47 provided the values of the axial derivatives can be determined.

2.65 The Calculation Of The Axial Derivatives

The methods of forward, backward and central differences were used in calculating the axial derivatives at points corresponding to the beginning, end and middle of the axis of symmetry respectively. For forward differences it can be shown that: [70]

$$\phi'_n = \frac{1}{h}\left(\Delta \phi_n - \frac{1}{2} \Delta^2 \phi_n + \frac{1}{3} \Delta^3 \phi_n + \dots \dots \right) \qquad 2.49$$

and:

$$\phi''_n = \frac{1}{h^2}\left(\Delta^2 \phi_n - \Delta^3 \phi_n + \frac{11}{12} \Delta^4 \phi_n + \dots \dots \right) \qquad 2.50$$

where h is the separation of the points along the axis of symmetry at which the values of potential, ϕ_n, are known and the Δ terms are given by:

$$
\begin{aligned}
\Delta \phi_n &= \phi_{n+1} - \phi_n \\
\Delta^2 \phi_n &= \phi_{n+2} - 2\phi_{n+1} + \phi_n \\
\Delta^3 \phi_n &= \phi_{n+3} - 3\phi_{n+2} + 3\phi_{n+1} - \phi_n \\
\Delta^4 \phi_n &= \phi_{n+4} - 4\phi_{n+3} + 6\phi_{n+2} - 4\phi_{n+1} + \phi_n \\
\Delta^5 \phi_n &= \phi_{n+5} - 5\phi_{n+4} + 10\phi_{n+3} - 10\phi_{n+2} + 5\phi_{n+1} - \phi_n \\
\Delta^6 \phi_n &= \phi_{n+6} - 6\phi_{n+5} + 15\phi_{n+4} - 20\phi_{n+3} + 15\phi_{n+2} - 6\phi_{n+1} + \phi_n
\end{aligned}
\qquad 2.51
$$

For central differences:

$$\phi'_n = \frac{1}{2h}\left((\phi_{n+1} - \phi_{n-1}) - \frac{1}{6}(\delta^2\phi_{n+1} - \delta^2\phi_{n-1}) + \frac{1}{30}(\delta^4\phi_{n+1} - \delta^4\phi_{n-1}) + \dots \right) \quad 2.52$$

and:

$$\phi''_n = \frac{1}{h^2}\left(\delta^2\phi_n - \frac{1}{12}\sigma^4\phi_n + \frac{1}{90}\sigma^6\phi_n + \dots \dots \dots \right) \qquad 2.53$$

where the δ terms are of the form:

$$\delta^2 \phi_n = \phi_{n+1} - 2\phi_n + \phi_{n-1}$$

$$\delta^4 \phi_n = \phi_{n+2} - 4\phi_{n+1} + 6\phi_n - 4\phi_{n-1} + \phi_{n-2}$$

$$\delta^6 \phi_n = \phi_{n+3} - 6\phi_{n+2} + 15\phi_{n+1} - 20\phi_n + 15\phi_{n-1} - 6\phi_{n-2} + \phi_{n-3}$$

$$\delta^2 \phi_n - \delta^2 \phi_{-n} = \phi_{n+2} - 2\phi_{n+1} + 2\phi_{n-1} - \phi_{n-2} \qquad 2.54$$

$$\delta^4 \phi_n - \delta^4 \phi_{-n} = \phi_{n+3} - 4\phi_{n+2} + 5\phi_{n+1} - 5\phi_{n-1} + 4\phi_{n-2} - \phi_{n-3}$$

$$\delta^6 \phi_n - \delta^6 \phi_{-n} = \phi_{n+4} - 6\phi_{n+3} + 14\phi_{n+2} - 14\phi_{n+1} + 14\phi_{n-1} - 14\phi_{n-2}$$
$$+ 6\phi_{n-3} - \phi_{n-4}$$

and finally, for backward differences:

$$\phi'_n = \frac{1}{h} \left(\nabla \phi_n + \frac{1}{2} \nabla^2 \phi_n + \frac{1}{3} \nabla^3 \phi_n + \frac{1}{4} \nabla^5 \phi_n + \frac{1}{6} \nabla^6 \phi_n + \dots \right) \qquad 2.55$$

and:

$$\phi''_n = \frac{1}{h^2} \left(\nabla^2 \phi_n + \nabla^3 \phi_n + \frac{11}{12} \nabla^4 \phi_n + \frac{5}{6} \nabla^5 \phi_n + \frac{137}{180} \nabla^6 \phi_n + \dots \right) \qquad 2.56$$

where the ∇ terms are given by:

$$\nabla \phi_n = \phi_n - \phi_{n-1}$$

$$\nabla^2 \phi_n = \phi_n - 2\phi_{n-1} + \phi_{n-2}$$

$$\nabla^3 \phi_n = \phi_n - 3\phi_{n-1} + 3\phi_{n-2} - \phi_{n-3}$$

$$\nabla^4 \phi_n = \phi_n - 4\phi_{n-1} + 6\phi_{n-2} - 4\phi_{n-3} + \phi_{n-4} \qquad 2.57$$

$$\nabla^5 \phi_n = \phi_n - 5\phi_{n-1} + 10\phi_{n-2} - 10\phi_{n-3} + 5\phi_{n-4} - \phi_{n-5}$$

$$\nabla^6 \phi_n = \phi_n - 6\phi_{n-1} + 15\phi_{n-2} - 20\phi_{n-3} + 15\phi_{n-4} - 6\phi_{n-5} + \phi_{n-6}$$

The accuracy of these equations and of the electron trajectory is governed by conditions similar to those considered previously. Since the values of potential are equally spaced by the comparatively small amount of 0.5 mm, the error involved in using the first two derivatives only was considered to be acceptable.

Programme Four was used to calculate the electron trajectories and the results are presented in the following pages.

2.7 The Electron Trajectories

The programme was first used to obtain the emitter position corresponding to optimum resolution and then to determine the effect of spherical and chromatic aberration.

2.7.1 The Optimum Resolution Trajectory

The trajectory of an electron starting from the vacuum level was considered and the contact potential difference between the emitter and the collector was assumed to be zero.

A value of 0.5 mm. was chosen for the critical separation of the electron from the axis of symmetry because it was thought that by starting at this small radial distance it would pass through the anode aperture paraxially. The trajectory of optimum resolution is a straight line because the electron is then moving radially through the retarding region instead of following an elliptical path. At this point it should be mentioned that the initial values for the electron trajectories are those corresponding to aperture X_1, therefore the paths in the field emission region to the left of X_1 are not considered in the present calculations. Further, except for those results in which the electrons follow a markedly elliptical path, initial slopes of $0'$, $15'$, and $30'$, the trajectories are plotted for the retarding region only because in this way they can be represented with greater accuracy. Fig. 2.11 shows the results of varying the initial slope from $0°$ to $2°$. It can be seen that the optimum resolution occurs at a value of slope equal to $54'$ which corresponds to an emitter position of 7.2 cm. to the left of the anode aperture. The good agreement between this value and that obtained by means of the approximate equations suggests that the value chosen for the initial separation did, in fact, correspond to a paraxial trajectory.

2.7.2 Spherical Aberration

It has been stated previously that spherical aberration can be reduced to a minimum by making aperture X_1 as small as possible. In order to investigate this effect, the initial slope corresponding to optimum resolution was used in conjunction with a range of values of the initial separation. Fig. 2.12 shows the results of these calculations, again the trajectories are plotted in the retarding region only. It can be

Fig. 2.11(a)

INITIAL SLOPE = 0'

COLLECTOR

ANODE

DISTANCE ALONG AXIS OF SYMMETRY FROM ANODE APERTURE (cm)

RADIAL SERARATION. (mm).
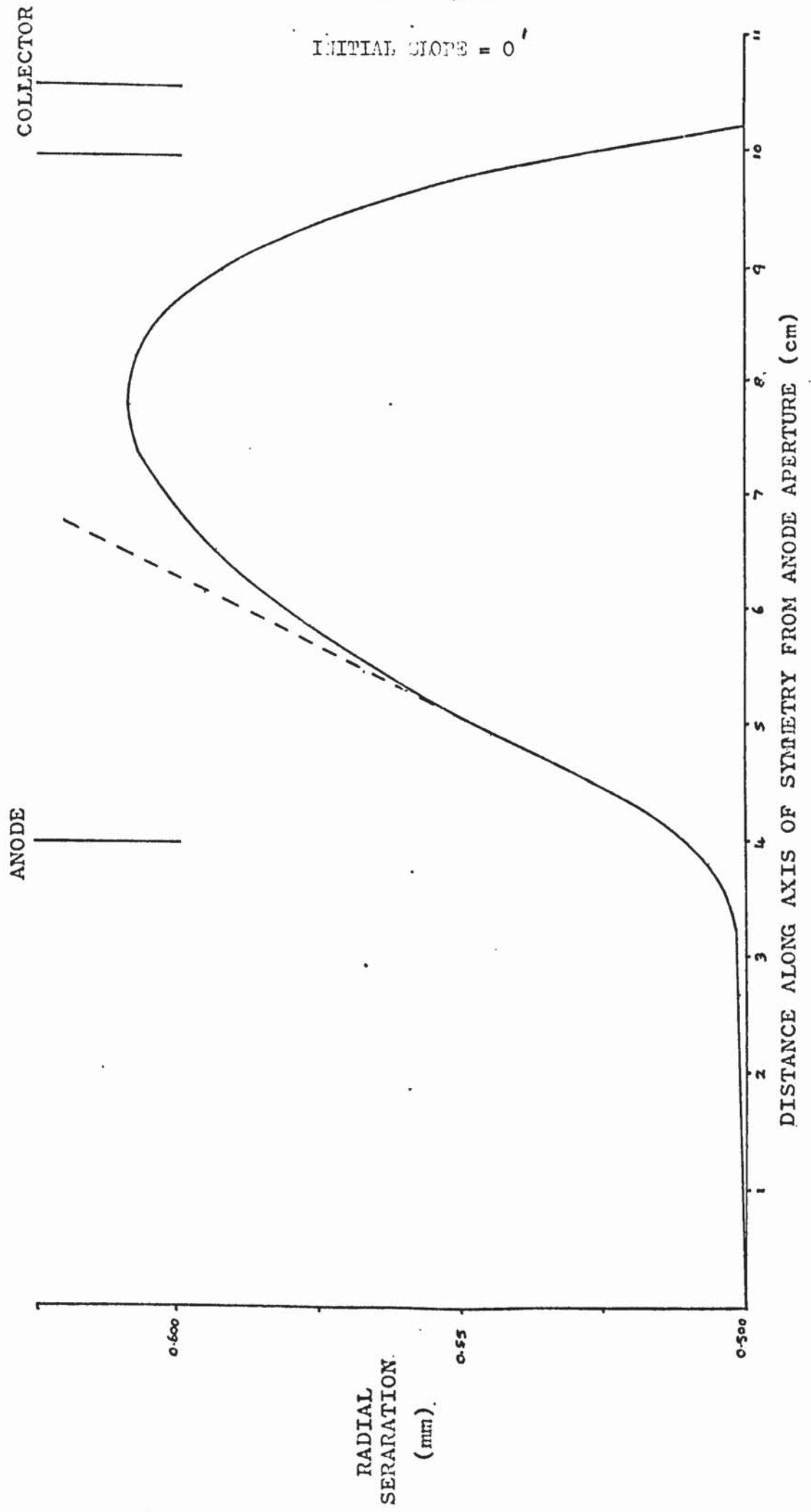
Fig. 2.11(b)

INITIAL SLOPE = 15'

COLLECTOR

ANODE

DISTANCE ALONG AXIS OF SYMMETRY FROM ANODE APERTURE (cm)

RADIAL SEPARATION (mm)
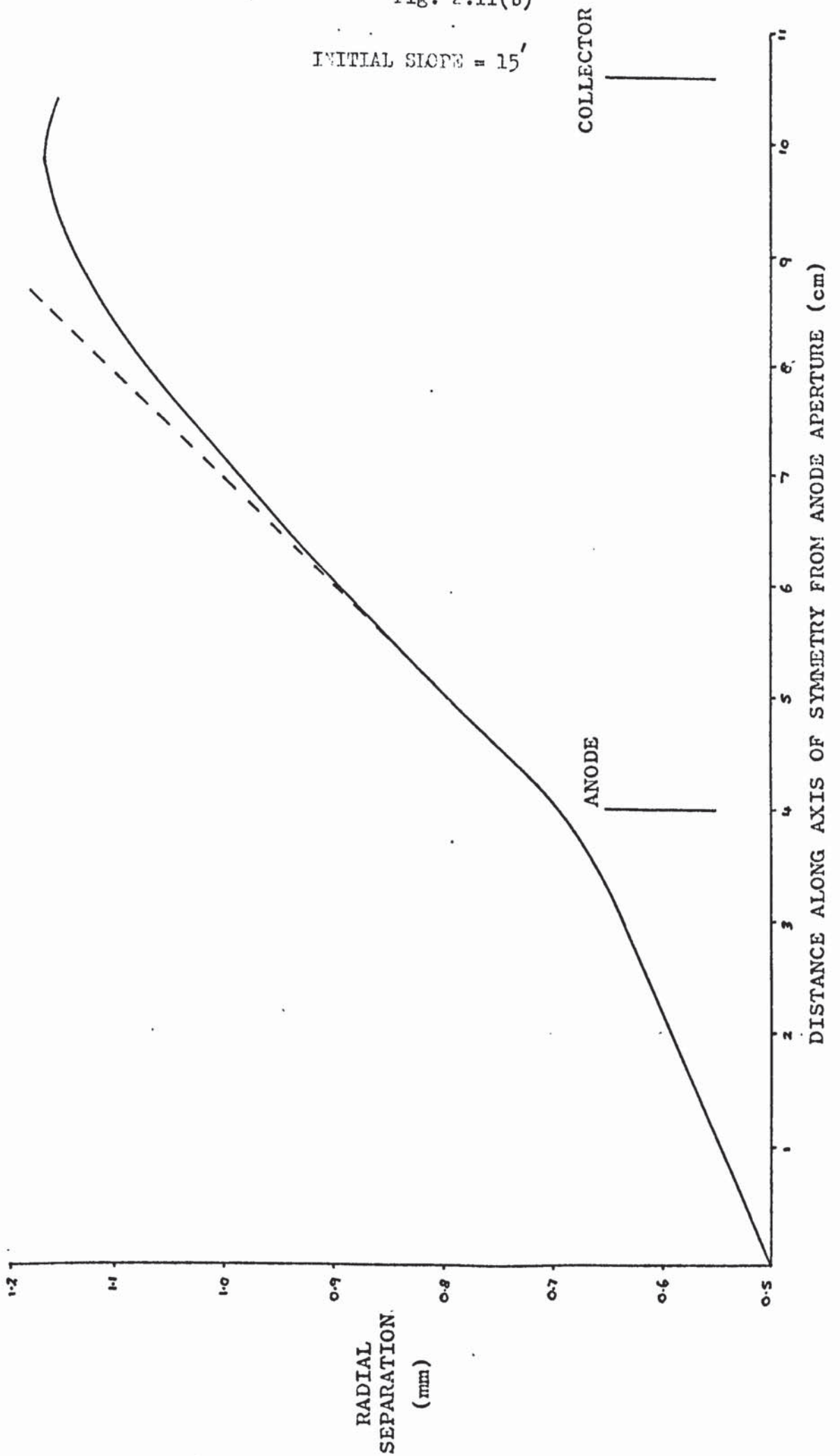
Fig. 2.11(c)
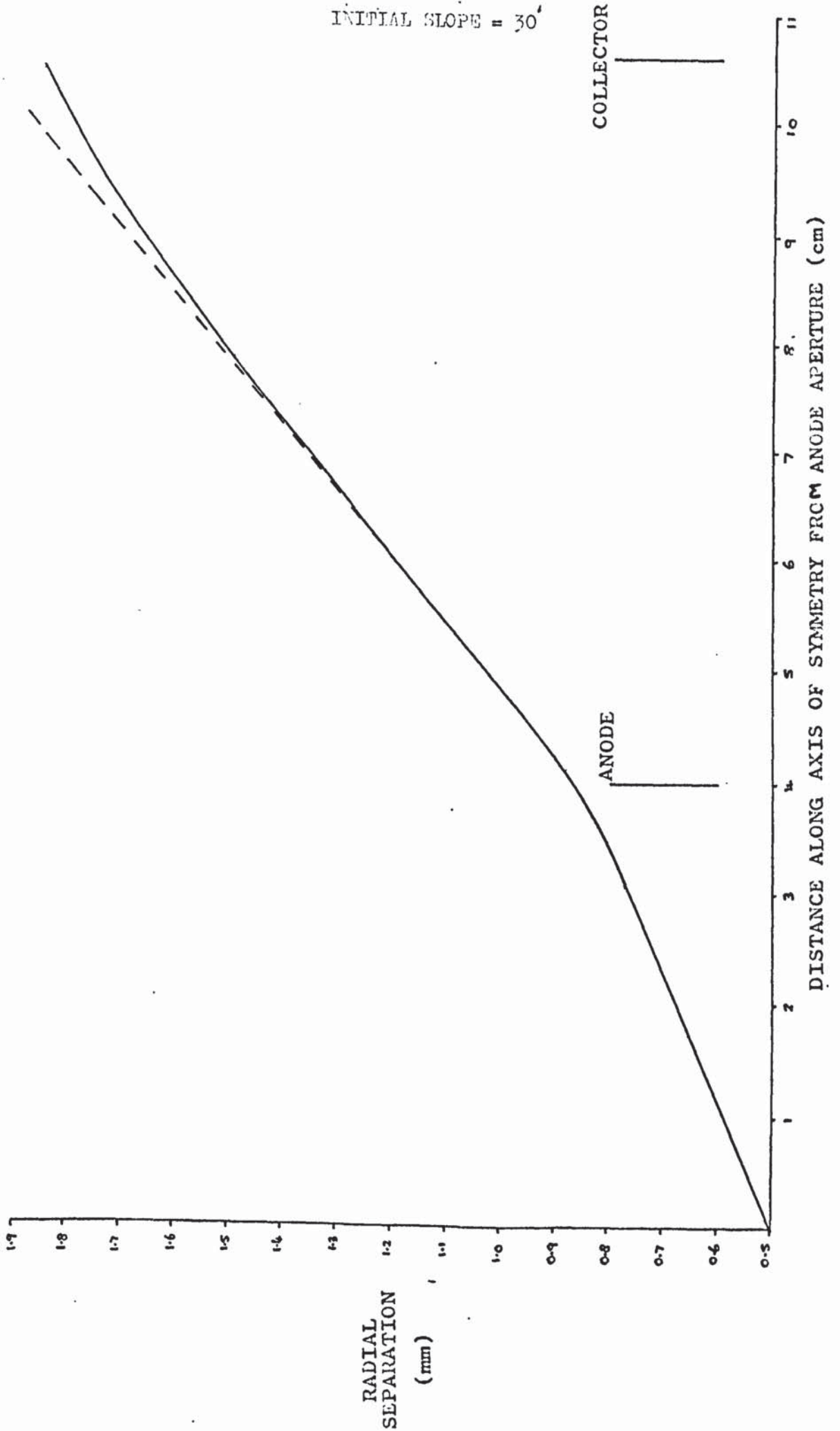
INITIAL SLOPE = 30'

Fig. 2.11(d)

INITIAL SLOPE = 45'

RADIAL
SEPARATION
(mm)

COLLECTOR

DISTANCE ALONG SYMMETRY AXIS FROM ANODE APERTURE (cm)

Fig. 2.11(c)

INITIAL SLOPE = 54′

RADIAL SEPARATION (mm)

COLLECTOR

DISTANCE ALONG SYMMETRY AXIS FROM ANODE APERTURE (cm)
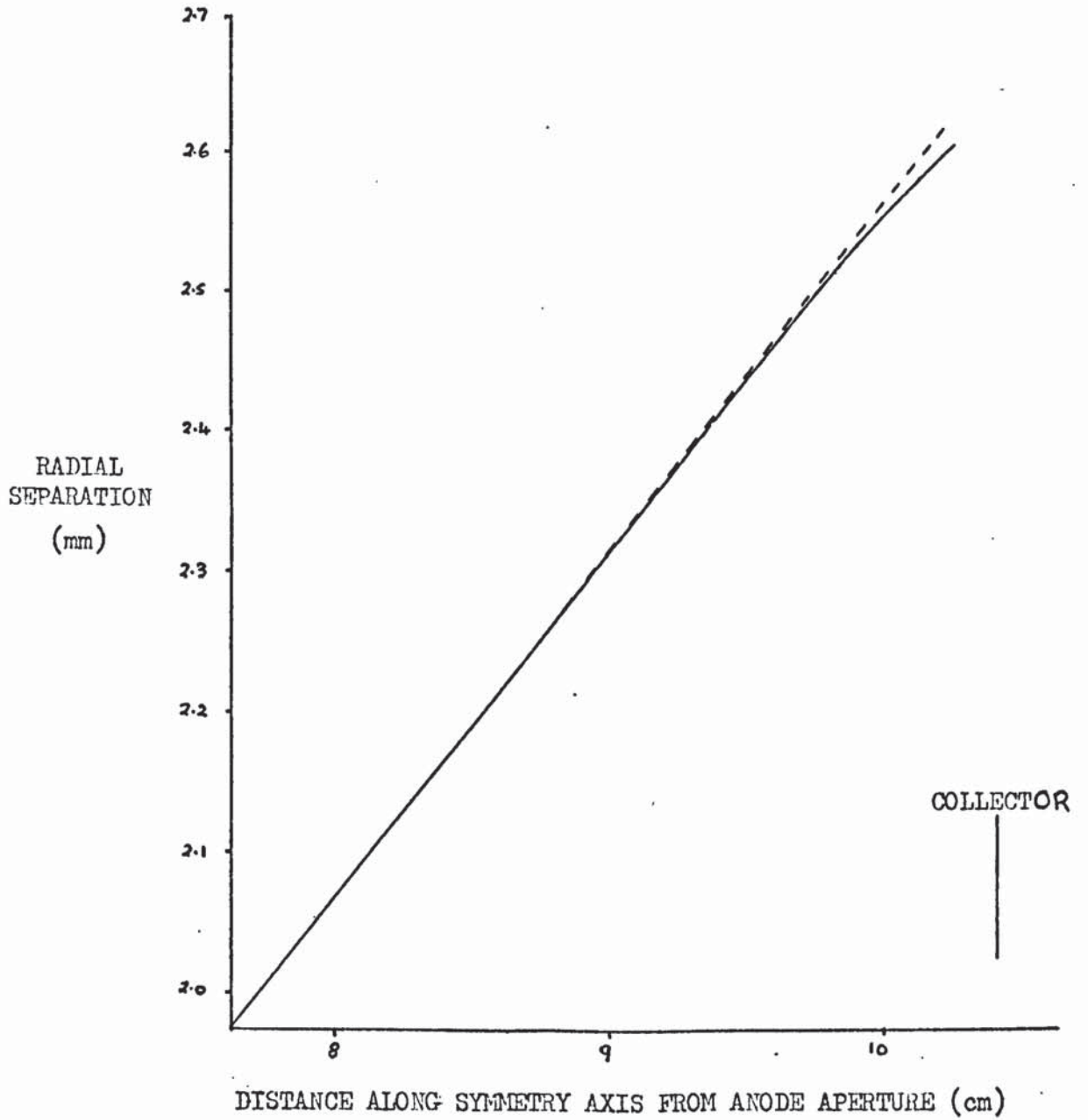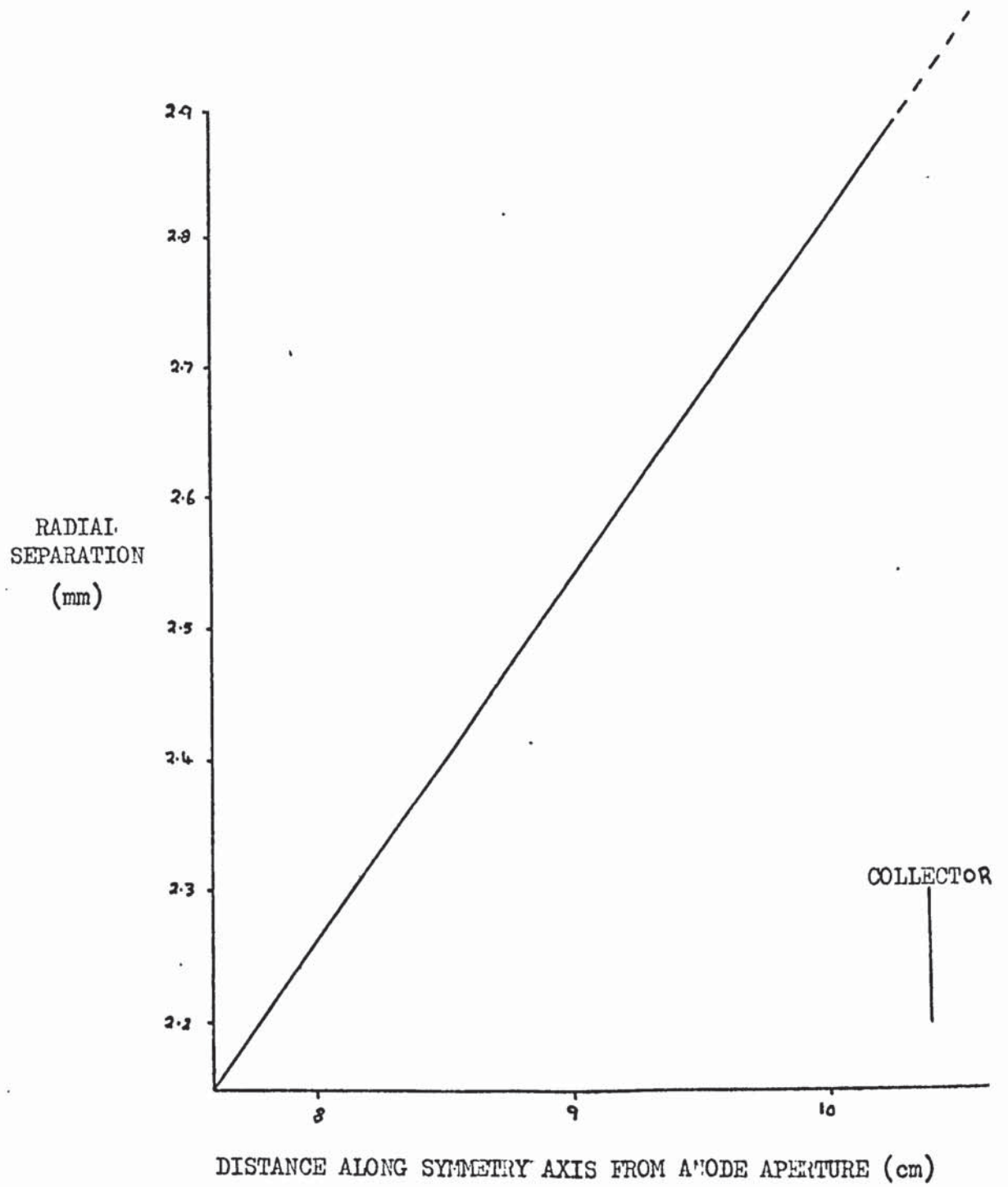
Fig. 2.11(f)

INITIAL SLOPE = 60'

RADIAL SEPARATION (mm)

COLLECTOR

DISTANCE ALONG SYMMETRY AXIS FROM ANODE APERTURE (cm)

Fig. 2.11(c)

INITIAL SLOPE = 120′

RADIAL SEPARATION (mm)

COLLECTOR

DISTANCE ALONG SYMMETRY AXIS FROM ANODE APERTURE (cm)
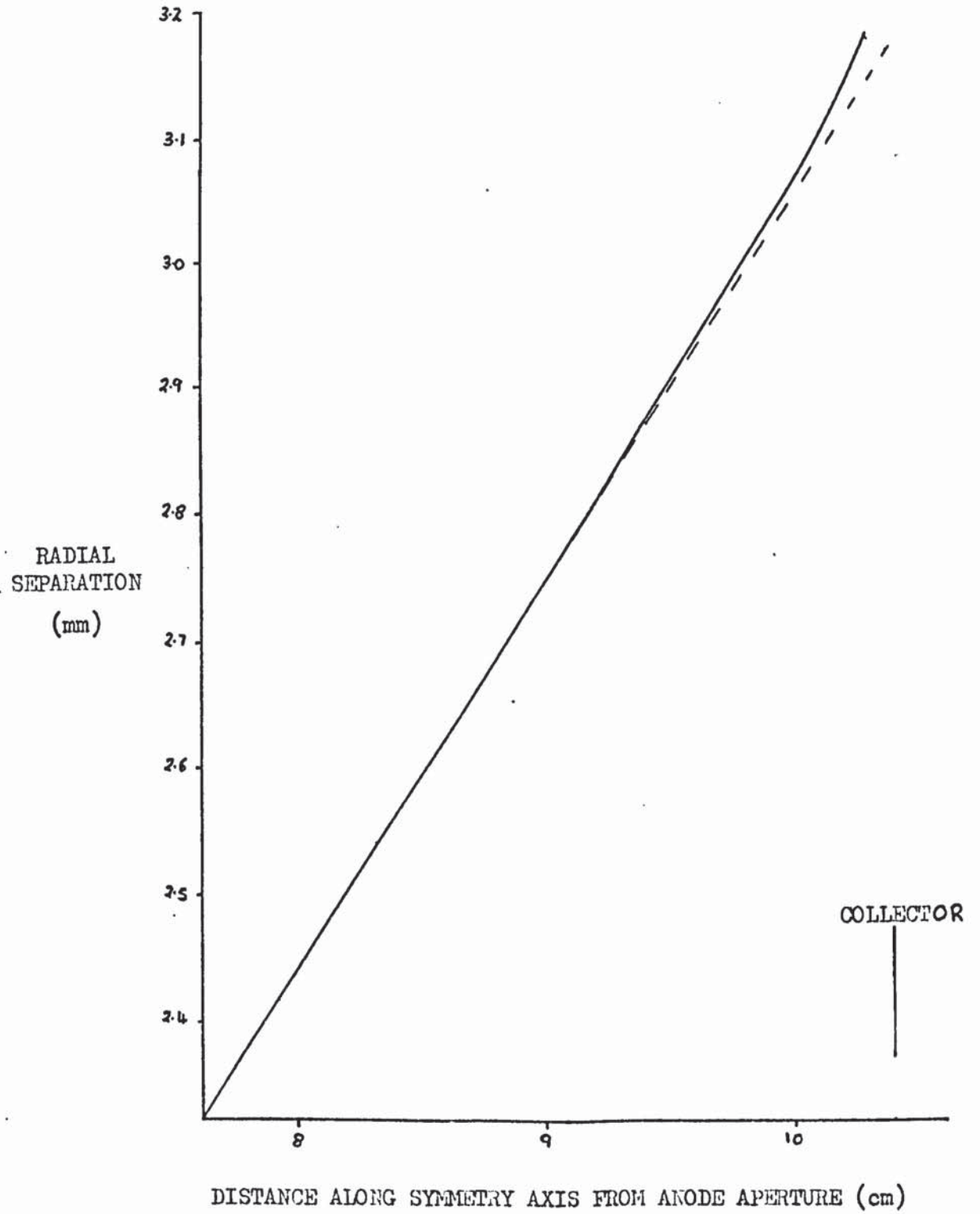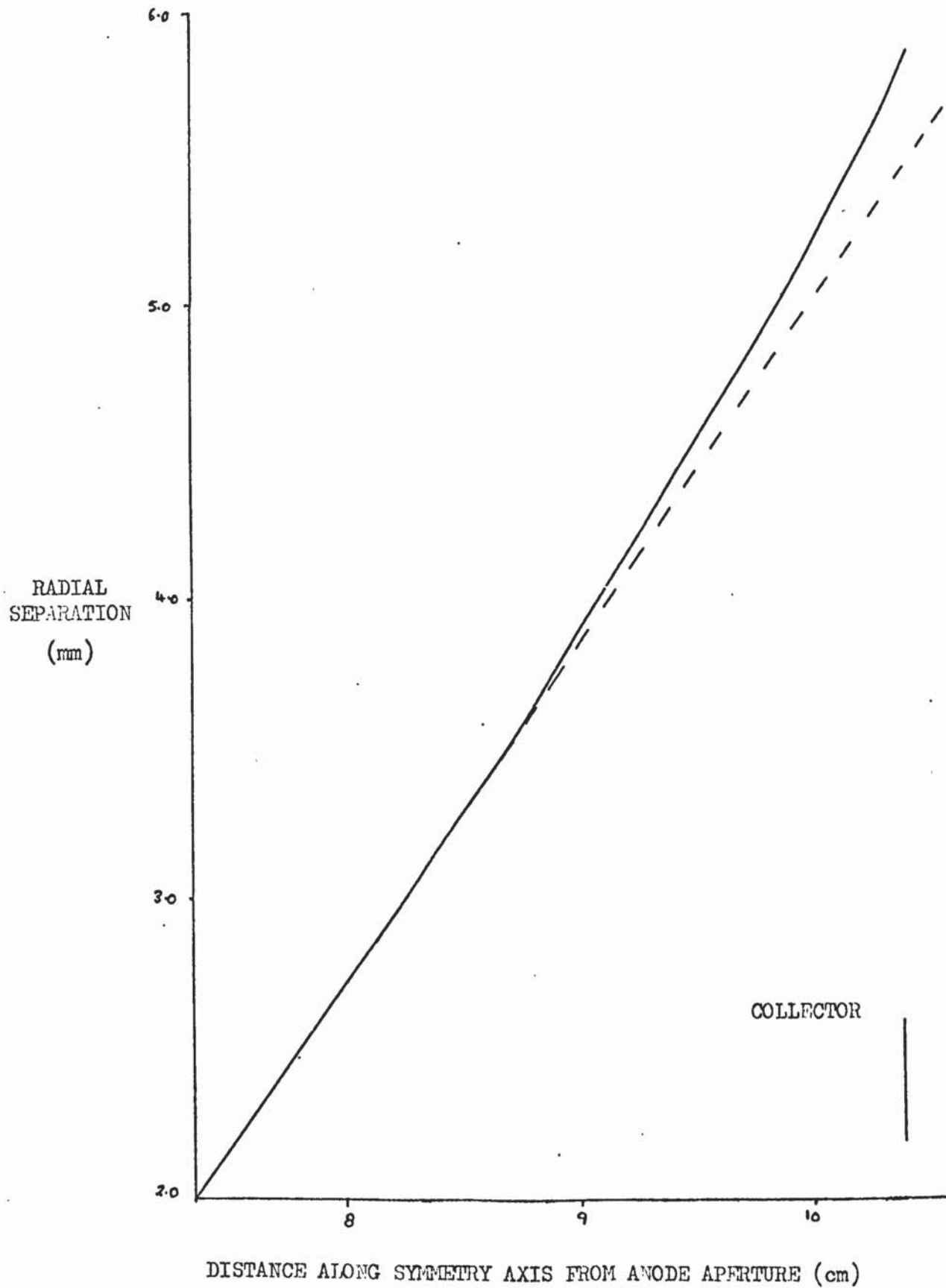
Fig. 2.12(a)

INITIAL SEPARATION = 0.25 mm
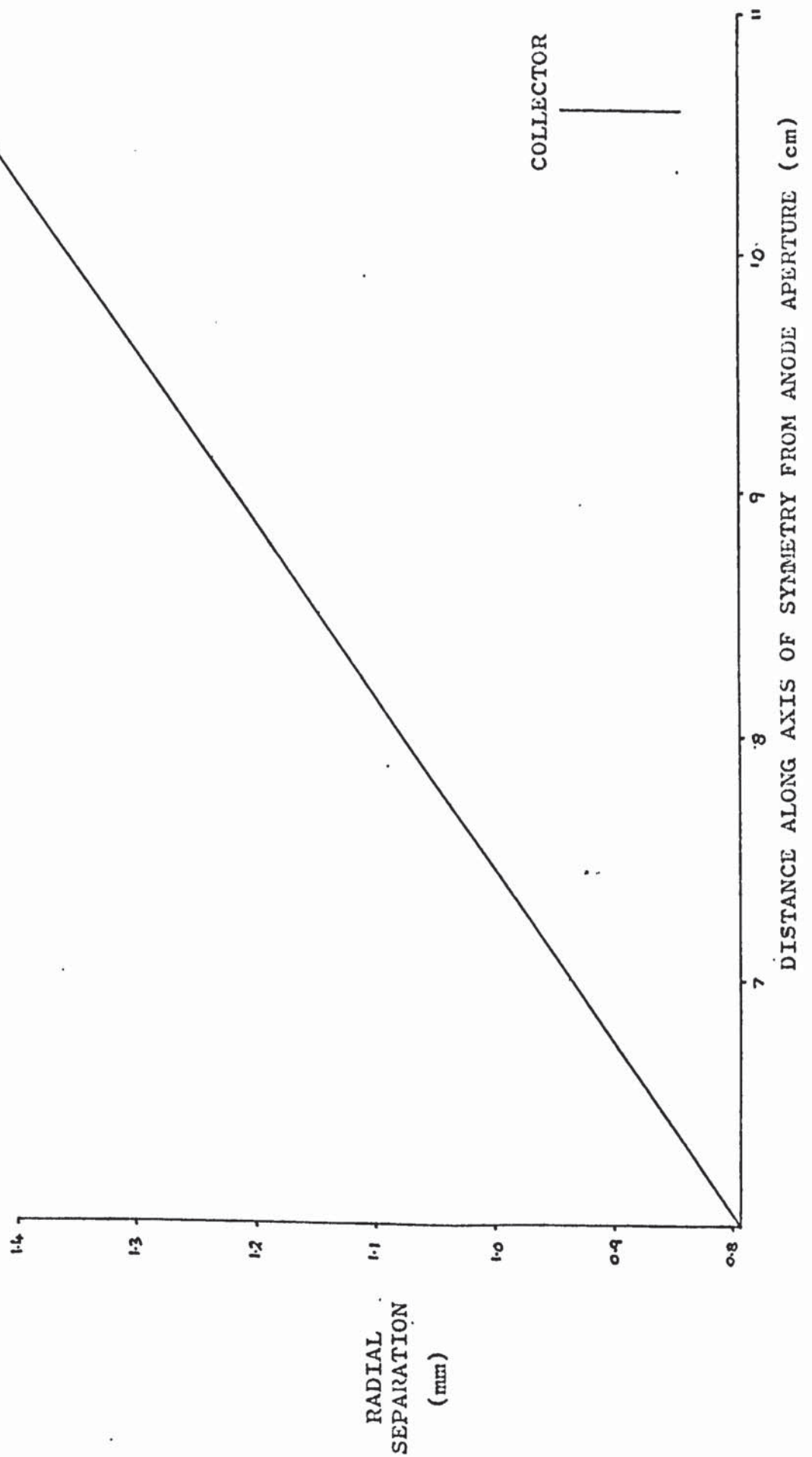
Fig. 2.12(b)

INITIAL SEPARATION = 0.5 mm
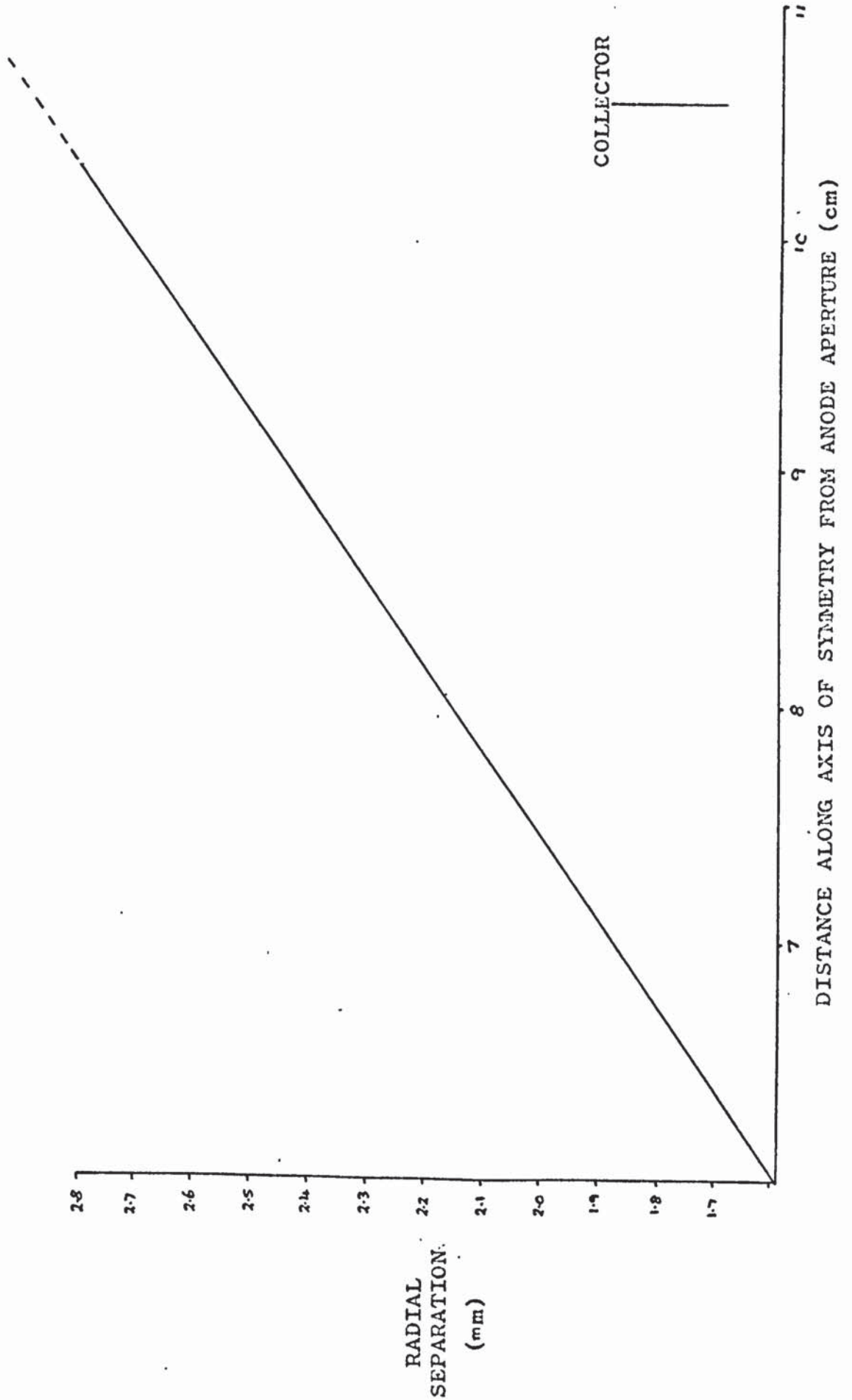
Fig. 2.12(c)

INITIAL SEPARATION = 0.75 mm
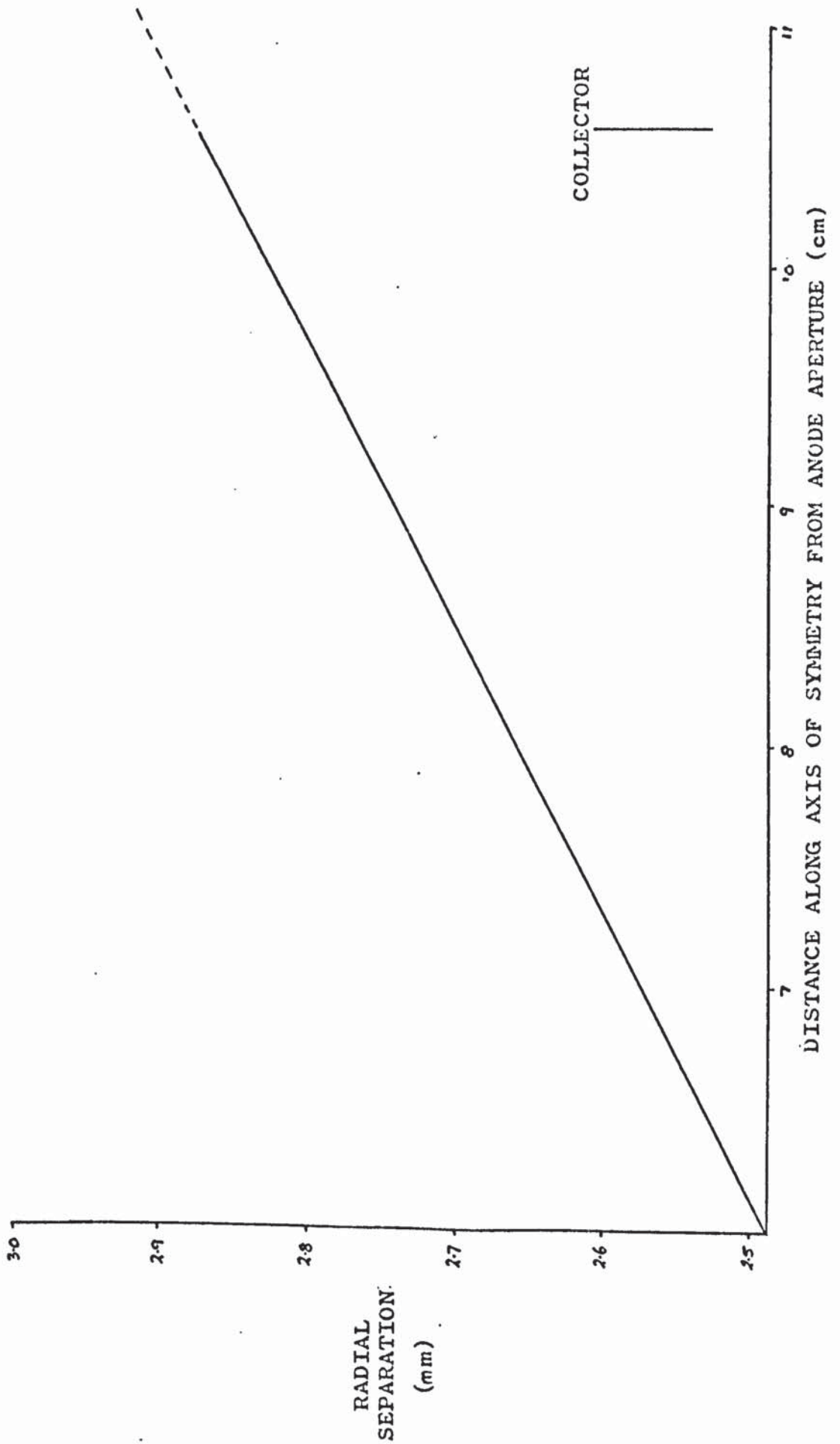
Fig. 2.12(d)

INITIAL SEPARATION = 1.0 mm

COLLECTOR

DISTANCE ALONG AXIS OF SYMMETRY FROM ANODE APERTURE (cm)

RADIAL SEPARATION (mm)
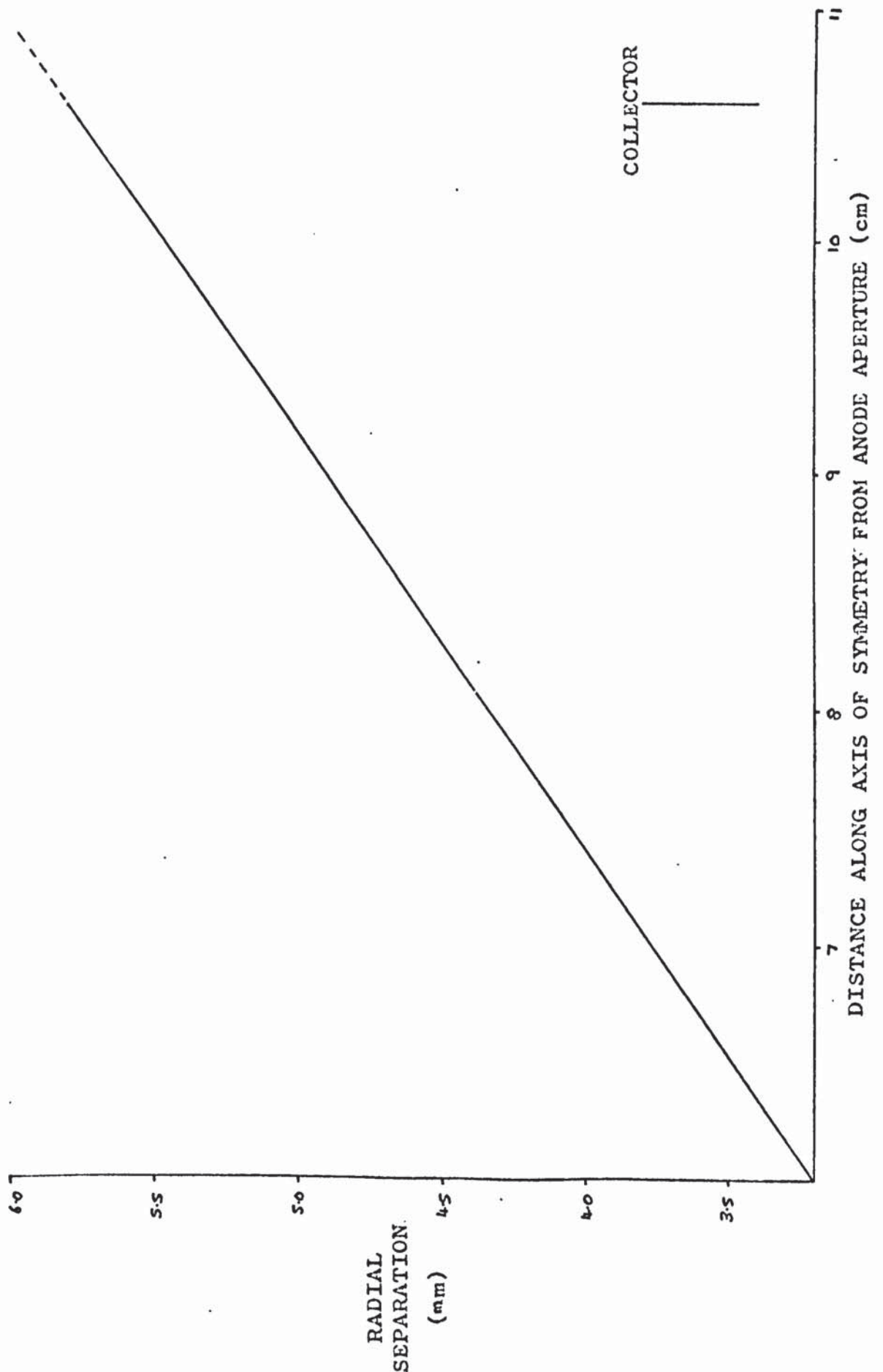
Fig. 2.12(e)

INITIAL SEPARATION = 1.25 mm
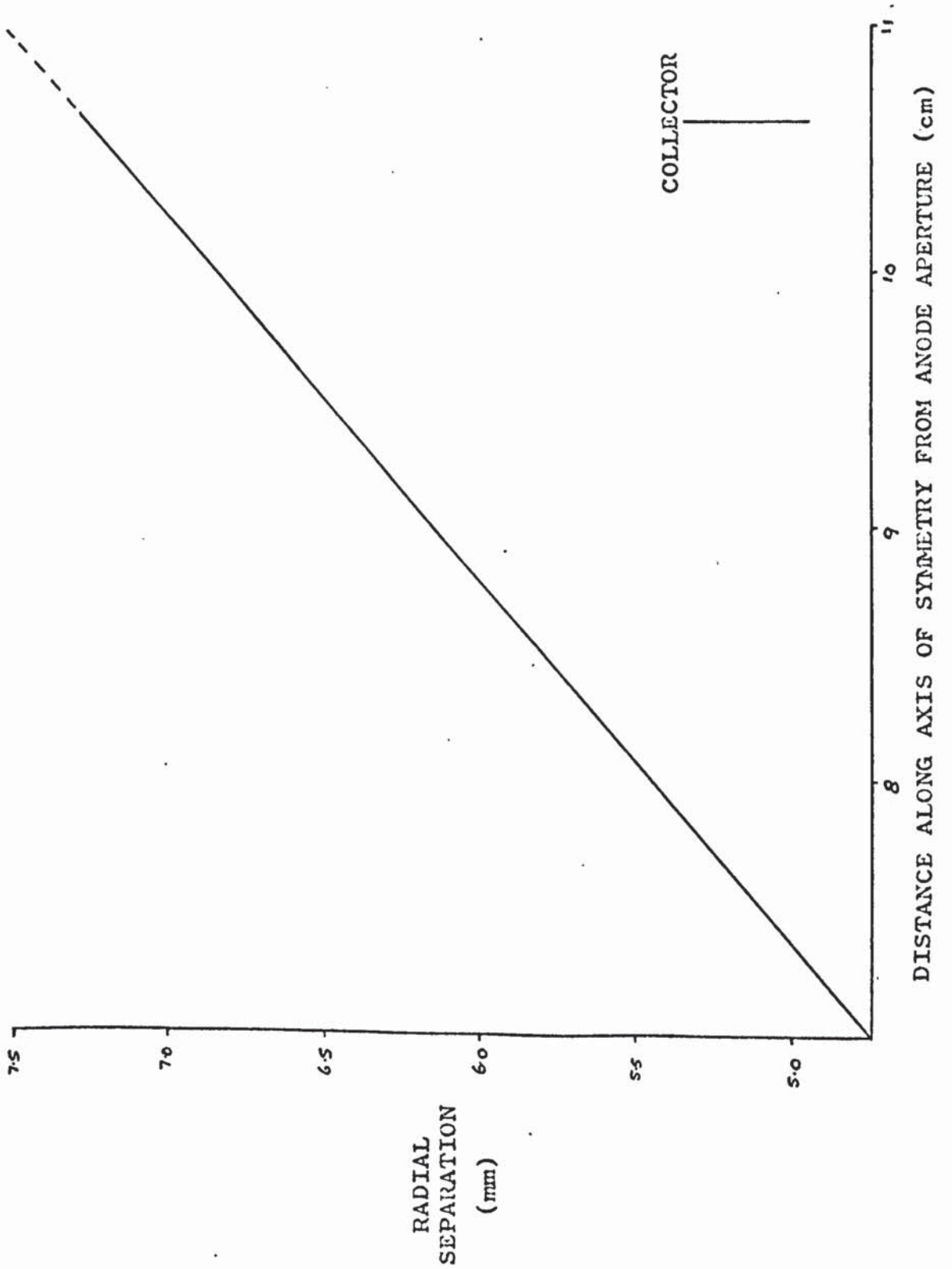
Fig. 2.12(f)

INITIAL SEPARATION = 1.50 mm
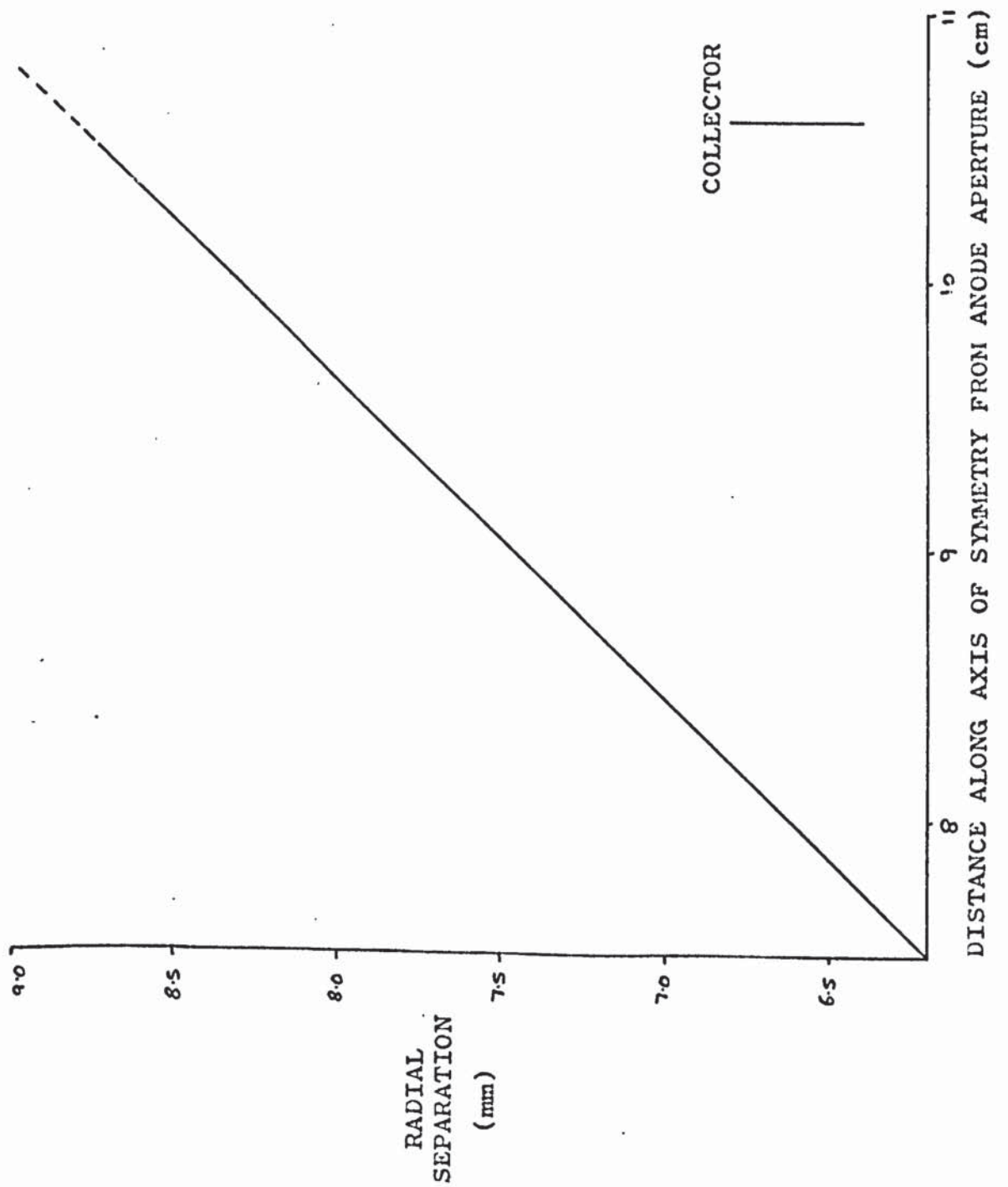
seen that there is no measurable effect over the range of initial separations considered, therefore aperture X_1 was given a radius of 1 mm. because this size also provided ample collected current.

2.7.3 Chromatic Aberration

The equation which describes the motion of an electron in an electro-static field is of the form [71]:

$$\ddot{r} = \frac{(1+\dot{r}^2)}{2\phi}\left(\frac{\partial\phi}{\partial r} - \dot{r}\frac{\partial\phi}{\partial z}\right) \qquad 2.58$$

where ϕ is the value of potential at some point (r, z) and $\dot{r}$ and $\ddot{r}$ are the first and second derivatives of r with respect to z.
In general, the values of potential used are the solutions of Laplace's equation for the particular system under consideration. It should be noted, however, that the value of ϕ used in equation 2.58 is strictly a measure of the total energy of the electron when it reaches (r, z) since, in deriving the equation, use is made of the identity:

$$\frac{1}{2}mv^2 = e\phi \qquad 2.59$$

where e and m are the charge and mass of the electron and v is its velocity at (r, z). When the solutions of Laplace's equation are substituted into equation 2.58 it follows that the motion of an electron which has entered the system with zero energy is described. If the electron enters the system with some other energy $\Delta\phi$, the velocity at (r, z) is given by:

$$\frac{1}{2}mv^2 = e(\phi + \Delta\phi) \qquad 2.60$$

and in order to describe the trajectory correctly, equation 2.58 must take the form:

$$\ddot{r} = \frac{(1+\dot{r}^2)}{2(\phi+\Delta\phi)}\left(\frac{\partial\phi}{\partial r} - \dot{r}\frac{\partial\phi}{\partial z}\right). \qquad 2.61$$

The gradients of potential do not change because the extra energy is a scalar additive constant.

Electrons leave the emitter and enter the analyser with a range of energies. Programme Four was therefore modified to investigate the effect of chromatic aberration by using the values of initial slope and separation corresponding to optimum resolution and by varying the value of $\triangle\phi$. Fig. 2.13 shows the result of these calculations. It can be seen that there is a definite effect because if the change of slope had been caused simply by the incremental increases in $\triangle\phi$ the graph would have been linear. The effect is, however, quite small because a change in $\triangle\phi$ of 0 to 10 volts produces a corresponding change in the final slope of only one degree.

2.8 Additional Factors Which Limit The Resolution Of The Analyser

2.8.1 Loss Of Collector Current

Fig. 2.14 shows the ideal retarding potential vs collector current plot for a perfect analyser in which a saturation region occurs at high values of collector-emitter potential difference when the energy distribution of the metal emitter no longer contributes extra electrons to the collected current. In practice, a plot similar to that shown in Fig. 2.15 is obtained.

Young and Müller [24] have suggested that the discrepancy may be caused by the production of secondary electrons at the collector which leads to a loss of collector current. If this suggestion is correct then a reflection coefficient R can be defined which is a function of the incident electron energy. A reflection coefficient is used instead of the usual secondary electron yield because it has been shown [72] that when electrons with energies of less than approximately 5 eV are incident on a target the secondary electrons produced are mainly elastically scattered primaries.

Consider Fig. 2.16. There are two electrons at energies E_1 and E_2 eV below the top of the potential barrier in the emitter. When the collector

Fig. 2.13
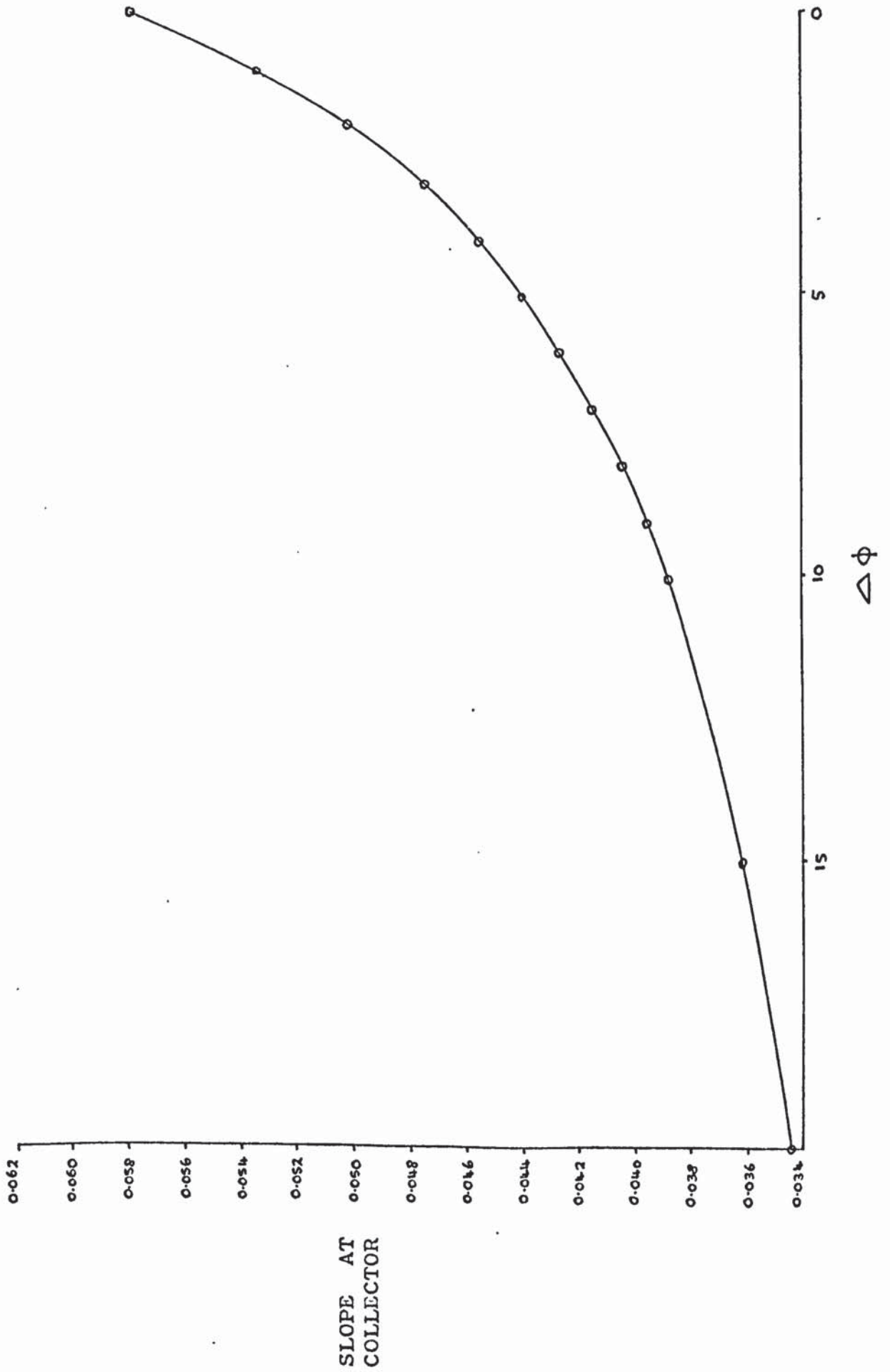
Fig. 2.14

IDEAL ANALYSER RESPONSE

COLLECTOR
CURRENT

COLLECTOR EMITTER POTENTIAL DIFFERENCE
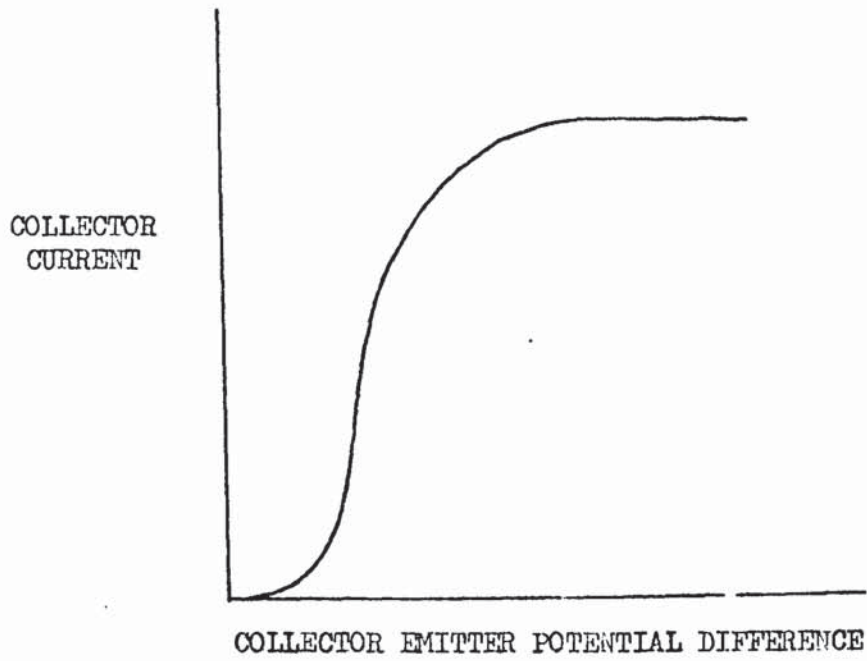
Fig. 2.15

EXPERIMENTAL ANALYSER RESPONSE

COLLECTOR
CURRENT

COLLECTOR EMITTER POTENTIAL DIFFERENCE
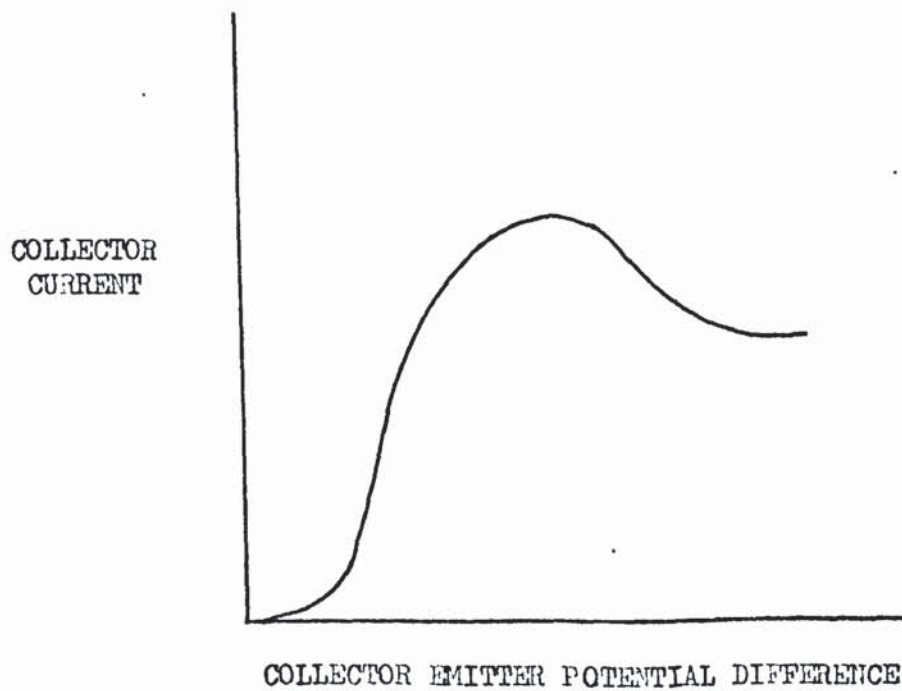
Fig. 2.16

(a) EMITTER COLLECTOR POTENTIAL DIFFERENCE = 0 VOLTS
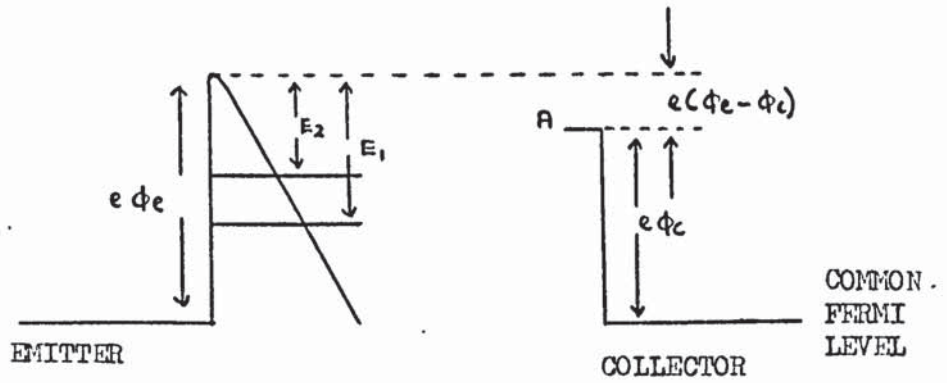
(b) EMITTER COLLECTOR POTENTIAL DIFFERENCE = V_2 VOLTS
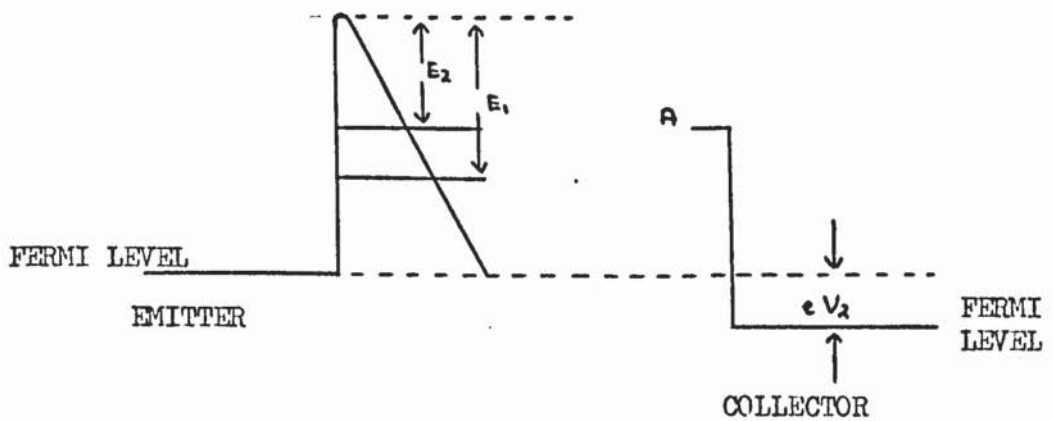
(c) EMITTER COLLECTOR POTENTIAL DIFFERENCE = V_1 VOLTS
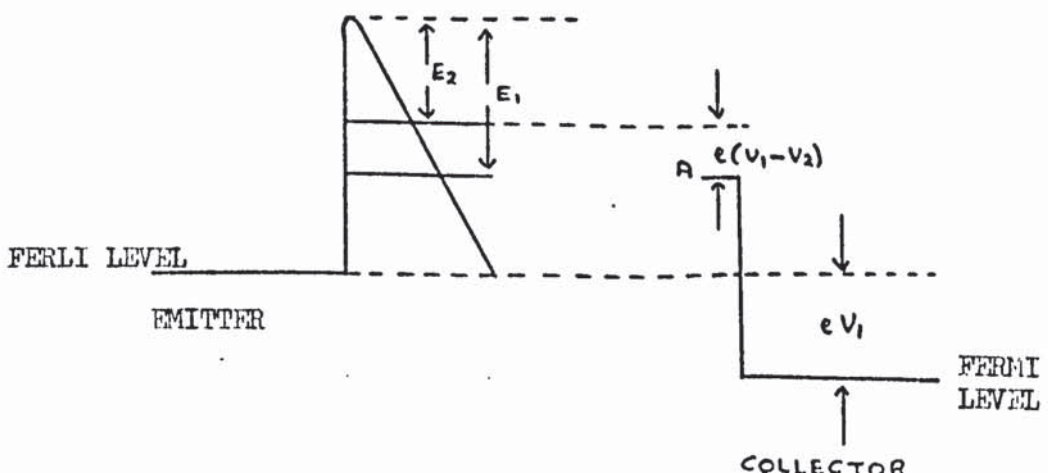
and emitter are in equilibrium their Fermi levels coincide and Fig. 2.16 shows that the electrons considered then have insufficient energy to surmount the potential barrier at the collector and are, therefore, not collected.

Suppose a potential difference of V_2 volts is applied between the emitter and the collector. It follows that the electron of energy E_2 can now just be collected and that its energy with respect to the top of the collector potential barrier, point A, will be zero. Similarly, when the collector-emitter potential difference is V volts the electron of energy E_1 eV is just collected and its energy with respect to A is also zero. The first electron will now, however, have more than enough energy for collection and will arrive at the collector with an energy of $e(v_1 - v_2)$ eV with respect to A. In the experiment the collector is maintained at earth while the emitter is made negative. It follows that when the collector-emitter potential difference is V_1 volts the reflection coefficient for the electron that was originally at E_2 eV below the potential barrier in the emitter is $R(ev_1 - ev_2)$ whereas the reflection coefficient for the other electron is $R(0)$. The measured collector current, $Im(v_1)$, at a collector-emitter potential difference of v_1 volts is, therefore, given by:

$$Im(v_1) = \int_{V_1}^{V_{onset}} \frac{dI}{dV}(\phi)\left(1 - R(ev_1 - e\phi)\right)d\phi \qquad 2.62$$

where $I(\phi)$ is the true collector current when the emitter-collector potential difference is ϕ volts. V onset is the value of emitter-collector potential difference which corresponds to the onset of collector current, therefore, $v_1 > $ v onset. It follows from equation 2.62 that:

$$\frac{d\,Im(v_1)}{dV} = \frac{d\,I(v_1)}{dV} - R(0)\frac{dI(v_1)}{dV} \qquad 2.63$$

and that the measured slope is related to the true slope by $R(0)$ which

- 44 -

is a constant for a given collector material.

Equation 2.63 shows that $d I_m(V_1)/dV$ is negative when $R(0)$ is greater than unity which could only occur if true secondary electrons are formed. This seems unlikely because not only are the energies of the primaries too small [72] but also the production of true secondary electrons is independent of the angle of incidence of the primary electrons at low energies [73] and experiment shows that the current loss can be removed by careful emitter positioning. It therefore seems improbable that energy dependent reflection of electrons at the collector or the production of secondary electrons are the cause of the loss of collected current shown in Fig. 2.15.

The form of the experimental results suggests a reflection coefficient which increases as the collector emitter potential difference increases and which reaches a maximum value where the measured characteristics show a constant collector current. Such a reflection coefficient would be a function of the collector-emitter potential difference only so that all electrons which form a given collector current would have the same value of R. The measured collector current would then be given by:

$$I_m(v) = I(v) (1 - R(v))$$

2.64

and it follows that:

$$\frac{d I_m(v)}{dv} = \frac{d}{dv} I(v) - \frac{d}{dv} (I(v) R(v))$$

2.65

Equation 2.65 shows that as $I(v)$ approaches its saturation value $I_s(v)$, a value of $dI(v)/dV$ occurs such that $dI_m(v)/dV$ is zero when the number of electrons gained by the collector is just balanced by the number being reflected. As $dI(v)/dV$ becomes smaller, $dI_m(v)/dV$ becomes negative and reaches its maximum negative value when $dI(v)/dv$ equals zero. The collector-emitter potential difference which corresponds to the saturation of $I(v)$ can therefore be determined from the slope

of the measured characteristics. For subsequent values of v:

$$\frac{d \, Im(v)}{dv} = - Is(v) \, \frac{d \, R(v)}{dv} \qquad\qquad 2.66$$

and equations 2.64 and 2.66 can be used to obtain Is(v) if the form of R(v) is known.

Although Müller [24] suggested that the loss of current in his experimental results was caused by secondary electrons he used a correction term which was similar to that described by equation 2.64. It was of the form:

$$T(v) = A + B \, exp \, (-k \, \Delta v) \qquad\qquad 2.67$$

where k is a constant and A and B are shown in Fig. 2.16.

$$\Delta v = V - V_{onset} \qquad\qquad 2.68$$

and

$$I(v) = Im(v)/T(v) \qquad\qquad 2.69$$

Equation 2.69 has since been used by many investigators but, unfortunately, if the experimental results are fairly close to the theoretical values, the correction makes A very small with respect to B so that the experimental result then becomes only a small fraction of the corrected result. This is obviously unsatisfactory and since virtually any form for R(v) can be used provided it is compatible with equations 2.64 and 2.66, it is more realistic to use the experimental results which come directly from the analyser and by following the arguments of the previous paragraph to decide whether the correlation between theory and experiment is good or otherwise. This procedure will be adopted in the following chapters.

It should be noted that the conclusions of the previous paragraph are independent of the form of R(v), therefore the total width of the energy distributions measured by the analyser can be obtained without the use of a correction factor.

2.8.2 Anomalous Broadening With Current Density

Many investigators [74] have reported the existence of energy broadening in electron beams of various current densities and recently Zimmermann[75] has reviewed the problem in great detail. He has shown that the broadening is caused by coulomb interactions between the electrons and that, although these interactions are usually weak, they can lead to drastic changes in the distribution of electron energy amongst the various degrees of freedeom in the beam.

The idealised analyser shown in Fig. 2.17 was used for the calculations and it was found that, for a divergent beam at some point in the field free region A - A, the energy spread ΔE could be related to the energy spread ΔEo at anode (1) by the relation:

$$(\Delta E)^2 = (\Delta E_o)^2 + C j R <E>^{1/2} (kT)^{-1} \qquad 2.70$$

where R is the radius of the anode aperture, j is the current density, $<E>$ is the average value of the energy components in the direction of the beam and C is a function of the beam parameters.

In obtaining values of ΔE, Zimmermann assumed that the electrons were accelerated up to anode (1) by a plane homogeneous field and that interactions did not occur during this process. The energy spread at anode (1) was therefore caused only by the thermal energies which the electrons attained at the cathode. It was further assumed that the current density in the beam was approximately constant during acceleration. The initial energy distribution used was that corresponding to thermionically emitted electrons and it was found that for a cathode temperature of 1500°K and a current density of 1×10^{-4} Amp. cm^{-2}. ΔE was approximately equal to 0.1 at a distance of 10 cm from anode (1) in the field free region. The relationship between ΔE and the half width is somewhat difficult to determine because it depends on the shape of the measured distribution. In all the examples quoted by Zimmermann, however,

Fig. 2.17

IDEALISED ANALYSER
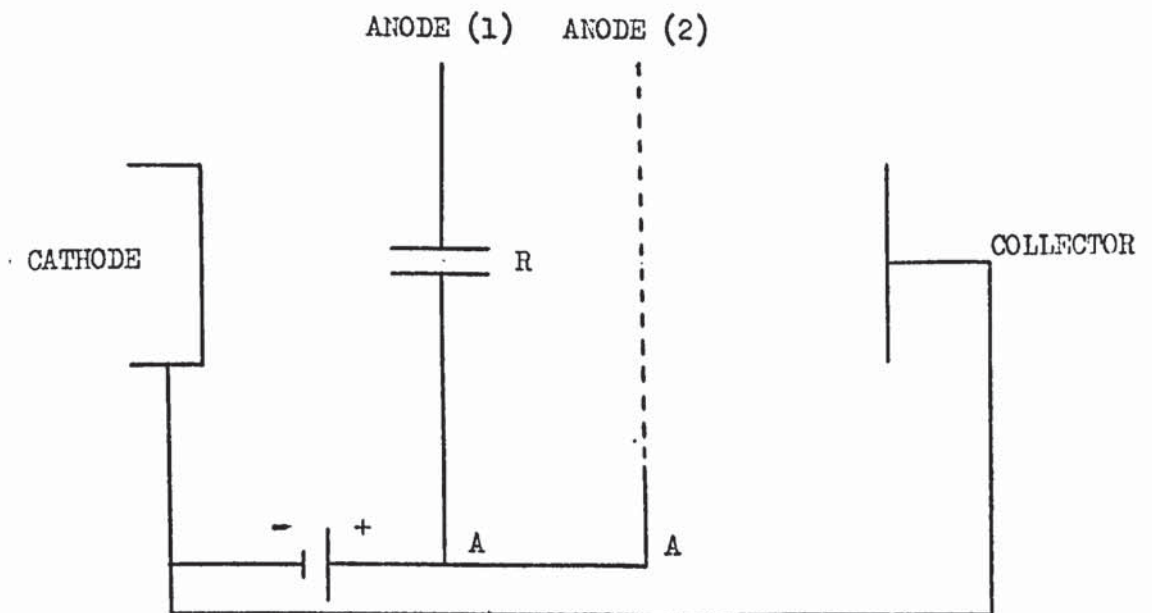
is approximately twice ΔE.

Although the energy broadening of a thermionic rather than a field emission source was considered it seems probable that the effect is also significant in the latter case for the following reasons:

(1) ΔE_o was calculated by assuming a constant current density and by ignoring interactions in the acceleration field. In the present analyser the acceleration field is not constant and the initial current density can be as high as 10^6 Amp. cm.$^{-2}$. Although the effect diminishes with beam diameter the current density at the anode is still approximately 10^{-7} Amp. cm.$^{-2}$ so that, although the subsequent effect in the field free space is smaller, it seems probable that ΔE_o would be larger than the values calculated by Zimmermann.

(2) The temperature of the field emitters used in the present work was approximately five times smaller than that of the thermionic emitter considered by Zimmermann. The broadening is, however, caused by interactions in the beam and not by the initial energy spread, therefore this effect would not seem too important.

Because the theory of anomalous broadening has not been applied to field emitters it was decided to investigate the effect experimentally. The results of these investigations will be presented in Chapter 5.

o o 0 o o

CHAPTER 3

The Theory Of Field Emission

3.1 Introduction

During the development of the theory of field emission, the two main
aspects of the phenomenon which were considered were first, the total
emitted current and second, the energies of the field emitted electrons.
This has been true of the theories corresponding to both metals and
semiconductors and reflects the chronological order in which experiments
on these materials were performed. The same order will be adhered to
in the rest of this chapter.

3.2 The Theory Of Field Emission From Metals

3.3 The Fowler Nordheim Equation

Although Fowler and Nordheim [9] [10] were the first to formulate the
field emission effect in terms of quantum mechanics, a deriviation
similar to that of Good and Müller [76] will be given in the following
pages because the work on energy distributions, which will be described
later, depends on the same definitions of the zero energy. This
definition is different from that used by Fowler and Nordheim.
It is supposed that the potential barrier at the surface of the metal
is deformed by the electric field, thus leading to a much narrower
barrier and the possibility of electron tunnelling [9]. The shape of
the barrier also depends on the image force which the electron
experiences as it escapes from the metal. Fig. 3.1 shows the effect
of these forces and the boundary conditions of the problem, where F is
the applied field, e is the charge on the electron, ϕ is the work
function, ϵ the Fermi level and Ea the bottom of the conduction band
of the emitter. All energies are measured with respect to the same zero
as the effective potential energy $U(x)$.
The Schröedinger equation, which must be solved in order to find the
transmission probability of an electron with an energy "component" Ex

Fig. 3.1
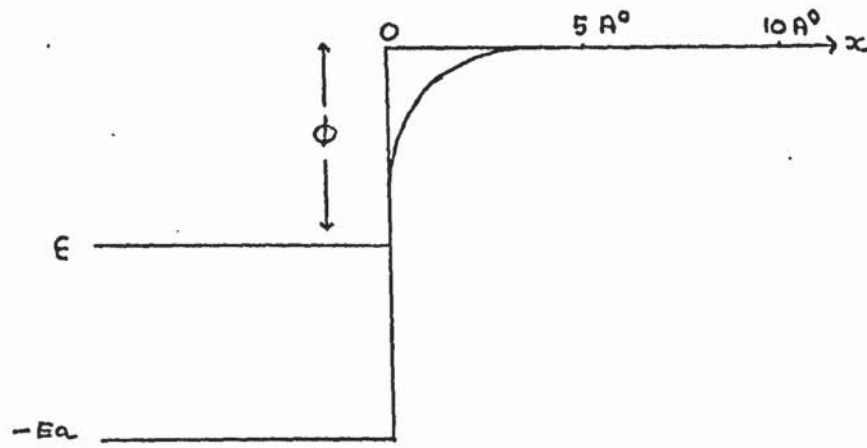
(a) BARRIER WITH NO FIELD APPLIED
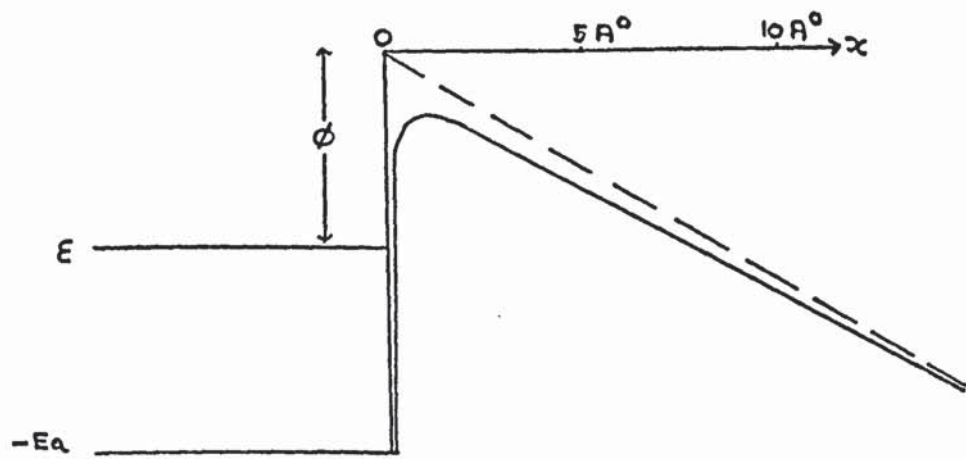
(b) BARRIER WITH FIELD APPLIED

FOR CASE (b):-

$$U(x) = \text{POTENTIAL ENERGY}$$

$$U(x) = -Wa \qquad \text{FOR } x < 0$$

$$U(x) = -eFx - \frac{e^2}{4x} \qquad \text{FOR } x > 0$$

in the direction normal to the barrier, is of the form:

$$\frac{\partial^2 \phi}{\partial x^2} + \left(\frac{2m}{\hbar^2}\right)\left(E_x + \frac{e^2}{4x^2} + e\, F x\right)\phi = 0 \qquad 3.1$$

where $\hbar$ is Planck's modified constant and m is the mass of the electron. This equation has been solved rigorously by Nordheim [10] and approximately by Good and Müller [76] who used the W.K.B approximation which is valid only for electrons whose energies are much smaller than the energy corresponding to the maximum height of the barrier. Since this corresponds to the energies of field emitted electrons, the use of the approximation is permissible. The details of the solution of equation 3.1 will not be considered here since it is already well documented [10][76][77]. The transmission probability of the electron with energy "component" E_x is found to be of the form:

$$D(E_x) = \exp\left(-\frac{4\sqrt{2m\,|E_x|^3}}{3\hbar e F}\, v(y)\right) \qquad 3.2$$

where $v(y)$ is a slowly varying function [76].

After finding the transmission probability, the next problem is to determine how many electrons with kinetic energy "components" E_x arrive normally at the barrier per unit area, per unit time, from the interior of the metal.

Fowler and Nordheim [9][10], like Sommerfeld, assumed that the conduction electrons in a metal behave like a gas of free particles which obey Fermi-Dirac statistics. It follows, therefore, that the probability of an electron state of energy E being filled is given by the expression:

$$F(E) = \left(\exp\left(\frac{(E-\epsilon)}{kT}\right) + 1\right)^{-1} \qquad 3.3$$

where k is Boltzmann's constant, T is the absolute temperature and ϵ is the Fermi energy. The actual number of available states can be

found from the fact that each volume, h^3, of phase space contains two states for particles of spin $\frac{1}{2}$; therefore, the number of available states in the momentum range $dP_x \, dP_y \, dP_z$ is given by:

$$ds = \frac{2V}{h^3} \, dP_x \, dP_y \, dP_z \qquad \qquad 3.4$$

where V is the volume of the metal. It follows that the number of electrons with momenta in the same range contained in volume V is:

$$dn = \frac{2V}{h^3} \, dP_x \, dP_y \, dP_z \frac{1}{\exp\left(\frac{E-\epsilon}{kT}\right)+1} \qquad 3.5$$

where $2mE = P_x^2 + P_y^2 + P_2^2$ 3.6

In order to find the number of electrons moving in the x direction with momenta within dP_x, the number per unit volume with momenta in the range $dP_x \, dP_y \, dP_z$ must be multiplied by the velocity in the x direction and the result integrated over all possible values of P_y and P_z. This leads to the expression:

$$dn_x = \int_{P_y=-\infty}^{\infty} \int_{P_z=-\infty}^{\infty} \frac{P_x}{m} \frac{2}{h^3} \, dP_x \, dP_y \, dP_z \frac{1}{\exp\left(\frac{E-\epsilon}{kT}\right)+1} \qquad 3.7$$

where v_x, the velocity in the x direction, equals P_x/m.

Since the x direction is the direction normal to the emitter surface, it follows that:

$$E_x = \frac{1}{2} m \, v_x^2 \qquad \qquad 3.8$$

and $\qquad\qquad m \, dE_x = P_x \, dP_x$ 3.9

Substitution of equations 3.6, 3.8 and 3.9 into equation 3.7 leads to an equation for $N(E_x) \, dE_x$, the number of electrons incident on the barrier per unit energy, per unit time, with normal "components" of kinetic energy in the range E_x to $E_x + dE_x$. The equation is of the form:

$$N(Ex)\,dEx = \frac{2}{h^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{dP_y\,dP_z}{\exp\left(\frac{Ex-\varepsilon}{kT} + \frac{P_y^2+P_z^2}{2mkT}\right)+1}\,dEx \qquad 3.10$$

Evaluation of the double integral by means of polar co-ordinates [76] leads to the result:

$$N(Ex)\,dEx = \frac{4\pi mkT}{h^3}\log\left(1+\exp-\left(\frac{Ex-\varepsilon}{kT}\right)\right) \qquad 3.11$$

It follows that $P(Ex)\,dEx$, the number of electrons that emerge from the metal per unit area, per unit time, within dEx, is of the form:

$$P(Ex)\,dEx = N(Ex)\,D(Ex)\,dEx \qquad 3.12$$

and that the total current density, I, can be found by integrating equation 3.12 over all possible energies i.e.

$$I = \int_{-Ea}^{\infty} P(Ex)\,dEx \qquad 3.13$$

Substitution of the expressions for $N(Ex)$ and $D(Ex)$ into equation 3.12 leads to the result:

$$P(Ex)\,dEx = \frac{4\pi mkT}{h^3}\exp\left(\frac{-4\sqrt{2m|Ex|^3}}{3\hbar eF}v\left(\frac{\sqrt{e^3F}}{|Ex|}\right)\right)\log\left(1+\exp-\left(\frac{Ex-\varepsilon}{kT}\right)\right)dEx \qquad 3.14$$

For electrons with energies in the neighbourhood of ε , the exponent of the transmission coefficient can be approximated by the first two terms of a power series expansion about $Ex = \varepsilon$. It follows that:

$$\frac{-4\sqrt{2m|Ex|^3}}{3\hbar eF}v\left(\frac{\sqrt{e^3F}}{|Ex|}\right) = -c + \frac{Ex-\varepsilon}{d} \qquad 3.15$$

where:

$$c = \frac{4\sqrt{2m\,\phi^3}}{3\hbar eF}v\left(\frac{\sqrt{e^3F}}{\phi}\right) \qquad 3.16$$

and

$$d = \hbar e F \left(2\sqrt{2m\phi} \; t\left(\frac{\sqrt{e^3 F}}{\phi} \right) \right)^{-1}$$ 3.17

$t(y)$ is a slowly varying function [76]. Equation 3.14, therefore, reduces to:

$$P(Ex)dEx = \frac{4\pi m}{h^3} \exp\left(-c + \frac{Ex-\epsilon}{d} \right) \cdot kT \log\left(1+\exp-\left(\frac{Ex-\epsilon}{kT} \right) \right) dEx \quad 3.18$$

3.3.1 Low Temperature Emission

For small values of T:

$$kT \log\left(1+\exp-\left(\frac{Ex-\epsilon}{kT} \right) \right) = 0 \qquad \text{when} \quad Ex > \epsilon$$

$$\text{and} \quad kT \log\left(1+\exp-\left(\frac{Ex-\epsilon}{kT} \right) \right) = \epsilon - Ex \qquad \text{when} \quad Ex < \epsilon$$

Substitution of these identities into equation 3.18 leads to the result:

$$P(Ex) = 0 \qquad \text{when} \quad Ex > \epsilon$$

$$\text{and} \quad P(Ex) = \frac{4\pi m}{h^3} \exp\left(-c + \frac{Ex-\epsilon}{d} \right)(\epsilon - Ex) \quad \text{when} \quad Ex < \epsilon$$

It follows, because there are no electrons above the Fermi level and because -Ea is usually far below the Fermi level, that the limits of integration in equation 3.13 can be changed to ϵ and $-\infty$. Therefore,

$$I = \frac{4\pi m}{h^3} \int_{-\infty}^{\epsilon} \exp\left(-c + \frac{Ex-\epsilon}{d} \right)(\epsilon - Ex) \, dEx \qquad 3.19$$

and after integration:

$$I = e^3 F \left(8\pi h\phi t^2\left(\frac{\sqrt{e^3 F}}{\phi} \right) \right)^{-1} \exp\left(-\frac{4\sqrt{2m}\phi^{3/2}}{3\hbar e F} v\left(\frac{\sqrt{e^3 F}}{\phi} \right) \right) \qquad 3.20$$

3.3.2 Emission At Higher Temperatures

Strictly speaking, equation 3.20 can only be applied when the emitter is at absolute zero. It is still approximately true for small temperatures but at higher temperatures the majority of the emitted electrons originate from energy levels well above the Fermi level and

the approximation used for the logarithmic term in equation 3.20 then becomes invalid. As the temperature is increased even more, electrons gain enough energy to escape from the metal by passing over the top of the potential barrier and Thermionic emission occurs. The approximation used in the transmission probability now also becomes invalid since Ex may be equal to or even greater than the maximum height of the barrier, $U(x)$ max.

There are therefore three different types of emission: field emission, at high fields and low temperatures, thermionic emission, at high temperatures and low fields and an intermediate Thermionic-Field or T-F range.

Provided electrons do not come from energy levels which are far above the Fermi level, equation 3.15 can still be used for the transmission probability. For such electrons, with $Ex > \epsilon$, the first approximation to the log term is of the form:

$$\log \left(1 + \exp - \left(\frac{Ex - \epsilon}{kT}\right)\right) \simeq \exp - \left(\frac{Ex - \epsilon}{kT}\right) \qquad 3.21$$

thus the expression for $P(Ex)$ becomes:

$$P(Ex) = \frac{4\pi m kT}{h^3} \exp \left(-c + (Ex - \epsilon)\left(\frac{1}{d} - \frac{1}{kT}\right)\right) \qquad 3.22$$

The exponential tail of the distribution above the Fermi level is shown quite clearly in this equation. Equation 3.22 can also be used in order to estimate the range of fields and temperatures for which the approximate value of $D(Ex)$ can be used. For this range, the exponential term must be such that $P(Ex)$ becomes appreciably smaller as the energy of the electrons approaches $U(x)$ max. This occurs when:

$$(U(x) \max - \epsilon)\left(\frac{1}{d} - \frac{1}{kT}\right) < 1 \qquad 3.23$$

Since $U(x)$ max $= -\sqrt{e^3 F}$ (4) and $\phi = -\epsilon$, inequality 3.23 reduces to:

$$\frac{1}{kT} > \frac{1}{d} + \frac{1}{\phi - \sqrt{e^3 F}} \qquad 3.24$$

For conditions experienced in practice, the right hand side of inequality 3.24 is much smaller than $1/d$ and can therefore be neglected. Substituting the expression for d and numerical values of the constants into inequality 3.24 leads to the result:

$$F > 8.83 \times 10^3 \, \phi^{1/2} \, T \qquad\qquad 3.25$$

which, for tungsten at room temperature, means that F must be greater than approximately 5×10^6 volt cm^{-1}. This is usually the case in practice. Thus, assuming inequality 3.25 is fulfilled, the total emitted current density can be found by integrating equation 3.18 over all possible energies i.e.

$$I(T) = e \int_{-\infty}^{\infty} P(Ex)dEx = \int_{-\infty}^{\infty} \frac{4\pi mKT}{h^3} \exp\left(-c + \frac{Ex - \epsilon}{d}\right) \log\left(1 + \exp - \left(\frac{Ex - \epsilon}{kT}\right)\right) dEx \quad 3.26$$

and:

$$I(T) = \frac{4\pi m e \, k^2 T^2}{h^3} e^{-c} \int_0^{\infty} X^{\frac{kT}{d} - 1} \log\left(1 + \frac{1}{X}\right) dX \qquad\qquad 3.27$$

where

$$X = \exp\left(\frac{Ex - \epsilon}{kT}\right).$$

The integral can be integrated by parts [76] and the final result is:

$$I(T) = I(0) \frac{(\pi kT)/d}{\sin((\pi kT)/d)} \qquad\qquad 3.28$$

where $I(0)$ is the total emitted current density at zero temperature of equation 3.20.

The case when inequality 3.25 is not fulfilled has also been considered in some detail. Dolan and Dyke[79], for example, investigated field emission under this condition by performing a numerical integration of their $D(Ex)$ curves and Cutler and Good [80] were able to find analytical expression for the higher order terms of the expansions used in obtaining equation 3.18. These correction terms will not be considered because

the need for them does not arise under most experimental conditions.

3.4 The Energy Distributions Of Field Emitted Electrons

3.4.1 The Normal Energy Distribution

The normal energy distribution, $P(Ex)$, is a measure of the number of electrons which have a given component of energy with the direction normal to the potential barrier (and, therefore, to the emitter surface). This distribution has already been derived in the previous section since the Fowler Nordheim equation measures the number of electrons leaving the emitter and this, in turn, depends on the normal components of their energy. The result is repeated below for convenience:

$$ P(Ex) = \frac{4\pi m kT}{h^3} \exp\left(-c + \frac{Ex-\epsilon}{d}\right) \log\left(1 + \exp-\left(\frac{Ex-\epsilon}{kT}\right)\right) \qquad 3.29 $$

3.4.2 The Total Energy Distribution

The total energy distribution, $T(E)$, is a measure of the number of electrons leaving the emitter with a given total energy. In deriving the expression for the total energy distribution it will again be assumed that the electrons obey Fermi-Dirac statistics and that the transmission probability depends only on the normal component of electron energy.

The total energy, E, of an electron is composed of three components of kinetic energy such that:

$$ Ex = E - \frac{Py^2}{2m} - \frac{Pz^2}{2m} \qquad 3.30 $$

where Ex is the component in the x direction. This direction will again be defined as normal to the potential barrier. For any given total electron energy, therefore, there is a range of values of Ex given by equation 3.30. There must, therefore, be a range of transmission probabilities also.

Suppose that $N(Ex,E)dEx \, dE$ is the number of electrons with total energies in the range E to E + dE whose x part of energy lies in the range Ex to Ex + dEx, incident upon the barrier per unit area. Then if

D(Ex) is the transmission probability, the number of electrons escaping with energies in this range is given by:

$$T(E_x, E) \, dEx \, dE = N(Ex, E) \, D(Ex) \, dEx \, dE \qquad 3.31$$

If all possible values of Ex consistent with a range of total energy from E to E + dE are taken into account, then the number of electrons with total energies in the range E to E + dE actually escaping through the barrier can be determined. This is the total energy distribution T(E), where:

$$T(E) = \int_{Ex} T(Ex, E) \, dEx \, dE \qquad 3.32$$

The expression for the transmission probability D(Ex) is already known, therefore, only N(Ex,E) dEx dE remains to be found.

If θ is the angle between the electron velocity vector and the normal to the surface and ϕ is the azimuthal angle, then the number of electrons with energies in the range E to E + dE incident between θ and $\theta + d\theta$ and between ϕ and $\phi + d\phi$ on unit area of the surface at x = 0, per unit time, is given by the product of the number of electrons arriving at θ per unit solid angle and the differential solid angle. This product is of the form: [23]

$$N(\Omega, E) \, d\Omega \, dE = \frac{n(E) \, dE \, |v| \cos\theta}{4\pi} \, \sin\theta \, d\theta \, d\phi \qquad 3.33$$

where n(E) is the number of electrons per unit volume with energies in the range E to E + dE and $|v|$, which equals $\sqrt{2E/m}$, is the magnitude of the electron velocity.

It is clear from equation 3.30 that:

$$Ex = \frac{1}{2} m v^2 \cos^2\theta \qquad 3.34$$

hence

$$- \frac{dEx}{m|v|} = |v| \sin\theta \cos\theta \, d\theta \qquad 3.35$$

Substitution of equation 3.35 into equation 3.33 leads to the result:

$$N(\Omega, E)\, d\Omega\, dE = -\frac{n(E)\, dE\, dEx}{4\pi \sqrt{2mE}}\, d\phi \qquad \text{3.36}$$

and integration over all ϕ gives:

$$N(Ex, E)\, dEx\, dE = -\frac{n(E)\, dEx\, dE}{2\sqrt{2mE}} \qquad \text{3.37}$$

$n(E)\, dE$ is the product of the number of available states per unit volume with energies in the range E to E + dE and the probability of these states being filled. The expression for the number of available states is of the form: [81]

$$Ns(E)\, dE = \frac{4\pi\, (2m)^{3/2}}{h^3}\, E^{1/2}\, dE \qquad \text{3.38}$$

and the probability of a state being filled is given by the Fermi function:

$$F(E) = \left(\exp\left(\frac{E-\varepsilon}{kT}\right) + 1 \right)^{-1} \qquad \text{3.39}$$

hence:

$$n(E)\, dE = \frac{4\pi\, (2m)^{3/2}\, E^{1/2}}{h^3 \exp\left(\frac{E-\varepsilon}{kT}\right) + 1}\, dE \qquad \text{3.40}$$

and finally:

$$N(Ex, E)\, dEx\, dE = \frac{-4\pi m}{h^3}\, \frac{dEx\, dE}{\exp\left(\frac{E-\varepsilon}{kT}\right) + 1} \qquad \text{3.41}$$

For a given value of E, Ex can vary from E to -Ea, therefore:

$$T(E)\, dE = \int_{Ex = E}^{Ex = -Ea} N(Ex, E)\, D(Ex)\, dEx\, dE \qquad \text{3.42}$$

When Ex equals -Ea the integrand is essentially zero so that -Ea can be replaced by $-\infty$. After performing the integration the result is:

$$T(E)\, dE = \frac{4\pi m d}{h^3} \exp\left(-c - \frac{\varepsilon}{d}\right)\, \frac{\exp\left(\frac{E}{d}\right)}{\exp\left(\frac{E-\varepsilon}{kT}\right) + 1}\, dE \qquad \text{3.43}$$

where c and d have already been defined by equations 3.16 and 3.17. This equation, 3.43, is the total energy distribution of electrons field emitted from a metal.

3.5 Comparison Between The Normal And Total Energy Distributions

The total energy distribution represents the distribution in total energy of field emitted electrons brought to a single potential outside the emitter whereas the normal energy distribution is the distribution of the "components" of electron energy normal to the emitter surface. The total energy distribution can therefore be expected to provide information concerning the origins of the field emitted electrons while the normal energy distribution should provide details of the transmission probabilities. Except for thin film tunnelling [82], where plane parallel electrodes are used, it has not proved possible to measure the normal energy distribution because immediately the electrons leave the emitter and are acted upon by electric fields, energy is transferred from the normal direction to other directions. The validity of the expression for the transmission probability has not, therefore, been proved experimentally but it has been accepted because the total energy distribution also contains the expression and experiment has shown this distribution to be correct.

3.6 The Theory Of Field Emission From Semiconductors

3.7 Theory Concerning The Total Emitted Current

In this section electron energies will be measured as shown in Fig.3.2 because this method is more convenient than the method used previously for metals. The effect of field penetration and the image force correction have been omitted for the sake of clarity. It can be seen that the Fermi energy, ϵ , is positive when the Fermi level is above the conduction band and negative when the Fermi level is below the conduction band. The work function, χ , is defined as the energy needed to remove an electron from the Fermi level and place it outside the semi-conductor and the energy of an electron at the bottom of the conduction band is given by:

Fig. 3.2

N TYPE SEMICONDUCTOR

(a) $\varepsilon > 0$

CONDUCTION BAND

FERMI LEVEL

DONOR LEVELS

VALENCE BAND

(b) $\varepsilon < 0$

CONDUCTION BAND

FERMI LEVEL
DONOR LEVELS
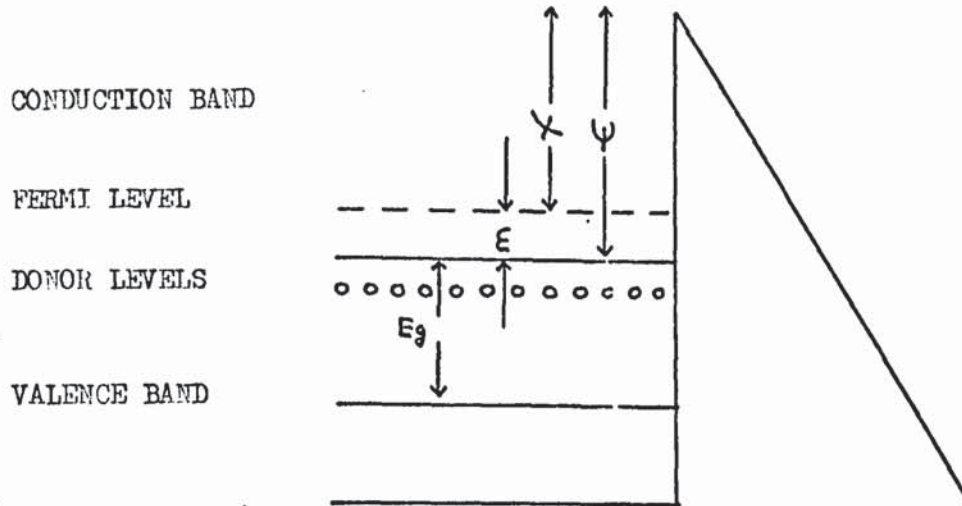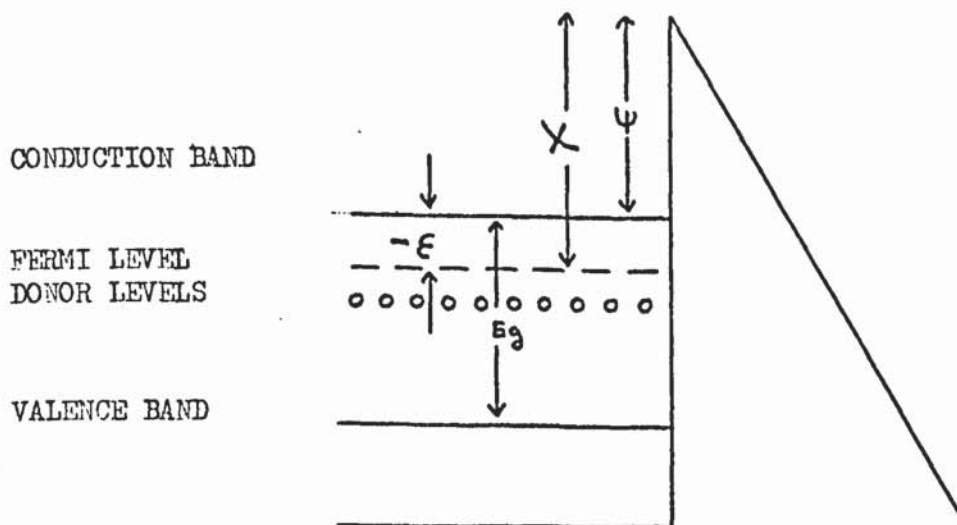
VALENCE BAND

$$\psi = \chi + \epsilon \qquad\qquad 3.44$$

It is assumed that Eg, the band gap energy, is much larger than $|\epsilon|$ so that holes in the valence band need not be considered.

3.7.1 The Transmission Probability

The expression for the transmission probability derived by Nordheim [10] will be used in this chapter. Fig. 3.3 shows the potential barrier in terms of the energy definitions mentioned above. Ex is the "component" of electron energy normal to the barrier and F is the applied field. Nordheim [10] found that for a metal:

$$D(Ex) = \exp\left(-\frac{4k}{3F}\,(\psi - Ex)^{3/2}\,\phi\left(\frac{e\sqrt{F}}{(\psi - Ex)}\right)\right) \qquad 3.45$$

where:
$$k = \left(\frac{8m\pi^2}{h^2}\right)^{1/2} \qquad\qquad 3.46$$

and $\phi(y)$ is a tabulated function [10].

For a semiconductor, which behaves like a dielectric if its dielectric relaxation time is sufficiently large, the image force correction must be multiplied by the factor $(z - 1)/(z + 1)$, where z is the dielectric constant. This leads to the result [26]:

$$D(Ex) = \exp\left(-\frac{4k}{3F}\,(\psi - Ex)^{3/2}\,\phi\left(\left(\frac{z-1}{z+1}\right)^{1/2}\frac{e\sqrt{F}}{(\psi - Ex)}\right)\right) \qquad 3.47$$

3.7.2 The Effect Of Field Penetration

When an electric field is applied to the surface of a metal or a semi-conductor a surface charge is induced. This induced charge has little effect in the case of a metal because the number of additional charges involved is small compared to the number of available conduction electrons. For a semiconductor, however, the position is somewhat different because the number of electrons in the conduction band is relatively small under most conditions. In order to supply the surface charge, therefore, additional conduction electrons must be created by ionisation in the

Fig. 3.3

BOTTOM OF CONDUCTION BAND
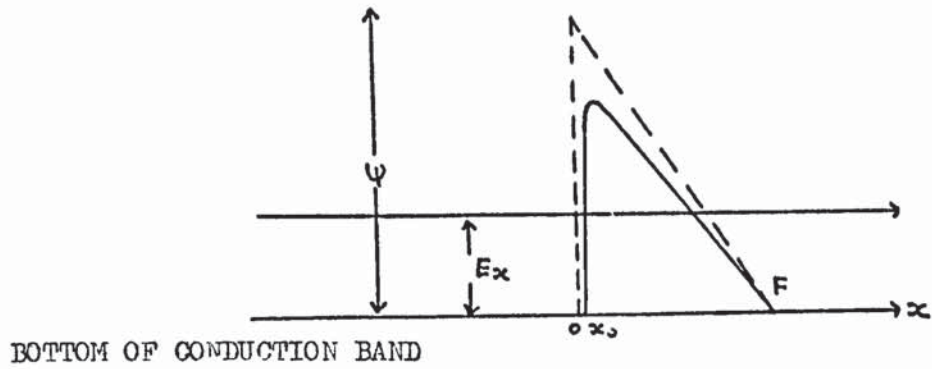
$$\dot{U}(x) = 0 \qquad\qquad \text{FOR } x < x0$$

$$U(x) = -Fx - \frac{e^2}{4x} \qquad\qquad \text{FOR } x > x0$$

$$\text{WHERE } \Psi = Fx0 + \frac{e^2}{4x0}$$

Fig. 3.4
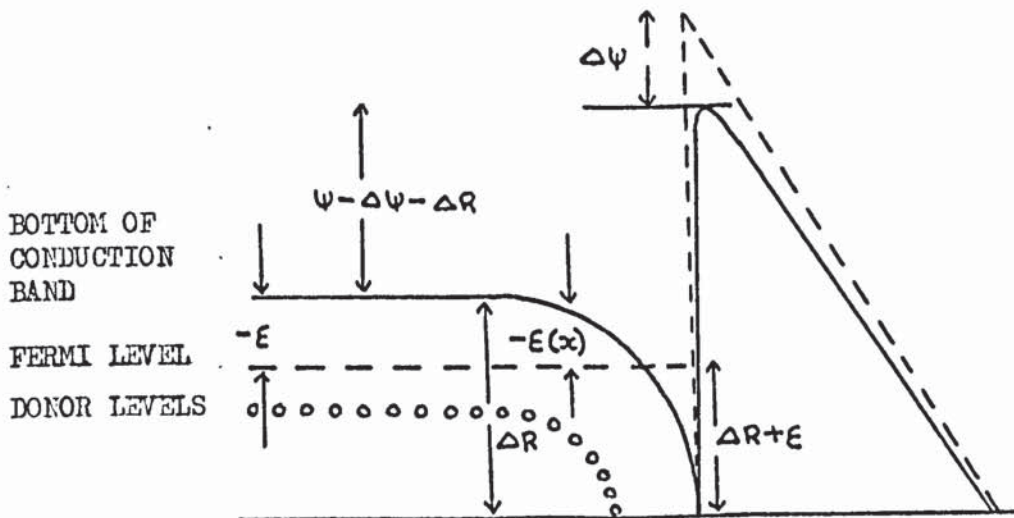
surface region when the semiconductor surface is made negative with respect to the other electrode in the system. It follows that the conduction band must dip down towards the Fermi level as the surface region is approached from the interior and, if the applied field, and hence the surface charge, is large, then the Fermi energy at the surface may even become positive. This degenerate case is shown in Fig. 3.4 where $-\varepsilon$ is the Fermi energy of the bulk material and $\triangle R$ is the maximum change in the Fermi energy.

Let $U(x)$ be the decrease in potential caused by field penetration at a distance x from the surface then:

$$U(0) = \triangle R \qquad\qquad U(x) \to 0 \text{ as } x \to \infty$$

$$\left(\frac{dU}{dx}\right)_{x=0} = \frac{F}{\varepsilon} \qquad\qquad \frac{dU}{dx} \to 0 \text{ as } x \to \infty \qquad\qquad 3.48$$

The density of electrons in the bulk conduction band is given by [83]:

$$n_c = \frac{4\pi (2m)^{3/2}}{h^3} \int_0^\infty \frac{E^{1/2}}{\exp\left(\frac{E-\varepsilon}{kT}\right) + 1} \, dE \qquad\qquad 3.49$$

since the Fermi energy is negative and energy is measured with respect to the bottom of the conduction band. In the surface region there is a decrease in potential of magnitude $U(x)$ at a distance x from the surface. It follows that the density of electrons in the conduction band at x is:

$$n_c(x) = \frac{4\pi}{h^3} (2m)^{3/2} \int_0^\infty \frac{E^{1/2}}{\exp\left(\frac{E - (U(x)+\varepsilon)}{kT}\right) + 1} \, dE \qquad\qquad 3.50$$

The number of ionised donors in the surface region is given by:

$$n_i = n \exp\left(\frac{U(x)}{kT}\right) \qquad\qquad 3.51$$

Poisson's equation for the surface region is, therefore, of the form:

$$\frac{d^2 U(x)}{dx^2} = -\frac{4\pi n e^2}{z}\left(\exp\left(\frac{U(x)}{kT}\right) - \frac{4\pi (2m)^{3/2}}{h^3} \int_0^\infty \frac{E^{1/2}}{1+\exp\left(\frac{E-(U(x)+\varepsilon)}{kT}\right)} dE\right) \quad 3.52$$

and by using the substitutions $x = E/kt$, $y = (U(x) - \epsilon)/kT$, the equation can be reduced to the more convenient expression:

$$\frac{d^2 U(x)}{dx^2} = \frac{-4\pi n e^2}{z}\left(\exp\left(\frac{U(x)}{kT}\right) - \frac{4}{\sqrt{\pi}}\cdot\frac{N_0}{n}\, F_{1/2}\left(\frac{U(x)-\epsilon}{kT}\right)\right) \qquad 3.53$$

where:
$$N_0 = (2\pi m\, kT)^{3/2} \qquad 3.54$$

and
$$F_{1/2}(y) = \int_0^\infty \frac{x^{1/2}}{1 + \exp(x-y)}\, dx \qquad 3.55$$

Integration of equation 3.53 subject to the boundary conditions in 3.48 leads to the result [26]:

$$F^2 = 2\left(\frac{z\,kT}{x_0}\right)^2\left(-1 + \exp-\left(\frac{\Delta R}{kT}\right) + \left(\frac{x_0}{X_0}\right)\frac{2}{\sqrt{\pi}}\int_{\epsilon/kT}^{\frac{\epsilon+\Delta R}{kT}} F_{1/2}(y)\,dy\right) \qquad 3.56$$

where
$$x_0 = \left(\frac{z\,kT}{8\pi n e^2}\right) = X_0\left(\frac{2\,N_0}{n}\right)^{1/2} \qquad 3.57$$

Equation 3.56 can be used to describe degenerate and non degenerate surfaces in the following way:

Degenerate Surface

In this case the conduction band dips below the Fermi level and for extreme degeneracy it is evident from Fig. 3.4 that:

$$\epsilon + \Delta R \gg kT \qquad 3.58$$

and that
$$\Delta R \gg \epsilon \qquad 3.59$$

The integral $F_{1/2}(y)$ (cf. equation 3.55) is very small except for values near the surface when y approaches $(\epsilon + \Delta R)/KT$. This fact can be used to obtain an approximate value for the integral

$$\int_{\epsilon/kT}^{\frac{\epsilon+\Delta R}{kT}} F_{1/2}(y)\,dy \qquad 3.60$$

which occurs in equation 3.56 because, near the surface:

$$F_{1/2}(y) \simeq \int_0^\infty x^{1/2}\, dx \qquad 3.61$$

Also, since $(\epsilon + \Delta R)/KT$ is much greater than ϵ/kT, the limits of integral 3.60 are essentially the same as those of integral 3.61, therefore integral 3.60 is approximately:

$$\int_0^{\frac{\epsilon+\Delta R}{kT}} \frac{2}{3} x^{3/2} \, dx = \frac{4}{15} \left(\frac{\epsilon + \Delta R}{kT} \right)^{5/2} \qquad 3.62$$

By substituting equation 3.62 into equation 3.56 and ignoring the small ionic terms, the result:

$$\epsilon + \Delta R \simeq \nu F^{4/5} \qquad 3.63$$

is obtained where:

$$\nu = \frac{1}{z^{2/5}} \left(\frac{225}{\pi^4 2^{11}} \frac{h^6}{m^3 e^4} \right)^{1/5} \qquad 3.64$$

Thus, by combining equations 3.63 and 3.58, it is evident that strong degeneracy occurs when:

$$F \gg \left(\frac{16}{15\sqrt{\pi}} \right) \frac{z \, kT}{X_0} \qquad 3.65$$

At room temperatures, this equation reduces to:

$$F \gg z^{1/2} \, 1.2 \times 10^6 \, \text{VOLT.CM}^{-1}$$

which, for cadmium sulphide, means that F must be greater than 4×10^6 volt cm^{-1}.

Non Degenerate Surface

For a strongly non degenerate surface the condition:

$$- (\Delta R + \epsilon) \gg kT \qquad 3.66$$

prevails. This can again be seen by reference to Fig. 3.4. Under condition 3.66 it can be shown that [26] equation 3.56 reduces to:

$$\sinh \left(\frac{\Delta R}{2kT} \right) = \frac{F x_0}{2z kT} \qquad 3.67$$

Therefore, by using equations 3.66 and 3.67, it is seen that the surface region is far from degenerate when:

$$F \ll (z \, kT)/X_0 \qquad 3.68$$

- 63 -

which reduces to:

$$F \ll Z^{1/2} \cdot 1 \cdot 5 \times 10^6 \text{ VOLT. CM.}^{-1} \qquad 3.69$$

at room temperature.

The conditions derived above for the fields necessary to cause degeneracy and non degeneracy in a semiconductor apply to an ideally perfect semi-conductor surface only. In a real semiconductor, surface states can also play an important part in determining the magnitude of the emitted current density.

3.7.3 The Effect Of Surface States

Consider the following experiment. A parallel plate capacitor is made by separating a thin layer of n type germanium from a metal plate by means of an insulator. When a potential difference is applied to the capacitor, the conductivity of the germanium will change if the charge induced in it is free to move. Shockley and Pearson [84] performed this experiment and found that only about one tenth of the induced charge caused a change in conductivity. This result was explained by Bardeen [85] who supposed that the immobile fraction of charge resides in electronic states which are peculiar to the surface of the semi-conductor. These states, which can even exist in the normally forbidden gap, may be caused by adsorbed atoms or as a consequence of the sudden departure of the crystal potential from periodicity at the surface. Electrons from the conduction band of a neutral semiconductor crystal will tend to fill the surface states until a potential drop is built up in the surface region such that the highest filled state is at an energy level corresponding to the Fermi level of the bulk material. This then leads to a region of depleted conductivity under the surface, as shown in Fig. 3.5.

Fig. 3.6 depicts the more general case of field emission from a semi-conductor with both donor and acceptor states. The donor states are assumed to be at an energy level E_d and the acceptor states at an energy

Fig. 3.5

(a) BEFORE EQUILIBRIUM

CONDUCTION BAND

DONOR LEVELS

VALENCE BAND

EMPTY SURFACE
STATE LEVELS
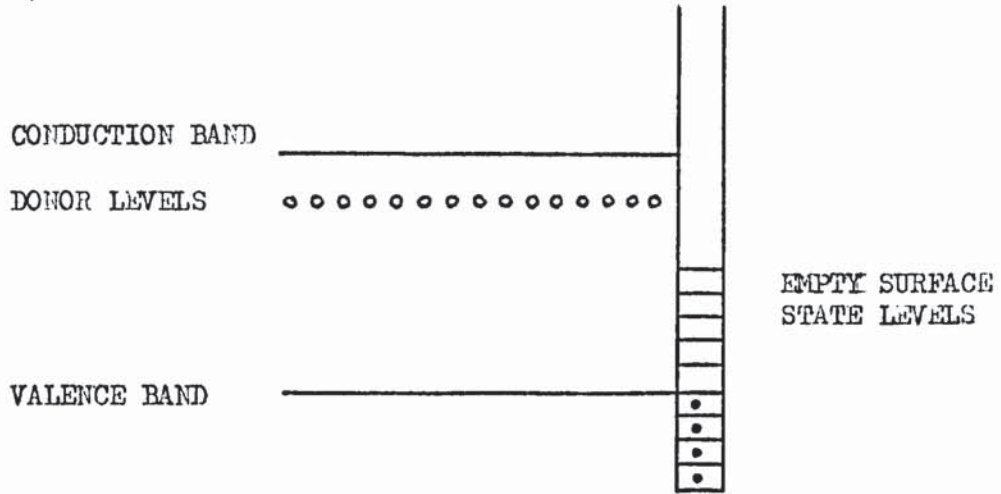
(b) EQUILIBRIUM ESTABLISHED

V_0

CONDUCTION BAND

FERMI LEVEL

DONOR LEVELS

VALENCE BAND

x_0

+ IONISED DONORS

• FILLED SURFACE
STATE LEVELS
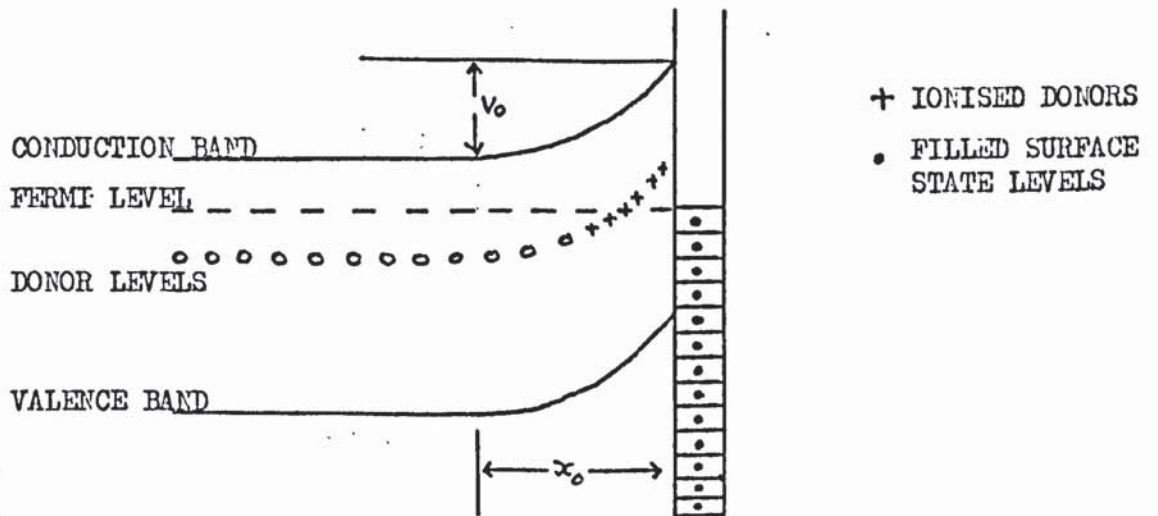
Fig. 3.6

(a) NO APPLIED FIELD

(b) WITH APPLIED FIELD

level Ea below the conduction band respectively. Let there be Nd donors and Na acceptors per unit area of semiconductor surface. The donor states are neutral when occupied by electrons and the acceptor states are neutral when occupied by holes. The following definitions will also be used:

(1) For no applied Field   ndo = the number of electrons on donor states.

pao = the number of holes on acceptor states.

E_0/z = The electric field at the top of the barrier caused by the surface states.

(2) When Field applied   nd = the number of electrons on donor states

pa = the number of holes on acceptor states

E/z = the electric field at the top of the barrier caused by surface states.

From Gauss' theorem:

$$\frac{F+E}{4\pi z e^2} = -(Nd - nd) + (Na - pa) \qquad 3.70$$

and when F = 0:

$$\frac{E_0}{4\pi z e^2} = -(Nd - nd_0) + (Na - pao) \qquad 3.71$$

If exp $(Eg/2kT)$ is much greater than exp $((\phi_0 - \epsilon)/kT)$ then it is obvious from Fig. 3.6 that the acceptor states will be occupied by electrons. In this case, pao = pa = 0 and equations 3.70 and 3.71 can be combined to form the expression:

$$\frac{F+E}{E_0} = \frac{(nd/Nd) + \beta}{\alpha + \beta} \qquad 3.72$$

where $\qquad \alpha = ndo/Nd \qquad\qquad 3.73$

and $\qquad \beta = (Na - Nd)/Nd \qquad\qquad 3.74$

The value of nd can be determined by considering the dynamic equilibrium of the donor states. If exchange of electrons with the conduction band is considered, an expression of the form:

$$en\left(\frac{kT}{2\pi m}\right)^{1/2} \exp\left(\frac{-\phi_0}{kT}\right) \exp\left(\frac{V}{kT}\right) p(Nd - nd) = \frac{e\,nd}{\tau} \qquad 3.75$$

- 65 -

can be obtained [26], where p is the average trapping cross section of the donor states and γ is the lifetime of the electrons in them. When F = 0:

$$en\left(\frac{kT}{2\pi m}\right)^{1/2} \exp\left(-\frac{\phi_0}{kT}\right) p\left(Nd - ndo\right) = \frac{e\ ndo}{\gamma} \qquad 3.76$$

Division of equation 3.75 by equation 3.76 leads to the result:

$$\exp\left(\frac{V}{kT}\right) = \frac{nd}{ndo} \quad \frac{(1-\alpha)}{1-\left((\alpha nd)/ndo\right)} \qquad 3.77$$

hence:

$$\frac{nd}{Nd} = \left(1 + \left(\frac{1-\alpha}{\alpha}\right)\exp\left(-\frac{V}{kT}\right)\right)^{-1} \qquad 3.78$$

and finally:

$$\frac{(\alpha+\beta)}{E_0}(F+E) = \left(1 + \left(\frac{1-\alpha}{\alpha}\right)\exp\left(\frac{-V}{kT}\right)\right)^{-1} + \beta \qquad 3.79$$

There is also a relationship between E and V which depends on the model used for the barrier.

3.8 The Emission Current Density

3.8.1 Emission With No Surface States And No Field Penetration

Although this is a rather idealised case, it is useful in that it provides a reference current density against which the current densities corresponding to emission from a semiconductor with surface states, or with field penetration, can be compared.

The electrons in the conduction band obey Maxwell Boltzmann Statistics provided the bottom of the conduction band is sufficiently above the Fermi level. In this case, by following arguments similar to those of section 3.3, it can be shown that the number of electrons with "components" of energy normal to the barrier in the range Ex to Ex Ex incident on the barrier per unit area, per unit time, is given by:

$$n_x(Ex) = \frac{n}{kT}\left(\frac{kT}{2\pi m}\right)^{1/2}\exp\left(\frac{-Ex}{kT}\right) \qquad 3.80$$

where n is the number of electrons per unit volume in the conduction band. The total emitted current density is, therefore:

- 66 -

$$I = e \int_0^\infty n_x(E_x) \, D(E_x) \, dE_x \qquad 3.81$$

By substituting equations 3.80 and 3.74 into equation 3.81, the final result is:

$$I = ne \left(\frac{kT}{2\pi m}\right)^{1/2} \exp\left(-\frac{4k}{3F}\psi^{3/2} \phi\left(\left(\frac{z-1}{z+1}\right)^{1/2}\frac{e\sqrt{F}}{\psi}\right)\right) \qquad 3.82$$

3.8.2 Emission With Field Penetration

If penetration of the applied field does not lower the conduction band sufficiently to invalidate Maxwell Boltzmann Statistics, then equation 3.82 can still be used to obtain the emitted current density provided n is multiplied by exp $(\Delta R/kT)$, where ΔR is given by equation 3.67. When the conduction band is lowered so that the surface of the semiconductor becomes degenerate, $n_x(E_x)$ takes the value corresponding to a metal because the electrons now obey Fermi Dirac Statistics. Under these conditions, Fig. 3.7 shows that the expression for the emitted current density can be obtained by using methods similar to those employed by Fowler and Nordheim[9][10] and by replacing the work function by $\psi - (\Delta R + \mathcal{E})$, which equals $\psi - \nu F^{4/5}$, and the Fermi energy by $\Delta R + \mathcal{E}$, which equals $\nu F^{4/5}$ (c.f. equation 3.63).

The equation for the emitted current density is of the form[26]:

$$I = \frac{e}{8\pi h} \frac{F^2}{(\psi - \nu F^{4/5})} \exp\left(\frac{-4k(\psi - \nu F^{4/5})^{3/2}}{3F}\right) \phi(u) \times$$

$$\left(1 - \left(1 + \frac{2k\nu F^{4/5}(\psi - \nu F^{4/5})^{1/2}}{F}\right)\exp\left(\frac{2k\nu F^{4/5}(\psi - \nu F^{4/5})^{1/2}}{F}\right)\right) \qquad 3.83$$

where $\quad u = \left(\frac{z-1}{z+1}\right)^{1/2}\frac{e\sqrt{F}}{(\psi - \nu F^{4/5})} \quad$ and $\quad F \ll \frac{2k}{3}(\psi - \nu F^{4/5})\phi$

3.8.3 Emission With Surface States

Electron emission from a semiconductor with surface states can be treated in a way analogous to that of emission with field penetration when the lowering of the conduction band is too small to invalidate Maxwell

Fig. 3.7

(a) FOWLER NORDHEIM MODEL OF A METAL

ϕ

FERMI LEVEL

E_F

BOTTOM OF CONDUCTION BAND
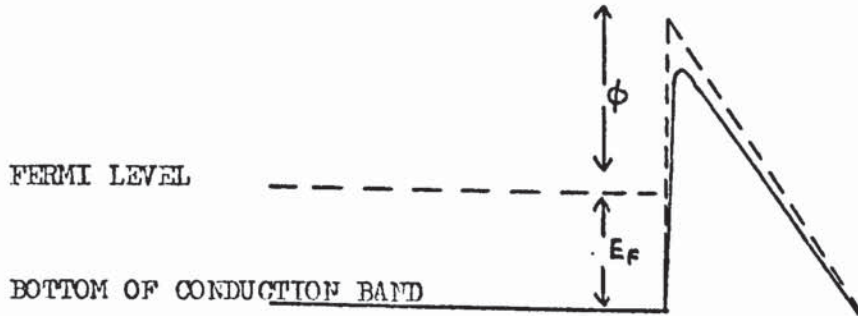
ϕ = WORK FUNCTION

E_F = FERMI ENERGY

(b) DEGENERATE SEMICONDUCTOR

$(\Psi - (\Delta R + \varepsilon))$ Ψ

FERMI LEVEL $-\varepsilon$

ΔR

$\Delta R + \varepsilon$

BOTTOM OF CONDUCTION BAND
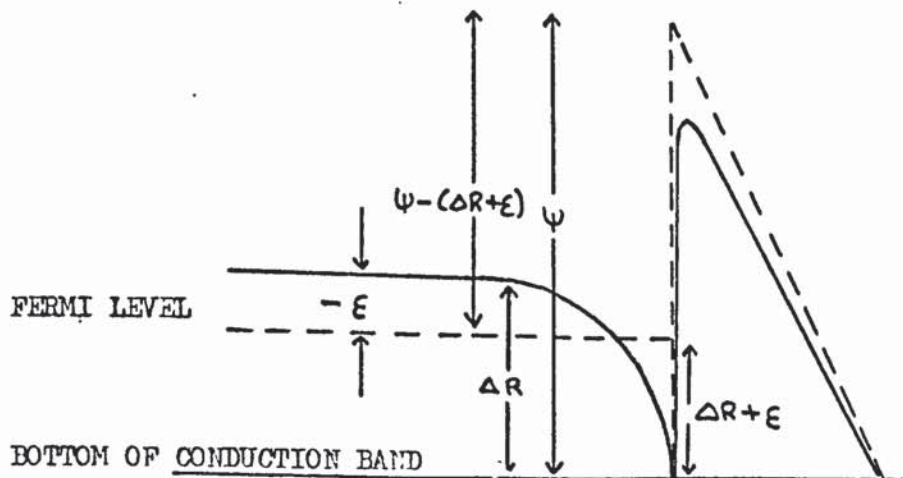
Boltzmann Statistics. The quantity $(\phi_0 - V)$ is a measure of the bending of the conduction band and is, therefore, a measure of the number of electrons per unit volume in the surface region. Equation 3.81 can still be used to obtain the total emitted current density but nx(Ex) now takes the form:

$$n_x(E_x) = \frac{n}{kT} \exp\left(\frac{V - \phi_0}{kT}\right)\left(\frac{kT}{2\pi m}\right)^{1/2} \exp\left(-\frac{E_x}{kT}\right) \qquad 3.84$$

where n is the number of electrons per unit volume in the surface region when the conduction band is flat. It follows that:

$$I = I_0 \exp\left(\frac{V}{kT}\right) \exp\left(\frac{-4k}{3F}\psi^{3/2}\phi\left(\left(\frac{z-1}{z+1}\right)^{1/2}\frac{e\sqrt{F}}{\psi}\right)\right) \qquad 3.85$$

where:
$$I_0 = en \exp\left(-\frac{\phi_0}{kT}\right)\left(\frac{kT}{2\pi m}\right)^{1/2} \qquad 3.86$$

3.8.4 Numerical Results

Stratton[26] has computed numerical values of the total field emitted current density for silicon carbide under the conditions described by equations 3.82, 3.83 and 3.85. His results are shown in Fig. 3.8. For a high density of surface states, there is a large internal potential drop which causes the conduction band to bend upwards in the surface region. The current at low fields is, therefore, small because the quantity $(V - \phi_0)$ in equation 3.85 is large and negative.

When the applied field is increased, the conduction band starts to fall as V becomes larger. The current, therefore, increases rapidly until V equals ϕ_0 when the conduction band in the surface region is flat. The emission is then governed by equation 3.82 and the applied field just balances the field, Fo, caused by the surface states.

At even higher fields penetration occurs and the conduction band starts to bend downwards. For small bending, the emitted current density is given by equation 3.82 multiplied by $\exp(\Delta R/kT)$, where ΔR is given by equation 3.67 but with the reduced field F - Fo replacing F. When

Fig. 3.8
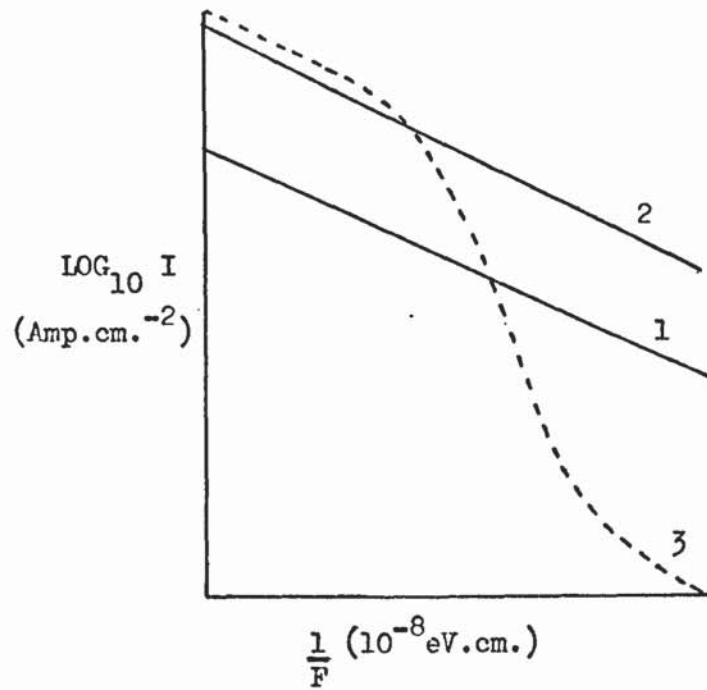
1. EMISSION WITH IMAGE FORCE BUT WITH NO SURFACE STATES AND NO FIELD PENETRATION

2. EMISSION WITH IMAGE FORCE AND FIELD PENETRATION BUT WITH NO SURFACE STATES

3. EMISSION WITH IMAGE FORCE AND SURFACE STATES BUT WITH NO FIELD PENETRATION

F - Fo is large, equation 3.83 must be used in which F - Fo again replaces F except in the expression corresponding to the transmission probability. The above theory is not strictly true because the expressions for the band bending caused by field penetration were derived by assuming that the Fermi level remains constant throughout the emitter and this is only true under conditions where zero or negligible conduction current flows. When the conduction current is appreciable, the energy level diagram is modified as shown in Fig. 3.9. The definitions are essentially the same as those in the previous work except that $U(x)$ is the distance from the top of the barrier to the bottom of the conduction band and $\phi(x)$ is the distance from the top of the barrier to the Fermi level. $\varepsilon(x)$ is again the distance from the bottom of the conduction band to the Fermi level and εs is the value of this function at the surface while $\varepsilon\infty$ is the value in the bulk of the material. There are also donor and acceptor levels at energies εd and εa below the conduction band respectively.

The emitted current density from the conduction band is again given by:

$$J = \frac{e}{8\pi h} \frac{F^2}{\phi} \exp\left(-\frac{4k\phi^{3/2}}{3F}\phi(u)\right)\left(1-\left(1+\frac{2k\phi^{1/2}\varepsilon_S}{F}\right)\left(\exp\left(\frac{2k\phi^{1/2}\varepsilon_S}{F}\right)\right)\right) \qquad 3.87$$

when $\varepsilon_S > 0$ and by:

$$J = ne\left(\frac{kT}{2\pi m}\right)^{1/2} \exp\left(-\frac{4k\psi^{3/2}}{3F}\phi(u)\right)\exp\left(\frac{\varepsilon_S}{kT}\right) \qquad 3.88$$

when $\varepsilon_S < 0$. As in previous work:

$$u = \left(\frac{z-1}{z+1}\right)^{1/2}\frac{e\sqrt{F}}{\phi} \quad \text{and} \quad \phi = \psi - \varepsilon_S$$

but now the degeneration parameter, ε_S, has a value appropriate to the passage of conduction current instead of $\varepsilon + \triangle R$, which was the value derived from equations 3.63 and 3.67 for the constant Fermi level case.

Poisson's equation for the emitter is of the form:

Fig. 3.9

(a) P TYPE SEMICONDUCTOR
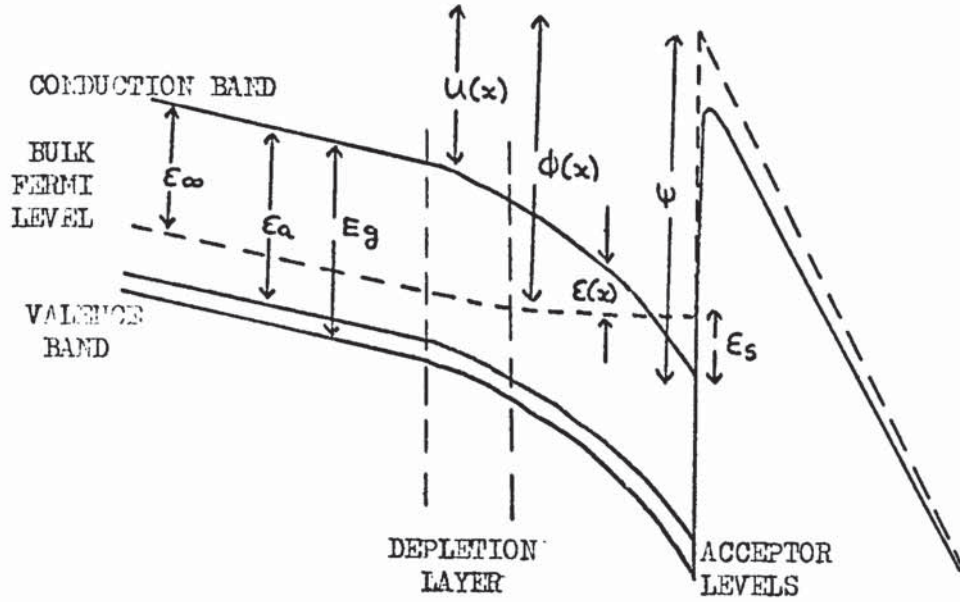
(b) N TYPE SEMICONDUCTOR
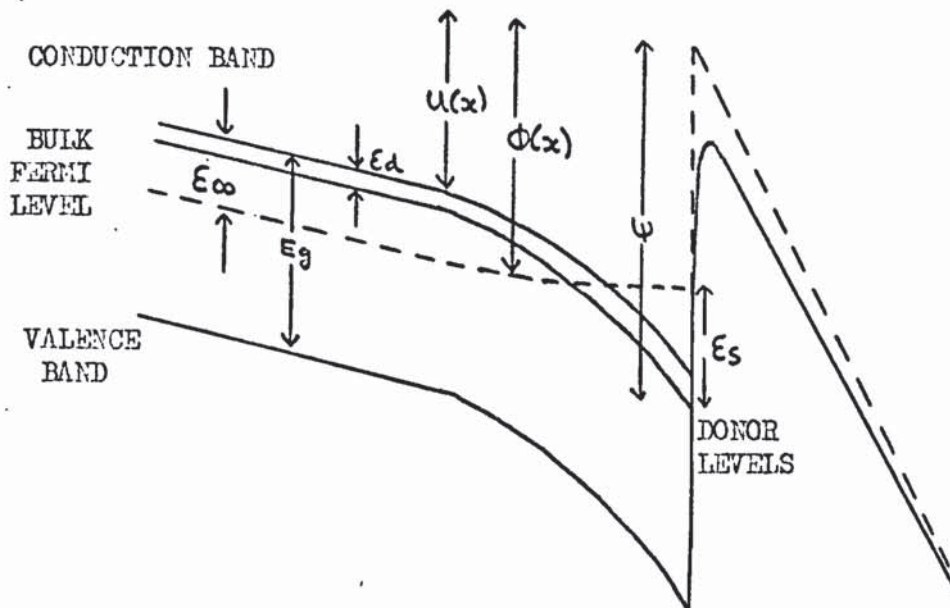
$$\frac{d^2 u(x)}{dx^2} = -\frac{4\pi e^2}{z} \left(n - p + N^- - N^+ \right) \qquad\qquad 3.89$$

where n, p, N^- and N^+ are the concentrations of electrons, holes and charged impurity centres. The conduction current density is given by by:

$$J = - \left(\mu_n n + \mu_p p \right) \frac{d\phi(x)}{dx} \qquad\qquad 3.90$$

where μ_n and μ_p are the mobilites of the electrons and holes respectively.

It is clear, by considering the energy levels in Fig. 3.9 in conjunction with the arguments presented in reference 77, that:

$$n = A F_{1/2} \left(\frac{\varepsilon}{kT} \right) \qquad\qquad 3.91$$

and

$$p = \alpha A F_{1/2} \left(\frac{-\varepsilon + Eg}{kT} \right) \qquad\qquad 3.92$$

where Eg is the energy gap, α equals $(mp/mn)^{3/2}$, $A = 4\pi (2 mn kT/h^2)^{3/2}$, mn and mp are the effective masses of the electrons and holes and $F_{\frac{1}{2}}$ is the integral defined by equation 3.55. The concentrations of ionised donors and acceptors are given by:

$$N^+ = Nd \left(1 + 2\exp\left(\frac{\varepsilon - \varepsilon d}{kT} \right) \right)^{-1} \qquad\qquad 3.93$$

and

$$N^- = Na \left(1 + 2\exp\left(\frac{\varepsilon a - \varepsilon}{kT} \right) \right)^{-1} \qquad\qquad 3.94$$

where Nd and Na are the concentrations of donor and acceptor sites. In the bulk material, the gradient of the bottom of the conduction band with respect to distance, dU/dx, is the same as the gradient of the Fermi level with respect to distance, therefore:

$$\left. \frac{dU}{dx} \right|_{x\to-\infty} = \left. \frac{d\varepsilon(x)}{dx} \right|_{x\to-\infty} = \frac{J}{\mu_n n\infty + \mu_p p\infty} \qquad\qquad 3.95$$

where n and p are the bulk concentrations of electrons and holes.

Near the surface, however, the gradient of the bottom of the conduction band is affected by field penetration and is not, therefore, the same as that of the Fermi level. If the internal field in this region is $E(x)$, it follows that:

$$\left. \frac{du}{dx} \right|_{x \to 0} = e\, E(x)_{x \to 0} = -\frac{e\,F}{z} \tag{3.96}$$

Let $y(x) = \varepsilon(x)/kT$ then, because expressions 3.91, 3.92, 3.93 and 3.94 are functions of $\varepsilon(x)/kT$, they can be rewritten in the form:

$$Q(y) = e\,(n - p + N^{-} - N^{+}) \tag{3.97}$$

and

$$b(y) = n + \frac{\mu_p}{\mu_n}\,p \tag{3.98}$$

Equations 3.89 and 3.90 can then be expressed in terms of $y(x)$ instead of x. If boundary condition 3.95 is used and the field strength in the sample is designated $E(y)$, it can be shown that [86]:

$$e^2 E(y) - \frac{J^2}{\mu_n^2\, b^2(y\infty)} = \frac{8\pi e\,kT}{z}\,\big(H(y) + G(y)\big) \tag{3.99}$$

and, by using boundary condition 3.96, expression 3.99 can be rewritten:

$$\frac{e^2 F^2}{z^2} - \frac{J^2}{\mu_n^2\, b^2(y\infty)} = \frac{8\pi e\,kT}{z}\,\big(H(ys) + G(ys)\big) \tag{3.100}$$

where:

$$ys = \varepsilon s / kT = y(x)\big|_{x \to 0} \tag{3.101}$$

$$y\infty = y(x)\big|_{x \to \infty} \tag{3.102}$$

$$H(y) = \int_{y\infty}^{y} Q(y)\,dy \tag{3.103}$$

and

$$G(y) = J \int_{y\infty}^{y} Q(y)\,\frac{dy}{e\,\mu_n\, b(y)\, E(y) - J} \tag{3.104}$$

The dependence of the space co-ordinate x on $y(x)$ follows from equation 3.90 and is of the form:

- 71 -

$$x = kT \int_{y}^{ys} \frac{e\,\mu_n\, b(y)}{e\,\mu_n\, b(y)\, E(y) - J}\, dy \qquad 3.105$$

3.8.5 The Zero Conduction Current Approximation

It should be noted that the zero conduction current approximation is used when the Fermi level is constant throughout the material and that its use does not imply that the emitted current density is zero. When the Fermi level is constant, equations 3.99 and 3.100 reduce to:

$$e^2\, E^2(y) = \frac{8\pi e\, kT}{z}\, H(y) \qquad 3.106$$

and
$$\frac{e^2\, F^2}{z^2} = \frac{8\pi e\, kT}{z}\, H(ys) \qquad 3.107$$

and $H(y)$ takes the form:

$$H(y) = e\, A \left(\frac{2}{3}\, F_{1/2}(y) + \alpha\, \frac{2}{3}\, F_{3/2}(-y - \omega g) + \right.$$

$$\left. \frac{Na}{A} \ln\left(1 + \frac{1}{2} \exp(y - \omega a) \right) + \frac{Nd}{A} \ln\left(1 + \frac{1}{2} \exp(\omega d - y) \right) \right) \qquad 3.108$$

Where $\omega g = Eg/kT$, $\omega a = \epsilon a/kT$ and $\omega d = \epsilon d/kT$. It can be shown [86] that for not very strongly doped semiconductors, the first part of equation 3.108 is much larger than the others, therefore, from equations 3.107 and 3.108:

$$F^2 = 8\pi z\, A\, kT\, \frac{2}{3}\, F_{3/2}(ys) \qquad 3.109$$

Degeneration starts to occur when the conduction band touches the Fermi level i.e. when $\epsilon(x)$, and hence $y(x)$, equals zero. At the surface, $y(x) = ys$, therefore, if the temperature is expressed in degrees absolute and the value $ys = 0$ is substituted into equation 3.109, an expression for the applied field corresponding to the onset of degeneracy can be obtained. It is of the form:

$$Fd = 1.44 \times 10^6 \, z^{1/2} \left(\frac{mn}{m}\right)^{3/2} \left(\frac{T}{300}\right)^{5/4} \text{VOLT.cm.}^{-1} \qquad 3.110$$

It is interesting to note that the values of Fd calculated in reference (86) are considerably smaller than the fields necessary for appreciable field emitted current density which suggests that degeneracy occurs in most field emitting semiconductors.

3.8.6 Emission From A P Type Semiconductor

When a p type semiconductor is subjected to the action of a high electro static field a depletion layer is formed just before the inversion layer at the surface (see Fig. 3.9 (a)). In this layer, the concentration of mobile carriers, and hence the conductivity, is much smaller than in the rest of the emitter so that the conduction and emission currents are both limited.

Fig. 3.9 (a) shows that the Fermi level is approximately half way between the conduction and valence bands in the depletion layer, therefore, $\varepsilon(x) = Eg/2$. This value of $\varepsilon(x)$ occurs at a value of y given by [86]:

$$ym = -\frac{\omega_g}{2} + \ln \sqrt{a} \qquad 3.111$$

where
$$a = \left(\frac{\mu_p}{\mu_n}\right)\left(\frac{m_p}{m_n}\right)^{3/2} \qquad 3.112$$

The corresponding effective mobile carrier density, $b(ym)$, can be found by substituting equation 3.111 into equation 3.98, whereupon:

$$bm = b(y_m) = A\sqrt{\pi a} \exp\left(-\frac{\omega_g}{2}\right) \qquad 3.113$$

It follows from equation 3.104 that there must be a limiting value of the current density at all points in the emitter given by the condition:

$$J \le e \mu_n \, b(y) \, E(y) \qquad 3.114$$

otherwise the integral in equation 3.104 would become infinite. In the depletion region, therefore:

$$J \leq e \mu_n \, b \, (y_m) \, E \, (y_m)$$ 3.115

and, because $E(y) \leq F/z$

$$J \leq e \mu_n \, b(y_m) \frac{F}{z} = e \mu_n \, A \sqrt{\pi a} \, \exp \left(-\frac{\omega_g}{2} \right) \frac{F}{z}$$ 3.116

The emitted current density will consequently rise until it reaches the value given by equation 3.116 and should then saturate.

3.8.7 The Effect Of The Internal Field

So far the effect of the internal field on the carrier mobility has not been considered. In general, this can be represented by the relation:

$$\mu(E) = \mu_0 \left(\frac{E_0}{E} \right)^{\sigma}$$ 3.117

where the exact value of σ depends on the machanism which actually restricts the mobility, E is the field above which the relevant mechanism occurs and μ_0 is the value of the mobility corresponding to E0.

For most semiconductors with internal fields in the range 10 to 100 volt. cm^{-1} the mobility is restricted by the generation of optical phonons and $\sigma = 1$ whereas, for weaker fields, scattering by acoustic phonons occurs and $\sigma = \frac{1}{2}$. Since calculation shows that the internal fields in a p type semiconductor may exceed 10^3 volt. cm^{-1} at the transition to the saturation region, it may be assumed that $\sigma = 1$.

Equation 3.116 then becomes:

$$J \leq e \, A \sqrt{\pi a} \, (\mu_0 E_0) \exp \left(-\frac{\omega_g}{2} \right)$$ 3.118

and the limiting value of J does not depend on the field strength. Under these conditions the transition should be abrupt and the saturation current density should approximately equal the limiting current density. When the relationship between σ and E is weaker than that given by

equation 3.117, the transition is more gradual because the limiting current density may then be field dependent.

3.8.8 Emission From An N Type Semiconductor

In an n-type semiconductor the concentration of mobile carriers decreases smoothly from a maximum value at the surface to a minimum value, $n\infty$ in the bulk material (see Fig. 3.9(b)). In general, the equation for the limiting current density is of the form:

$$J_{lim} = \frac{e \, \mu_n \, b(y) \, F}{z} \qquad\qquad 3.119$$

where $b(y)$, the effective mobile carrier density, is given by equation 3.98. Since the majority carriers in an n type semiconductor are electrons, it follows from equation 3.98 that $b(y)$ can be replaced by $n\infty$ in equation 3.119, thus:

$$J_{lim} = \frac{e \, \mu_n \, n\infty \, F}{z} \qquad\qquad 3.120$$

If $\mu(E)$ is not taken into account, the emitted current will saturate and the value of the saturation current density will be proportional to the bulk carrier concentration. This, in turn, is given by the equation [86]:

$$n\infty = \left(\frac{2\pi m k T}{h^2}\right)^{3/4} N_d^{1/2} \exp\left(\frac{-\mathcal{E}_d}{kT}\right) \qquad \text{for } T < T_o \qquad 3.121$$

$$n\infty = N_d \qquad\qquad\qquad\qquad \text{for } T > T_o \qquad 3.122$$

$$T_o = 750 \left(\frac{m}{m_n}\right)\left(\frac{N_d}{10^{18}}\right)^{2/3} \qquad\qquad 3.123$$

when $T < T_o$, therefore, the saturation current density depends on temperature and when $T > T_o$ it is independent of temperature. Baskin et.al.[86] have performed a numerical analysis of the equations used in this section. Their results show that for a p type semiconductor the transition to the saturation region is very dependent on $\mu(E)$ and for an n type semiconductor the magnitude of the saturation current is strongly dependent on the bulk carrier concentration. They were also

able to obtain, as a function of the applied field, values of the width of the depletion layer in a p type semiconductor and of the electron concentration in the surface of an n type semiconductor.

3.9 The Total Energy Distribution Of The Field Emitted Electrons

It will again be assumed that the x direction is the direction perpendicular to the potential barrier and that the probability of escape depends only on the "component" of electron energy in this direction. Then if:

$$E_\perp = \frac{1}{2m} \left(P_y^2 + P_z^2 \right) \qquad 3.124$$

it follows that the escape probability is $D(E - E_\perp)$. The component of electron velocity in the x direction is given by:

$$V_x = \frac{\partial E}{\partial P_x} \qquad 3.125$$

and if $F(E)$ is the electron energy distribution function, it will be obvious from arguments similar to those presented in section 3.3 that:

$$P_T(E)\,dE = \frac{2}{h^3} F(E) \int_E^{E+dE} D(E-E_\perp) \frac{\partial E}{\partial P_x}\,dP_x\,dP_y\,dP_z \qquad 3.126$$

where $P_T(E)$ is the total energy distribution. Equation 3.126 can be rewritten in the form:

$$P_T(E) = \frac{2}{h^3} F(E) \iint_{P_y\,P_z} D(E-E_\perp)\,dP_y\,dP_z \qquad 3.127$$

The integral must be performed over all the possible values of P_y and P_z which, together with P_x, form the total momentum P_T. In the case of free electrons, P_T is the momentum which corresponds to the total energy value E since the electron energy is at all times proportional to the square of the electron momentum. The solution of equation 3.127 can be obtained by changing to the polar co-ordinates θ and $P_\perp$ in a plane perpendicular to the x direction. Fig. 3.10 shows that for free electrons and any value of θ, $P_\perp$ varies from 0 to P_T as $E_\perp$ varies from 0 to E. The value of momentum, P_T, may not correspond to the same value of

- 76 -

Fig. 3.10

SURFACE OF CONSTANT MOMENTUM
IN MOMENTUM SPACE
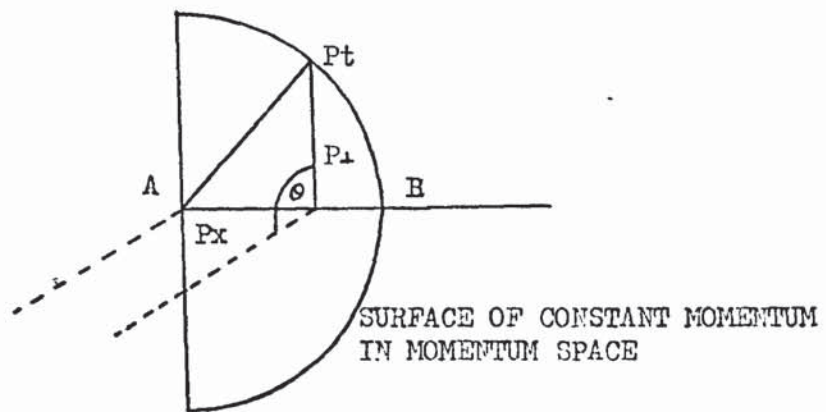
AT A, Px = 0, P⊥ = Pt
AT B, Px = Pt, P⊥ = 0

electron energy, E, in the case of electrons in the conduction band of a semiconductor because the square of electron momentum is then not necessarily proportional to the electron energy. It follows that as $P\perp$ varies from 0 to PT, $E\perp$ varies from 0 to some energy value $Em(E, \Theta)$ which is a function of Θ also because the energy-momentum relationship may be different in different directions in the material. The effective mass, mn, of electrons in the conduction band is a measure of the relationship between energy and momentum and from the definition of effective mass [87], it is obvious that:

$$Em(E, \Theta) = \frac{P_T^2}{2m_n} \qquad 3.128$$

whereas for free electrons:

$$E = P_T^2 / 2m \qquad 3.129$$

therefore:
$$Em(E, \Theta) = \left(\frac{m}{m_n}\right) E \qquad 3.130$$

After changing to polar co-ordinates equation 3.127 becomes

$$P_T(E) = \frac{2}{h^3} F(E) \int_0^{2\pi} d\Theta \int_0^{P_T} D(E - E\perp) \cdot P\perp \, dP\perp \qquad 3.131$$

and it can be seen from equation 3.124 that $mdE = P\perp \, dP\perp$ therefore:

$$P_T(E) = \frac{2m}{h^3} F(E) \int_0^{2\pi} d\Theta \int_0^{Em(E,\Theta)} D(E - E\perp) \, dE\perp \qquad 3.132$$

Equation 3.132 can be rewritten in the form:

$$P_T(E) = \frac{2m}{h^3} F(E) \left(\int_0^{2\pi} d\Theta \int_0^{E} D(E - E\perp) dE\perp - \int_0^{2\pi} d\Theta \int_0^{E - Em(E,\Theta)} D(E - E\perp) dE\perp \right) \qquad 3.133$$

so that:

$$P_T(E) = kF(E)\left(\int_0^E D(Ex)\,dEx - \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^{E-Em(E,\theta)} D(Ex)\,dEx \right) \qquad 3.134$$

where $\quad K = \dfrac{4\pi m}{h^3} \qquad$ and $\quad D(Ex) \equiv D(E-E\perp) \qquad 3.135$

3.9.1 The Transmission Probability D(Ex)

The expression for the transmission probability can be obtained by means of the W.K.B. approximation [88]. It is of the form:

$$D(Ex) = \exp(-B(Ex)) \qquad 3.136$$

where $\qquad B(Ex) = 2\left(\dfrac{(2m)^{1/2}}{\hbar}\right)\int (\phi(x) - Ex)^{1/2}\,dx \qquad 3.137$

and $\phi(x)$ is the barrier potential energy measured with respect to the bottom of the conduction band. Equation 3.137 is not generally used in conjunction with other equations because the mathematics involved becomes prohibitive. Instead, it is usual to expand it about some convenient energy, say ε', by means of a Taylor Series, and to use this expansion in place of the full expression. After this process B(Ex) becomes:

$$B(Ex) = b(\varepsilon') - (Ex - \varepsilon')c(\varepsilon') + (Ex - \varepsilon')^2 a(\varepsilon') + \ldots \qquad 3.138$$

where $\quad b(\varepsilon') = B(\varepsilon'),\ c(\varepsilon') = -\dot{B}(\varepsilon')$ and $a(\varepsilon') = \ddot{B}(\varepsilon')/2 \qquad 3.139$

and the quantities a, b and c, corresponding to a barrier which includes the image force correction, are given by [45]:

$$a = 2.56\left(\theta^{1/2}(F/10^7)(1 - (wi/\theta)^2)\right)^{-1} v(wi/\theta) \qquad 3.140$$

$$b = 6.83\,\theta^{3/2}(10^7/F)\,v(wi/\theta) \qquad 3.141$$

and
$$c = 10.25 \, \Theta^{1/2} (10^7/F) \, t \, (\psi_i/\Theta)$$
3.142

where
$$\Theta = \psi - \epsilon'$$
3.143

ψ is the electron affinity of the emitter and v and t are tabulated functions [89]. The depression of the barrier height, ψ_i, can be calculated from the expression:

$$\psi_i = 1.2 \, (F/10^7)^{1/2} (z-1)^{1/2} (z+1)^{-1/2}$$
3.144

Substitution of equation 3.138 into equation 3.134 leads to the result:

$$P_T(E) = K \, F(E) \frac{e^{-b}}{2\pi} \int_0^{2\pi} d\Theta . \int_{E-E_m(E,\Theta)}^{E} e^{c(E_x - \epsilon')} dE_x$$
3.145

provided the quadratic and higher order terms in equation 3.138 can be ignored. The major part of the integral must come from Ex values near to E, therefore, it is permissible to choose ϵ' equal to E. It then follows, after integration of equation 3.145, that:

$$P_T(E) = K \, F(E) \frac{e^{-b(E)}}{c(E)} \left(1 - \frac{1}{2\pi} \int_0^{2\pi} e^{-c(E)E_m(E,\Theta)} d\Theta \right)$$
3.146

In the case of degenerate emission when the Fermi energy is positive and the Fermi level is above the bottom of the conduction band it is evident that most of the electrons come from energy levels near to the Fermi level. It follows that E, and hence ϵ', can be set equal to the surface Fermi energy ϵ_s. Equation 3.146 then takes the form [46][47]

$$P_T(E) = \frac{K \, F(E)}{c(\epsilon_s)} e^{-(b(\epsilon_s)+c(\epsilon_s)\epsilon_s)} e^{c(\epsilon_s)E} \left(1 - e^{-r_c c(\epsilon_s)E} \right)$$
3.147

where
$$r_c = m/m_n$$
3.148

Similarly, for non degenerate emission when the Fermi energy is negative,

the expansion should be made about the bottom of the conduction band because most of the emitted electrons will come from near this level. In this case, equation 3.146 becomes:

$$P_T(E) = \frac{K \; F(E)}{c(o)} e^{-b(o)} e^{c(o)E} \left(1 - e^{-r_c \; c(o)E} \right)$$

3.149

3.10 Application Of The Theory To Cadmium Sulphide

Section 3.7.2 shows that the bottom of the conduction band bends below the Fermi level in the surface of a degenerate semiconductor by an amount E_S, which equals $E + \triangle R$, and that:

$$E + \triangle R = \nu F^{4/5}$$

3.150

where

$$\nu = \frac{1}{z^{2/5}} \left(\frac{225}{\pi^4 2^{11}} \frac{h^6}{m^3 e^4} \right)^{1/5}$$

3.151

Equation 3.150 is valid provided:

$$E_S \gg kT$$

3.152

and

$$F \gg z^{1/2} \; 1.2 \times 10^6 \; \text{VOLT. cm.}^{-1}$$

3.153

For cadmium sulphide this infers that F must be greater than 4×10^6 volt cm^{-}. The expression for the value of applied field, FD, at which the onset of degeneracy occurs was derived in section 3.8.5 and is of the form:

$$FD = 1.44 \times 10^6 \times z^{1/2} \left(\frac{m_n}{m} \right)^{3/2} \left(\frac{T}{300} \right)^{5/4} \text{VOLT. cm.}^{-1}$$

3.154

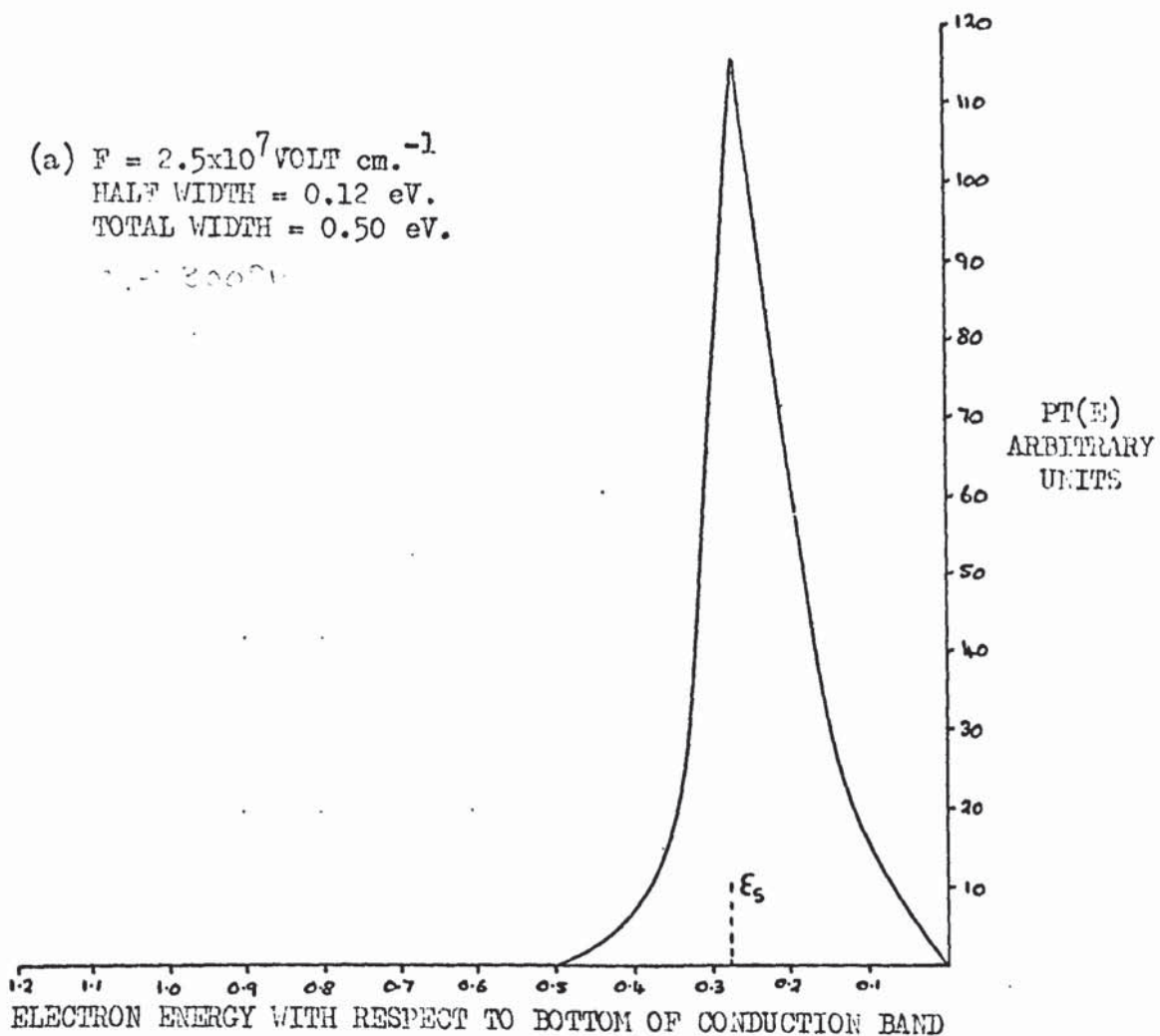Equation 3.154 gives a value of 5×10^5 volt cm^{-1} for FD in the case of cadmium sulphide.

The fields used in the experiments were estimated from the radius of the emitter, the applied potentials and the usual correction factor [33].

Fig. 3.11

(a) F = 2.5x10^7 VOLT cm.$^{-1}$
    HALF WIDTH = 0.12 eV.
    TOTAL WIDTH = 0.50 eV.

PT(E)
ARBITRARY
UNITS

ε_s

ELECTRON ENERGY WITH RESPECT TO BOTTOM OF CONDUCTION BAND

(b) F = 5x10^7 VOLT cm.$^{-1}$
    HALF WIDTH = 0.17 eV.
    TOTAL WIDTH = 0.65 eV.

PT(E)
ARBITRARY
UNITS

ε_s

ELECTRON ENERGY WITH RESPECT TO BOTTOM OF CONDUCTION BAND

Fig. 3.11

(c) F = 7.7x10^7 VOLT cm.$^{-1}$
HALF WIDTH = 0.24 eV.
TOTAL WIDTH = 0.90 eV.

PT(E)
ARBITRARY
UNITS

ELECTRON ENERGY WITH RESPECT TO BOTTOM OF CONDUCTION BAND

(d) F = 1x10^8 VOLT cm.$^{-1}$
HALF WIDTH = 0.28 eV.
TOTAL WIDTH = 1.00 eV.

PT(E)
ARBITRARY
UNITS

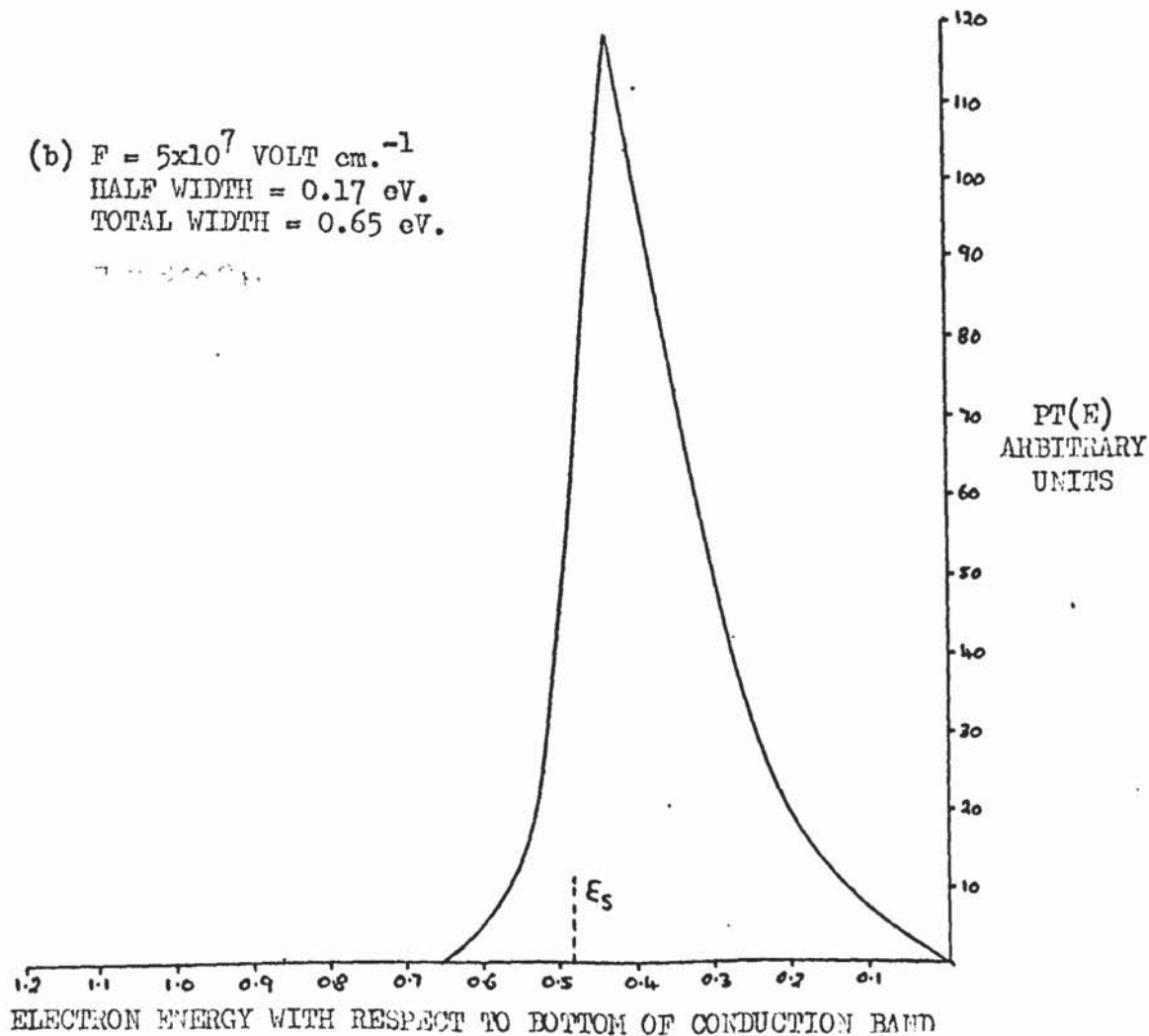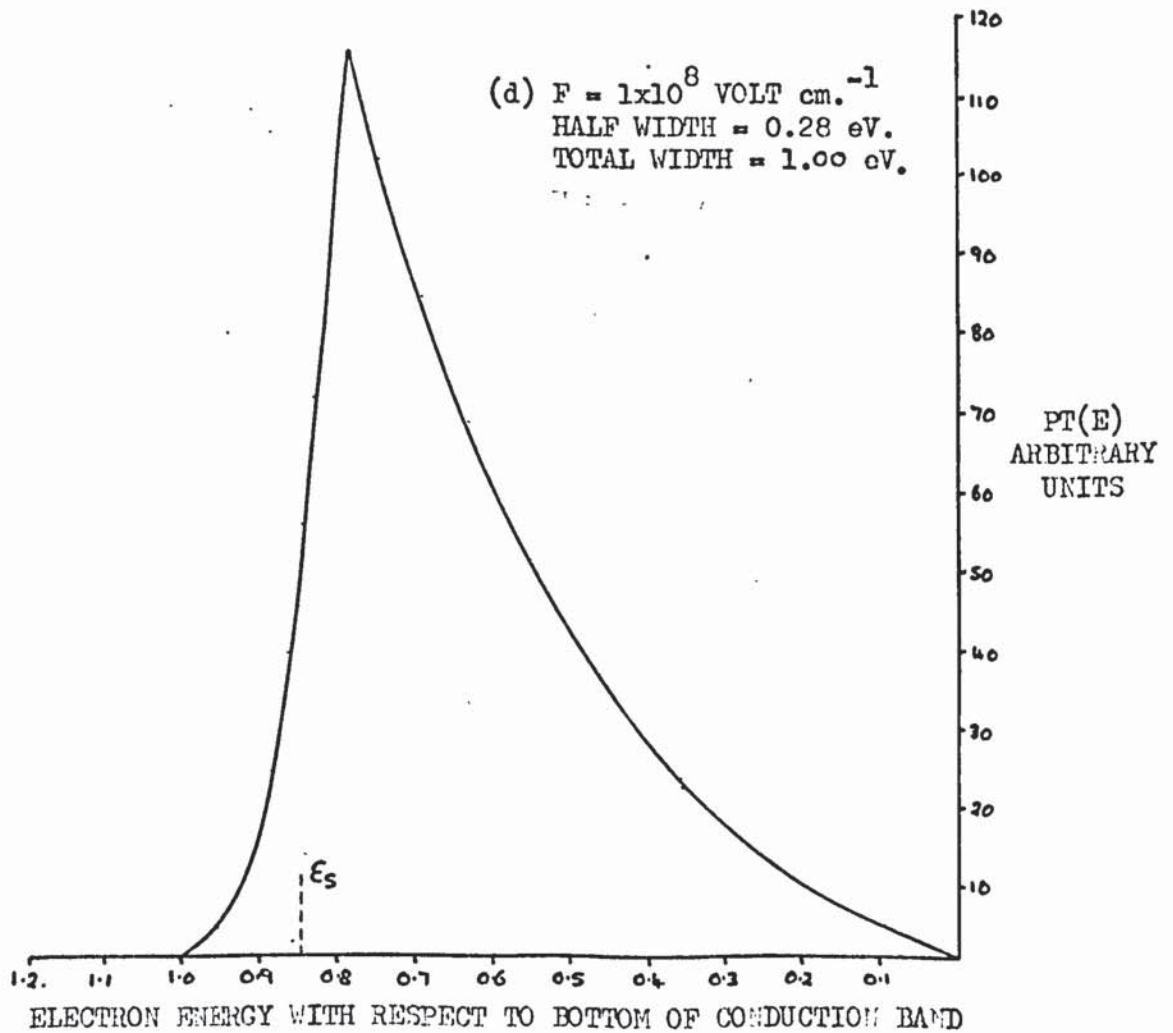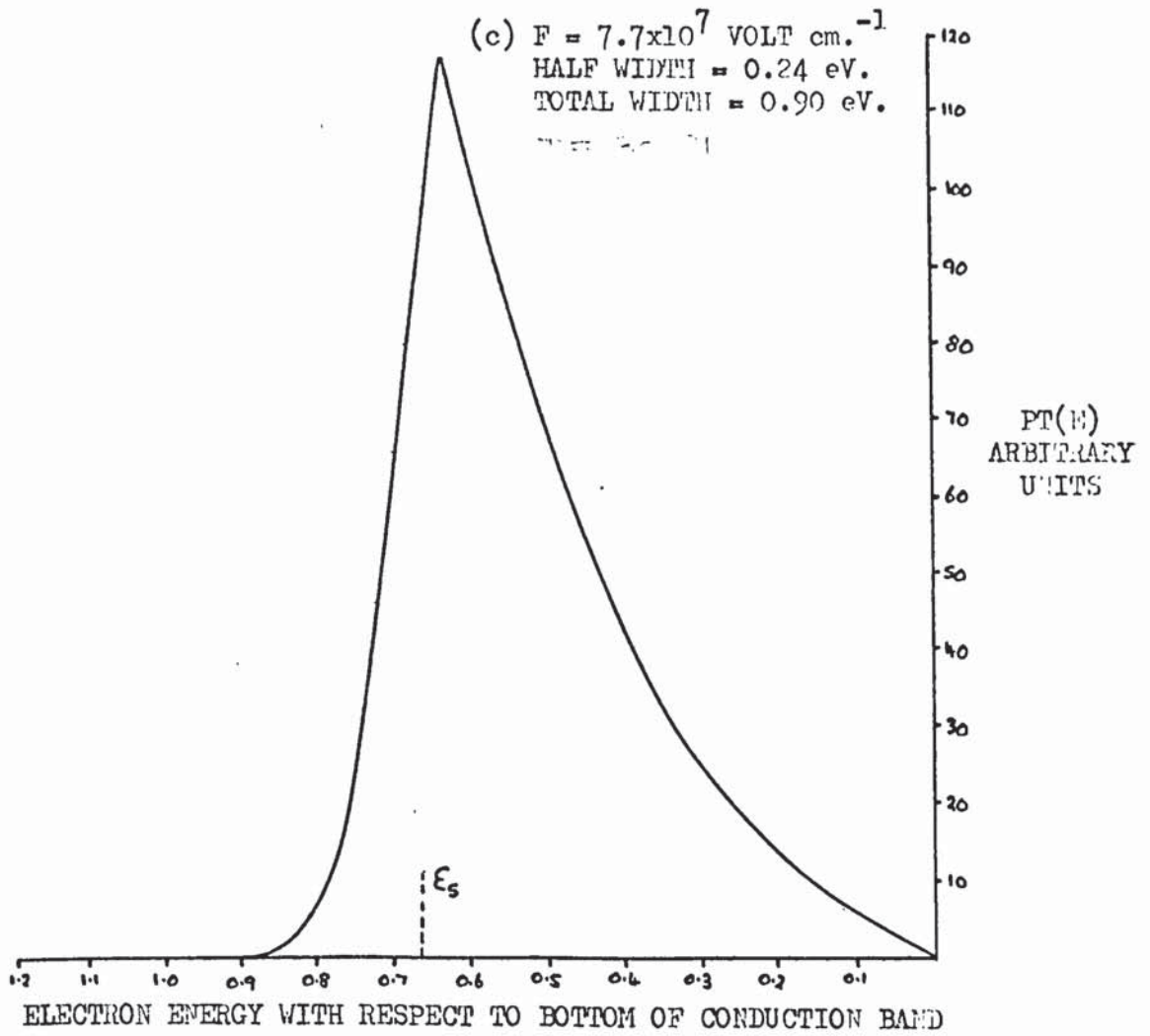ELECTRON ENERGY WITH RESPECT TO BOTTOM OF CONDUCTION BAND

They were found to be of the order of 5×10^7 volt cm^{-1}. It follows that the surface of the cadmium sulphide was degenerate and that equation 3.152, with the free electron mass replaced by the cadmium sulphide effective electron mass, can be used to obtain $\mathcal{E}$ s as a function of F. Equation 3.147 was then used to derive the theoretical energy distribution. The expansion about the Fermi level described in section 3.9.1 is not valid above fields of the order of 2×10^8 volt cm^{-1} so that this limitation, in conjunction with inequality 3.153, restricted the theoretical results to applied values in the range 1×10^7 volt cm^{-1} to 1×10^8 volt cm^{-1}. The relevant parameters for cadmium sulphide were inserted in equations 3.150 and 3.147 and the results of the calculations are presented in Fig. 3.11.

It should be noted that only "pure" field emission has been considered i.e. the effect of surface states has been ignored, and that the equations used are strictly true only for the "zero" current approximation case.

o o O o o

CHAPTER 4

Experimental Techniques

4.1 Sample Preparation

4.1.1 Tungsten Emitters

Tungsten wire of diameter 0.005" was electrolytically etched in an aqueous 1.N solution of sodium hydroxide. It was found that the wire etched preferentially at the liquid-air interface thus forming a long thin "neck". When this neck became thin enough the weight of the wire below the surface of the electrolyte caused it to break and leave a very sharp tip. Unfortunately, the electrolyte continued to etch and the tip was, therefore, blunted.

Initially the current in the solution was monitored by means of an ammeter and when it fell quickly, which was caused by the breaking of the "neck", it was switched off by hand. Although this method gave fairly reproducible tips of good shape, their diameter was approximately five times larger than could be used for field emission experiments. It was, therefore, decided to use electronic instead of manual switching. Fig. 4.1(a) shows the sensing circuit and Fig. 4.1(b) the sensing and controlling circuits. RT is the resistance of the tungsten wire. When the tip is being etched the current in the circuit is given by:
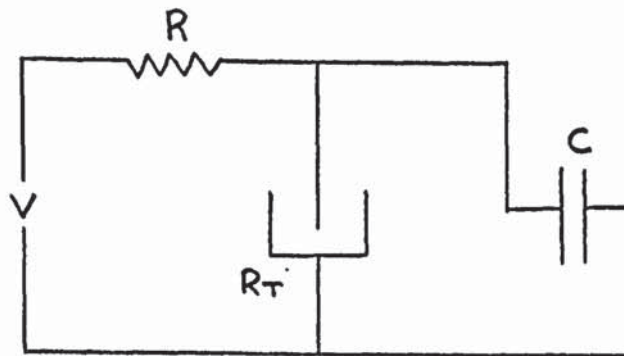
$$I = \frac{V}{R + R_t} \qquad 4.1$$

where V is the applied voltage. The potential drop across the resister is, therefore:
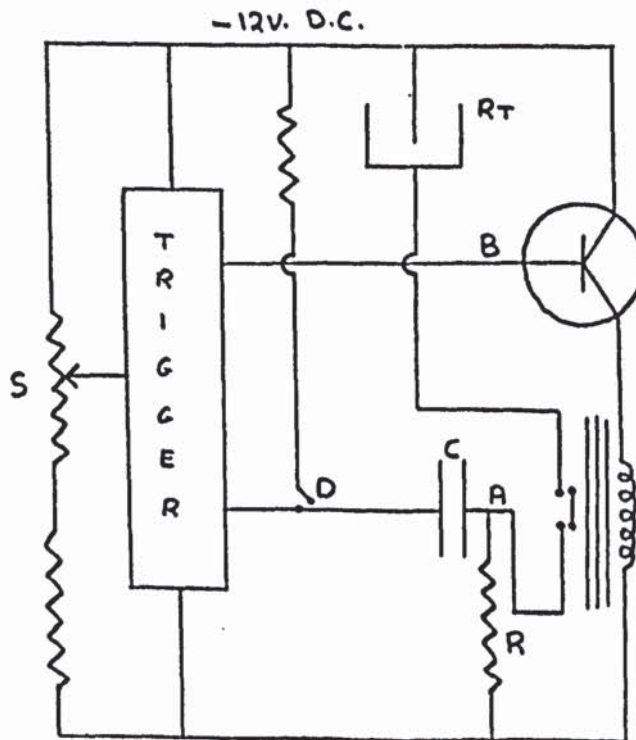
$$V_R = \frac{V \cdot R}{R + R_T} \qquad 4.2$$

When the wire breaks, the value of RT increases by a factor of at least a hundred so that VR falls to a very small value. The consequent voltage pulse is fed to the controlling circuit via the capacitor C.

Fig. 4.1(a)

THE SENSING CIRCUIT

Fig. 4.1(b)


THE SENSING AND CONTROL CIRCUIT

The sensing and controlling circuits are a.c. rather than d.c. coupled
in order to suppress small transient voltages caused by variations in
the etching current which occur, for example, because of bubble formation.
Under normal conditions the steady etching in the sensing circuit is
approximately 4 mA. so that the potential at A is -8.8 volts. When the
"neck" breaks, the current falls to 50μA. leading to a potential at
A of -0.1 volts. The positive pulse of height 8.7 volts is then fed, via C,
to the base of the first transistor in the Schmitt trigger which is
designed to trigger when the base of this transistor is more positive
than -1.15 volts. The base is set at -2 volts by potentiometer S so that the
8.7 volt pulse from C switches the Schmitt trigger and a pulse is fed
to B, the base of the transistor which controls the relay. The relay
then opens and the etching current is switched off. The delay between
the breaking of the "neck" and the etching current being switched off
is less than a millisecond. In order to reset the circuit for the next
sample a small negative pulse is fed to the base of the first transistor
in the Schmitt trigger by closing switch D.

Fig. 4.2 shows scanning electron micrographs of tungsten emitters that
were obtained with this device. The tips are very reproducible and the
success rate is approximately 70%.

4.1.2 Cadmium Sulphide Emitters

The cadmium sulphide emitters were prepared from single crystal 0001
oriented samples which were 1 mm by 1 mm by 1 cm long. Emitters are
usually made from this material by etching with hydrochloric acid which
is then followed by polishing with hot phosphoric acid because the
hydrochloric acid tends to etch preferentially and thus leaves a surface
on which there may be several microtips instead of one[51]. Although
phosphoric acid etching is much slower than that of hydrochloric acid
it was decided that most of the etching should be performed with
phosphoric acid in order to ensure the production of a smooth, well-
defined emitter.

Fig. 4.2
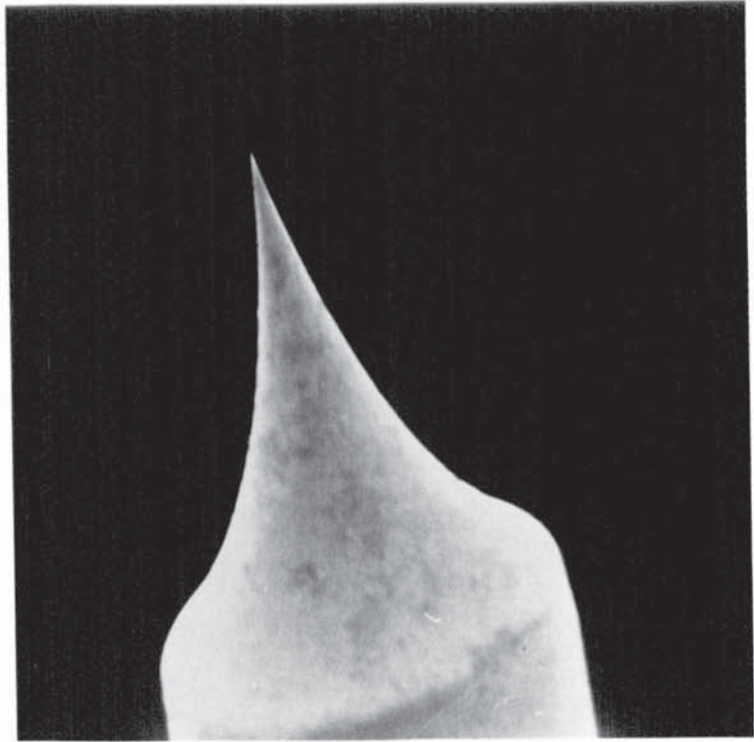
1cm. = 8 micron.
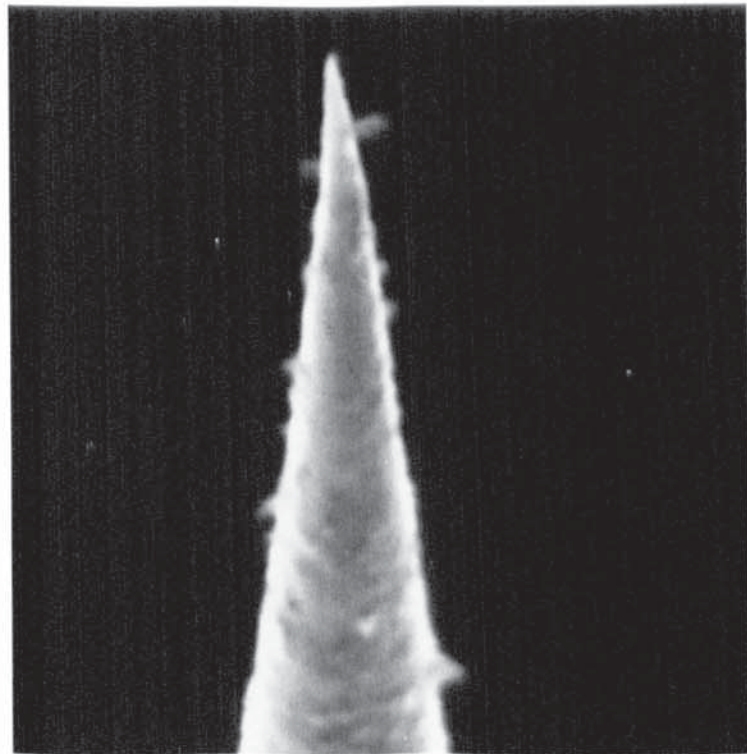
1cm. =,8200 A°

The cadmium sulphide samples were mounted as shown in Fig. 4.3 because this type of holder could not only be used during the etching and cleaning processes but could also be used in the analyser. Sodium silicate was employed as a glue because it was found to be virtually inert to hot phosphoric acid fumes.

After the sodium silicate had dried, the sample was mounted vertically with the tungsten rod of the sample holder connected to a small electric motor. It was known that rotation inhibits preferential etching, therefore the tip end was rotated in a beaker of concentrated hydrochloric acid until the diameter of the crystal was reduced to about 250 microns over approximately half its length. When this diameter was attained, the sample was removed from the hydrochloric acid, washed in distilled water and thoroughly dried.

The final etching was performed with hot phosphoric acid. Initially the acid was heated by placing a small drop of it on a loop of resistance wire through which a current was being passed. This method was not successful, however, because the hot acid attacked the wire. The problem was overcome by threading the resistance wire through a fine glass capillary tube before forming it into a loop. The glass then served as protection for the wire and a loop was made which was small enough to support the drop of acid by surface tension. The sample was attached to a vernier microscope support so that it could be moved up and down by small amounts through the hot acid. In this way the diameter of the shank was reduced from 250 microns to approximately 10 microns.

Finally, the sample was placed in such a position that part of it protruded through the bottom of the drop of acid. Preferential etching then occurred at the acid-air interface and by carefully watching the tip through a X12 microscope while controlling the etching rate by means of the current flowing through the resistance wire, it was possible to remove the sample when the protruding part just fell off.

- 84 -

Fig. 4.3

Cd.S. EMITTER

AQUADAG

CRYOGENIC TUBE

SODIUM SILICATE

SPOT WELD

TUNGSTEN ROD
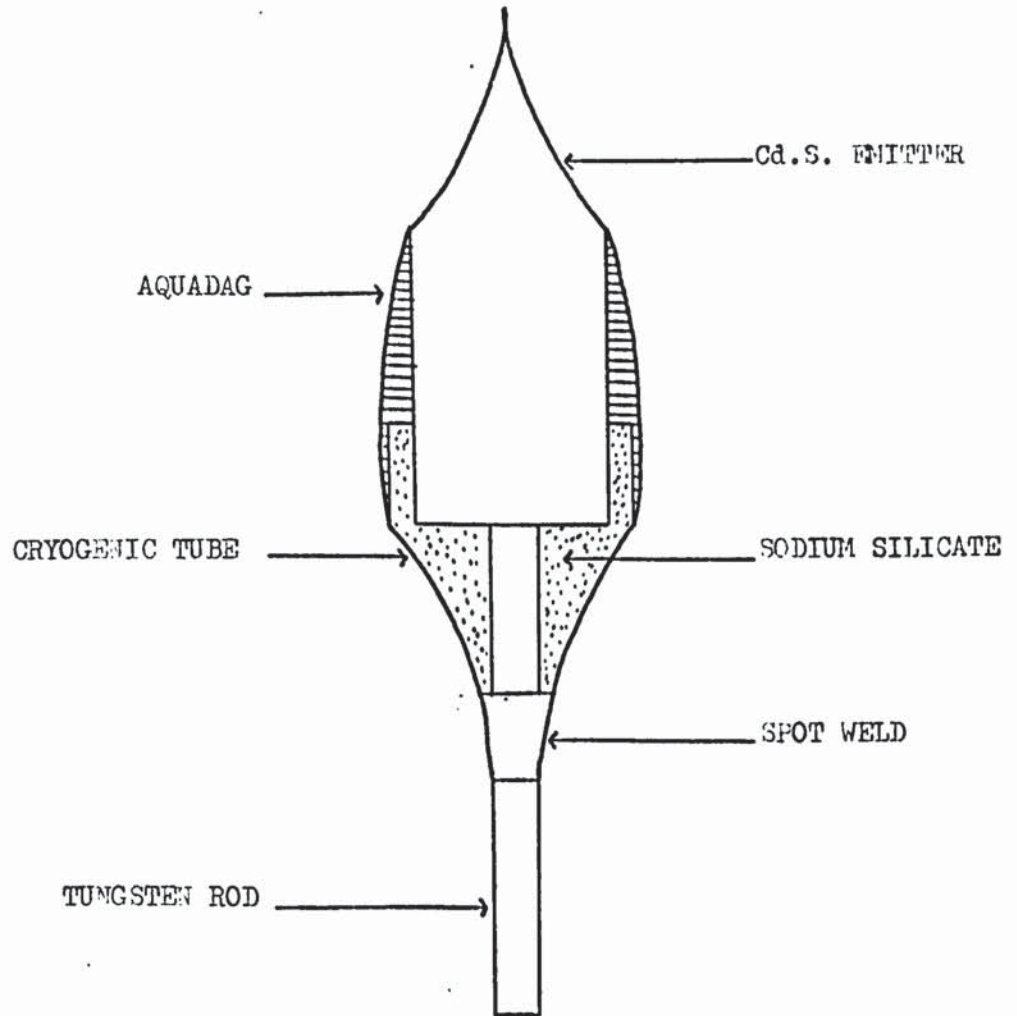
EMITTER HOLDER

The resulting field emitter was then carefully washed in distilled water and thoroughly dried. Good electrical contact between the cadmium sulphide and the holder was ensured with a small drop of Aquadag colloidal graphite suspension and the sample was stored in an evacuated dessicator until needed.

Although this method of tip making is time consuming and requires a certain "knack", it can be seen from the scanning electron micrographs in Fig. 4.4 that it produces fairly smooth, well defined emitters. The success rate is fairly good also because from twelve samples made in this way, seven proved to be good field emitters.

4.2 The Vacuum System

4.2.1 General Considerations

The Fowler Fordheim equation[9] [10]. shows that the emitted current density is sensitive to the work function of the emitter which, in turn, is dependent on the presence of absorbed gas on the emitter surface. It follows that in order to obtain stable emission from an initially clean sample, the experiments must be performed at high or ultra high vacuum when the time to form a monolayer of gas is large compared to the time needed to perform an experiment.

A constant work function is especially important in the measurement of the total energy distribution because the experimental determination of this quantity is correct only if the contact potential difference between the emitter and the collector remains constant (cf. equation 2.3). In the case of a semiconductor the attainment of stable emission is even more important because, not only can absorbed gas give rise to surface states, but also, in the case of the high resistance semi-conductors like cadmium sulphide, a large voltage drop may appear along the shank of the emitter caused by the passage of the emission current. If this current fluctuates, the potential difference between the end of the emitter and the collector will also fluctuate which, according to equation 2.4, will cause error in the measured total energy distribution.

Fig. 4.4
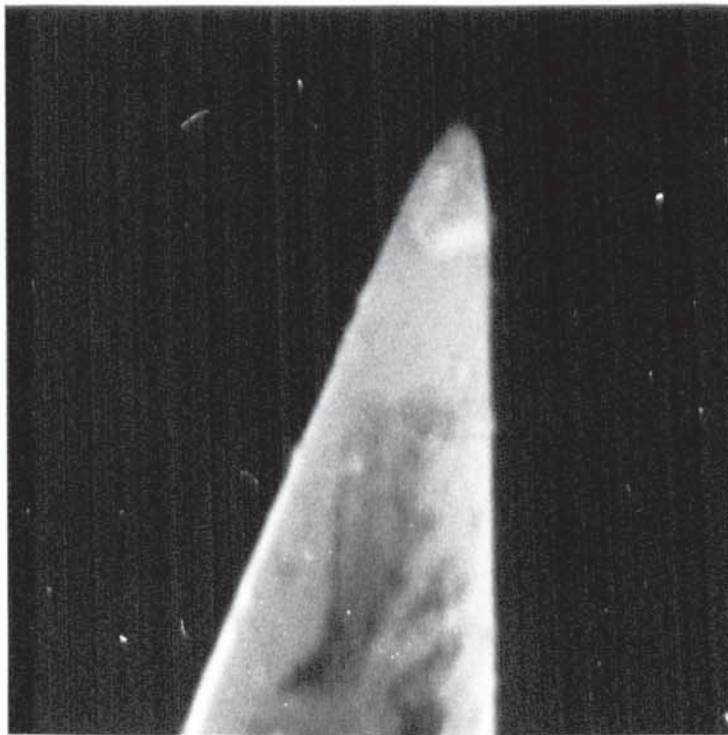
1cm. = 8300 A°

1cm. = 6500 A°

Classical gas kinetic theory shows that the number of gas molecules
per second which strike unit area of the enclosure containing them is
given by:

$$z = p \left(2\pi m k T \right)^{-1/2} \qquad\qquad 4.3$$

where m is the mass of a molecule, p is the gas pressure, k is
Boltzmann's constant and T is the temperature. For nitrogen gas at a
pressure of 1×10^{-9} torr and room temperature, equation 4.3 yields a
value for z of 4×10^{11} impacts $sec^{-1} cm^{-2}$. If a sticking probability
of unity is assumed and a monolayer is taken to be equal to 2.5×10^{14}
molecules cm^{-2}, then it takes 6.3×10^2 seconds for a monolayer to
form.

In addition to the number of impacts given by equation 4.3, it has been
shown that there are extra gas molecular impacts caused by polarisation
of the gas molecules in the presence of the emission field [90]. The
equation for the number of impacts per unit area, per unit time now
becomes:

$$z_F = z \left(\frac{1 + \alpha F^2}{2kT} \right) \qquad\qquad 4.4$$

where   is the polarisability of the molecule under consideration, F
is the electric field strength and Z is given by equation 4.3. For a
field of 1×10^8 volt cm and nitrogen gas, there is a 2.7 fold increase
in the number of impacts per unit time per unit area so that the mono-
layer formation time becomes approximately 2.5×10^2 seconds. It would
seem that even at 1×10^{-9} torr there is little time in which to perform
experiments. It must be remembered, however, that the highest possible
value of the sticking probability has been assumed and that the electric
field term depends on the square of the electric field. The value of
electric field chosen to illustrate equation 4.4 is somewhat higher than
that normally used in field emission work, therefore, in practice, the
polarisation term would be smaller than 2.7.

4.2.2 The Vacuum Equipment

The vacuum equipment is shown in Fig. 4.5. A conventional Edwards type EO2 water cooled oil diffusion pump was used which was backed first by an Edwards type EO1 water cooled oil diffusion pump and secondly by an Edwards type ED50 two stage rotary pump. The second diffusion pump served two purposes, it slightly increased the pumping speed and also reduced backstreaming of oil from the main diffusion pump to the analyser chamber [91]. Backstreaming was further reduced by the use of a liquid nitrogen cooled trap and a water cooled baffle on the main diffusion pump. Oil from the rotary pump was trapped in a liquid nitrogen cooled backing line trap which was made of copper and filled with aluminium shavings and sintered glass discs [92]. In order to obtain the highest vacuum possible, Edwards type 705 silicon oil was used in the diffusion pumps and the rotary pump was filled with Edwards number 16 oil.

The vacuum system was protected against mains failure by means of a magnetic isolation/air admittance valve which was placed in the backing line. This valve was designed to close immediately the electrical supply to the system was interrupted, thus maintaining the vacuum in the diffusion pumps and preventing the hot oil from "cracking" by exposure to air at atmospheric pressure. At the same time, the air admittance valve on the rotary pump side of the backing line opened so that a pressure differential could not develop across the pump. This prevented rotary pump oil being drawn into the backing line. The system was further protected by a metal bellows which was placed in series with the cooling water supply to the diffusion pumps. The bellows was connected to a micro-switch so that when the cooling water flow rate fell below some predetermined level, the electrical supply to the heaters on the diffusion pumps was switched off, thereby preventing over heating. In order to obtain the highest possible vacuum the system was baked in an oven. Heat loss from the oven was minimised by making its sides

- 87 -

Fig. 4.5
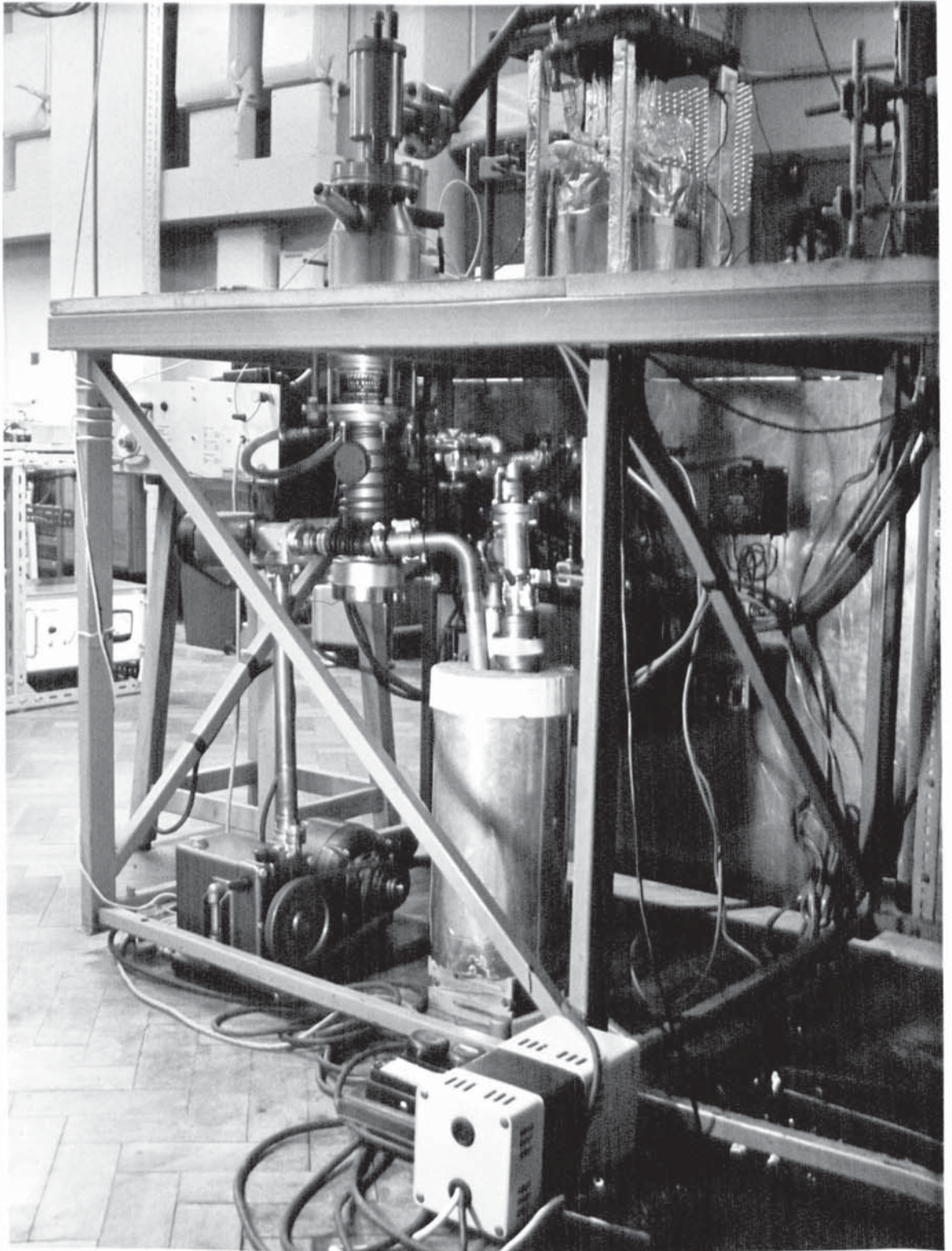
from 1" thick fibre glass sheet sandwiched between inner and outer walls of 16 s.w.g. Duralumin. The system was usually heated to 200°C over a weekend and then left to cool slowly for two days before experiments were performed. A higher temperature and a faster heating and cooling rate were available but they were not used because the glass-metal seals in the analyser were very susceptible to thermal shock. All seals on the bakeable side of the vacuum system were made with aluminium rings or copper gaskets between Vacuum Generators flanges while those on the non-bakeable side were made with indium wire. Conventional Edwards O ring seals were used in the backing line. The pressures in the vacuum system were measured with a Pirani gauge at low vacuum and with a Mullard IOG12 ionisation gauge at high vacuum. The table below describes a typical run of the vacuum system:

Time	Backing Pressure	Analyser Pressure
0	Atmospheric	Atmospheric
	Switch on rotary pump	
5 min.	10 microns	10 microns
	Put liquid nitrogen in backing line trap	
10 min.	$2\frac{1}{2}$ microns	$2\frac{1}{2}$ microns
	Switch on main diffusion pump	
20 min.	$2\frac{1}{2}$ microns	6×10^{-7} torr
	Switch on small diffusion pump	
50 min.	$2\frac{1}{2}$ microns	3×10^{-7} torr
	Bake system for two days at 200°C	
	Lift oven at time = 0 again	
0	$2\frac{1}{2}$ microns	5×10^{-9} torr
	Outgass gauge for 20 min.	
20 min.	$2\frac{1}{2}$ microns	8×10^{-10} torr
	Fill high vacuum trap with liquid nitrogen	

When the analyser was first used after being pumped down, the pressure

usually rose to approximately 3×10^{-9} torr because the field emitted electrons heated the phosphor screen on the anode, thus liberating trapped gas. To remove this problem, the screen was continuously bombarded with electrons field emitted from a tungsten tip for 130 minutes, after which the pressure in the analyser could usually be maintained at 1×10^{-9} torr or less while experiments were being performed.

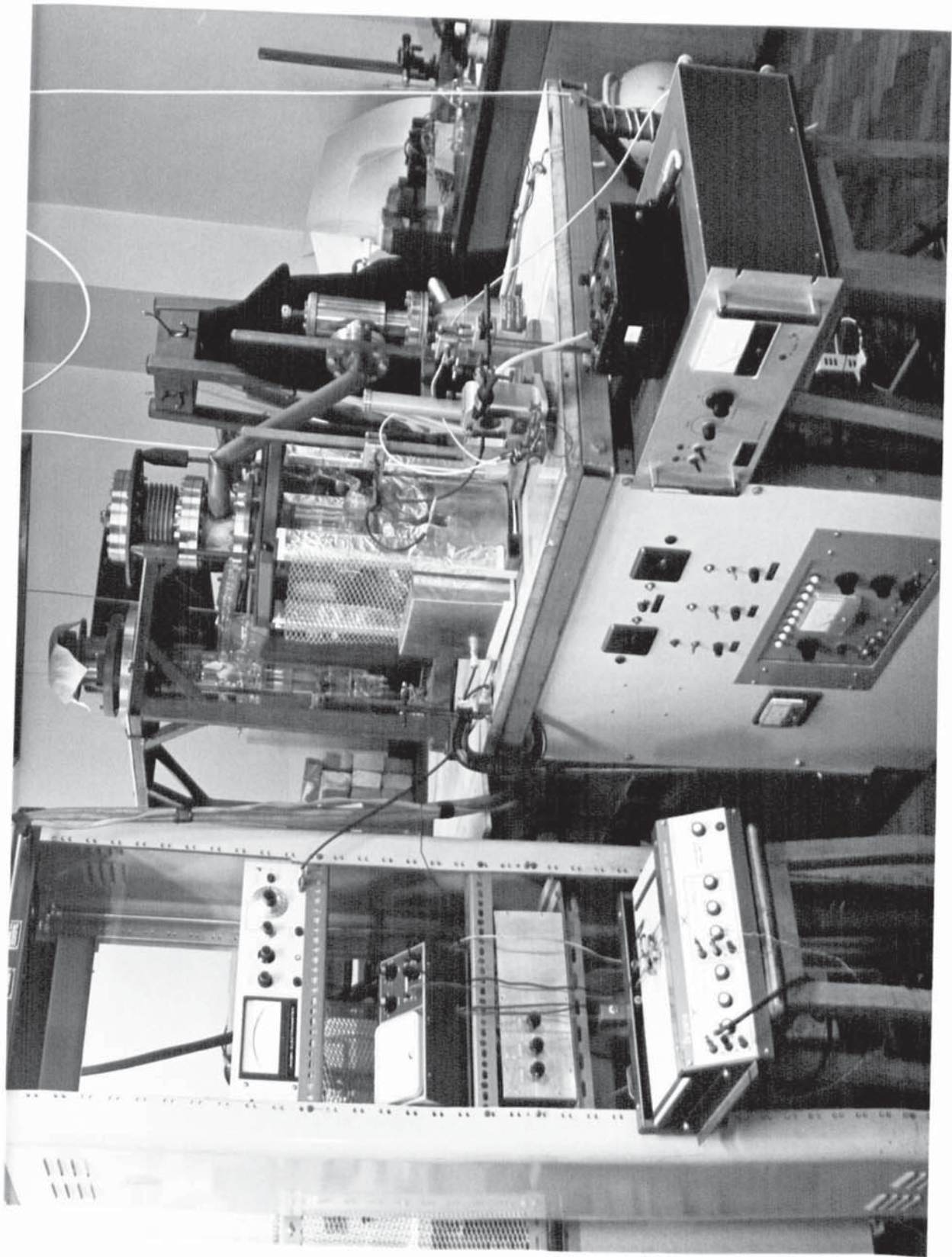## 4.3 The Analyser And Measuring Circuits

A general view of the analyser and measuring circuits is shown in Fig. 4.6

4.3.1 The Fabrication Of The Analyser

The analyser is shown schematically in Fig. 4.7. The collector, which also forms the outer vacuum wall, was made by depositing a tin oxide coating [93] on the inside of a 5 litre pyrex glass flask. The anode was spun on a lathe from a piece of 1/16" thick oxygen free high conductivity copper and the first aperture, which was in the top of the anode, was made 3 mm long to ensure that off axis electrons could not enter the collector region. In order to eliminate the high order electron lens aberrations the second aperture was given radiused edges[94]. All electrical connections to the inside of the analyser were made from nickel wire because this metal makes excellent spot welds to the tungsten wire used for the lead throughs. The tip moving screws which, in conjunction with the bellows, allow lateral movement of the emitters, were made of mild steel and had ball bearing contacts which were lubricated with powdered molybdenum disulphide. In this way the possibility of siezure during breakout was avoided.

The phosphor screen was made from Levy West type ZDB/F3 phosphor by allowing the phosphor powder to settle on the anode from a solution of phosphor and acetone. A few drops of sodium silicate were also added to the solution to act as a binder. The screen was made as thin as possible in order to reduce the amount of trapped gas and before it was deposited, the anode was chemically cleaned and electropolished in order to remove

Fig. 4.6

Fig. 4.7

BELLOWS

MOVEMENT SCREW

CRYOSTAT

MAIN SUPPORT

GLASS METAL SEALS

H.T. LEADTHROUGH

TIP WHEEL
ASSEMBLY
ANODE WITH
PHOSPHOR SCREEN
TIN OXIDE COLLECT

COLLECTOR CURRENT
LEADTHROUGH
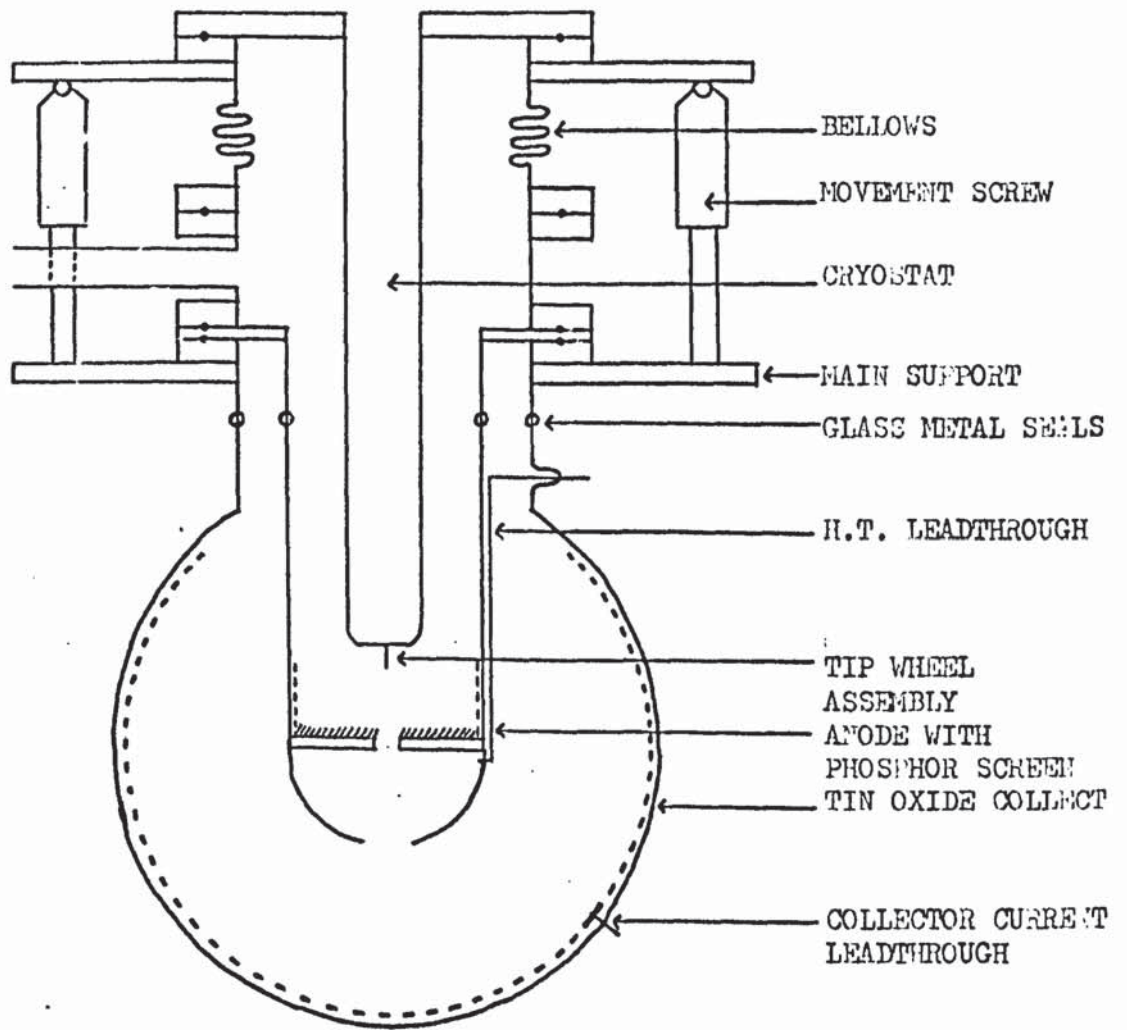
SCHEMATIC OF ANALYSER

small protusions which could lead to field emission and, in severe cases, to electrical breakdown.

Before the analyser was used all insulation resistances were measured. The glass between the high voltage lead through and the collector was thoroughly cleaned of tin oxide with a solution of powdered zinc in hydrochloric acid until the ideal value of the surface leakage resistance for clean glass was obtained. The outside of the collector was also thoroughly cleaned before use to eliminate the possibility of leakage from the high voltage lead through to the collector current lead through and a small drop of Aquadag colloidal graphite suspension was placed on the inside of the collector to ensure good electrical contact between the collector current lead through and the tin oxide coating.

4.3.2 The Tip Wheel Mechanism

Cadmium sulphide emitters are very fragile because the material is so brittle. It was, therefore, decided to design a mechanism which would make it possible to place six emitters in the analyser at one time, then, if one failed to work, the equipment would not have to be dismantled and rebaked before another could be tried.

The device in its final form is shown in Fig. 4.8 while Fig 4.9 shows an exploded view of the components used in its construction. All moving parts were made of silver or molybdenum because, not only do these materials possess good thermal and electrical conductivities, but also there is little friction between them under ultra high vacuum conditions. Fig 4.8 shows that the mechanism consists mainly of three wheels. The wheel on the left has location grooves which locate on a spring loaded lug, the wheel in the middle holds the emitters and the wheel on the right is a star wheel.
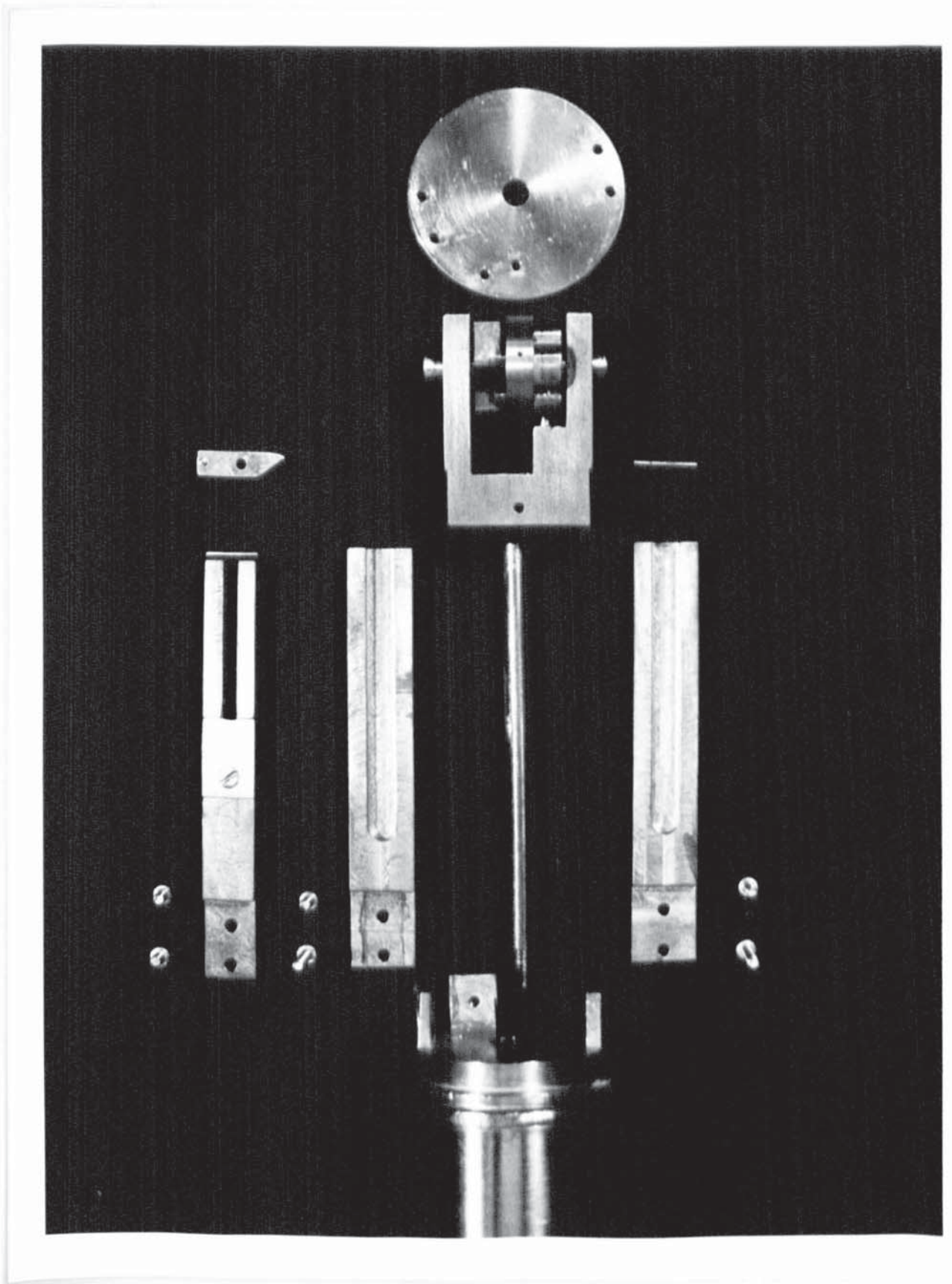
Initially the metal block in which the wheels are located is at the bottom of its movement and an emitter protrudes through the hole in the baseplate. When the next emitter is needed, the metal block is lifted so that the first emitter is withdrawn until it is clear of the baseplate.

Fig. 4.8
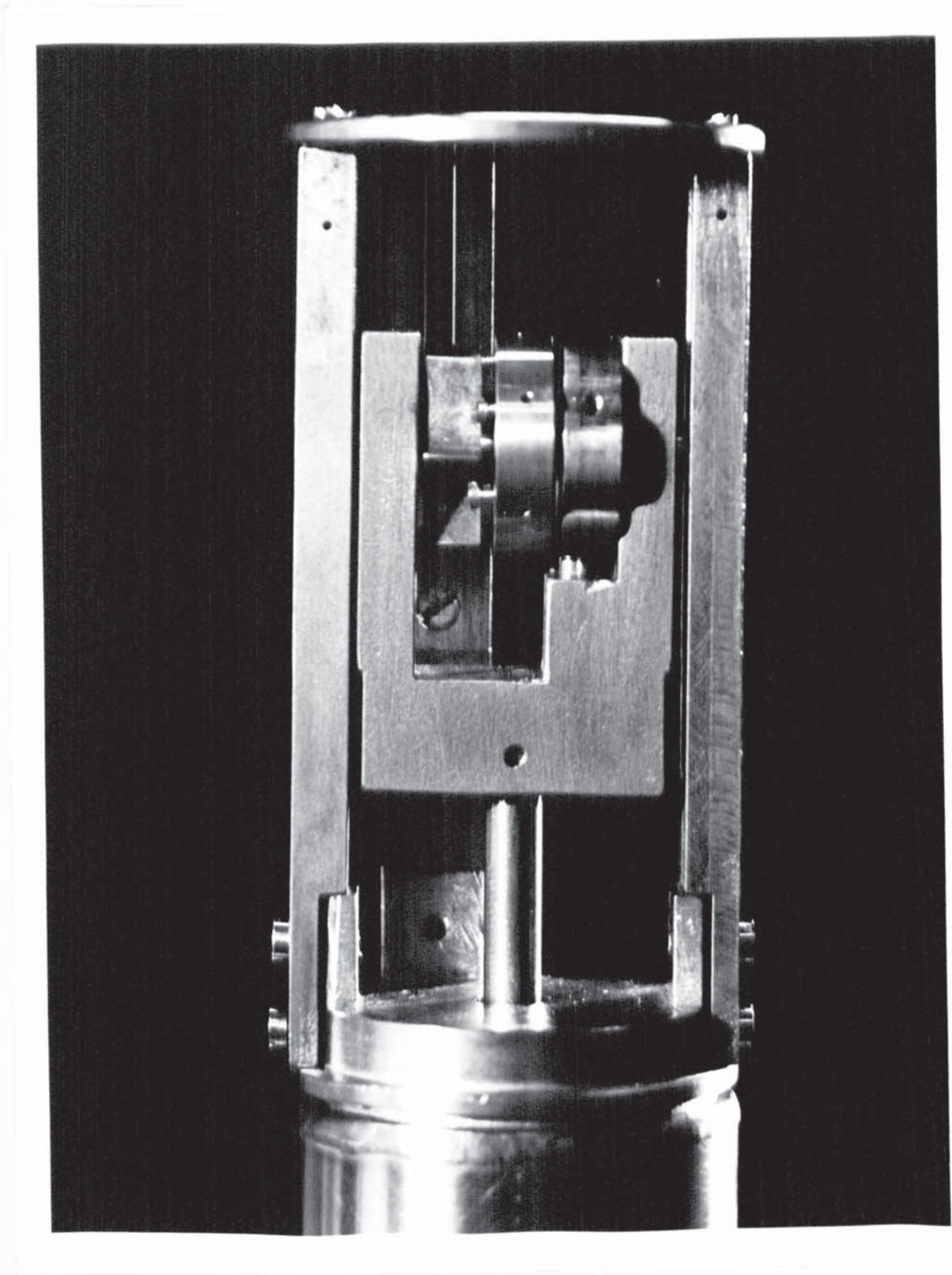
TIP WHEEL ASSEMBLY

Fig. 4.9

TIP WHEEL COMPONENTS

As the lifting motion is continued, the star wheel engages on the ratchet and all three wheels rotate because they are on the same spindle. Finally, the spring loaded lug clicks home on the location wheel and the next emitter points downwards.

The metal block is then moved back towards the baseplate. When the star wheel reaches the ratchet it pushes it aside, any tendency to rotation being prevented by the spring loaded lug. The ratchet falls back to its original position under the influence of gravity after the star wheel has passed and the metal block again comes to rest on the baseplate. It follows that the new emitter now protrudes through the hole and that it is vertical in at least one plane because of the location of the spring loaded lug on the left hand wheel. While the analyser is being operated the baseplate is given the same potential as the emitters which are not being used. In this way they are screened from the field produced by the anode and field emission occurs from the protruding tip only.

Linear motion is conveyed from the outside to the inside of the vacuum system via an edge welded stainless steel bellows. The vacuum joints in the tip wheel mechanism were vacuum brazed by Vacuum Generators and the component parts were thoroughly degreased and baked to 400°C under vacuum before being used in the analyser.

4.3.3 The Measuring Circuit

The measuring circuit is shown in Fig. 4.10. A Brandenburg type 800 power supply was used for the high voltage and the collector current was measured by means of a Keithly type 610 BR electrometer. The retarding potential was supplied from a battery via the 10 turn helical potentiometer B and the linear potentiometers A_1 and A_2. Retarding potential vs collector current curves were monitored on a Bryans type 2000 X-Y recorder, the Y plates being driven by the 3 volt analogue output of the electrometer, while measurements of the total emitted current were made with a mirror galvanometer which had a sensitivity

Fig. 4.10
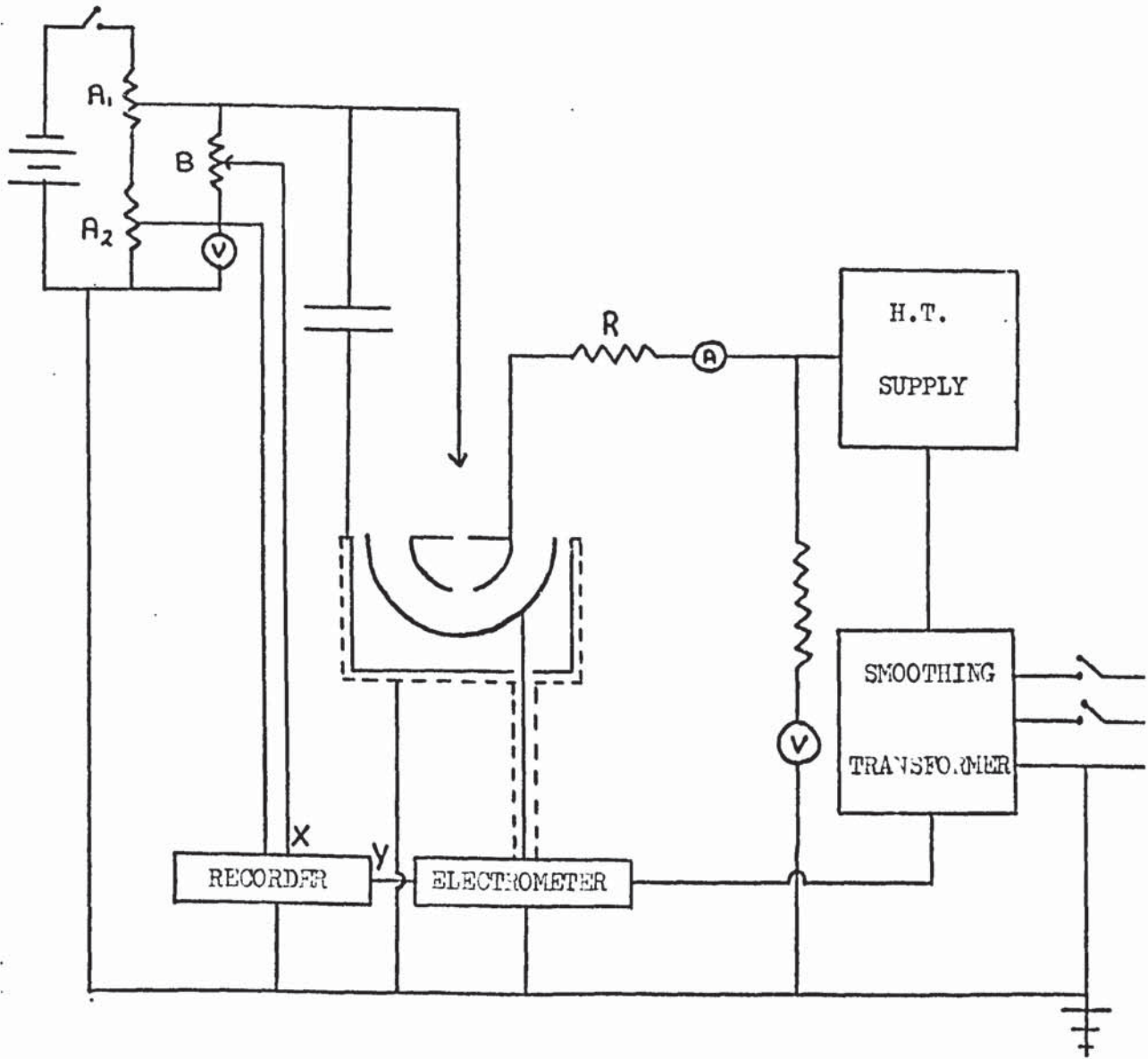
THE MEASURING CIRCUIT

of approximately 6.7×10^{-10} Amp mm^{-1}. The mains supply to the high voltage supply and the electrometer was obtained from a stabilising transformer in order to improve the regulation of these instruments. When a tungsten emitter was used in the analyser it was found that no collector current could be recorded until a retarding potential of approximately -4 volts had been applied to the emitter and that most of the information concerning the energy distribution occurred between retarding potential values of -4 and -5.5 volts. It proved impossible, therefore, to monitor the collector current vs retarding potential curves at high sensitivity because the initial -4 volts then used up all the X movement on the recorder. This effect was especially noticeable in the case of cadmium sulphide where, in some cases, a voltage drop across one emitter of tens of volts had to be overcome before collection took place. Potentiometers A_1 and A_2 were, therefore, included so that the voltage scan corresponding to the energy distribution could be put across B, and hence across the recorder, while the initial voltage could be dropped across A_1 and measured by the valve voltmeter V. In this way it proved possible to obtain a sensitivity of 60 mV cm^{-1} for the 1.5 volt scan.

The high voltage was measured by means of a model 8 AVO with a 10 K.V. series multiplier. The multiplier was placed as shown in Fig. 4.10 in order to avoid the formation of a potential divider with the $10^8\Omega$ resistor R which was included in the circuit to protect the high voltage supply in the event of large instantaneous breakdown currents. Capacitor C was used to prevent pickup on the emitter and its value was chosen so that it did not appreciably increase the time constant of the retarding potential circuit. In order to ensure that this was the case, two collector current vs retarding potential curves were taken. The first was obtained by increasing the retarding potential until the end of the scan had been reached and the second by starting at the end of the scan and decreasing the retarding potential. No appreciable

hysteresis effect was observed between the two curves.

The magnetic fields in the vicinity of the analyser were measured with a Bell type 120 gauss meter and it was found that a vertical field of 2 gauss and transverse fields of 0.7 and 0.5 gauss were present. A careful investigation of these fields showed that the transverse components were caused by the welds in the feet of the analyser support. The feet were, therefore, placed on a sheet of mu metal in order to trap the magnetic flux and, although the vertical field was hardly affected, the transverse fields were reduced in value to 0.02 and 0.3 gauss.

An electron with momentum p which is moving normally to a magnetic field of strength B follows a curved path and it can be shown [95] that the radius of curvature, σ , of the path is given by:

$$\sigma = \frac{e\,B}{P} \qquad\qquad 4.5$$

Where e is the charge on the electron. Values of the product Bσ for various electron energies are listed below [96]:

Electron Energy (K.ev.)	B. (Gauss cm.)
1	110
2	160
3	190
4	215
5	238
10	350

The maximum transverse field of 0.3 gauss therefore leads to a path with a radius of curvature in excess of 300 cm for 1 K.ev. electrons. It follows that the transverse magnetic fields have negligible effect in the accelerating region above the anode because the distance from the anode to the emitter is only 3 cm and also because the electrons reach their terminal velocity within a few tip radii [97] after leaving the emitter. The same cannot be said of the retarding region where the

electrons are brought almost to rest before collection.

The investigation of the vertical field was performed with an external coil. The coil was used to obtain a retarding potential collector current curve for tungsten with the vertical field reduced to zero and another retarding potential vs collector current tungsten curve with the vertical field doubled. There was no measurable difference between the two curves, therefore, the vertical magnetic field was ignored. A similar experiment with a bar magnet showed that a transverse field in the retarding region of only 6 gauss seriously affected the results. It was decided, therefore, to remove the transverse magnetic fields with mu metal shielding.

A cylinder of inner radius a and outer radius b, made from a material with permeability μ, can be shown [98] to possess a shielding ratio g which is given by:

$$g = \frac{1}{4\mu} \left((\mu+1)^2 - \frac{a^2}{b^2} (\mu-1)^2 \right)$$ 4.6

where g is the ratio of the magnetic field before shielding to the magnetic field after shielding. For large values of μ, equation 4.6 reduces to:

$$g \simeq \frac{\mu}{4b^2} (b^2 - a^2)$$ 4.7

Mu metal has a maximum permeability of 100,000 and at 20 gauss this value falls to approximately 20,000 [99]. The value of permeability must, therefore, be nearer 100,000 since the fields to be shielded are not greater than 0.3 gauss.

The shield was made from 0.5 mm thick mu metal sheet. Thin sheet was chosen because it was easy to form into a cylinder and because of the limited space available. The outside diameter was 12 cm . Substitution of these values into equation 4.7 leads to a shielding ration of approximately 200. It was not possible to measure the transverse fields at the collection with the shield in place because the gauss meter probe

was physically too large but it was decided that the shielding was ample because the field values quoted were maximum values and the transverse fields were found to decrease towards the collector. This decision was supported by a repeat of the experiment with the magnet when it was found that the 6 gauss transverse field had very little effect on the retarding potential vs collector current curves when the shield was in place.

The collected current was usually in the range 1×10^{-11} Amp to 1×10^{-9} Amp therefore care had to be taken to ensure than any spurious currents did not exceed 1×10^{-11} Amp . The main cause of extraneous current was leakage across the glass between the high voltage lead-through and the collector. This could have been eliminated by the use of a guard ring[24], but the extra work involved in making the collector did not seem worthwhile. The largest leakage current experienced so far has not exceeded 5×10^{-13} Amp and this can be eliminated by careful adjustment of the electrometer zero settings, therefore, the decision to ignore the guard ring seems to have been correct. The leakage resistance of the screened cables was measured before use and found to be greater than $5 \times 10^{12} \Omega$ therefore, there could be no loss of collector current through them. They did, however, give rise to a capacity effect which was minimised by using cables which were as short as possible. The anode and collector also acted as a capacitor and when the high voltage supply to the anode was switched on there was a large surge of current through the electrometer as this capacitance charged up. The electrometer was protected from the surge by switching to a high current range which also reduced the time constant of the circuit, thereby making the effect less noticeable.

A more serious problem was that of the capacity between the collector and the metalware of the system. The retarding potential could not be isolated from the metalware because of the way in which the tip wheel mechanism had been designed so that it induced a current in the collector

as it was varied. The magnitude of this current was minimised by
earthing the magnetic shield and the maximum current then recorded was
1×10^{-12}Amp for a fast 0 - 10 volt change of retarding potential.
The maximum current was a function of the rise time of the retarding
potential and this was small because of reasons mentioned earlier,
therefore, it was possible to eliminate the problem.

4.4 Sample Cleaning

The method of d.c. field desorption was used to clean the samples
because the usual method of heating was experimentally difficult.
Although the method of a.c. field desorption enables the emission to
be viewed as the sample is being cleaned, it was not used because it
has been shown to cause premature fracture in some semiconductors [38].
The usual procedure was to look at the initial emission pattern, which
usually consisted of a few random spots of light, and then to reverse
the polarity and leave the sample for a few moments. This process was
then repeated at increasing values of the reverse field until symmetry
started to appear in the image.

When a sample was cleaned in this way it was usual to obtain stable
emission for perhaps 5 minutes, after which time the pattern became less
clear and the emission less stable. The time of stable emission usually
increased with the number of experiments performed on a sample, caused
probably by it's becoming cleaner and also by the improvement in the
vacuum as the phosphor screen became outgassed.

4.5 The Method Of Tip Positioning

The electrons in the retarding region have their highest probability of
being collected when they arrive at the collector normally because all
their energy is then available to overcome the potential barrier at the
collector. It has already been shown that the analyser has its maximum
resolution when normal incidence occurs, therefore it is correctly
adjusted when, for any given value of retarding potential, the collected
current is a maximum.

There is also a geometrical factor to be taken into account because more current can pass through the anode aperture as the emitter moves nearer to it. The samples were, therefore, placed in the theoretically ideal position, the current was maximised by watching the electrometer and adjusting the positioning screws, and an energy spectrum was taken. This procedure was repeated above and below the ideal position and the half widths of the distributions obtained were compared. The position corresponding to the minimum half width was then assumed to be the position of optimum resolution.

o o O o o

CHAPTER 5

PRESENTATION OF RESULTS

5.1 Introduction

The performance of the analyser was investigated before it was used
to measure the energy distribution of electrons field emitted from
cadmium sulphide so that the possible sources of error mentioned in
Chapter 2 could be isolated. Tungsten emitters were used for this
purpose because they were easy to make and gave stable emission for
long periods. In addition, the energy distribution of electrons field
emitted from tungsten is well documented and could, therefore, be used
as a measure of the resolution of the analyser.

5.2 The Results From Tungsten

5.2.1 The Dependence Of The Half Width Of The Distribution On Emitter
Position

This experiment was performed in order to ensure that the theoretical
optimum emitter position calculated in Chapter 2 was correct. During
the experiment the emission current density was kept constant and the
energy distribution was measured with the emitter at different axial
distances from the fluorescent screen.

It was found that the shape of the collector-emitter potential difference
vs collector current curve depended to some extent on the collector-
emitter potential difference at which the collector current was maximised.
An example of this effect is shown in Fig. 5.1. Such an effect is to be
expected if, as was suggested in Chapter 2, the reflection coefficient
is a function of the collector-emitter potential difference instead of
being a function of electron energy. The experiment was, therefore,
repeated and the collector current corresponding to different emitter-
screen distances was maximised at the same position on the collector-
emitter potential difference vs collector current curve. The position
chosen was that corresponding to the collector current saturation (X in

Fig. 5.1

CURVE A MAXIMISED AT X
CURVE B MAXIMISED AT Y

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

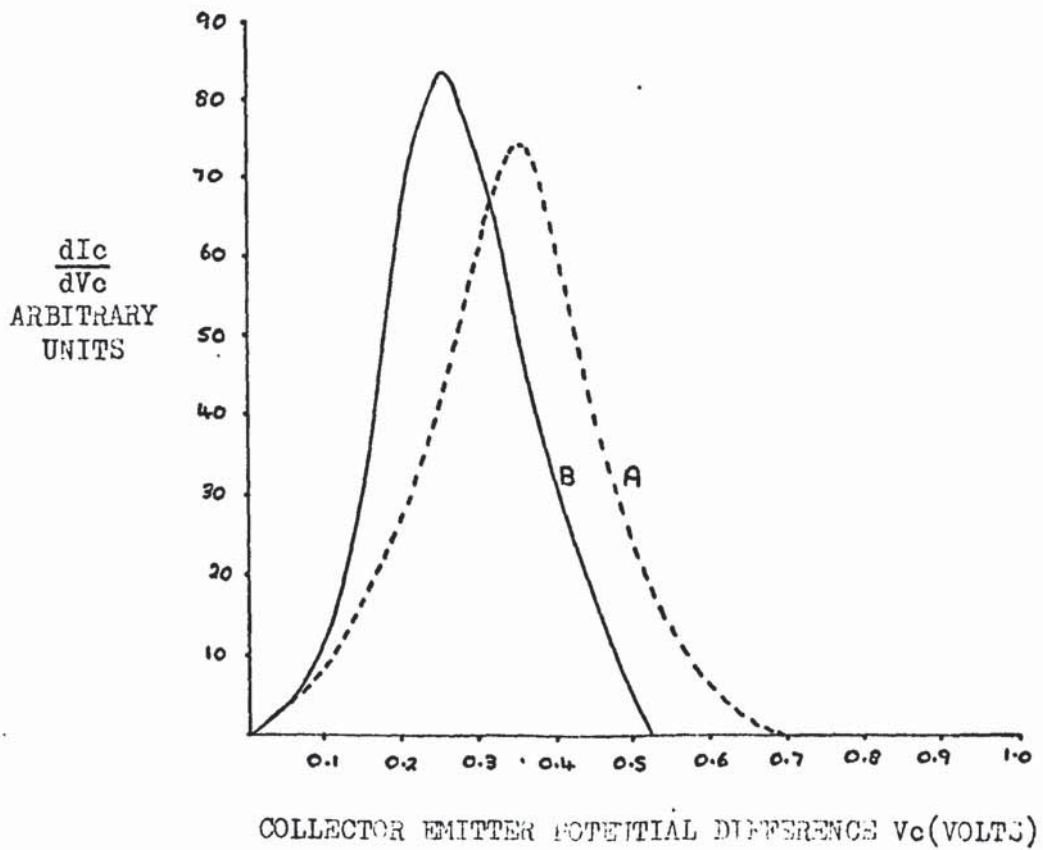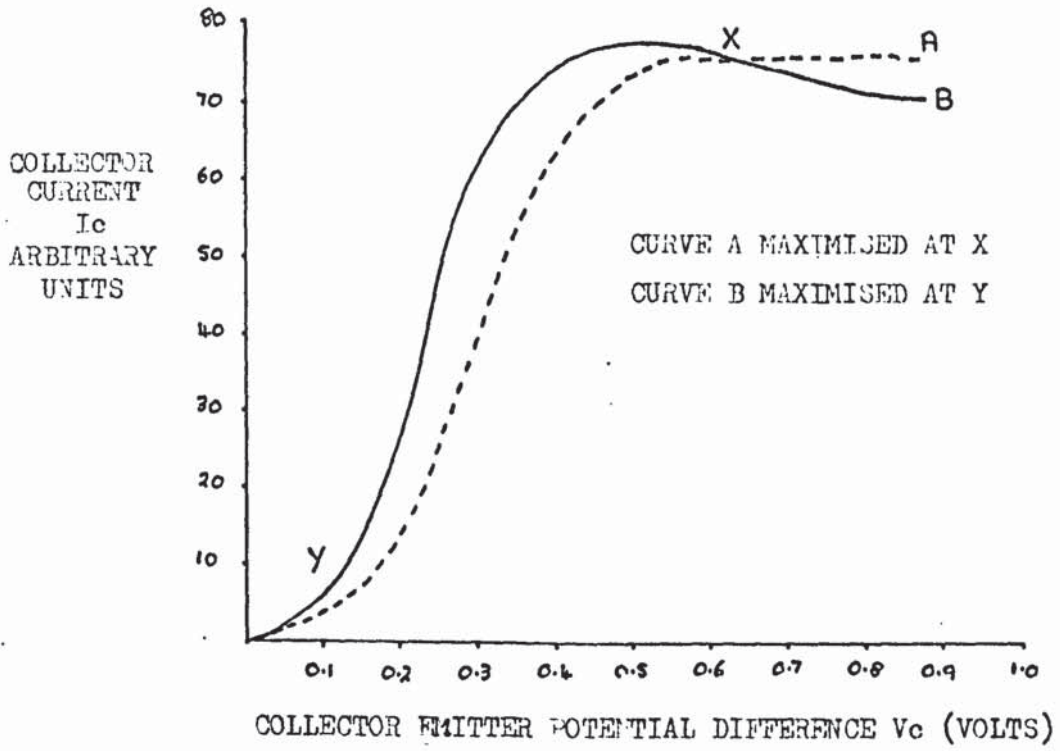COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.1) because this was the easiest position to determine for successive runs.

The results of the second experiment are shown in Fig. 5.2. At each emitter screen distance at least four distributions were measured and the distribution with the smallest half width was used. For the sake of clarity, the emitter-collector potential difference relative to the value corresponding to the onset of collector current is quoted. The derivatives of the experimental curves were obtained numerically by means of the approximate relationship:

$$\left(\frac{dI_c}{dV_c}\right)_{V_c=V} = \frac{I_c(v+h) - I_c(v-h)}{2h} \qquad 5.1$$

Great care was used in chosing h, the separation of the points on the curve, because too small a value leads to a large amount of scatter in the values of dIc/dVc and too large a value leads to a loss of accuracy. The optimum value was chosen by a process of trial and error.

Fig. 5.2 shows that the smallest half width, and hence the greatest resolution, occurred at an emitter-aperture distance of 7.3 cm $\pm$ 1 mm. This distance was determined by adding the constant aperture-screen distance to the variable screen-emitter distance, which was measured by means of a Vernier telescope. The accuracy of the emitter-screen distance was limited to $\pm$ 1 mm because the etched shank of the emitter, which was approximately 1 mm long, could not be resolved by the telescope. Section 2.7.1 showed that the theoretical optimum resolution position is 7.2 cm. The agreement between theory and experiment is, therefore, very good which suggests that the equations used in Chapter 2 accurately describe the electron trajectories in the analyser.

5.2.2 The Dependence Of The Half Width Of The Distribution On Emission
Current Density

In Chapter 2 it was suggested that the half width of the measured energy distribution could depend on the emission current density. An experiment was therefore performed to investigate this effect.
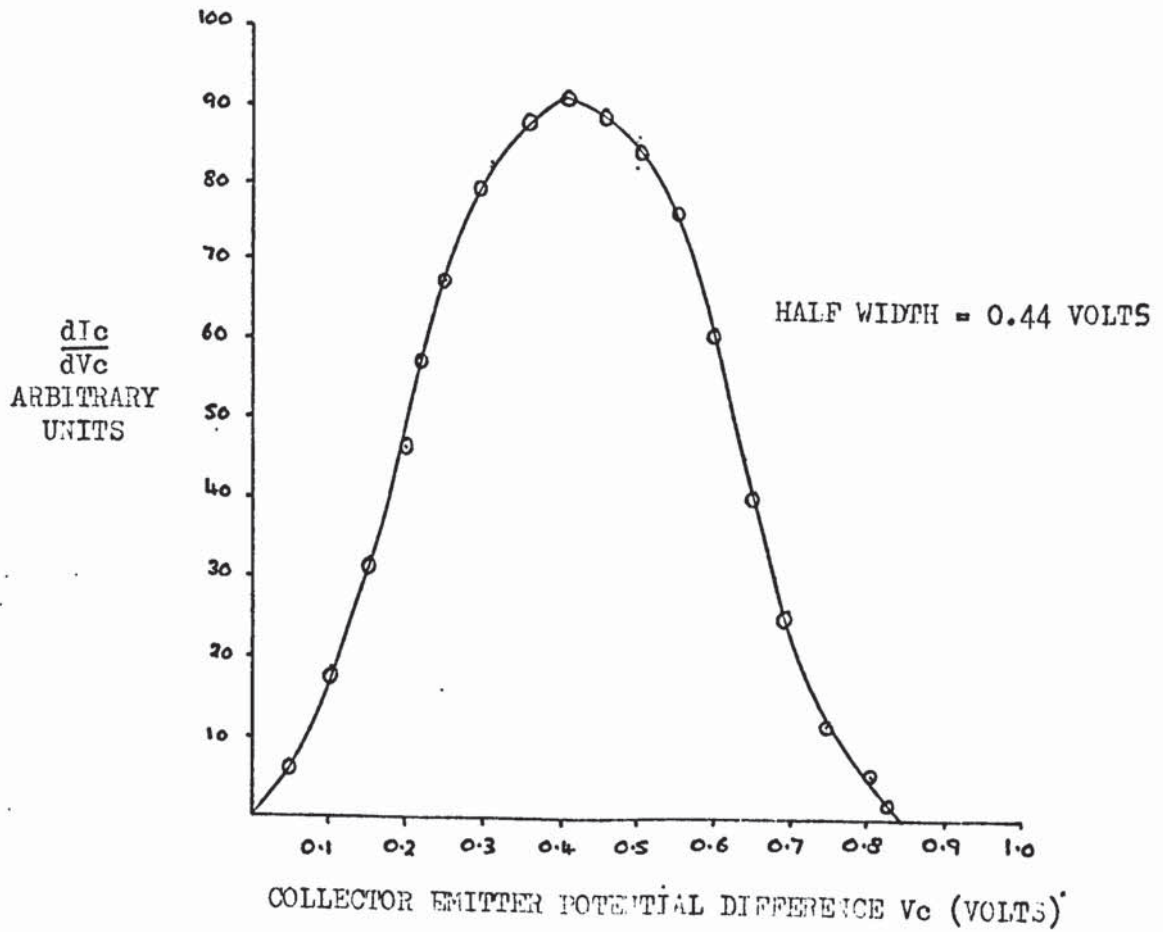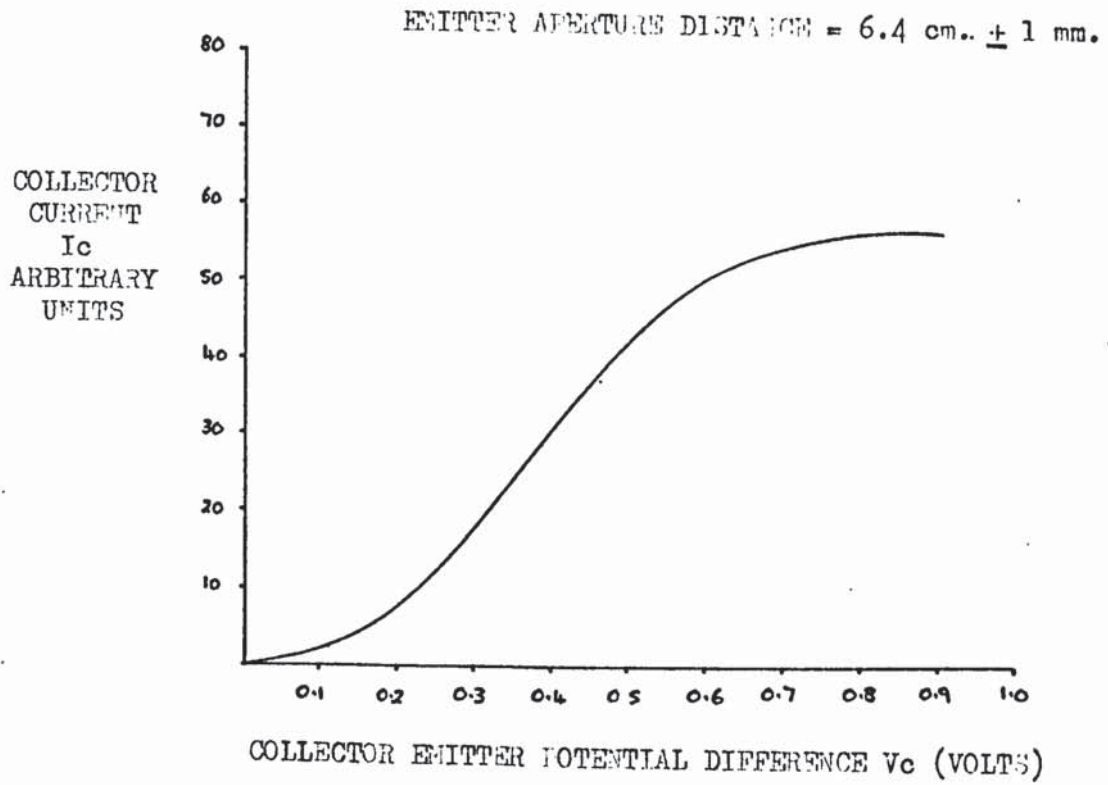
Fig. 5.2(a)

EMITTER APERTURE DISTANCE = 6.4 cm. ± 1 mm.

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

HALF WIDTH = 0.44 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.2(b)

EMITTER APERTURE DISTANCE = 6.7 cm. $\pm$ 1 mm.

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\frac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 0.31 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.2(c)

EMITTER APERTURE DISTANCE = 7.0 cm. ± 1 mm.

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 0.29 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.2(d)

EMITTER APERTURE DISTANCE = 7.3 cm. ± 1 mm.

COLLECTOR CURRENT Ic ARBITRARY UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH = 0.23 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.2(e)

EMITTER APERTURE DISTANCE = 7.6 cm. ± 1 mm.

COLLECTOR CURRENT Ic ARBITRARY UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH = 0.27 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)
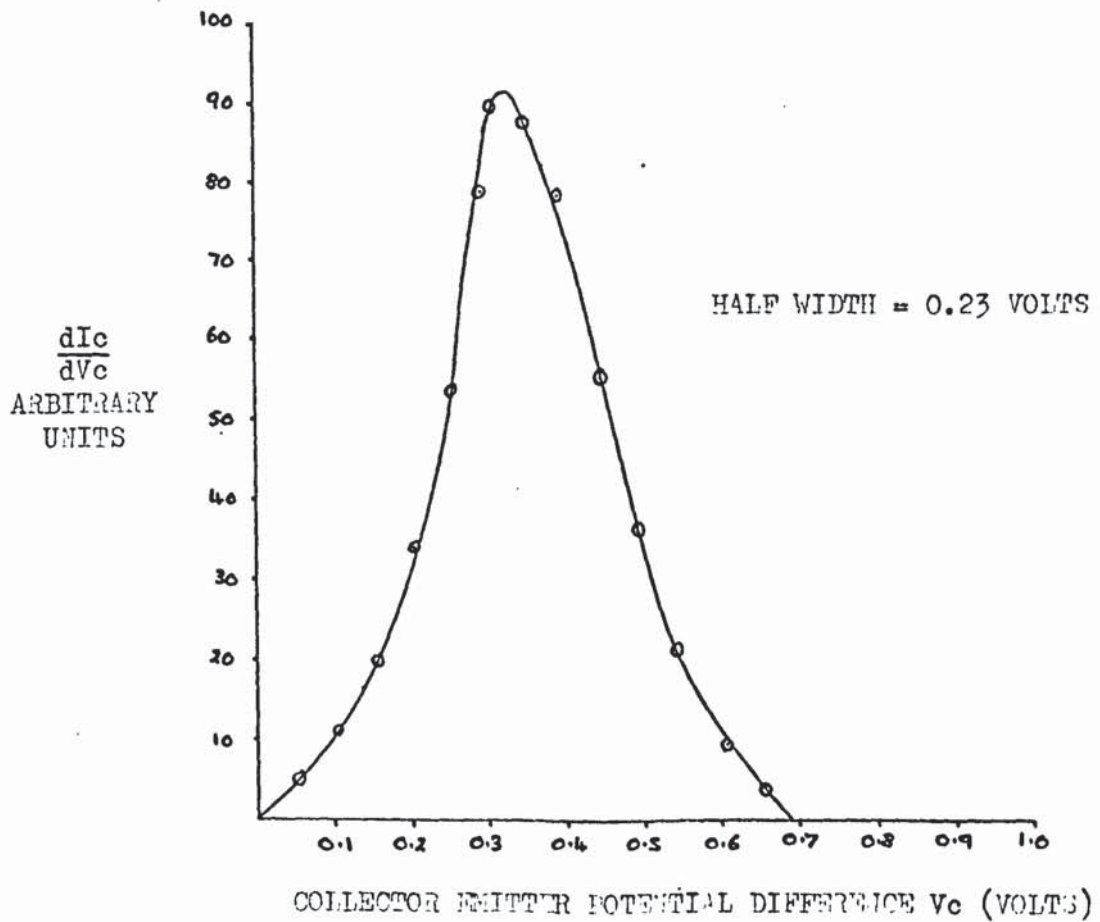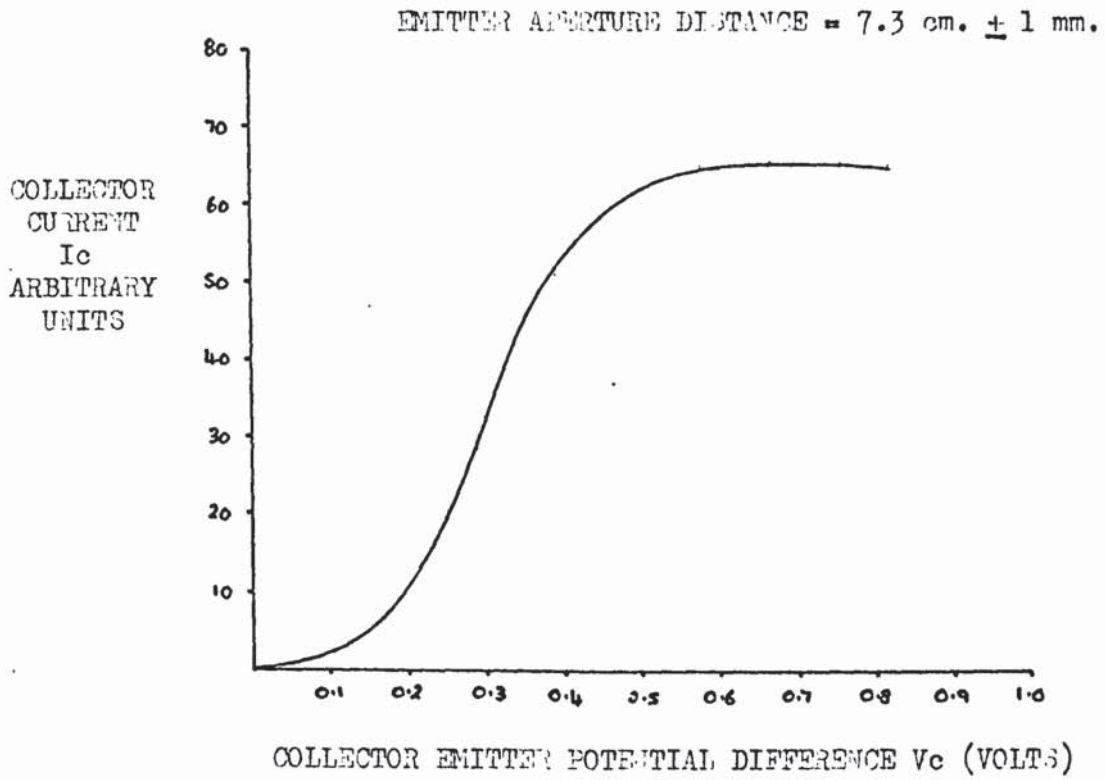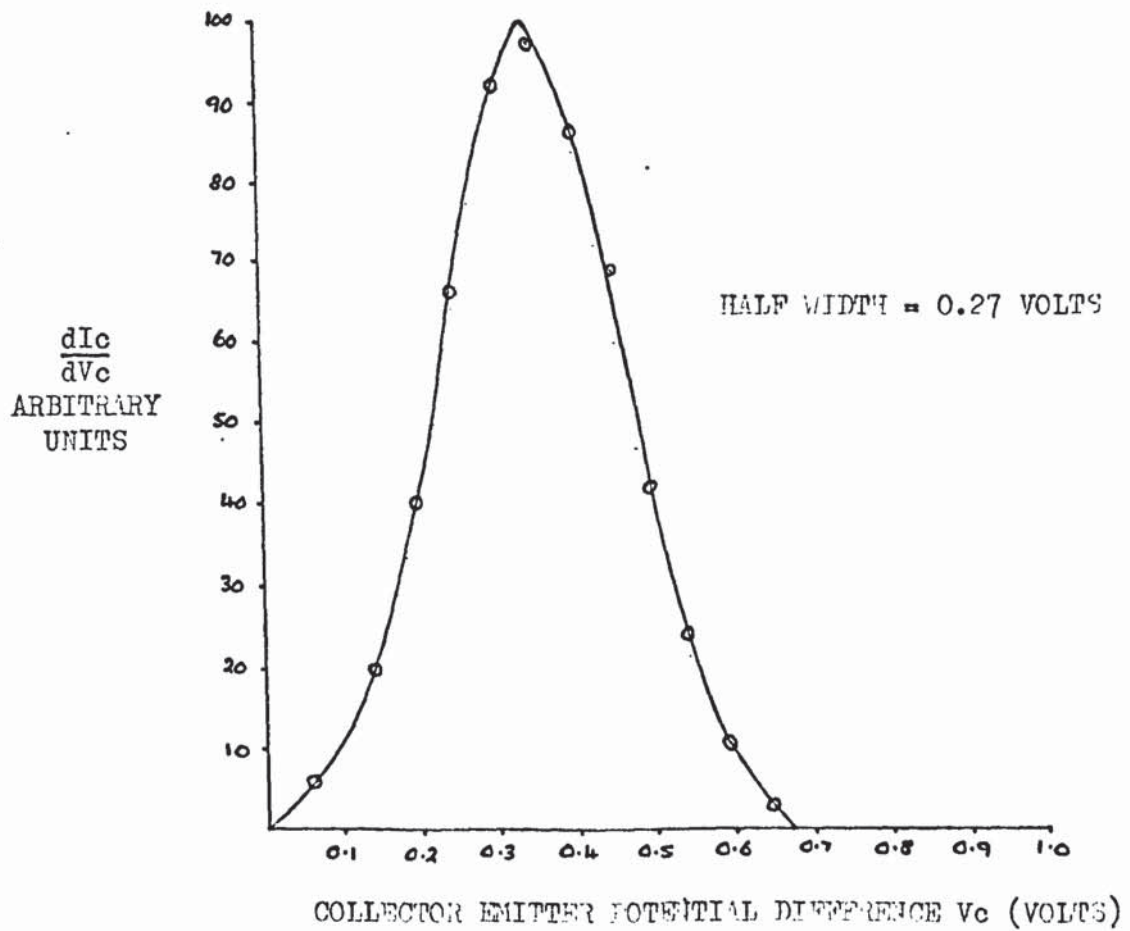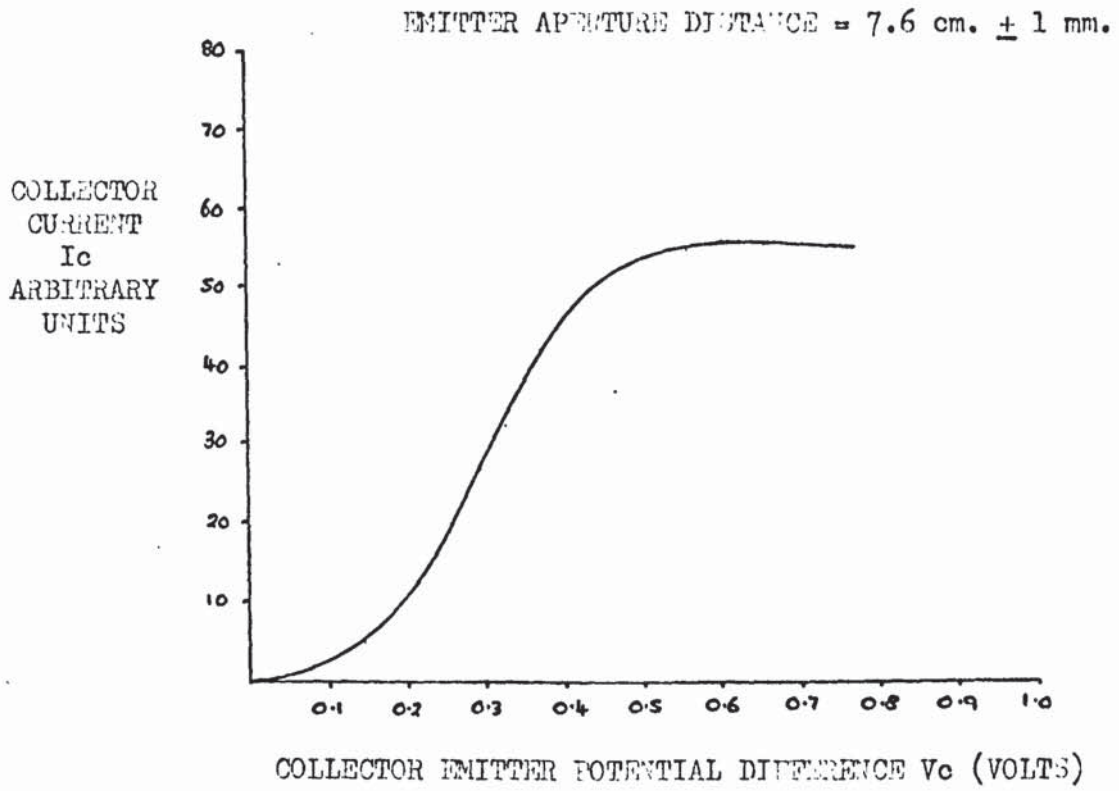
The collector current was maximised at the same point on the collector current vs collector-emitter potential difference curve that Young and Müller [24] used and the emission current density was then increased by increasing the voltage between the emitter and the anode. Fig. 5.3 shows the results of the experiment. It can be seen that the half width of the measured energy distribution depends markedly on the emission current density and that the effect is not caused by the loss of collected current at saturation because this is similar for all results.

5.2.3 Conclusions From The Tungsten Results

The preceeding paragraphs show that accurate results cannot be obtained with this type of analyser unless a great deal of care is taken over positioning the emitter and the emission current density is made as small as possible. The positioning consists of two processes which will be termed coarse and fine, the coarse positioning has already been described in section 5.2.1.

Fig. 5.1 shows the dependence of the distribution on fine positioning and, with the arguments of section 2.8.1, suggests that the true energy distribution cannot be measured unless fine positioning is performed at every value of collector current. To prove this point, and to determine the resolution of the analyser, the energy distribution from tungsten was measured with the collector current maximised just beyond the onset and the emission current density as small as possible. The result of this experiment is shown in Fig. 5.4 and it can be seen that the agreement with Young and Müller's [24] corrected result is very good except at high values of collector-emitter potential difference. Fig. 5.1, which was obtained with the emitter in the same coarse position as that of Fig. 5.4, and at low emission current density, shows that when the collector current is maximised at its saturation value the agreement at high collector-emitter potential differences improves whereas the agreement near the onset deteriorates.

The half width of the measured distribution in Fig. 5.4 is almost in

Fig. 5.3

COLLECTOR CURRENT Ic ARBITRARY UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

A. MAXIMUM COLLECTOR CURRENT = 1×10^{-11} Amp.
B. MAXIMUM COLLECTOR CURRENT = 2×10^{-10} Amp.
C. MAXIMUM COLLECTOR CURRENT = 5×10^{-10} Amp.

$\frac{dIc}{dVc}$ ARBITRARY UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.4

COLLECTOR
CURRENT
I_c
ARBITRARY
UNITS

MAXIMUM COLLECTOR CURRENT $= 1 \times 10^{-11}$ Amp.

COLLECTOR EMITTER POTENTIAL DIFFERENCE V_c (VOLTS)

—— EXPERIMENT AT $300°$K
- - - MÜLLER $300°$K CORRECTED
EXPERIMENTAL CURVE

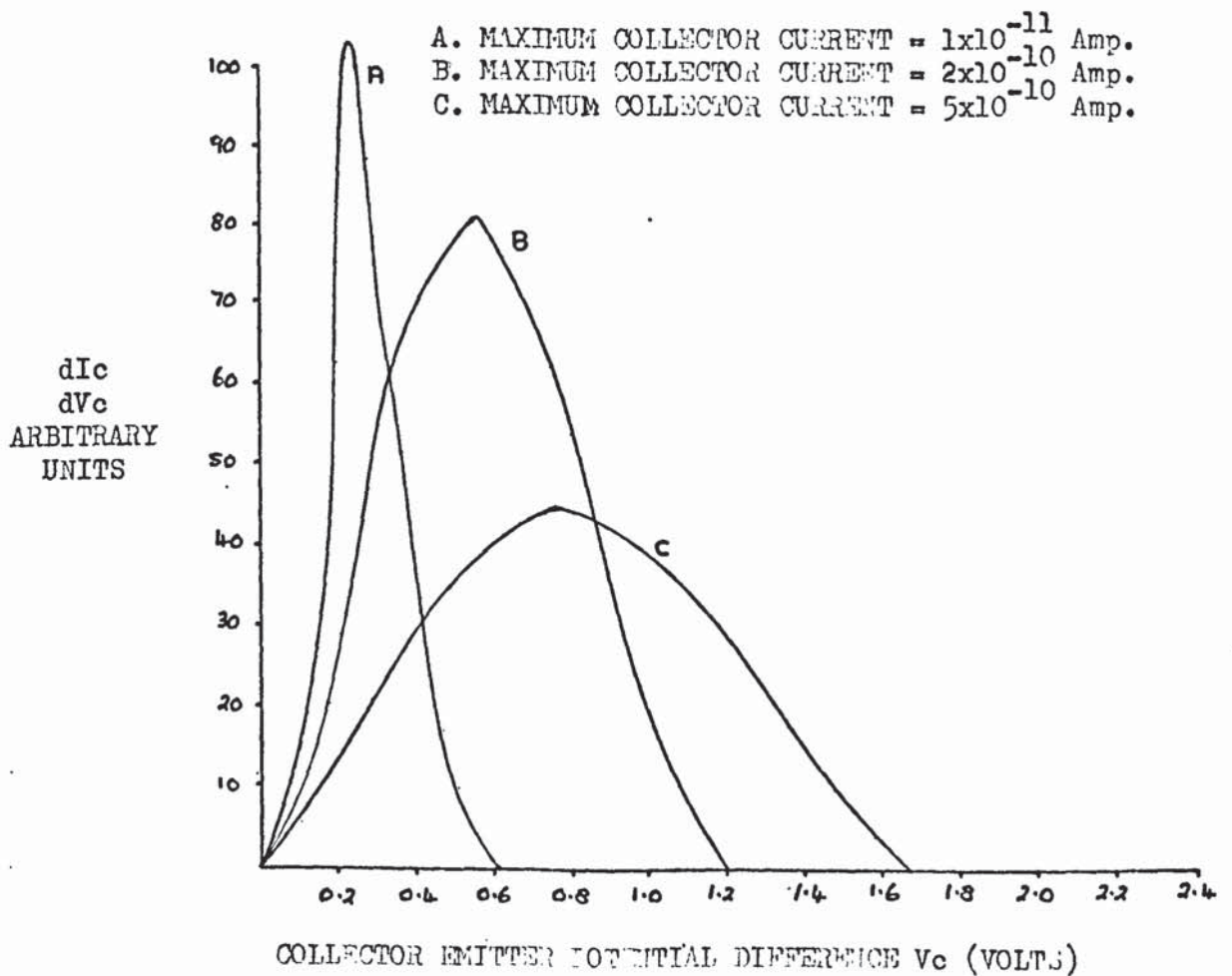$\dfrac{dI_c}{dV_c}$
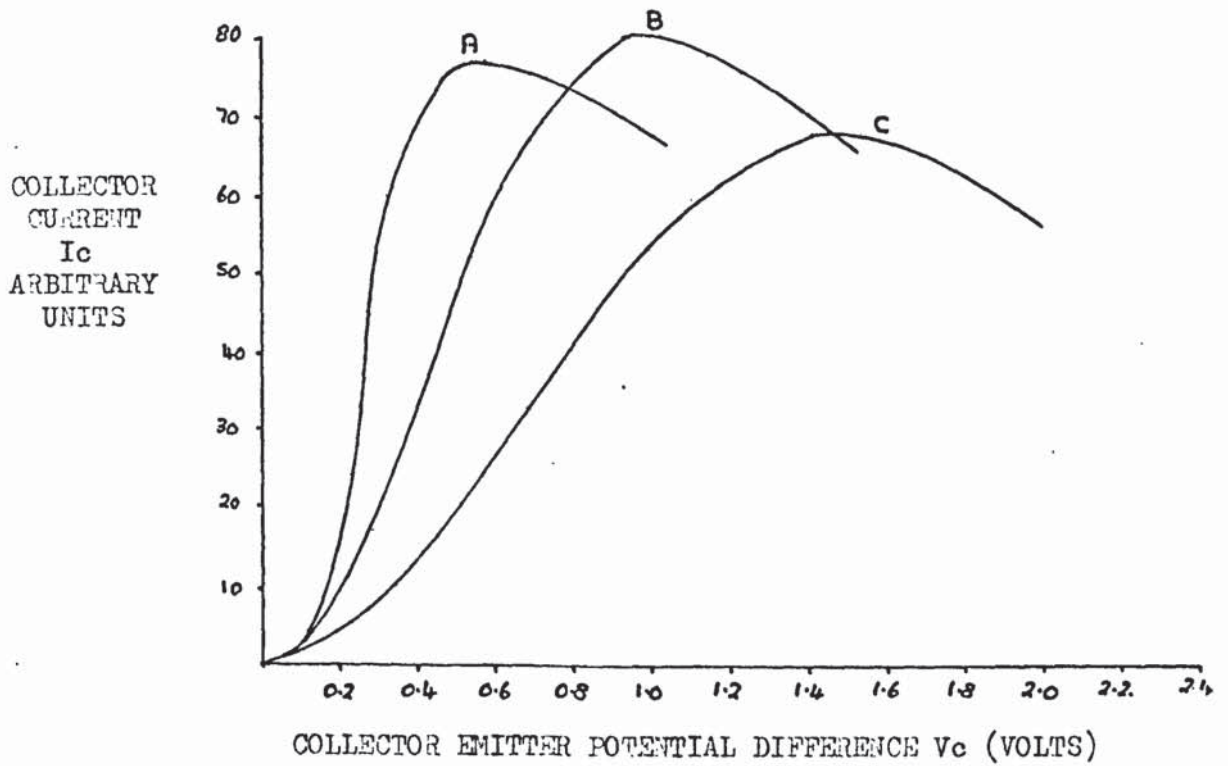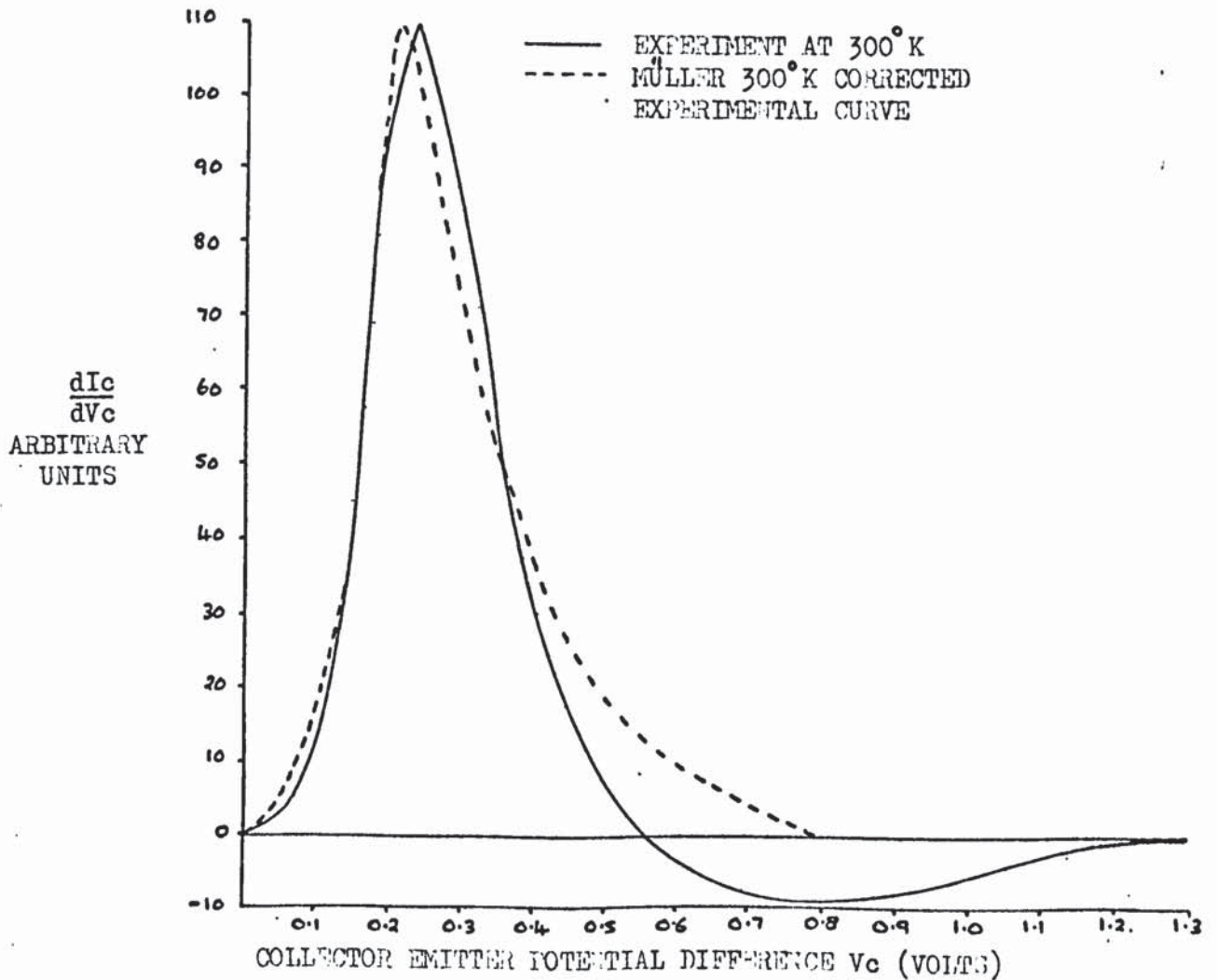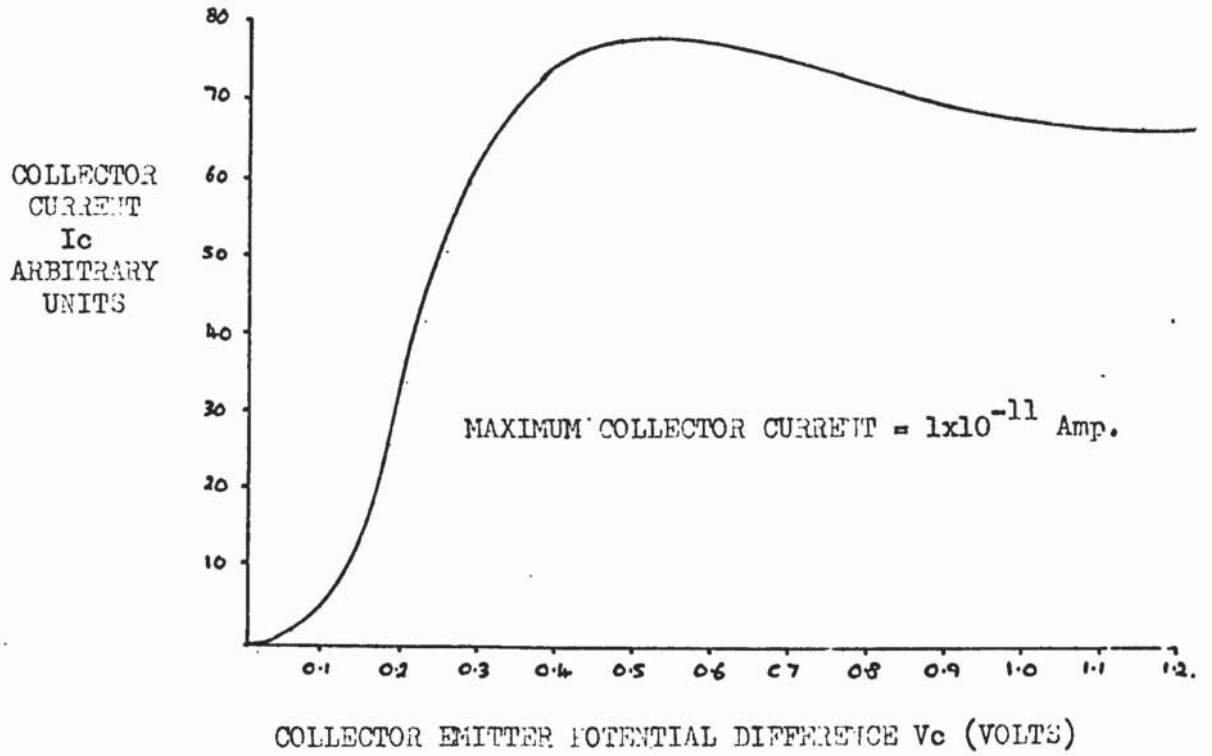ARBITRARY
UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE V_c (VOLTS)

exact agreement with Young and Müller's [24] result and the full width
of the distribution which, from the arguments of section 2.8.1 occurs
when $dIm(v)/dVc$ takes its maximum negative value, is also in good
agreement with the value obtained by Young and Müller [24]. The small
separation between the peaks of the distribution would dissapear if the
Young and Müller [24] correction were applied to the results. The
resolution of the analyser must, therefore, be at least as good as that
of Young and Müller's [24] analyser which, they claim, is of the order
of 0.03 eV.

5.3 The Results From Cadmium Sulphide

The cadmium sulphide results were much more difficult to obtain than
those from tungsten for the following reasons:

(1) Non-Stable Emission Current

When the anode voltage was applied to a cadmium sulphide emitter for
the first time the emission pattern usually consisted of a few random
spots of light which were very unstable and which were caused by large
fluctuations in the emission current density. These fluctuations were
removed in the following way:

After the initial field emission voltage had been applied the anode
polarity was reversed and a negative voltage of the same magnitude was
used for a short time. The polarity was then reversed again so that
field emission occurred and it was usually found that more random spots
of light had appeared and that the onset of emission occurred at a
slightly lower anode voltage. The anode voltage was then increased in
steps of approximately 500 volts and the above process was carried out
after each step until a voltage of approximately one and a half times
the initial voltage had been reached, when the field desorbing voltage
alone was increased. In this way, as the voltage increased, more spots
began to appear until the pattern started to quiver i.e. changed rapidly
between two stable configurations. At higher anode voltages the
quivering ceased and the random spots of one of the stable configurations

coalesced into one large spot which subsequently started to show structure. Finally, a pattern, which was recogniseable between different samples, was obtained and the emission current stabilised.

This procedure was repeated several times on some emitters before stable emission was achieved and thus occupied a great deal of experimental time. Even when the stable pattern was obtained it could sometimes be lost if too high an anode voltage was used or if the sample was subjected to intense illumination, but it usually returned when the voltage and/or the illumination were reduced. This phenomenon made it especially difficult to obtain results at high current density because in some cases the entire procedure had to be carried out many times in order to obtain the stable pattern over the range of anode voltages used.

(2) Potential Drop Along The Emitter

Although the emission currents drawn from the cadmium sulphide samples were small, typically in the range 1×10^{-7} Amp to 5×10^{-10} Amp , the high resistance of the emitters led to a potential drop along their length which in some cases amounted to tens of volts.

Fig. 5.5 shows the emitter-collector circuit in schematic form, the anode has been omitted for the sake of clarity. R_e is the emitter resistance and I_e is the emission current. It is obvious that point B is $I_e R_e$ volts positive with respect to point A and that the collector is higher in energy that the emitter tip. Point A of the emitter has, therefore, to be increased in energy before collection is possible and this is achieved by means of potentiometer D. When a negative potential of $I_e R_e$ volts is applied at A the energy difference between the emitter tip and the collector is zero and any further negative potential placed on A enables electrons to be collected and the energy distribution to be measured.

Although the potential drop effect described above could easily be removed, it unfortunately caused another more difficult problem. The

Fig. 5.5
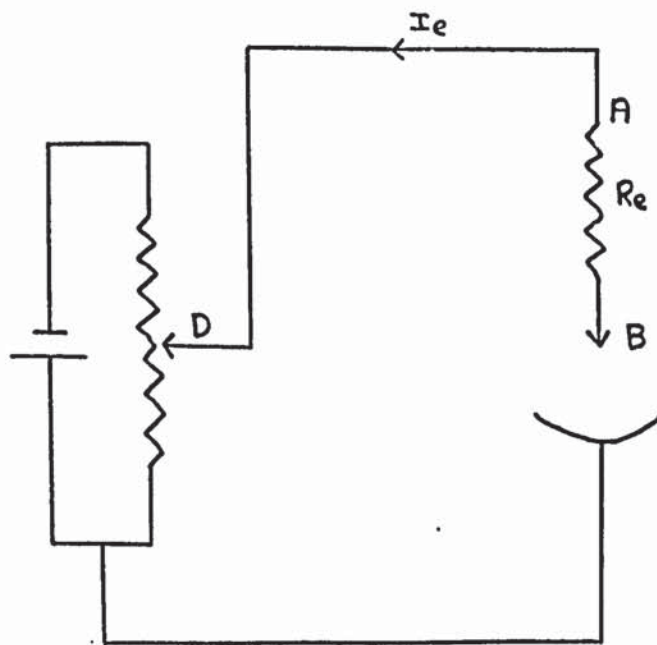
SCHEMATIC OF
EMITTER – COLLECTOR
CIRCUIT

energy distribution from cadmium culphide is approximately 3 volts wide so that the collector current rises from zero to its saturation value when the potential at A is changed by 3 volts. If the current through the emitter is not stable the potential drop along the emitter also changes, therefore, the potential difference between the emitter tip and the collector may not change in the same way as the potential at A. In the lower resistance samples this effect was not too severe but in the high resistance samples, where the potential drop could be as high as 70 volts, even a 1% variation in the emission current could lead to a change in potential difference of several volts between the collector and the emitter tip. It will, therefore, be appreciated that a great deal of difficulty was experienced in obtaining meaningful results from the high resistance samples and that the lower resistance samples also caused the same problem when they were used under conditions of zero illumination.

(3) Sample Positioning

The fluctuations in the potential drop along the emitter made it impossible in most cases to maximise the collector current at a collector-emitter potential difference half way between the onset value and the value corresponding to saturation. Most of the results were, therefore, taken with the collector current maximised at the saturation point where the effect was not so severe and there was thus a loss in resolution at the onset, as described in section 5.2.1. The initial high resistance results also suffered a further loss of resolution because they were obtained before the broadening caused by high current density had been discovered. This loss of resolution occurred because when the experiment described in section 5.2.1 was first performed the emission current was increased as the emitter was moved away from the screen so that the maximum collector current always had the same value. This proved more convenient for taking measurements. The change in resolution caused by different emitter positions plus the change caused

by the varying current density, therefore, led to an incorrect emitter position being used.

As experience was gained in the use of the analyser it proved possible to obtain results in spite of the problems described above and in the final set of experiments very stable emission currents could be achieved for fairly long periods. In order to ensure that the energy distributions were accurate each was measured at least four times and the distributions presented in the following pages are the best of those obtained.

5.3.1 The High Resistance Cadmium Sulphide Results

Figs. 5.6.1 to 5.6.5 show the energy distributions obtained from the high resistance (10^7 ohm. cm) samples. Results 5.6.1 to 5.6.4 were taken at low resolution, as described in the previous section, whereas result 5.6.5 was obtained during the final set of experiments and was, therefore, taken at high resolution. A collimated beam of white light from a tungsten filament lamp was used to illuminate the samples. It can be seen that two peaks were present in all but the result from sample five and that the second peak could be depressed by illumination in the case of sample one and removed by a combination of illumination and field desorption in samples two and three. After the same treatment sample four still showed evidence of the second peak. The result from sample five was obtained after the process described in section 5.3 had been performed on it, and at high illumination which reduced its resistance and hence reduced the voltage drop along its length. The half widths and overall widths of the energy distributions from samples one to four are not meaningful because of reasons already mentioned but it is believed that the existence of the second peak is real because it occurred in three different emitters and also because no mechanism can be proposed which relates the phenomenon to the resolution of the analyser.

5.3.2 The Lower Resistance Cadmium Sulphide Results

Figs. 5.7.1 to 5.7.4 show the results obtained from the lower resistance

Fig. 5.6.1(a)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS
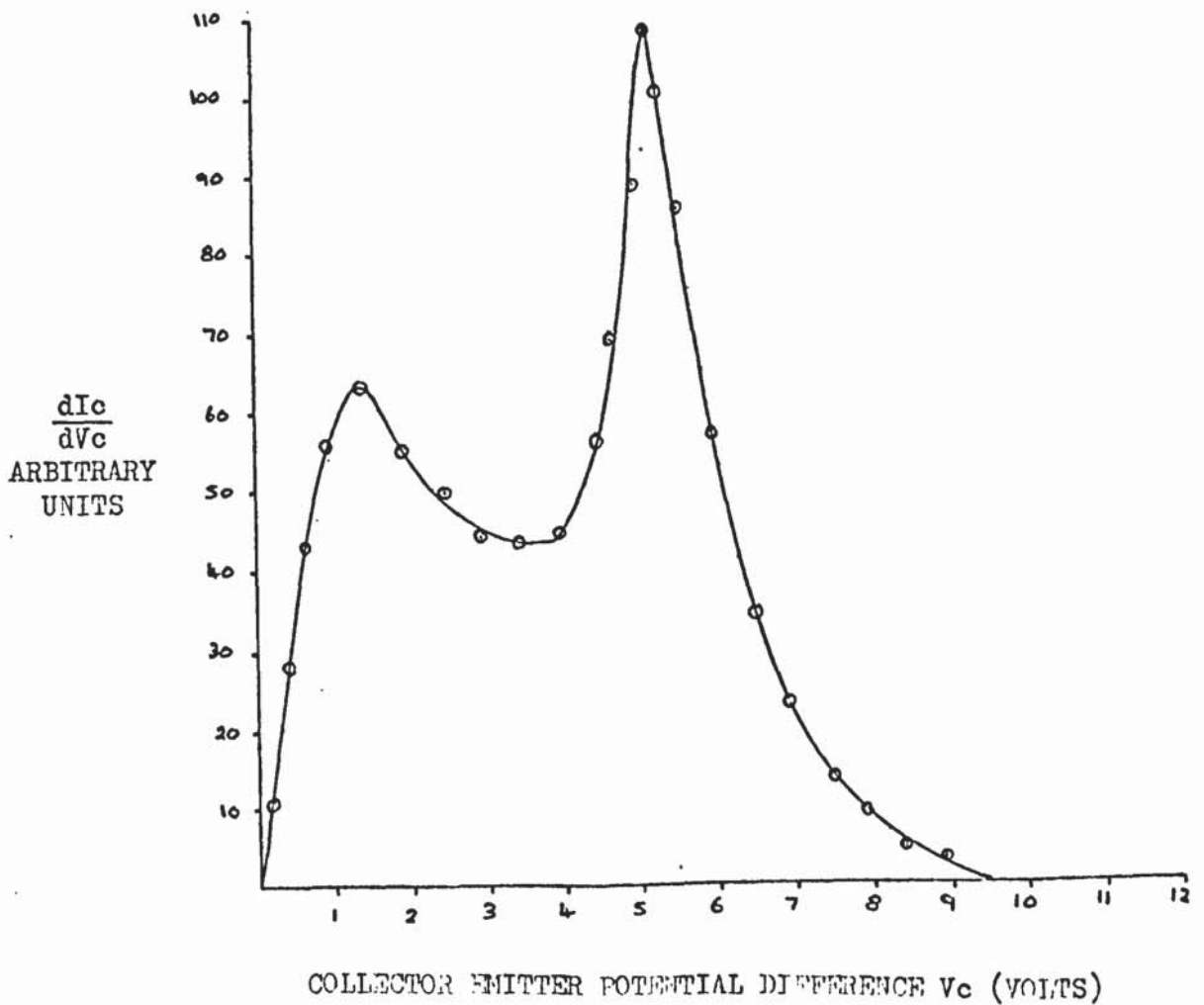
Cd.S. EMITTER 1

MAXIMUM COLLECTOR CURRENT = 4.5×10^{-10} Amp.
ONSET VOLTAGE = 70 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.1(b)

Cd.S. EMITTER 1

MAXIMUM COLLECTOR CURRENT = 2×10^{-9} Amp.
ONSET VOLTAGE = 40 VOLTS
MAXIMUM LIGHT

COLLECTOR CURRENT Ic ARBITRARY UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.2(a)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 2

ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\frac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH OF FIRST PEAK = 1.1 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.2(b)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 2

MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 0.9 VOLTS
TOTAL WIDTH = 4.25 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.3(a)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 3

ONSET VOLTAGE = 60 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\frac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH OF FIRST PEAK = 1.3 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.3(b)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 3

MAXIMUM COLLECTOR CURRENT = 1x10^{-10}Amp.
ONSET VOLTAGE = 58 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 1.3 VOLTS
TOTAL WIDTH = 3.6 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.4(a)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 4

ONSET VOLTAGE = 35 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH OF FIRST PEAK = 1.0 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.6.4(b)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 4

ONSET VOLTAGE = 42 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\frac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH = 0.8 VOLTS
TOTAL WIDTH = 3.45 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig.5.6.5

Cd.S. EMITTER 5

MAXIMUM COLLECTOR CURRENT = 1.1×10^{-10} Amp.
ONSET VOLTAGE = 57 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS).

HALF WIDTH = 0.65 VOLTS
TOTAL WIDTH = 3.0 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS).

(10^6 ohm cm) cadmium sulphide emitters.

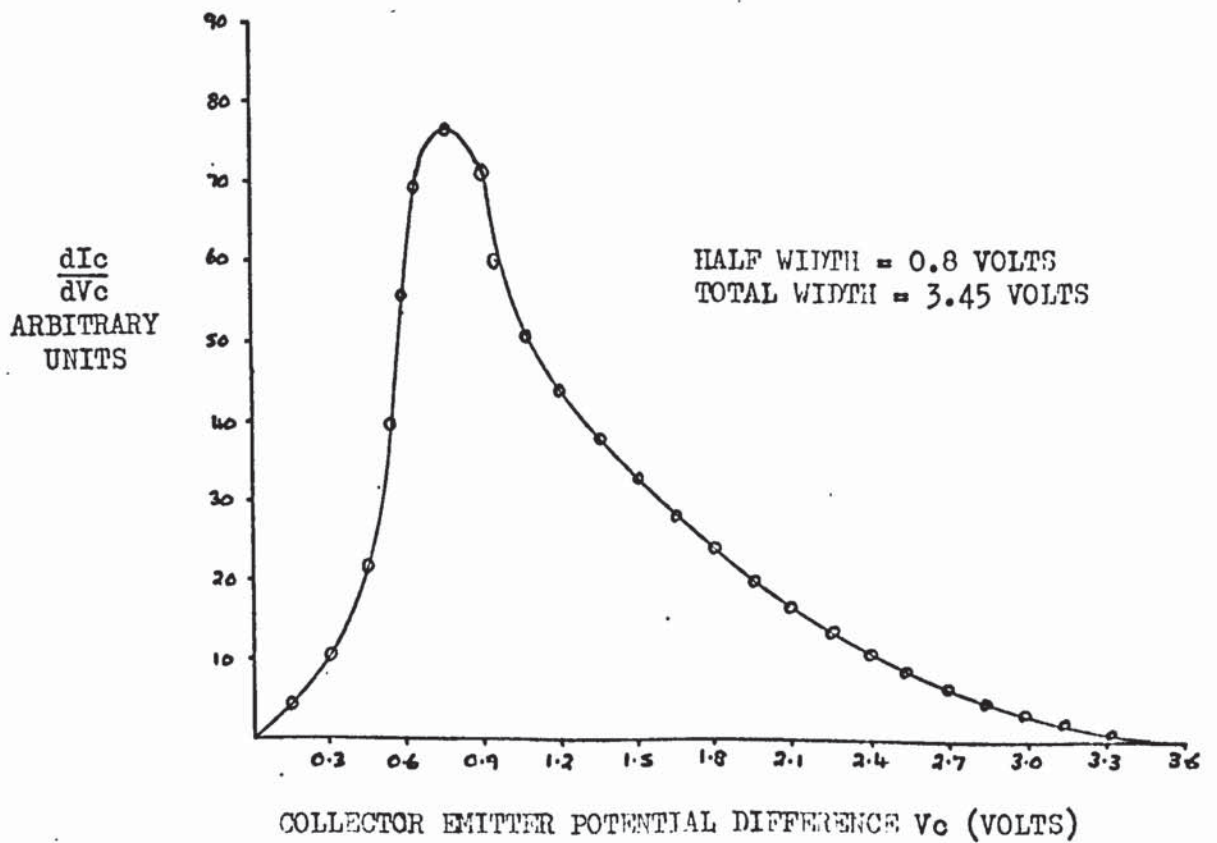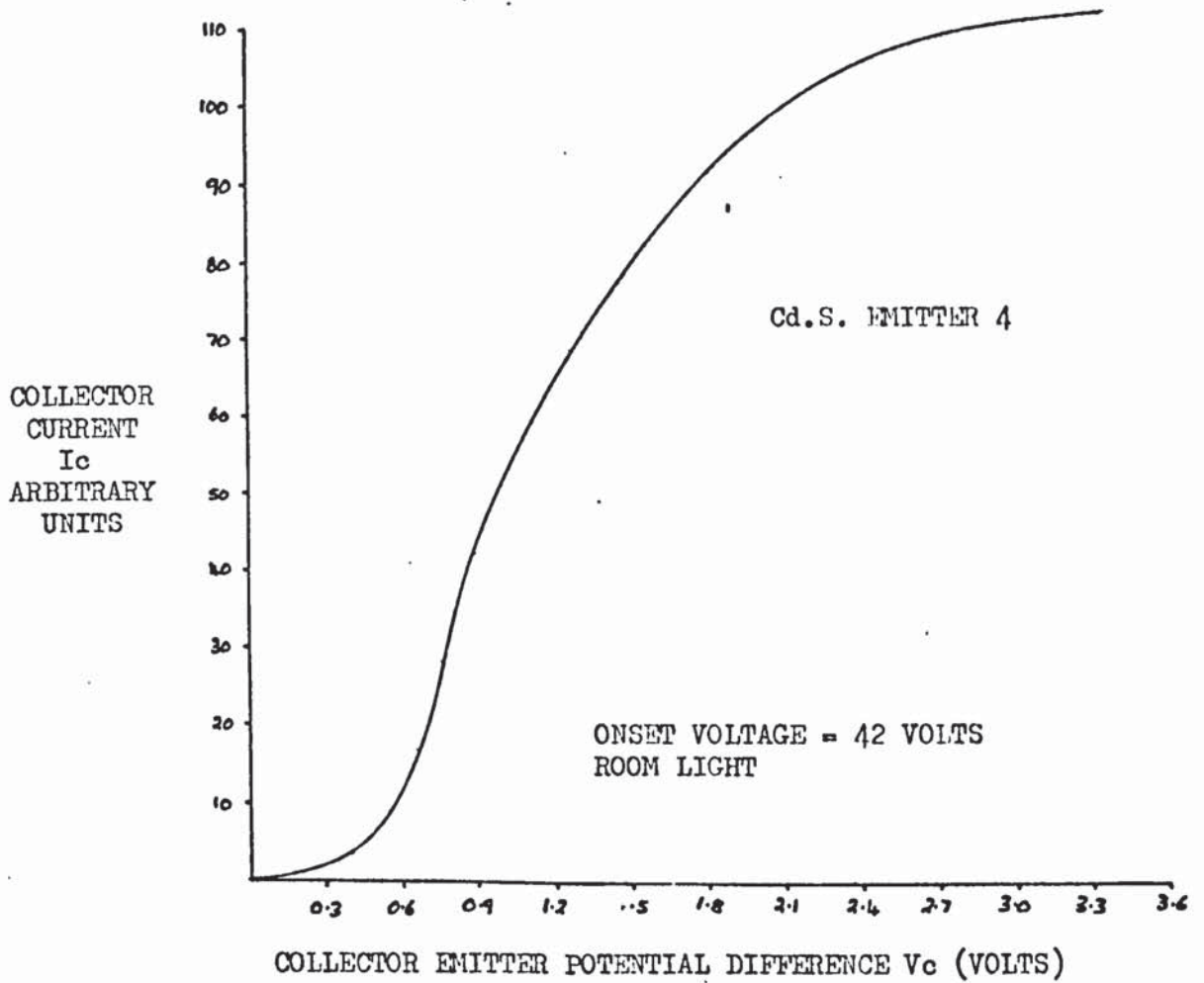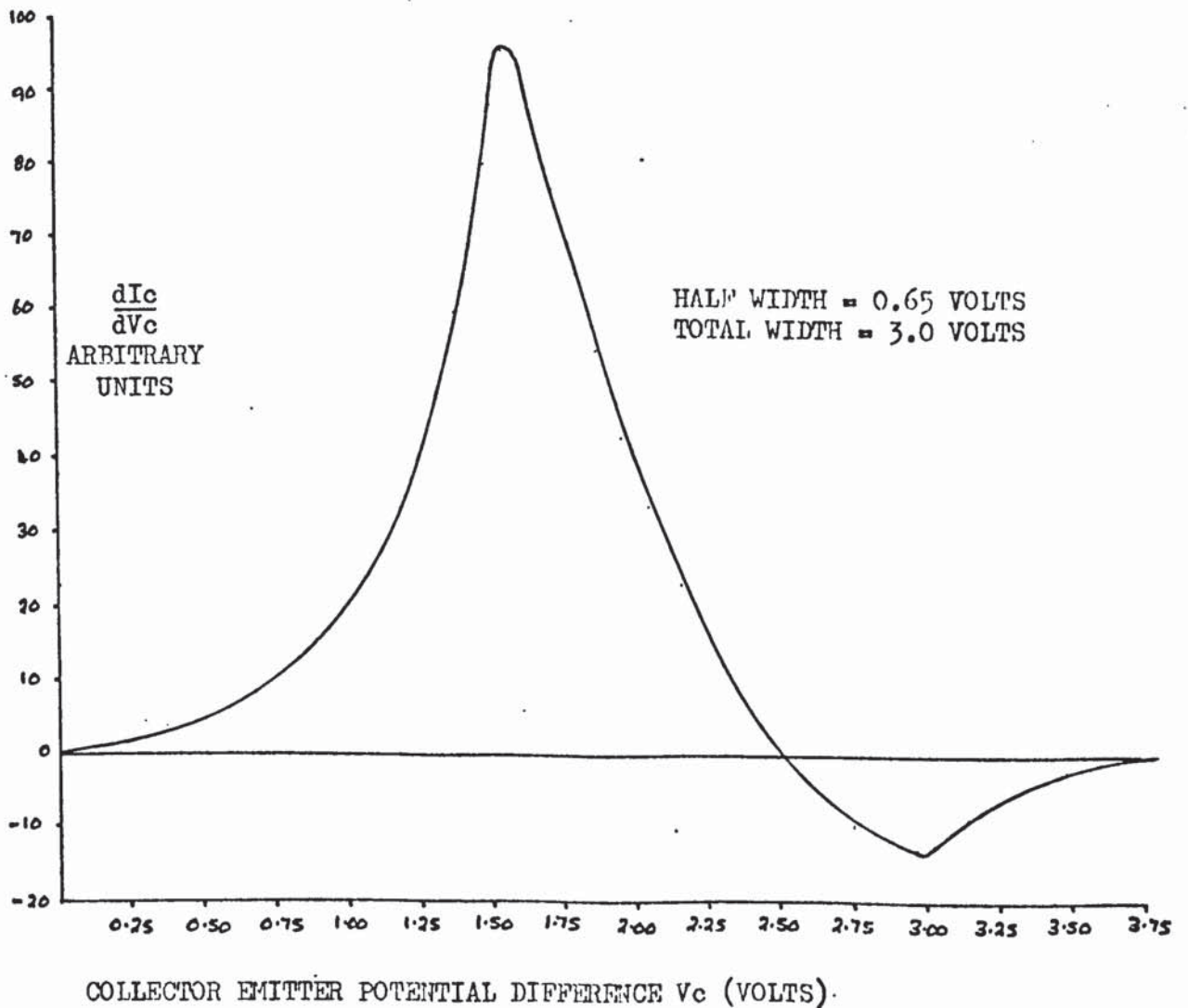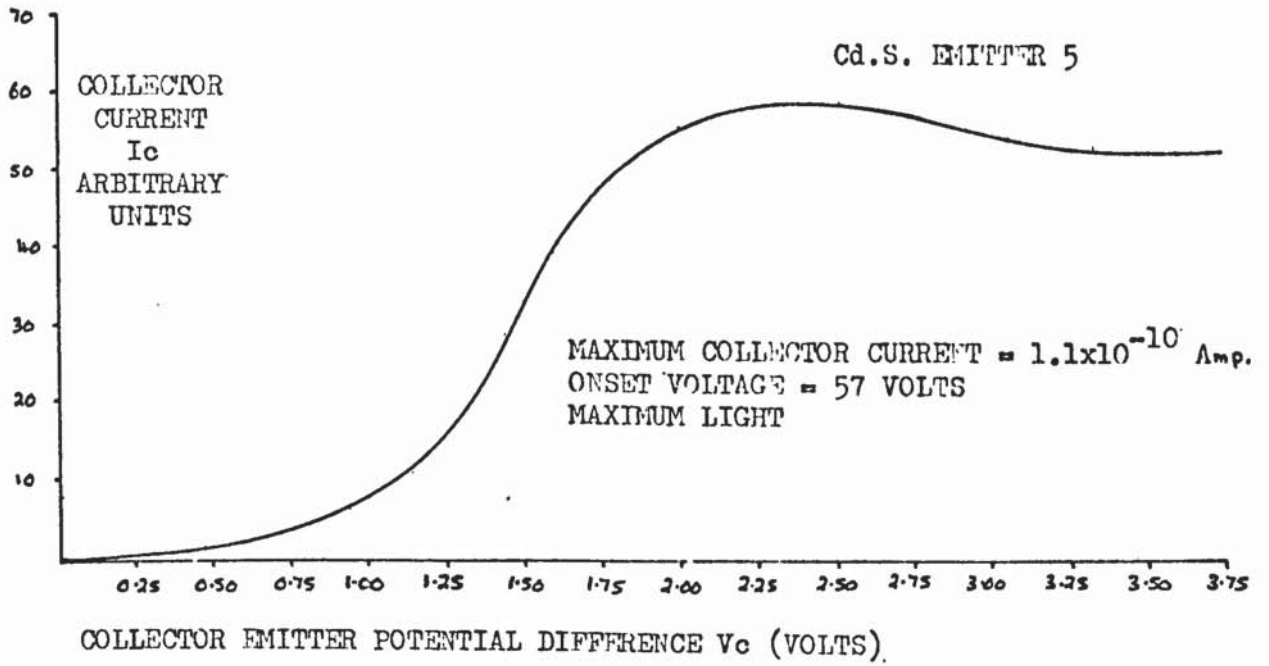A range of anode voltages and light intensities were applied to the emitters and the corresponding onset voltages and maximum collector currents were noted. Each result was taken at the same emitter-screen distance as that of the best tungsten result and in many instances it proved possible to maximise the collector current at a collector-emitter potential difference half way between the onset value and the value corresponding to saturation. The total widths of the energy distributions were obtained by the method described in section 2.8.1 and stable emission was achieved by following the procedure of section 5.3.(1).

A second peak was not observed in any of the distributions. The results shown in Figs. 5.7.1 to 5.7.4 were obtained after stable emission had been achieved and table 5.1 presents the experimental data in a more convenient form. The maximum collector current is in amperes and the half width, total width and onset voltage are quoted in electron volts.

TABLE 5.1

Emitter 6

Light	Maximum Collector Current	Onset Voltage	Half Width	Total Width
Max.	1.1×10^{-11}	10.5	0.57	1.80
Max.	2×10^{-11}	11.0	0.65	2.00
Max.	1.7×10^{-10}	14.5	0.70	2.23
Max.	8.5×10^{-10}	17.0	1.56	4.20
Min.	1×10^{-11}	16.0	0.65	2.20
Min.	1.5×10^{-10}	22.0	1.10	3.25
Min.	2.8×10^{-10}	32.0	1.50	4.00

Emitter 7

Light	Maximum Collector Current	Onset Voltage	Half Width	Total Width
Max.	3.2×10^{-11}	9.0	0.58	1.88
Max.	2.8×10^{-10}	18.0	0.82	3.10
Max.	6.5×10^{-10}	27.0	1.60	3.90
Min.	1×10^{-11}	32.0	0.67	1.80

Fig. 5.7.1(a)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 6

MAXIMUM COLLECTOR CURRENT = 1.1×10^{-11} Amp.
ONSET VOLTAGE = 10.5 VOLTS
MAXIMUM LIGHT

COLLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 0.57 VOLTS
TOTAL WIDTH = 1.80 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.1(b)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 6

MAXIMUM COLLECTOR CURRENT = 2×10^{-11} Amp.
ONSET VOLTAGE = 11 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 0.65 VOLTS
TOTAL WIDTH = 2.0 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.1(c)

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.1(d)

Cd.S. EMITTER 6

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

MAXIMUM COLLECTOR CURRENT = 8.5×10^{-10} Amp.
ONSET VOLTAGE = 17 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 1.56 VOLTS
TOTAL WIDTH = 4.2 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.2(a)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 6

MAXIMUM COLLECTOR CURRENT = 1×10^{-11} Amp.
ONSET VOLTAGE = 16 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\frac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH = 0.65 VOLTS
TOTAL WIDTH = 2.2 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.2(b)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 6

MAXIMUM COLLECTOR CURRENT = 1.5×10^{-10} Amp.
ONSET VOLTAGE = 22 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 1.1 VOLTS
TOTAL WIDTH = 3.25 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.2(c)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 6

MAXIMUM COLLECTOR CURRENT = 2.8×10^{-10} Amp.
ONSET VOLTAGE = 32 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH = 1.5 VOLTS
TOTAL WIDTH = 4.0 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.3(a)

COLLECTOR CURRENT Ic ARBITRARY UNITS

Cd.S. EMITTER 7

MAXIMUM COLLECTOR CURRENT = 3.2×10^{-11} Amp.
ONSET VOLTAGE = 9 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS

HALF WIDTH = 0.58 VOLTS
TOTAL WIDTH = 1.88 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.3(b)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 7

MAXIMUM COLLECTOR CURRENT = 2.8×10^{-10} Amp.
ONSET VOLTAGE = 18 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS).

$\dfrac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 0.82 VOLTS
TOTAL WIDTH = 3.1 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.3(c)

COLLECTOR
CURRENT
Ic
ARBITRARY
UNITS

Cd.S. EMITTER 7

MAXIMUM COLLECTOR CURRENT = 6.5×10^{-10} Amp.
ONSET VOLTAGE = 27 VOLTS
MAXIMUM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\frac{dIc}{dVc}$
ARBITRARY
UNITS

HALF WIDTH = 1.6 VOLTS
TOTAL WIDTH = 3.9 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.7.4(a)

COLLECTOR CURRENT Ic ARBITRARY UNITS (y-axis)

Cd.S. EMITTER 7

MAXIMUM COLLECTOR CURRENT = 1×10^{-11} Amp.
ONSET VOLTAGE = 30 VOLTS
ROOM LIGHT

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

$\dfrac{dIc}{dVc}$ ARBITRARY UNITS (y-axis)

HALF WIDTH = 0.67 VOLTS
TOTAL WIDTH = 1.80 VOLTS

COLLECTOR EMITTER POTENTIAL DIFFERENCE Vc (VOLTS)

Fig. 5.8

EMITTER 6
ROOM LIGHT

Fig. 5.9

EMITTER 7
ROOM LIGHT

A careful study of TABLE 5.1 leads to the following conclusions:

(1) For a given light intensity the half width and total width increase as the current and potential drop increase.
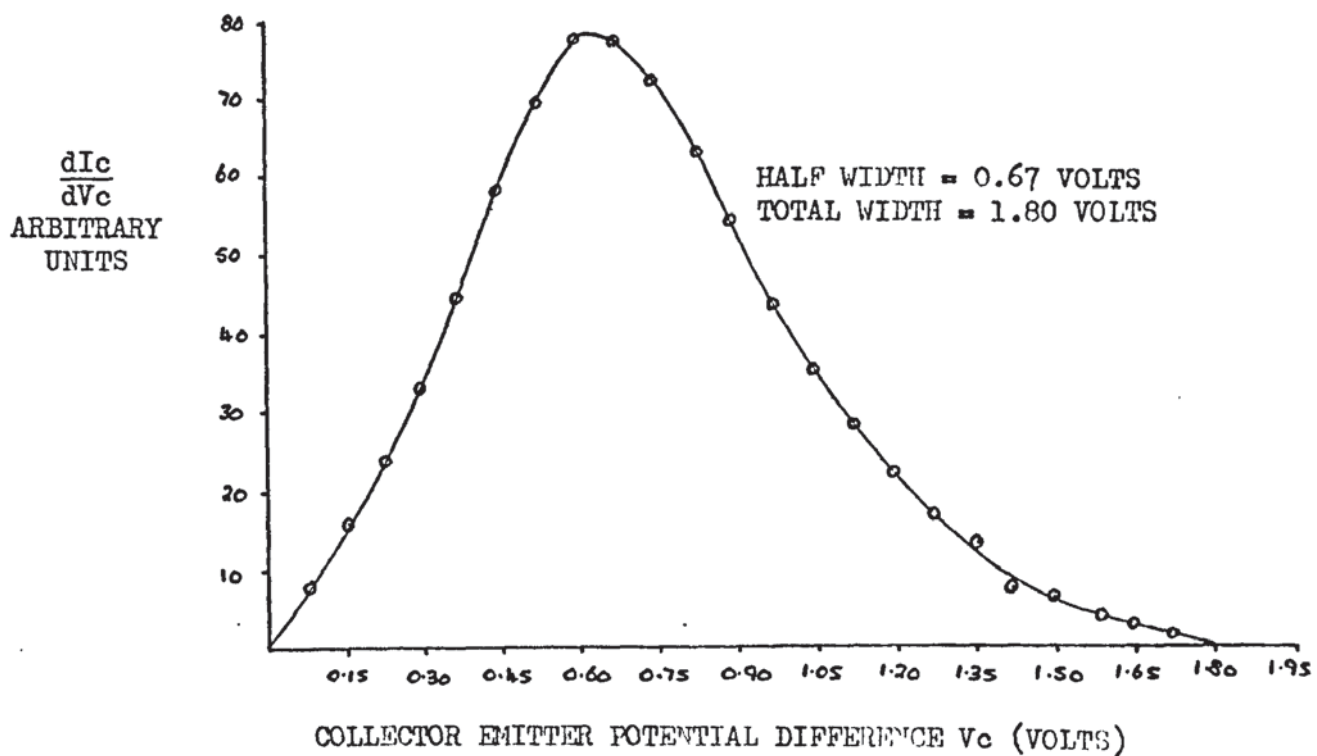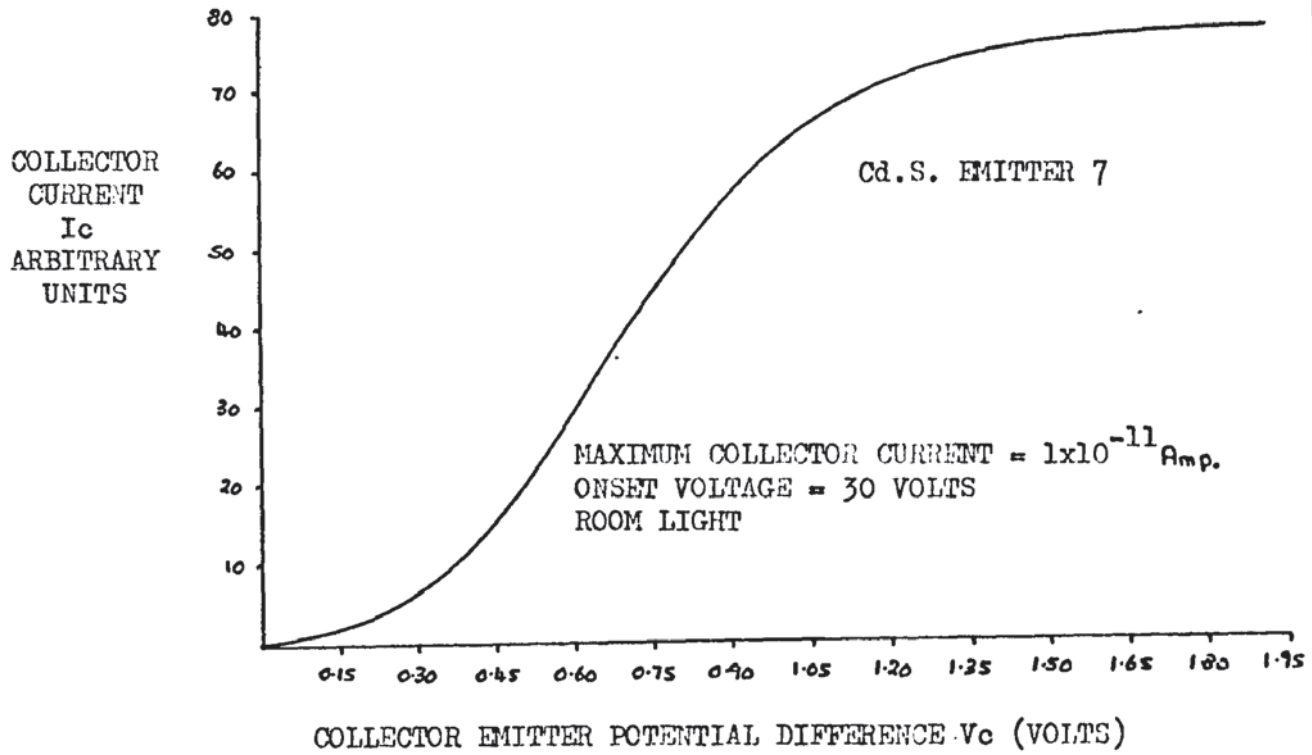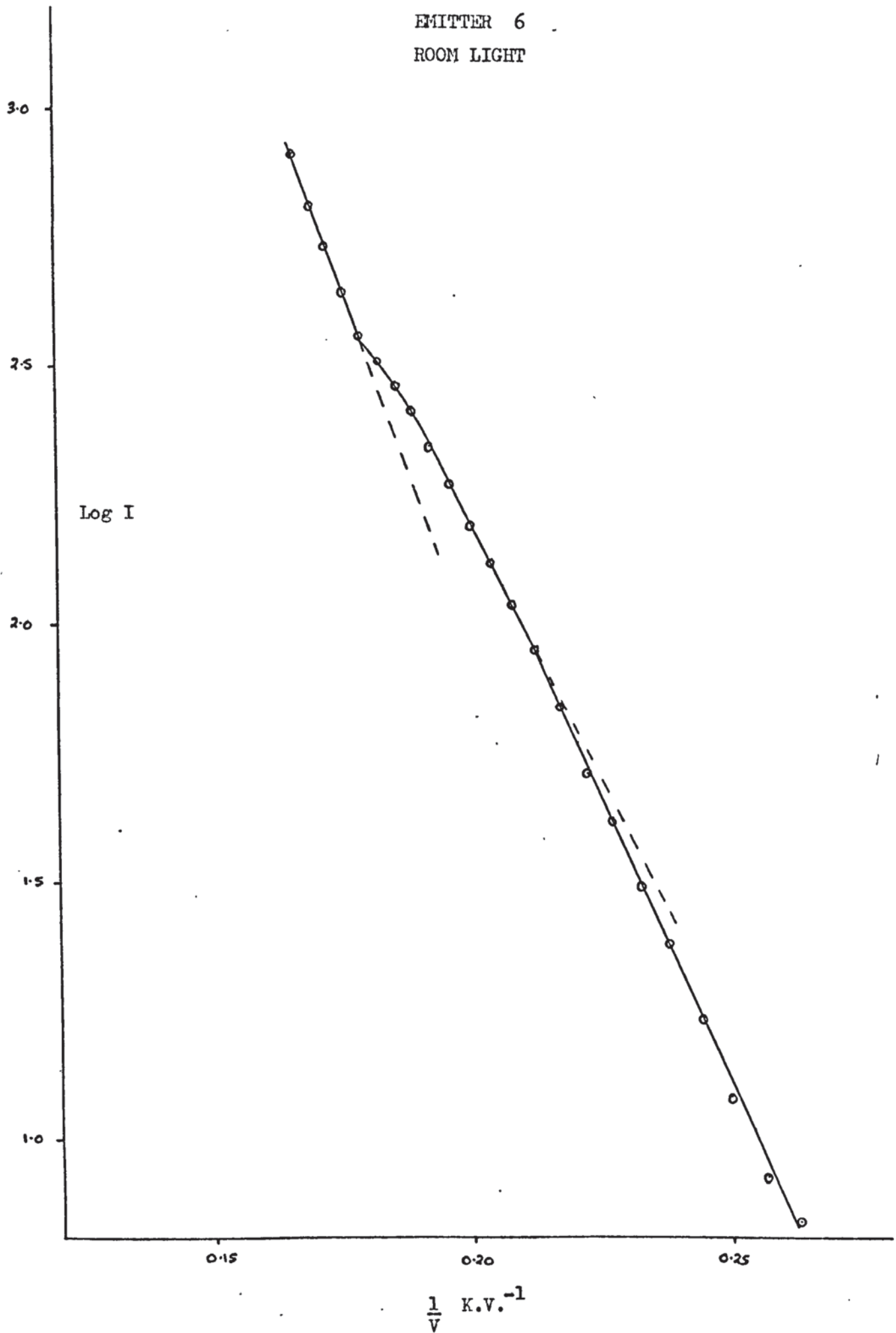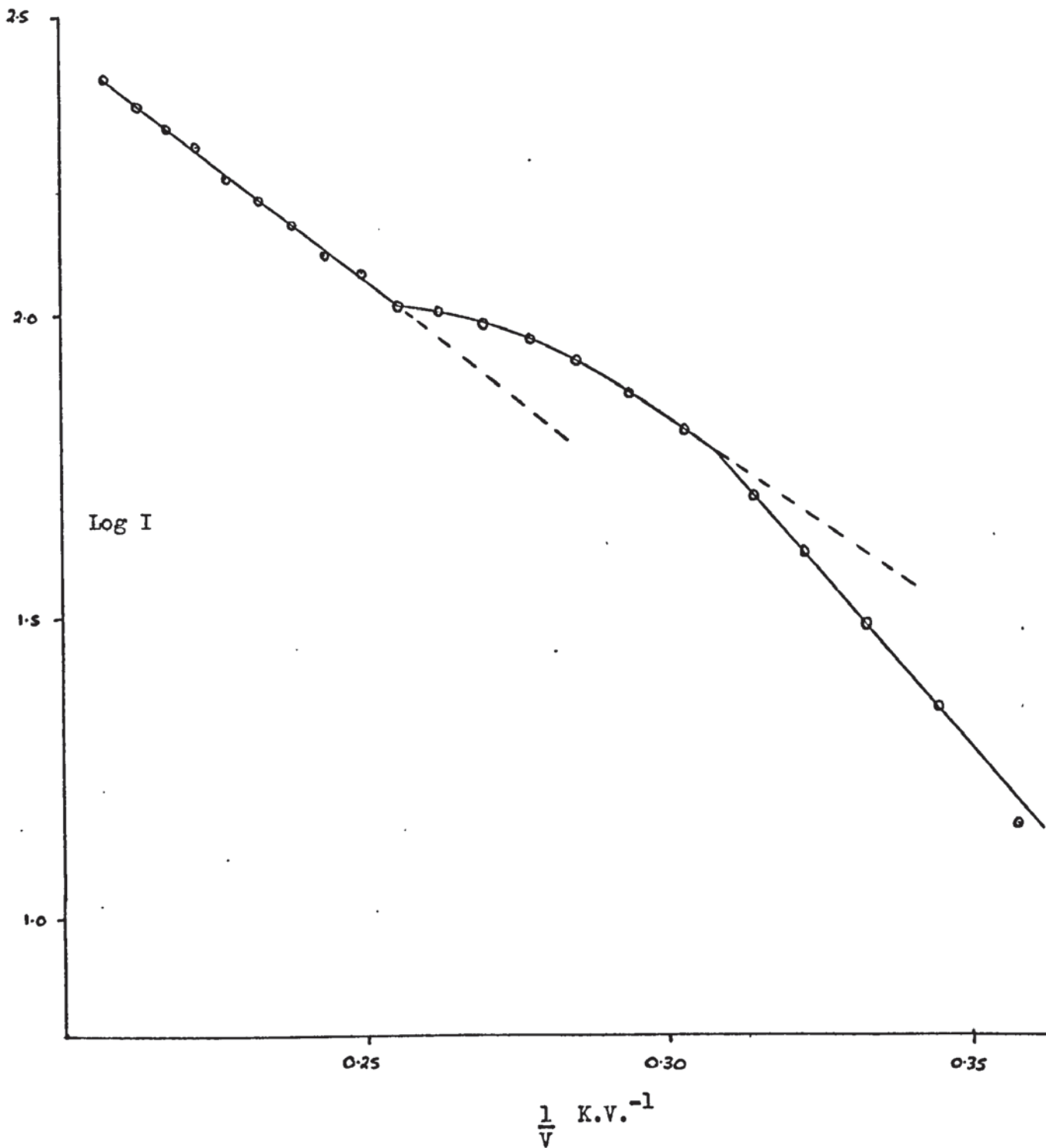
(2) For a given current the half width, total width and potential drop are smaller with maximum light than with minimum light.

(3) For a given potential drop there is a smaller current with minimum light than there is with maximum light and the half width and total width are also smaller.

It should be noted that the onset voltage is not strictly equal to the potential drop because it contains small contributions from the contact potential difference between the emitter and the collector and also from surface effects [44].

These results, in conjunction with those of section 5.2.2 suggest that the half width and total width are independent of the potential drop along the sample and depend only on the magnitude of the emission current. If this is so, it also follows that the half width and total width are smaller with illumination than without.

5.3.3 The Fowler Nordheim Plots

An attempt was made to obtain Fowler Nordheim plots from all the emitters but, because of the unstable nature of the emission current, this proved possible for the lower resistance samples only. The results are shown in Figs. 5.8 and 5.9. An attempt to measure the onset voltage as a function of anode potential was also made and, although results were obtained, the fluctuations in the potential drop along the sample made it impossible to obtain accurate values for the onset voltage. These results will, therefore, not be presented.

o o O o o

CHAPTER 6

Discussion Of Results

6.1 Comparison Between Theoretical And Experimental Results

Fig 6.1 shows the theoretical energy distribution for an applied field
of 5×10^7 volt cm^{-1} compared with the experimental distribution from
one of the well cleaned emitters which had been placed in the optimum
resolution position. It can be seen that the agreement between theory
and experiment is poor. Not only are the half widths and total widths
different but also the sharp onset of current at the Fermi level does
not exist in the experimental result. The only similarity is at the
low energy side of the distribution where $N(E)$ approaches zero.
The poor agreement at the onset of collection cannot have been caused
by lack of resolution because the tungsten results showed that the
analyser is capable of measuring the sharp onset of current at the
Fermi level of a tungsten emitter. Further, the experimental result
in Fig. 6.1 was obtained from an emitter which had been positioned by
maximising the collector current at a value of collector-emitter potential
difference half way between the values corresponding to onset and
saturation. The results of section 5.2.3 show that this position gives
good resolution at the onset of collector current.
It also seems doubtful that the large experimental values of half width
and total width could have been caused by lack of resolution because
the distributions obtained from tungsten and the theoretical work on
the analyser both suggest that the cadmium sulphide emitter was placed
in the optimum resolution position. The theoretical work also shows
that the effects of chromatic and spherical aberration are too small
to have caused such a large discrepancy. It is improbable that the
poor agreement between theory and experiment was caused by energy
broadening with current density because the maximum collected current
used for the result in Fig. 6.1 was of the same order of magnitude as

Fig. 6.1
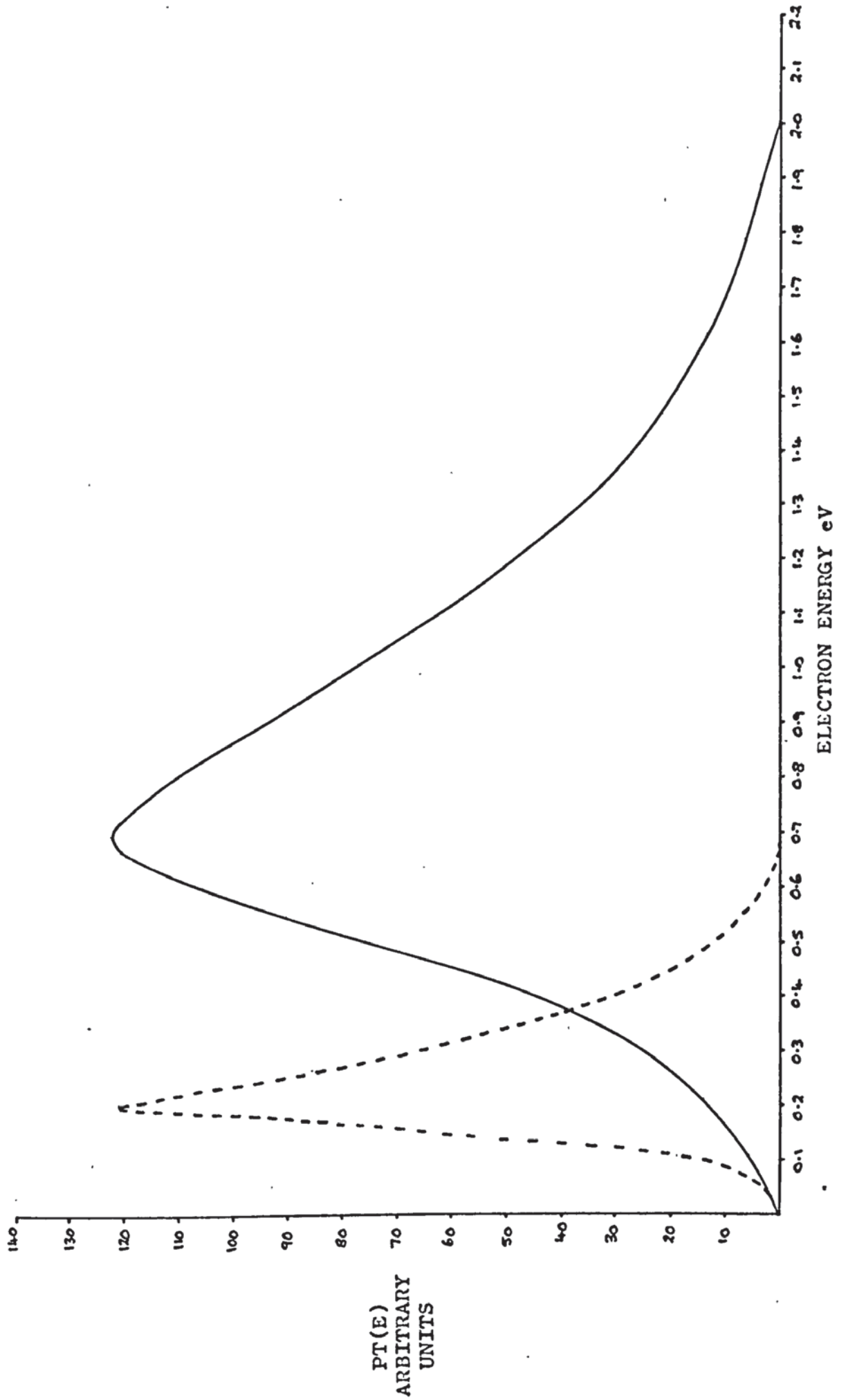
that used in the accurate tungsten results.

Another possible source of experimental error is the potential drop along the sample which fluctuates as the emission current fluctuates and leads to the problems described in section 5.3. This could cause serious errors in the measured energy distribution. The result in Fig. 6.1 was obtained after stable emission had been achieved, therefore, it is thought that the effect was not serious in this emitter. The arguments of the previous paragraphs show that the poor agreement between theory and experiment is not caused by the analyser and is, therefore, a property of the emitter.

6.2 The Hot Electron Effect

The broad energy distributions obtained from cadmium sulphide have been attributed by Shcherbakov and Sokol'skaya [40] to the emission of hot electrons. It will be shown that this explanation is doubtful, especially when it is considered in conjunction with the results of section 5.2.2. The above authors found that for cadmium sulphide, the log I_a vs. $1/V$ characteristics (where I_a is the emission current and V is the voltage applied to the emitter) were non linear and that the form of the non linearity could not be explained in terms of the theory [26] which was in existence at that time. It was also found that a large voltage drop could appear along the emitter, but that this voltage was too small a fraction of V to enable the error in measuring V to be the cause of the non linearity. Elison [100] had already suggested that additional charge carriers could be created by the application of an electric field to a high resistance semiconductor, therefore, Shcherbakov and Sokol'skaya assumed that the field generated by the potential drop along the emitter gave the electrons extra energy which was sufficient to cause ionisation and to create further conduction electrons. The subsequent increase in the electron concentration then enhanced the emission current and caused the bending of the log I_a vs $1/V$ curves.

In a later paper [52] the same authors considered the effect in more

detail. They used a low resolution retarding potential analyser with which they measured the collector-emitter potential difference corresponding to the onset of collector current as a function of the emission current. This potential difference equals $\triangle$vk, the resistive potential drop along the emitter, provided $\triangle$vk is much greater than the contact potential difference between the emitter support and the emitter plus the voltage drop in the space charge region caused by field penetration [44]. From these values they determined R, the mean resistance of the emitter, as a function of $\triangle$vk. It was found that the value of R remained constant until some definite value of $\triangle$vk when it decreased rapidly.

They then assumed that the decrease in R was caused by the production of additional conduction electrons. This effect was measured by defining Ro and No as the resistance and electron concentration in the region of constant resistance. The ratio N/No, where N is the increased electron concentration, was then determined from the ratio R/Ro. Since N/No was known as a function of $\triangle$ vk and Ia was also known as a function of $\triangle$ vk, it then proved possible to obtain the relationship between Ia and N/No. The emission current should depend on the electron concentration and should increase as the latter increases. From the relationship N/No = f(Ia), it was found that this did, in fact, occur. Although the results of the above experiment seem to confirm the hypothesis of carrier generation by hot electrons, there are several factors which make this conclusion doubtful. In the first place, Shcherbakov and Sokol'skaya did not use a well defined emitter but instead used an emitter which consisted of several microtips. The values of $\triangle$ vk, therefore, refer to one microtip whereas the values of Ia are composed of contributions from many microtips. It should also be noted that the value of Ia depends on the electron concentration in the surface region whereas the ratio N/No refers to the bulk concentration because it is determined from the mean resistance of the sample.

Shcherbakov and Sokol'skaya conceded these points but claimed that independent measurements, of which no details were given, showed that the emission current increased with the carrier concentration and that the carrier concentration increased exponentially with the voltage drop along the emitter. They do not mention at which point along the emitter the carrier concentration was measured but, if it was measured in the surface region, where it should have been measured in order to validate their hypothesis, the results can be interpreted in a different manner. The current through the emitter, and hence the voltage drop along it, increase with the applied field. At the same time, the electron concentration in the surface region increases because of field penetration and will increase rapidly since the Fermi function is exponential in form. Therefore, if the concentration was measured in the surface region it would be expected to vary with the potential drop. Although these arguments do not disprove the hypothesis that electron heating was the cause of the effects noted by Shcherbakov and Sokol'skaya, they do cast doubts upon its validity.

Shcherbakov and Sokol'skaya next[51] used an electron energy analyser in order to measure the energy distribution of electrons field emitted from cadmium sulphide. Their design was similar to that of the present analyser and their results from tungsten showed that their equipment possessed fairly good resolution. It was found that the half widths of the measured cadmium sulphide distributions increased with $\triangle vk$ and this was attributed to hot electrons. Again the hot electron hypothesis is attractive because such electrons would possess a high escape probability and would leave the emitter at energy levels well above the Fermi level, thus broadening the distributions. The hypothesis can be criticised on the grounds that the energy distribution would be expected to broaden with applied field, as shown in section 3.10, and at the same time the corresponding increase in current would cause the potential drop to increase. The half width could, therefore, be measured as a function of $\triangle vk$ but this fact is not in itself proof of the existence of hot

- 110 -

electrons.

It was shown in section 5.2.2 that the half width and total width of the energy distribution obtained from a tungsten emitter also increases with increased emission current and the reasons for this were outlined in section 2.8.2. Section 5.3.2 showed that the half widths of the distributions obtained from the cleaned cadmium sulphide emitters was not a function of Δvk but was, instead, a function of the emission current. It follows that the increase in half width with applied field, emission current or potential drop cannot be determined uniquely and that, once again, the existence of hot electron emission cannot be proved by the results of Shcherbakov and Sokol'skaya.

6.3 Possible Causes Of The Broadened Energy Distributions

Fig. 6.1 suggests that the lack of agreement at the low energy end of the distribution may be caused by electrons which originate from energy levels below the bottom of the conduction band while at the high energy end the discrepancy may be the result of electron emission from energy levels far above the Fermi level. Possible reasons for the existence of these extra low and high energy electrons are discussed below.

6.3.1 The Extra Low Energy Electrons

Many investigators have reported the existence of double peaks in the energy distributions obtained from various semiconductors. Authur [53], for example, noted the effect in germanium field emitters. He found that the existence of the second peak depended on the position of the emission pattern relative to the anode aperture of his Young and Müller type analyser and that the second peak dissapeared when the emitter was subjected to field evaporation. Previous work [38] had shown this to be an efficient method of cleaning germanium field emitters. After evaporation the half width of the measured distribution was still approximately twice as large as the theoretical value but it became smaller when the tip was partially annealed. Authur interpreted his results in terms of emission from a degenerate conduction band and

supposed that the second peak was caused by electrons from energy levels below the bottom of this band. He could not, however, attribute the effect to emission from surface states because the surface state distribution was too small and the adsorption of approximately one monolayer of oxygen had little effect on the measured distribution. Hughes and White [55] found similar effects in their work on field emission from gallium arsenide. The emitters, which were cleaned by d.c. field desorption in the presence of hydrogen [56], gave energy distributions which consisted of a stable peak and an unstable peak. They were able to show that the stable peak was consistent with valence band emission and that the unstable peak was caused by electrons which came from energy levels above the top of this band. The existence of the second peak was again found to depend on the emitter position and they suggested that it was caused by emission from surface states because it was known that the density of these states can be very high on gallium arsenide. This argument is also supported by the surface disorder which can arise when a semiconductor is subjected to d.c. field desorption [38]. The measurements in both of the above experiments were performed on low resistance samples so that there were no large potential drops along the emitters. Hot electron emission could not, therefore, be the cause of the poor agreement that was found to exist between theory and experiment.

The present results are similar in many ways to the examples given above. A second peak was found which could be removed by d.c. field desorption and in some of the early results the existence of this second peak was found to depend on tip positioning. It should be noted that the tip positioning referred to in this section is accomplished by lateral rather than axial movement, therefore, it seems possible that the second peak is caused by emission from surface states associated with contamination because the second peak would be present in results obtained from a contaminated part of the emitter whereas it would

disappear if the results were taken from a clean part of the emitter surface. When the emitter is completely clean, lateral movement should not produce a second peak and this was found to be true in the case of the clean cadmium sulphide samples. In addition to the above results it was found that the relative heights of the peaks from one of the cadmium sulphide emitters could be changed by illumination and that in the case of the cleaned emitters the half widths were smaller with illumination than without for similar values of emission current. Haas and his co-workers [101] have investigated the interaction of oxygen with the surface of cadmium sulphide crystals. They explained their results in terms of the potential difference which exists across the double layer at the surface. This double layer consists of a negatively charged outer layer, which is caused by chemisorbed oxygen, and a positively charged inner layer caused by ionised donors. It was supposed that incident light creates an electron-hole pair and that the hole then migrates towards the surface under the action of the field generated by the double layer. The hole then combines with a chemisorbed oxygen atom and liberates it, thus reducing the potential difference across the surface region. They found; by measuring the potential drop, that an oxygen free surface could not be obtained until a bake out temperature of at least $400^{\circ}C$ had been used. In explaining their results they also used the fact that ultra high vacuum bake out produces bulk and surface defects in cadmium sulphide which can trap holes [102].

The investigations of Haas and his co-workers suggest that the appearance of excess low energy electrons in the cadmium sulphide results can be explained in the following way:

When a cadmium sulphide emitter has been made and placed in the vacuum system it is probably covered with a layer of contamination which is left behind after the etching process. It is possible that oxygen is also present because the work of Haas et. al. showed that when chemically cleaned cadmium sulphide was outgassed oxygen, and compounds of oxygen,

were among the evolved products. The vacuum system, including the emitter, is then baked and the obvious contamination, such as water vapour, is removed. Chemisorbed oxygen will not be removed because the system was usually baked to $200^{o}C$ only, for reasons given in section 4.2.2. The method of d.c. field desorption is then used to clean the emitter. When the desorbing field is applied the energy levels in the surface of the emitter are as shown in Fig. 6.2.
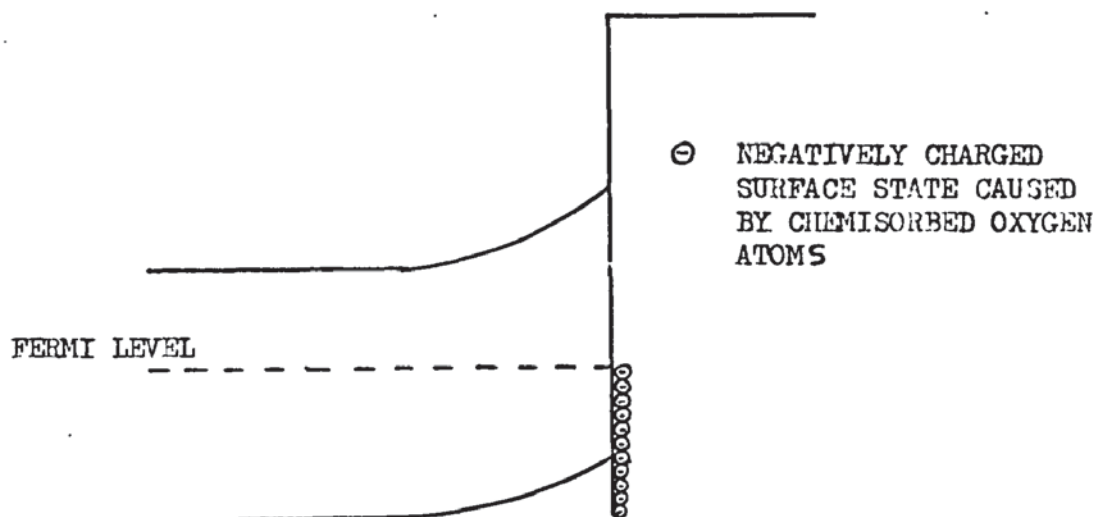
Before the field is applied, surface states capture electrons from the conduction band and become filled up to the Fermi level. This causes a lack of electrons in the surface region and the conduction band bends upwards. When the field is applied, more positive charges are needed to sustain it and these are supplied initially by ionisation of the neutral, gross surface contamination. As the applied field becomes higher, however, the particles of ionised surface contamination are attracted by the anode and stripped from the emitter. Finally, the field penetrates the surface region which then becomes positively charged. The conduction band bends upwards even further when this condition is reached.

When high enough fields have been applied to the sample all the gross contamination will have become ionised and will, therefore, have been removed. The oxygen atoms, which cause the surface states, will, however, still be present because they must lose two electrons before they can be stripped from the sample. The first electron can be removed by shining light on the sample when an electron-hole pair is created. The hole will move towards the surface under the action of the field in the surface region and will then neutralise a chemisorbed oxygen atom. This atom can then be removed by the same process that removed the gross contamination.
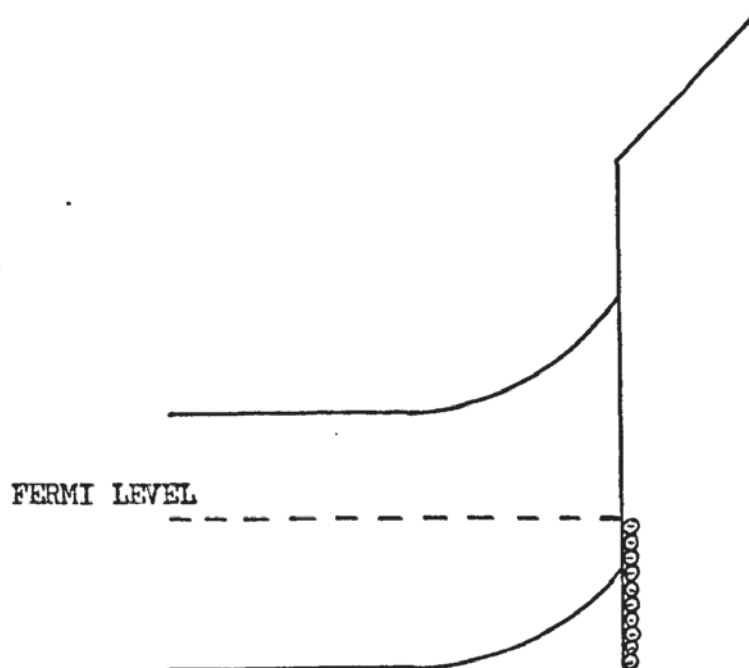
The results with double peaks were obtained before the cleaning procedure had been perfected, therefore, it seems possible that they were characteristic of a grossly contaminated surface. This hypothesis is

Fig. 6.2

(a) NO APPLIED FIELD

⊖   NEGATIVELY CHARGED
SURFACE STATE CAUSED
BY CHEMISORBED OXYGEN
ATOMS

FERMI LEVEL

(b) WITH APPLIED FIELD

FERMI LEVEL

supported by the work of Nicolaou and Modinos [103] who have shown that a second peak can be expected under conditions where a surface layer of foreign atoms gives rise to a double potential barrier. They have also shown that the second peak diminishes as the thickness of the surface layer is decreased which suggests that the results from the supposedly clean cadmium sulphide results may still contain electrons from the second peak because, although this peak can no longer be resolved, it can still cause the low energy side of the measured energy distribution to be broadened. This mechanism would also explain the disappearance of the second peak in the results from emitters 2 and 3 and the partial disappearance of the peak in the result from emitter 4 after additional field desorption.
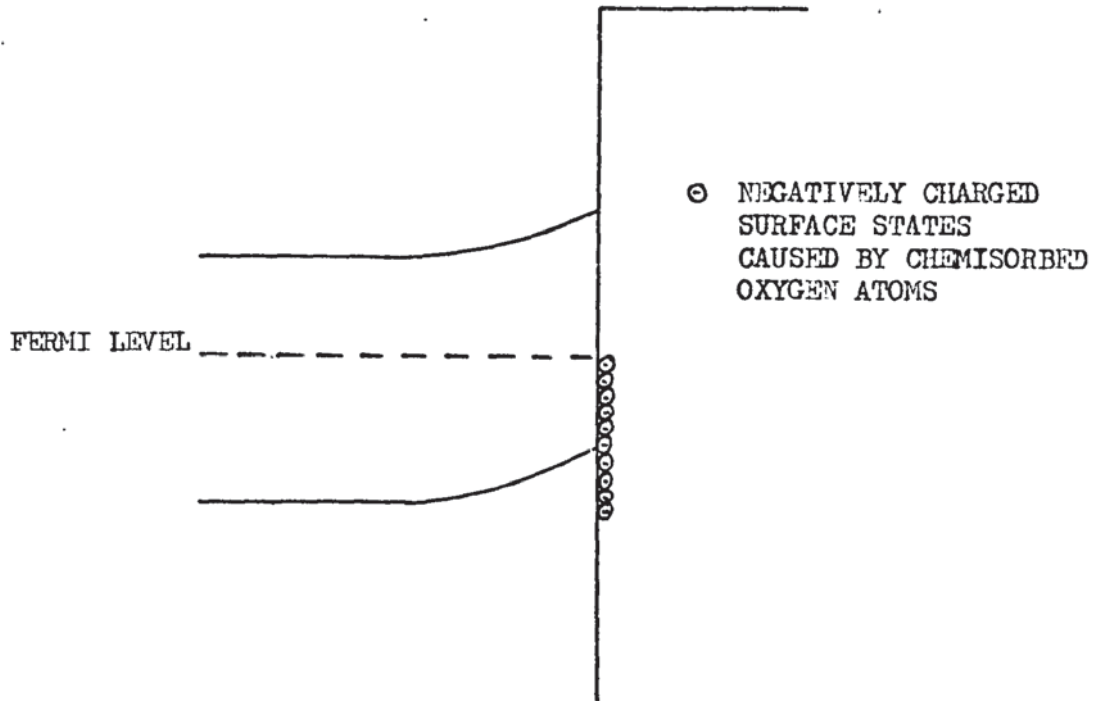
When the emission field is applied the energy levels in the emitter are as shown in Fig. 6.3. Before the field is applied there is a lack of electrons behind the surface for reasons already explained and when the field is applied negative charges are needed to sustain it. These are initially supplied by the surface states but, as the field is increased, field penetration occurs and the induced negative charges cause the conduction band to bend downwards. Electron-hole pairs can still be created by illumination but the holes will only migrate towards the surface provided the conduction band is not depressed below the flat band condition.

If the applied field is small enough and if the illumination is high enough, chemisorbed oxygen atoms can still be neutralised by holes. Although the neutralised atoms are then not bound to the surface they may not escape because some of them may be polarised in the high emission field and thus remain in the vicinity of the emitter. This may explain why the emission pattern sometimes became very unstable when the emitter was illuminated and why the pattern then consisted of small dots of light which appeared to be in random motion.
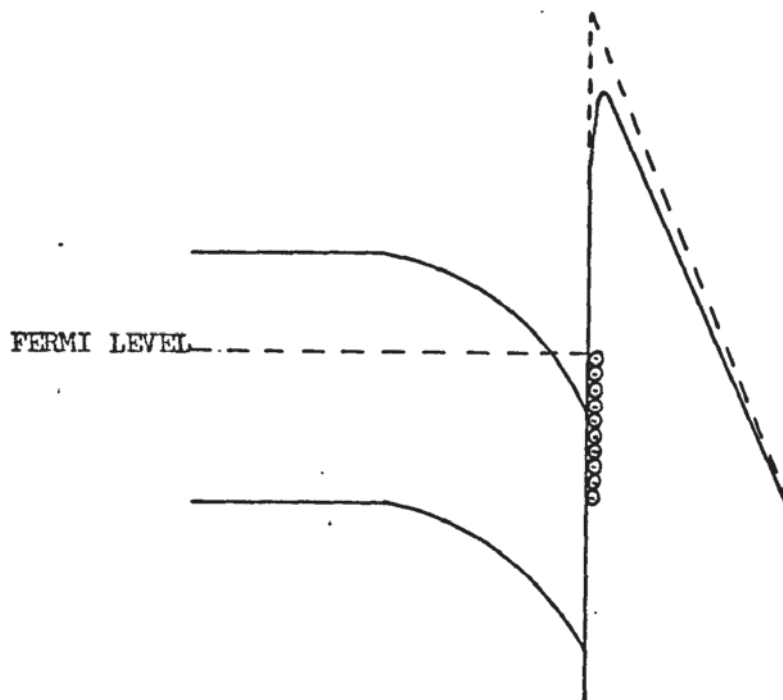
The presence of chemisorbed oxygen on the emitter surface may also

Fig. 6.3

(a) NO APPLIED FIELD

⊖ NEGATIVELY CHARGED
    SURFACE STATES
    CAUSED BY CHEMISORBED
    OXYGEN ATOMS

FERMI LEVEL

(b) WITH APPLIED FIELD

FERMI LEVEL

explain the effect of illumination on the energy distributions. Thus, the release of oxygen by illumination in the case of emitter 1 would cause the second peak to be depressed because the emitter surface would be less contaminated and the same process in the nearly clean samples would simply cause the distribution to become narrower which, in fact, occurs. This explanation must be treated with caution, however, since it is almost certain that the results from emitters 1, 6 and 7 were taken when the surface was degenerate, therefore, it is difficult to understand the mechanism whereby the holes still migrate towards the surface.

6.3.2 The Extra High Energy Electrons

The non linearity of the Fowler Nordheim plots obtained from high resistance n type semiconductors has recently been explained by Baskin et. al. [43] without recourse to the mechanism of hot electron emission. They showed that the effect is caused by the inability of these materials to generate sufficient mobile carriers during field emission which leads to a saturation of the emission current. The results of section 5.3.3 suggest that such an effect occurred in samples 6 and 7. It is interesting to note that there are more extra high energy electrons present in the distribution obtained from emitter 5 than there are in the results obtained from the lower resistance samples.

The theoretical energy distribution in Fig. 6.1 was calculated from the work of Stratton which ignores the effect of the current flowing through the emitter. The main difference between Stratton's work and the work of Baskin et. al. is that Stratton calculates the electron concentration in the surface region from the field penetration alone whereas Baskin et. al. consider the surface electron concentration to be the result of two mechanisms, the field penetration and the ability of the emitter material to generate mobile carriers.

When an electric field is applied to the surface of a semiconductor the surface charge density needed to sustain the field is supplied by

ionisation in the surface region. As the field is increased, electrons in the surface region leave the emitter by tunnelling through the potential barrier and are replaced by electrons from the interior. In Stratton's theory, the electron concentration available for emission is assumed to be the concentration which has been induced by field penetration, replenishment of the surface region is not considered. Fig. 6.4 shows that this leads to a set of electrons with energies which are distributed about the Fermi level in the surface region in a manner similar to that of a metal.

In a real semiconductor, however, electrons which leave the surface region must be replaced by electrons from the bulk. It is obvious from Fig. 6.5 that electrons from the interior of the emitter will have higher energies than electrons at the Fermi level in the surface region. Although there must exist some mechanism whereby a number of these electrons can lose energy and occupy the vacant lower energy levels, it seems possible that there could be an increase in the number of electrons with energies above the surface Fermi level over the number predicted by Stratton's theory. This mechanism could, therefore, explain the extra high energy electrons which are observed in experimental results but are not predicted by theory. A detailed explanation would require a complete reformulation of Stratton's work.

6.4. Suggestions For Future Work

The tentative theories given in the preceeding section cannot strictly be treated separately because in a real semiconductor there is probably some interaction between the effects they attempt to describe. One of the effects must, therefore, be removed so that the other one can be investigated alone. It is obvious that the phenomena attributed to contamination should be removed first, by obtaining a truly clean emitter. If, in addition, the ideas used in the double peak theory are employed as a guide to obtaining a clean emitter, and such an emitter results, evidence will have been obtained in support of this

Fig. 6.4
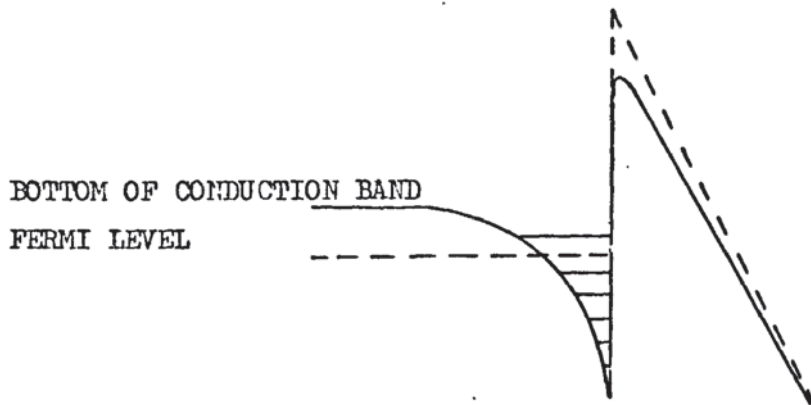
BOTTOM OF CONDUCTION BAND

FERMI LEVEL

**(a) STRATTON MODEL OF SEMICONDUCTOR
WITH DEGENERATE SURFACE**

(b) FOWLER NORDHEIM MODEL OF A METAL

FILLED
ENERGY
LEVELS
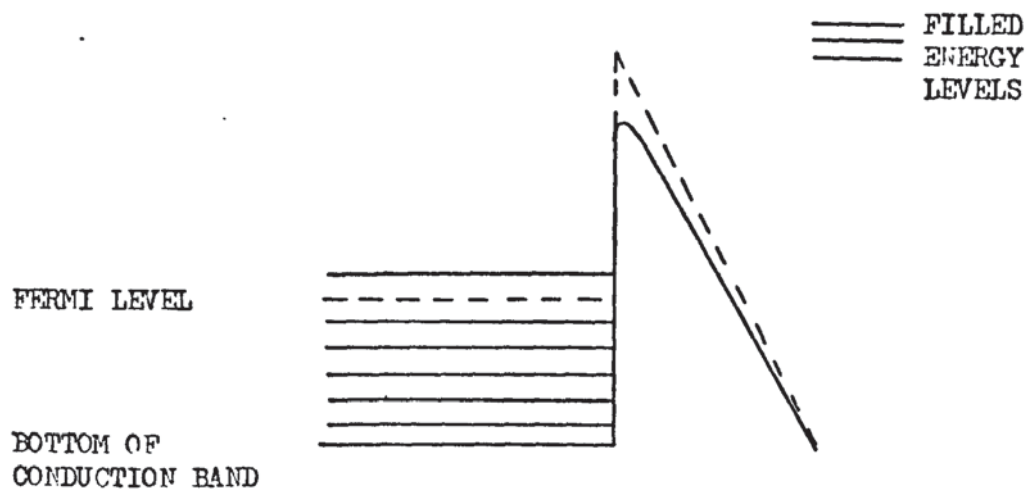
FERMI LEVEL

BOTTOM OF
CONDUCTION BAND

Fig. 6.5

BASKIN et.al. MODEL OF
DEGENERATE SEMICONDUCTOR

A   CONDUCTION BAND

          FERMI LEVEL

                                              B
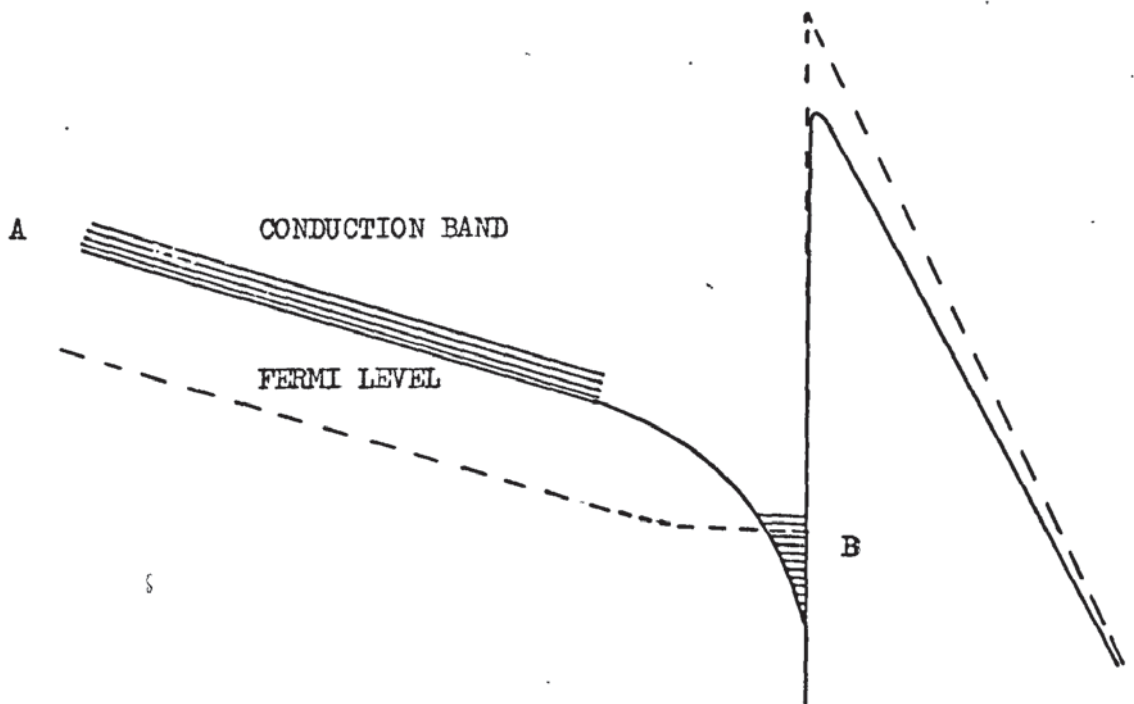
A. ELECTRONS GENERATED IN THE BULK
B. ELECTRONS GENERATED BY FIELD PENETRATION
≡ FILLED ENERGY LEVELS

theory. When emission from a clean emitter has been achieved the second theory can be investigated.

6.4.1 The Production Of A Clean Emitter

The cleaning of field emitters is usually accomplished by heating or by field desorption, sometimes both processes are used together. In the case of a semiconductor heating sometimes leads to a change in the doping, therefore, it is not an adviseable method to use. Recently Marien and Loosveldt [104] have suggested that cadmium sulphide can be cleaned by hydrogen assisted field evaporation but it has not been possible to attempt this method in the present work because of the extra experimental equipment needed.

The theory of section 6.3.1 suggests that the removal of the most tenacious contaminant, chemisorbed oxygen, can be achieved by field desorption if a large number of holes is present in the surface region. Although electron-hole pairs can be created in the emitter, the holes may become trapped in defects and, therefore, may not move towards the surface. In order to obtain a clean emitter by this method, therefore, an experimental method must be employed which inhibits the production of defects.

It is well known that rough mechanical handling can create defects in cadmium sulphide, therefore, no mechanical shaping should be used during the fabrication of the emitters. Heat also leads to the formation of defects. Unless another etchant is found, heating during the preparation of the emitters is unavoidable and heating during the vacuum system bake-out is also, unfortunately, a problem which cannot be removed. Good vacua can, however, be obtained with modern ultra high vacuum systems with moderate baking, in some cases temperates as low as 150°C can be used.

It was suggested in section 6.3.1 that light can assist in the removal of oxygen from cadmium sulphide. It is, therefore, obvious that the emitter should be subjected to intense light during field desorption

and that the frequency of the light should be variable so that the maximum rate of electron-hole pair creation can be obtained. Once the surface contamination has been removed in this way it should not be allowed to return. This could be partially achieved by the use of a modern ultra high vacuum system employing electrostatic ion pumps which would remove all the sources of contamination inherrent in oil or mercury diffusion pump systems.

6.4.2 Experiments With A Clean Emitter

If a clean emitter can be obtained by the methods mentioned above it will be possible to investigate the origin of the extra high energy electrons by means of the following experiments.

An energy distribution must be obtained from each of a set of emitters with different resistivities and in order to make the results comparable the same value of total emitted current should be used every time. In addition, the value of collector-emitter potential difference corresponding to the zero in the low energy tail of the energy distribution and the potential drop at various points along the emitter must be noted. These potential drops can be measured by means of ohmic gallium contact [105] and a digital voltmeter. The results of the above experiments will give the slope at onset, which can be characterised by the difference in the value of collector-emitter potential difference corresponding to the onset and the maximum of the energy distribution, as a function of the emitter resistivity and the potential drop in the surface region caused by field penetration. It should also prove possible from these results to obtain a measure of the excess energy of electrons from the bulk over the energy of electrons in the surface region. If, in addition, Fowler Nordheim plots are taken for each sample, evidence for or against the ideas presented in section 6.3.2 can be obtained.

It may be difficult to perform the above experiments with the present analyser because good stability is needed in order to measure the values of collector-emitter potential difference which correspond to the onset

of collector current, and hence to find the onset of the energy

distribution, and also because high resolution, which is necessary

to measure any change in the slope at onset, is available only at small

values of total emission current. Unfortunately, at these values of

current the signal to noise ratio of the analyser starts to become

a problem. If the instability is caused by contamination then a good

vacuum system and a clean emitter should remove it but the signal to

noise ratio will not be improved because its value is determined by

the measuring circuits used. The problem of resolution could probably

be removed by using a Plummer [106] type analyser. Not only does this

analyser possess extremely good resolution and a high value of signal

to noise ratio but also energy distributions from virtually any part

of the emitter surface can be measured with the same high precision

whereas in the present analyser the maximum resolution is available

along the axis of symmetry only. The Plummer analyser also yields the

energy distribution directly so that graphic differentiation is not

necessary and it does not suffer from a loss of collector current at

saturation. The problem of fine positioning mentioned in section 5.2.3,

therefore, no longer exists.

A plummer type analyser and a clean emitter would also make a detailed

investigation of the contamination problem possible. Even a partially

cleaned emitter could be used because the resolution does not vary with

lateral movement of the tip, therefore, a realistic comparison between

clean and contaminated areas is possible. Although with the present

analyser the second peak was found to appear and disappear with lateral

movement, this is a coarse effect caused by relatively clean and heavily

contaminated areas. A more refined experiment to compare the energy

distributions from clean and slightly contaminated areas is not possible

because of the variable resolution of the analyser.

If, in addition to the Plummer analyser, a mass spectrometer head is

used it should be possible to investigate the relationship between the

low energy side of the energy distribution and the presence of oxygen and other gases in the vacuum system. It may even be possible to determine the energy levels at which the contamination resides and, therefore, to identify it.

o o 0 o o

```
'BEGIN'
'REAL' R1,H,K,A1,A2,A3,A4,A0;
'REAL' 'ARRAY' R[0:34],S[0:51],V[1:868,1:5];
'INTEGER' M,N,X,Y,P,Q,T,TT,U,ZK,ZZ,TO;
'BEGIN'
SELECT INPUT(3);
R1:=READ;
'FOR' N:=0 'STEP' 1 'UNTIL' 34 'DO' R[N]:=READ;
'FOR' M:=0 'STEP' 1 'UNTIL' 51 'DO' S[M]:=READ;
'FOR' M:=1 'STEP' 1 'UNTIL' 31 'DO'
'BEGIN'
T:=M;
H:=S[M+1]-S[M];
K:=S[M]-S[M-1];
V[T,1]:=(H*(R1*R1))/((H+K)*(R1*R1+(2*H*K)));
V[T,2]:=-1;
V[T,3]:=(K*(R1*R1))/((H+K)*(R1*R1+(2*H*K)));
V[T,4]:=(2*H*K)/((2*H*K)+R1*R1);
V[T,5]:=0;
NEWLINE(1);
PRINT(T,3,0);SPACE(4);
PRINT(M,2,0);SPACE(2);
PRINT(N,2,0);
'END';
N:=1;
Y:=1;
ZZ:=0;
ZK:=1;
X:=31;
Q:=1;
TT:=9;
T:=31;
P:=20;
U:=51;
HERE:'FOR' M:=Y 'STEP' 1 'UNTIL' X 'DO'
'BEGIN'
A1:=(1/(S[M]-S[M-1]))*((R[N]-R[N-1])*(1-((R[N]-R[N-1])/(2*R[N])))+
(R[N+1]-R[N])*(1+((R[N+1]-R[N])/(2*R[N]))));
A2:=(1/(R[N+1]-R[N]))*((S[M+1]-S[M-1])*(1+((R[N+1]-R[N])/(2*R[N]))));
A3:=(1/(S[M+1]-S[M]))*((R[N]-R[N-1])*(1-((R[N]-R[N-1])/(2*R[N])))+
(R[N+1]-R[N])*(1+((R[N+1]-R[N])/(2*R[N]))));
A4:=(1/(R[N]-R[N-1]))*((S[M+1]-[M-1])*(1-((R[N]-R[N-1])/(2*R[N]))));
A0:=((R[N]-R[N-1])*((1/(S[M]-S[M-1]))+(1/(S[M+1]-S[M])))+
(1/(R[N]-R[N-1]))*(S[M+1]-S[M-1]))*(1-((R[N]-R[N-1])/(2*R[N])))+
((R[N+1]-R[N])*((1/(S[M]-S[M-1]))+(1/(S[M+1]-S[M])))+(1/(R[N+1]-R[N])
))*(S[M+1]-S[M-1]))*(1+((R[N+1]-R[N])/(2*R[N])));
V[T+M,1]:=A4;
V[T+M,2]:=A1;
V[T+M,3]:=-A0;
V[T+M,4]:=A3;
V[T+M,5]:=A2;
NEWLINE(1);
PRINT(T+M,3,0);SPACE(4);
PRINT(M,2,0);SPACE(2);
PRINT(N,2,0);
```

```
'END';
M:=X;
T:=T+M;
'IF' T=448 'THEN' N:=N-1;
N:=N+1;
'IF' N 'LE' 5 'THEN' 'GOTO' HERE;
'IF' N=16 'THEN' 'GOTO' HOME;
'IF' N 'GE' 28 'THEN' 'GOTO' PLACE;
'LF' N>16 'THEN' 'GOTO' ANYWHERE;
'IF' T=186 'THEN'
'BEGIN'
X:=20;
'GOTO' HERE;
'END';
'IF' T=206 'THEN'
'BEGIN'
Y:=22;
X:=31;
N:=N-1;
T:=T-21;
'GOTO' HERE;
'END';
'IF' T=216 'THEN'
'BEGIN'
Y:=1;
N:=7;
X:=20;
'GOTO' HERE;
'END';
'IF' T=236 'THEN'
'BEGIN'
Y:=22;
X:=31;
T:=T-21;
N:=N-1;
'GOTO' HERE;
'END';
'IF' ZZ=1 'THEN' 'GOTO' THERE;
Q:=Q+1;
P:=P+1;
Y:=Q;
X:=P;
T:=T-(Q-1);
ZZ:=1;
'GOTO' HERE;
THERE:Y:=Q+21;
X:=31;
N:=N-1;
'IF' T=448 'THEN' N:=N+1;
T:=T-(Y-1);
ZZ:=0;
'GOTO' HERE;
HOME:Y:=9;
T:=T-8;
X:=29;
'GOTO' HERE;
ANYWHERE:'IF' N 'LE' 25 'THEN'
'BEGIN'
```

```
TT:=TT+1;
U:=U-1;
T:=T-(TT-1);
Y:=TT;
X:=U;
'GOTO' HERE;
'END';
'IF' N=26 'THEN'
'BEGIN'
Y:=20;
T:=T-(Y-1);
X:=41;
'GOTO' HERE;
'END';
'IF' N=27 'THEN'
'BEGIN'
Y:=23;
T:=T-(Y-1);
X:=40;
TT:=24;
U:=40;
'GOTO' HERE;
'END';
PLACE:'IF' N 'LE' 33 'THEN'
'BEGIN'
TT:=TT+1;
U:=U-1;
Y:=TT;
Y:=T-(Y-1);
Y:=TT;
T:=T-(Y-1);
X:=U;
'GOTO' HERE;
'END';
AGAIN:TO:=READ;
NEWLINE(1);
PRINT(V[TO,1],5,5);SPACE(3);NEWLINE(1);
PRINT(V[TO,2],5,5);SPACE(3);NEWLINE(1);
PRINT(V[TO,3],5,5);SPACE(3);NEWLINE(1);
PRINT(V[TO,4],5,5);SPACE(3);NEWLINE(1);
PRINT(V[TO,5],5,5);SPACE(3);NEWLINE(1);
ZK:=ZK+1;
'IF' ZK 'LE' 868 'THEN' 'GOTO' AGAIN;
'END';
'END';
****
```

```
'BEGIN'
'INTEGER' X,Y,K,Q,JM,LR,I,PIV,R,J,M,N;
'REAL' T;
'REAL' 'ARRAY' V[1:307],C[1:307,1:29];
'BEGIN'
M:=READ;
N:=READ;
Y:=1;
HERE:'FOR' X:=1 'STEP' 1 'UNTIL' Y+14 'DO'
C[Y,X-Y+(M+1)'/'2]:=READ;
'IF' Y 'LE' 14 'THEN'
'BEGIN'
Y:=Y+1;
'GOTO' HERE;
'END';
Y:=16;
Q:=2;
THERE:'FOR' X:=Q 'STEP' 1 'UNTIL' 2*Y-Q 'DO'
C[Y,X-Y+(M+1)'/'2]:=READ;
'IF' Y 'LE' 292 'THEN'
'BEGIN'
Y:=Y+1;
Q:=Q+1;
'GOTO' HERE;
'END';
Y:=294;
Q:=280;
ANYWHERE:'FOR' X:=Q 'STEP' 1 'UNTIL' 307 'DO'
C[Y,X-Y+(M+1)'/'2]:=READ;
'IF' Y 'LE' 306 'THEN'
'BEGIN'
Y:=Y+1;
Q:=Q+1;
'GOTO' ANYWHERE;
'END';
'FOR' K:=1 'STEP' 1 'UNTIL' 307 'DO'
'BEGIN'
V[K]:=READ;
'END';
LR:=(M+1)'/'2;
'FOR'R:=1 'STEP' 1 'UNTIL' LR-1 'DO'
'FOR'I:=1 'STEP' 1 'UNTIL' LR-R 'DO'
'BEGIN'
'FOR' J:=2 'STEP' 1 'UNTIL' M 'DO'
C[R,J-1]:=C[R,J];
C[R,M]:=C[N+1-R,M+1-I]:=0;
'END';
'FOR' I:=1 'STEP' 1 'UNTIL' N-1 'DO'
'BEGIN'
PIV:=I;
'FOR' R:=I+1 'STEP' 1 'UNTIL' LR 'DO'
'IF' ABS(C[R,1])>ABS(C[PIV,1]) 'THEN' PIV:=R;
'IF' PIV 'NE' 1 'THEN'
'BEGIN'
T:=V[I];
```

```
V[I]:=V[PIV];
V[PIV]:=T;
'FOR' J:=L 'STEP' 1 'UNTIL' M 'DO'
'BEGIN'
T:=C[I,J];
C[I,J]:=C[PIV,J];
C[PIV,J]:=T;
'END';
'END';
V[I]:=V[I]/C[I,1];
'FOR' J:=2 'STEP' 1 'UNTIL' M 'DO'
C[I,J]:=C[I,J]/C[I,1];
'FOR' R:=I+1 'STEP' 1 'UNTIL' LR 'DO'
'BEGIN'
T:=C[R,1];
V[R]:=V[R]-T*V[I];
'FOR' J:=2 'STEP' 1 'UNTIL' M 'DO'
C[R,J-1]:=C[R,J]-T*C[I,J];
C[R,M]:=0;
'END';
'IF' LR 'NE' N 'THEN' LR:=LR+1;
'END';
V[N]:=V[N]/C[N,1];
JM:=2;
'FOR' R:=N-1 'STEP' -1 'UNTIL' 1 'DO'
'BEGIN'
'FOR' J:=2 'STEP' 1 'UNTIL' JM 'DO'
V[R]:=V[R]-C[R,J]*V[R-1+J];
'IF' JM 'NE' M 'THEN' JM:=JM+1;
'END';
'FOR' K:=1 'STEP' 1 'UNTIL' 307 'DO'
'BEGIN'
NEWLINE(1);
PRINT(V[K],4,2);
'END';
'END';
'END';
****
```

```
'BEGIN'
'INTEGER' X,Y,K,Q,JM,LR,I,PIV,R,J,M,N;
'REAL' T;
'REAL' 'ARRAY' V[1:868],C[1:868,1:41];
'BEGIN'
M:=READ;
N:=READ;
Y:=1;
HERE:'FOR' X:=1 'STEP' 1'UNTIL' Y+20 'DO'
C[Y,X-Y+(M+1)'/'2]:=READ;
'IF' Y 'LE' 20 'THEN'
'BEGIN'
Y:=Y+1;
'GOTO' HERE;
'END';
Y:=22;
Q:=2;
THERE:'FOR' X:=Q 'STEP' 1 UNTIL' 2*Y-Q 'DO'
C[Y,X-Y+(M+1)'/'2]:=READ;
'IF' Y 'LE' 847 'THEN'
'BEGIN'
Y:=Y+1;
Q:=Q+1;
'GOTO' HERE;
'END';
Y:=849;
Q:=829:
ANYWHERE:'FOR' X:=Q 'STEP' 1 'UNTIL' 868 'DO'
C[Y,X-Y+(M+1)'/'2]:=READ;
'IF' Y 'LE' 867 'THEN'
'BEGIN'
Y:=Y+1;
Q:=Q+1;
'GOTO' ANYWHERE;
'END';
'FOR' K:=1 'STEP' 1 'UNTIL' 868 'DO'
'BEGIN'
V[K]:=READ;
'END';
LR:=(M+1)'/'2;
'FOR' R:=1 'STEP' 1 'UNTIL' LR-1 'DO'
'FOR' I:=1 'STEP' 1 'UNTIL' LR-R 'DO'
'BEGIN'
'FOR' J:=2 'STEP' 1 'UNTIL' M 'DO'
C[R,J-1]:=C[R,J];
C[R,M]:=C[N+1-R,M+1-I]:=0;
'END';
'FOR' I:=1 'STEP' 1 'UNTIL' N-1 'DO'
'BEGIN'
PIV:=I;
'FOR' R:=I+1 'STEP' 1 'UNTIL' LR 'DO'
'IF' ABS(C[R,1])>ABS(C[PIV,1]) 'THEN' PIV:=R;
'IF' PIV 'NE' I 'THEN'
```

```
'BEGIN'
T:=V[I];
V[I]:=V[PIV];
V[PIV]:=T;
'FOR' J:=1 'STEP' 1 'UNTIL' M 'DO'
'BEGIN'
T:=C[I,J];
C[I,J]:=C[PIV,J];
C[PIV,J]:=T;
'END';
'END';
V[I]:=V[I]/C[I,1];
'FOR' J:=2 'STEP' 1 'UNTIL' M 'DO'
C[I,J]:=C[I,J]/C[I,1];
'FOR' R:=I+1 'STEP' 1 'UNTIL' LR 'DO'
'BEGIN'
T:=C[R,1];
V[R]:=V[R]-T*V[I];
'FOR' J:=2 'STEP' 1 'UNTIL' M 'DO'
C[R,J-1]:=C[R,J]-T*C[I,J];
C[R,M]:=0;
'END';
'IF' LR 'NE' N 'THEN' LR:=LR+1;
'END';
V[N]:=V[N]/C[N,1];
JM:=2;
'FOR' R:=N-1 'STEP' -1 'UNTIL' 1 'DO'
'BEGIN'
'FOR' J:=2 'STEP' 1 'UNTIL' JM 'DO'
V[R]:=V[R]-C[R,J]*V[R-1+J];
'IF' JM 'NE' M 'THEN' JM:=JM+1;
'END';
'FOR' K:=1 'STEP' 1 'UNTIL' 868 'DO'
'BEGIN'
NEWLINE(1);
PRINT(V[K],4,2);
'END';
'END';
'END';
****
```

```
'BEGIN'
'REAL' H,R,RD,RDD,RDDD,NV,VR,VRR,VZ,VZZ,K1,K2,K3,K4,K5,K6;
'INTEGER' I;
'REAL' 'ARRAY' V[-1:210],D1[0:209],D2[0:209];
'BEGIN'
H:=0.05;V[-1]:=1000;'FOR' I:=0 'STEP' 1'UNTIL' 210 'DO' V[I]:=READ;
'FOR' I:=0 'STEP' 1 'UNTIL' 3 'DO'
'BEGIN'
K1:=V[I+1]-V[I];
K2:=V[I+2]-2*V[I+1];
K3:=V[I+3]-3*V[I+2]+3*V[I+1]-V[I];
K4:=V[I+4]-4*V[I+3]+6*V[I+2]-4*V[I+1]+V[I];
K5:=V[I+5]-5*V[I+4]+10*V[I+3]-10*V[I+2]+5*V[I+1]-V[I];
K6:=V[I+6]-6*V[I+5]+15*V[I+4]-20*V[I+3]+15*V[I+2]-6*V[I+1]+V[I];
D1[I]:=(1/H)*(K1-(1/2)*K2+(1/3)*K3-(1/4)*K4+(1/5)*K5-(1/6)*K6);
D2[I]:=(1/(H↑2))*(K2-K3+(11/12)*K4-(5/6)*K5+(137/180)*K6);
'END';
'FOR' I:=4 'STEP' 1 'UNTIL' 206 'DO'
'BEGIN'
K1:=V[I+1]-2*V[I]+V[I-1];
K2:=V[I+2]-4*V[I+1]+6*V[I]-4*V[I-1]+V[I-2];
K3:=V[I+3]-6*V[I+2]+15*V[I+1]-20*V[I]+15*V[I-1]-6*V[I-2]+V[I-3];
K4:=V[I+2]-2*V[I+1]+2*V[I-1]-V[I-2];
K5:=V[I+3]-4*V[I+2]+5*V[I+1]-5*V[I-1]+4*V[I-2]-V[I-3];
K6:=V[I+4]-6*V[I+3]+14*V[I+2]-14*V[I+1]+14*V[I-1]-14*V[I-2]+6*
    V[I-3]-V[I-4];
D1[I]:=(1/(2*H))*(V[I+1]-V[I-1]-(1/6)*K4+(1/30)*K5-(1/140)*K6);
D2[I]:=(1/(H↑2))*(K1-(1/12)*K2+(1/90)*K3);
'END';
'FOR' I:=207 'STEP' 1 'UNTIL' 210 'DO'
'BEGIN'
K1:=V[I]-V[I-1];
K2:=V[I]-2*V[I-1]+V[I-2];
K3:=V[I]-3*V[I-1]+3*V[I-2]-V[I-3];
K4:=V[I]-4*V[I-1]+6*V[I-2]-4*V[I-3]+V[I-4];
K5:=V[I]-5*V[I-1]+10*V[I-2]-10*V[I-3]+5*V[I-4]-V[I-5];
K6:=V[I]-6*V[I-1]+15*V[I-2]-20*V[I-3]+15*V[I-4]-6*V[I-5]+V[I-6];
D1[I]:=(1/H)*(K1+(1/2)*K2+(1/3)*K3+(1/4)*K4+(1/5)*K5+(1/6)*K6);
D2[I]:=(1/(H↑2))*(K2+K3+(11/12)*K4+(5/6)*K5+(137/180)*K6);
'END';
K1:=0;
'FOR' RD:=0.01512,0.01541,0.01570,0.01599,0.01628,0.03492'DO'
'BEGIN'
R:=0.05;
'FOR' I:=0 'STEP' 1 'UNTIL' 209 'DO'
'BEGIN'
NV:=V[I]-(((R↑2)/4)*D2[I])+K1;
VR:=(-R/2)*D2[I];
VZ:=D1[I];
VRR:=(-1/2)*D2[I];
VZZ:=D2[I];
RDD:=((1+RD↑2)/(2*NV))*(VR-(RD*VZ));
RDDD:=((VRR/(2*NV))*(RD-RD↑2+RD↑3-RD↑4))-((RD*VZZ*(1+RD↑2))/(2*NV))
      -((3*RDD*VZ*(1+RD↑2))/(2*NV));
PRINT(R,4,4);
```

```
SPACE(3);
PRINT(RD,4,5);
SPACE(3);
K3:=I*H;
'IF' I=0 'THEN' K3:=1;
K2:=R/K3;
PRINT(K2,4,5);
SPACE(3);
K2:=K2-RD;
PRINT(K2,4,5);
SPACE(3);
PRINT(NV,4,3);
SPACE(3);
PRINT(V[I],4,3);
SPACE(3);
PRINT(D1[I],5,3);
SPACE(3);
PRINT(D2[I],5,3);
NEWLINE(1);
R:=R+H*RD+((H↑2*RDD)/2)+((H↑3*RDDD)/6);
RD:=RD+H*RDD+((H↑2*RDDD)/2);
'END';
'END';
'END';
'END';
****
```

References

1. A. V. Crewe, D. N. Eggenburger, J. Wall and L. M. Welter: Rev. Sci. Instrum. 39, 576 (1968)

2. R. W. Wood: Phys. Rev. 5, 1 (1897)

3. J. E. Lilienfield: Phys. Z. 23, 506 (1972)

4. W. Schottky: Z. Physik. 14, 63 (1923)

5. B. S. Gossling: Phil. Mag. 1, 609 (1926)

6. R. A. Millikan and C. F. Eyring: Phys. Rev. 27, 51 (1926)

7. R. J. Piersol: Phys. Rev. 31, 441 (1928)

8. R. A. Millikan and C. C. Lauritsen: Phys. Rev. 33, 598 (1929)

9. R. H. Fowler and L. Nordheim: Proc. Roy. Soc. Lond., Ser. A 119, 173 (1928)

10. L. Nordheim: Proc. Roy. Soc. Lond., Ser. A 121, 626 (1928)

11. J. W. Beams: Phys. Rev. 44, 803 (1933)

12. L. R. Quarles: Phys. Rev. 48, 260 (1935)

13. D. H. Moore: Phys. Rev. 50, 344 (1936)

14. A. J. Ahearn: Phys. Rev. 50, 238 (1936)

15. E. W. Müller : Z. Physik. 102, 734 (1936)

16. A. Wehnelt and W. Schilling: Z. Physik. 98, 286 (1936)

17. E. W. Müller : Z. Physik. 106, 541 (1937). 108, 668 (1938)

18. R. Haefer: Z. Physik. 116, 604 (1940)

19. J. E. Henderson and R. E. Badgley: Phys. Rev. 38, 590 (1931)

20. J. E. Henderson, K. V. Mackenzie and R. K. Dahlstrom: Phys. Rev. 48, 484 (1935)

21. G. Richter: Z. Physik. 119, 406 (1942)

22. E. W. Müller: Z. Physik. 120, 261 (1943)

23. R. D. Young: Phys. Rev. 113, 110 (1959)

24. R. D. Young and E. W. Muller: Phys. Rev. 113, 115 (1958)

25. L. Apker and E. Taft: Phys. Rev. 88, 1037 (1952)

26. R. Stratton: Proc. Phys. Soc. (London). B 68, 746 (1955)

27. R. L. Perry: Office Of Naval Research Final Report, Project
NSONR - 72401, Authority NR072 - 171 (May 1957) p.13

28. M. I. Elison and G. F. Vasiliev: Radiotekh. i Elektron. 4, 728 (1959)

29. R. L. Perry: J. Appl. Phys. 32, 128 (1961)

30. F. G. Allen: J. Phys. Chem. Solids. 8, 119 (1959)

31. L. A. D'Asaro: J. Appl. Phys. 29, 33 (1958)

32. A. I. Klimin, B. N. Sedykh and I. L. Sokol'skaya: Soviet Physics
Solid State. 2, 1673 (1961)

33. R. Stratton: Phys. Rev. 125, 67 (1962)

34. R. Siewatz and M. Green: J. Appl. Phys. 29, 1034 (1958)

35. P. Handler: J. Phys. Chem. Solids. 14, 1 (1960)

36. F. G. Allen: (unpublished)

37. L. Ernst: Phys. Stat. Sol. 24, 177 (1967)

38. J. R. Arthur: J. Chem. Phys. Solids. 25, 585 (1964)

39. J. R. Arthur: J. Appl. Phys. 36, 3221 (1965)

40. G. P. Shcherbakov and I. L. Sokol'skaya: Soviet Physics Solid State.
3, 120 (1961)

41. H. Neuman: Z. Naturf. 23a, 204 (1968)

42. M. I. Elison, A. G. Zdan, V. F. Krapivin, Z. B. Lipkovski, V. N.
Lukski, and V. B. Sandomirsky: Radiotekn. i. Electron. 10, 1228
(1965)

43. L. M. Baskin, O. I. Lvov and G. N. Fursey: Phys. Stat. Sol. 47,
49 (1971)

44. R. Stratton: Phys. Rev. 135, A794 (1964)

45. E. L. Murphy and R. H. Good: Phys. Rev. 102, 1464 (1956)

46. R. Fischer: Phys. Stat. Sol. 2, 1088 (1962)

47. R. Fischer: Phys. Stat. Sol 2, 1466 (1962)

48. A. M. Russell: Phys. Rev. Letters 9, 417 (1962)

49. J. T. Law: J. Phys. Chem. Solids. 14, 9 (1960)

50. A. M. Russell and E. Litov: App. Phys. Letters 2. 64 (1963)

51. G. P. Shcherbakov and I. L. Sokol'skaya: Soviet Physics Solid State.
4, 2581 (1963)

52. G. P. Shcherbakov and I. L. Sokol'skaya: Soviet Physics Solid State. 4, 31 (1962)

53. J. R. Arthur: Surface Science. 2, 389 (1964)

54. J. R. Arthur: Tenth Field Emission Symposium (Baldwin-Wallace College, Berea, Ohio, 1963)

55. O. H. Hughes and P. M. White: Phys. Stat. Sol. 33, 309 (1969)

56. J. R. Arthur: J. Appl. Phys. 37, 3057 (1966)

57. W. A. Mc.Dowell, E. Braun and E. E. Donaldson: Bull Am. Phys. Soc. 11, 34A (1966)

58. A. J. Dekker: Solid State Physics. Macmillan. Page 212

59. P. Lukirsky: Z. Physik. 22, 351 (1924)

60. V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hiller and A. W. Vance: Electron Optics and the Electron Microscope. John Wiley and Sons Ltd. Page 444

61. V. K. Zworykin, G.A.Morton, E.G.Ramberg, J.Hillier and A. W. Vance: Electron Optics and the Electron Microscope. John Wiley and Sons Ltd. Page 443

62. L. Page and N. I. Adams: Principles of Electricity. Van Nostrand. Page 99.

63. Sir James Jeans: The Mathematical Theory of Electricity and Magnetism. Cambridge University Press. Page 225.

64. R. H. Galloway, H. McL. Ryan and M. F. Scott: Proc. I.E.E. 114, No. 6, 824 (1967)

65. G. Stephenson: Machematical Methods For Science Students. Longmans. Page 86.

66. V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier and A. W. Vance: Electron Optics and the Electron Microscope. John Wiley and Sons Ltd. Page 377

67. H. Thurnau: Collected Algorithms from C.A.C.M. Algorithm 195

68. S. C. Loh: Proc. I.E.E. 117, No. 3, 641 (1970)

69. F. Rodriguez: Collected Algorithms from C.A.C.M. Algorithm 189

70. A. D. Booth: Numerical Methods. Butterworths. Chapters 3 and 4

71. V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier and A. W. Vance: Electron Optics and the Electron Microscope. John Wiley and Sons Ltd. Page 401

72. L. N. Dobretsov and M. V. Gomoyunova: Emission Electronics. Keter Press Binding. Page 247

73. K. G. McKay: Advances in Electronics and Electron Physics. Vol. 1. Academic Press Inc. Page 76

74. J. Arol Simpson and C. E. Kuyatt: J. Appl. Phys. 37, 3805 (1966)

75. B. Zimmermann: Advances in Electronics and Electron Physics. Vol. 29. Academic Press Inc. Page 257

76. R. H. Good and E. W. Müller: Handbuch der Physik. Springer-Verlag. Berlin. 1956. Vol. 21. Page 181

77. S. C. Miller and R. H. Good: Phys. Rev. 91, 174 (1953)

78. L. W. Nordheim: Phys. Z. 30, 177 (1929)

79. W. W. Dolan and W. P. Dyke: Phys. Rev. 95, 327 (1954)

80. P. H. Cutler and R. H. Good: Bull. Amer. Phys. Soc. 1, 279 (1956)

81. A. J. Dekker: Solid State Physics. Macmillan. Page 213

82. H. Kanter and W. A. Feibelman: J. Appl. Phys. 33, 3580 (1963)

83. A. J. Dekker: Solid State Physics. Macmillan. Page 308

84. W. Shockley and G. L. Pearson: Phys. Rev. 74, 232 (1948)

85. J. Bardeen: Phys. Rev. 71, 717 (1947)

86. L. M. Baskin, O. I. Lvov and G. N. Fursey: Phys. Stat. Sol. (b) 47, 49 (1971)

87. A. J. Dekker: Solid State Physics. Macmillan. Page 248.

88. W. A. Hamson: Phys. Rev. 123, 85 (1961)

89. R. E. Burgess, H. Kroemer and J. M. Houston: Phys. Rev. 90, 515 (1953)

90. E. W. Müller and Tien Tzou Tsong: Field Ion Microscopy, Principles and Applications. Elsevier. Page 25

91. H. L. Caswell in Physics of Thin Films. Ed. Georg Hass. Academic Press. Page 24

92. F. B. Haller: Rev. Sci. Instrum. 35, 1356 (1964)

93. R. Gomer: Field Emission and Field Ionisation. Oxford University Press. Page 168.

94. R. Basset (Private communication)

95. H. V. Pilling: Concise Intermediate Physics. The English Universities Press Ltd. Page 183.

96. G. W. C. Kaye and T. H. Laby: Physical and Chemical Constants. Longmans. Page 187.

97. R. Gomer: Field Emission and Field Ionisation. Oxford University Press. Page 37.

98. L. Page, N. I. Adams: Principles of Electricity. D. Van Nostrand Company Inc. Page 125.

99. Handbook of Chemistry and Physics. Chemical Rubber Publishing Company. Page 2728.

100. M. I. Elison: Radiotekhnika. Elektronika. 4, 140 (1959)

101. K. J. Haas, D. C. Fox and M. J. Katz: J. Phys. Chem. Solids. 26, 1779 (1965).

102. H. Berger, K. W. Boer and E. H. Weber: Z. Phys. 158, 501 (1960)

103. N. Nicolaou and A. Modinos: J. Phys. C: Solid St. Phys. 4, 2859 (1971)

104. J. Marien and J. Loosveldt: Phys. Stat. Sol. (a). 8, 213 (1971)

105. R. W. Smith: Phys. Rev. 97, 1525 (1955)

106. E. W. Plummer and C. E, Kuyatt: Rev. Sci. Instrum. 43, 108 (1972)

o o O o o

ACKNOWLEDGEMENTS

o o O o o