The preparation and properties of luminiscent films on silicon

G. S. Edwards

THE PREPARATION AND PROPERTIES OF LUMINESCENT FILMS ON SILICON

by

G.S. Edwards B.Sc.

A thesis presented in candidature for the degree of Doctor of Philosophy in the University of Durham

May 1970

ABSTRACT

The interconnection problems in solid state electroluminescent displays, due to complex addressing and memory requirements, can be minimised by making a monolithic device in which light is produced in a thin film on silicon and controlled by underlying integrated circuitry. This thesis describes the preparation and properties of some luminescent films for this application.

Preliminary work was aimed at depositing epitaxial ZnS phosphor films on silicon but, due to problems of reproducibility and activation, this was not a success. Its termination, however, was prompted mainly by another reason; the discovery by the author of the phosphor willemite (Zn_2SiO_4:Mn) which seemed admirably suited for deposition as a thin luminescent film on silicon integrated circuitry. The main advantage was the novel method of fabrication which involved the conversion of an oxide layer on the silicon surface to willemite by reaction with an evaporated film of ZnF_2:Mn.

The films produced were composed of microcrystallites embedded in a matrix of unreacted oxide. They were brightly green cathodoluminescent. In an MIS structure the current was assymetrical with bias. For negative and low positive bias on the metal the films were highly insulating but with high positive bias relatively large currents flowed accompanied by weak green electroluminescence. A conduction model, based on capacitance-voltage, current-voltage and luminescence experiments, has been proposed. In this, mobile positive ions in the film discharge only when the cathode is metal

so that with positive bias, they accumulate in front of the silicon where field intensification occurs. Electrons then tunnel from the silicon to collision excite the luminescence of manganese centres inthe willemite.

The observation of d.c. electroluminescence in thin films of willemite together with the ease and reproducibility of the fabrication process and its apparent compatibility with silicon technology, is thought to be a significant step in the development of a monolithic display based on silicon.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. M.J. Morant for his supervision and for his guidance throughout the work and the preparation of this thesis. He wishes to thank the Science Research Council for financial support, Professor D.A. Wright for the use of his laboratory facilities and the workshop staff, headed by Mr. F. Spence, both for the construction of apparatus and for their unfailing cheer. He would also like to thank his colleagues in the Department for many useful discussions (particularly after six), Mr. P.G. Martin for help in the construction of the capacitance-voltage electronics and Dr. A.N. Rushby for operating the electron microscope in the Department. The use of the scanning electron microscopes at the Research Laboratories of J.J. Jobling, the School of Engineering Science at Bangor and the Department of Geology at Durham is gratefully acknowledged. Finally the author would like to thank his mother for so willingly typing this thesis.

CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 Electronic display systems

The ultimate object of the work described in this thesis is the production of a display device which is to be completely compatible with modern electronics. A display is essentially a device to convey information across the interface between a machine and an observer. The ideal display achieves optimum matching at the interface and consequently its design is determined not only by the requirements of the observer but also by the type of information presented to the interface.

Electronic displays range in complexity from simple indicator lamps and alpha-numeric devices to television receiver tubes and addressable computer output terminals. The prime requirement of any display is to convert the information into a visible form. The devices fall into two classes. In the first, electrical energy is converted directly into light output (i.e. active displays). In the second, the electrical energy is used to modulate an external light source (i.e. passive displays). A further requirement for any other than simple systems is the ability to show different types of information (e.g. numbers or letters) quickly, one after the other, on the same display area. This is usually achieved for alpha-numerics in one of three ways (a) by selectively lighting up dots in a 7 by 5 matrix, (b) by selectively lighting up elements in a seven bar arrangement, and (c) illumination of full character shapes

stacked one behind the other as in the gas discharge 'Nixie' tube. (Fig 1-1). Further sophistication necessitates the device having a memory so that it can retain the original display once it has been written without the continual re-addressing that would require excessively large bandwidths for the information link. Advanced computer output graphics also require displays to be addressable from both sides of the interface.

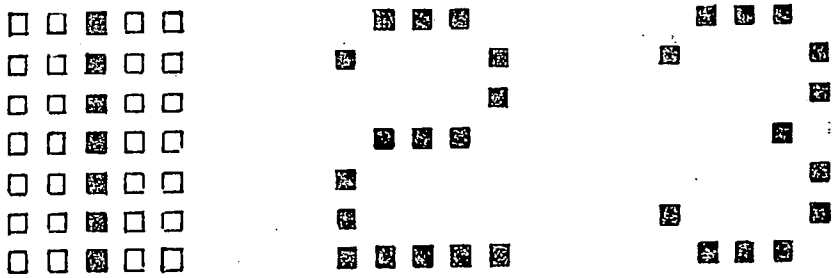
Many types of system have been proposed for electronic display. Some, like the cathode ray tube, have become almost universally accepted for certain applications such as television, whilst the potential of others has remained unrealised. A brief description of the main types of display will now be given.

1) Passive displays

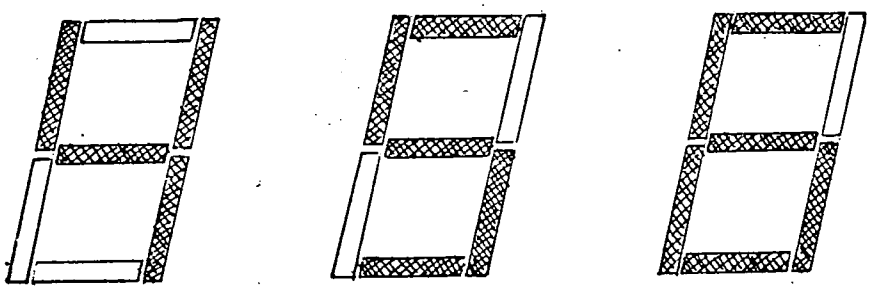
The modulation of light can be achieved by selective absorption or by scattering of the incident light. Displays can also be made by deflecting the light prior to incidence on a screen. Several physical processes have been proposed for such systems.

a) Photochromism Photochromism is the reversible change that occurs in the absorption spectrum of certain materials produced by the incidence of electromagnetic radiation. By using ultra-violet light to write the display and infra-red for erasure, practical displays can be made (Ref 1-1).

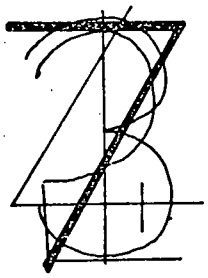
b) Electrochromism When materials change their absorption or scattering properties under the influence of an electric field they are said to be electrochromic. The Franz-Keldysh



a) 7 by 5 alpha-numeric display



b) 7 bar stylised numeric



c) coincident 'natural' shapes usually stacked as in a 'Nixie' tube

Fig 1-1 Some alpha-numeric display configurations

effect is an example of this in which small shifts in the absorption edge of certain materials (such as GaAs) are produced by an applied field, but these are usually too small for practical displays. However certain 'liquid crystals' characterised by long thread-like molecules, exhibit a form of electrochromism that has been used to achieve practical displays in thin film structures (Ref 1-2). The effect has been attributed to the movement of contaminant ions in the film disrupting the regular alignment of the long molecules so that selected areas can be made opaque. A similar effect can also be produced by thin films of certain magnetic materials which can be made to scatter light when suitably activated by electric fields.

c) Light deflection systems Light deflection systems utilize either movements of the optical components of the system or changes in the refractive index of the optical media induced either acoustically or electro-optically.

d) Mechanical display systems Mechanical display systems, mentioned here for the sake of completeness, include the more mundane but very useful types of display dependent on the physical movement of pointers and patterns usually achieved electromagnetically (e.g. clocks and meters).

2) Active displays The conversion of electrical energy into light can be accomplished in general by any of three processes:- gas discharges, the incandescence of solids and the luminescence of phosphors. All three are used in active displays.

a) Gas discharges Neon lamps and fluorescent lights have been used for many years for large scale displays and for indicators. Alpha-numeric display devices based on glow discharges are also used extensively in the form of 'Nixie' tubes. Several of the disadvantages of 'Nixie' tubes have now been eliminated in a new generation of glow discharge displays that utilize a dot matrix (Ref 1-3). These devices are potentially very cheap and it is also possible to incorporate storage facilities into them.

b) Incandescent displays Incandescent tungsten filaments are used for a great many display applications where their high brightness and cheapness are considerable assets.

c) Cathodoluminescent displays The cathode-ray tube has been developed to a high degree of sophistication brought about mainly by the consumer demand for television and the military demand for radar displays. CRT systems are now produced that incorporate all the features required for the more complex displays, but at a high price. Several of the other disadvantages of the CRT can be eliminated by cold-cathode thin film structures and these are being actively studied (Ref 1-4). However, the short lifetimes of these devices at the moment may prove to be a fundamental limitation.

d) Electroluminescent displays Considerable research effort has been devoted to the production of electroluminescent displays. ZnS phosphor powders have been used in the form of display panels for several years but they have always been relatively dim and have usually required high a.c. voltages preferably at frequencies

of a few kHz. Nevertheless a television display based on a ZnS powder phosphor has been made by Mitsubishi Ltd (Japan) although it has not yet appeared commercially.

More recently both powder layers and thin films (Ref 1-5) have been made to give out sufficient light to be plainly visible in a lighted room using only d.c. voltages of a few hundred volts. However, an 'electroforming' process is at present required to initiate the electroluminescence in these devices and probably because of this the working lifetimes are short. It has been proposed that this sort of device would be very suitable for large scale display in conjunction with thin film transistors.

In ZnS electroluminescence (referred to as EL in this thesis) is thought to be the result of excitation of luminescent centres by electrons that have been injected into the luminescent material as the result of localised high fields. The alternative form of EL shown by other materials, such as GaP, utilizes the recombination of excess carriers which have been injected from a forward biased p-n junction. The extensive study of this phenomenon and its attendant technology has resulted in light emitting diodes of GaP (giving red and green light) and GaAsP (red) being commercially available for use as indicator lamps. Alpha-numeric displays are also available in the form of 7 by 5 matrices of discrete diodes from Hewlett-Packard Ltd and as seven bar element displays from Monsanto Ltd. More recently, Hitachi Ltd (Japan) and Standard Telephone Laboratories Ltd., have produced a completely monolithic numeric display on a single chip of GaAsP. For direct compatibility with computer logic circuitry or

other digital input all these types of display incorporate a separate silicon integrated circuit decoder separately connected to the display element.

With the success of the GaP technology the basic research effort is now being devoted to the new generation of electroluminescent materials such as AlN and tertiary compounds such as InGaP which are potentially more efficient and capable of emitting light of several different wavelengths. However, this work may be forestalled by the development at Bell Laboratories of a new range of phosphors excited by infra-red radiation and thus capable of being used in conjunction with the very efficient GaAs infra-red emitters.

1.2 The disadvantages of existing display systems

Apart from the GaP type displays the main disadvantage of all the systems described above is that they are incompatible with solid state circuitry either because of their high drive voltage or large current consumption. The passive displays based on both photochromic and electrochromic devices have the additional disadvantages in that they are relatively unproven and they appear to suffer from a slow response and a short lifetime. Mechanical systems suffer equally from a slow response and they are also unreliable compared with completely electronic displays although they can be comparatively cheap. It is considered that light deflection systems may be too sensitive to misalignment to be practical as versatile working displays.

Among the active displays it is unfortunate that gas discharge devices require relatively high drive voltages because their cheapness and easy availability offer considerable advantages. Incandescent lamps can also be relatively cheap for large scale displays but they consume a comparatively large amount of power and generate unnecessary heat as well as having a relatively short life. The cathode-ray tube also has several disadvantages for display. Although it is physically large, electronically complicated, expensive and fragile, it will nevertheless probably reign supreme for many years because of its high degree of sophistication. Thin film cold cathodes may remove some but not all of the disadvantages of the cathode-ray tube.

In contrast with the display systems described above, GaAs and GaP diodes require only low drive voltages (1.7 V) and currents (10 mA) so that they are compatible with integrated circuitry. They are also very reliable and potentially cheap although at the moment their prices are not competitive with the more established types of indicator display. However, advanced displays require complex interconnection with electronic circuitry for addressing and possibly memory purposes. It is certain that these functions will be performed with silicon integrated circuitry for the foreseeable future so the interconnection problem is due to a fundamental incompatibility of the materials for light emission and for integrated circuitry. As displays grow in complexity the interconnection

problem increases and it at least sets a cost limit to the use of large scale displays of this type.

A completely monolithic display based on silicon would solve the interconnection problem and would thus have considerable advantages over existing displays. These advantages should further increase in the future as more and more information is presented for display from solid state devices. These considerations were the main reason for carrying out the work described in this thesis which is aimed at the eventual production of a device in which the light output from an electroluminescent film on silicon is controlled directly by circuitry within the silicon.

1.3 A silicon-based electroluminescent device

Silicon itself is not electroluminescent because the probability of non-radiative transitions occuring is much greater than the probability of radiative transitions. Although the reasons for this are not completely understood, a contributory factor is the detailed energy structure of the silicon and its impurity centres which show that a change in momentum must accompany the radiative recombination of an electron and a hole. Even if this were not so any EL from silicon would be in the infra-red because the energy gap is 1.1 eV whereas for visible EL a material with a band gap greater than about 1.7 eV is required.

An alternative EL mechanism to the direct recombination of electrons and holes in silicon is EL due to

avalanche breakdown. In this 'hot' electrons generated by very high fields in the silicon lose energy radiatively within the conduction band. The radiative process in such high fields is however very inefficient and consequently only very weak EL is possible, even if the high fields can be generated uniformly. Nevertheless Fairchild Semiconductors Ltd have developed avalanche luminescent diode arrays (18 by 32) for use as film markers where the very low light output can be tolerated.

Although silicon itself cannot emit light efficiently by recombination across the energy gap, it is possible in theory to inject carriers from the silicon into an electroluminescent material with which it is in intimate contact. Efficient minority carrier injection cannot occur however at a forward biased p-n heterojunction from the narrower energy gap material (Si in this case) into the wider gap phosphor because energy level considerations dictate that the injection efficiency will always be low due to the outflow of majority carriers from the phosphor (Ref 1-6). A high field injection mechanism, such as tunnelling, must therefore be utilised, and in order to attain the necessary high field the electroluminescent material must not only be relatively insulating but it must also be extremely uniform to prevent dielectric breakdown. Large voltages are not, however, required to produce the high fields if the luminescent films are very thin (less than about 1000 Å).
A possible device would then have a metal-insulator-silicon (MIS) structure with a transparent top contact in place of the metal to

allow the light to escape. The mechanism which is proposed for the radiative transitions of electrons injected from silicon into a suitable phosphor film is similar to the present theories of d.c. electroluminescence in ZnS:Mn phosphors. These depend on high field injection from conducting inclusions in the ZnS and the subsequent acceleration of the carriers to excite manganese luminescent centres either by resonant transfer with copper activator centres or by direct collision. The light output is then characteristic of the manganese ion in the host lattice as it relaxes back to its equilibrium level. A schematic energy band diagram of this mechanism in the proposed device is shown in Fig 1-2.

It should be noted that electroluminescent materials such as GaP which utilize forward biased junction injection and subsequent recombination of electron-hole pairs via impurity centres cannot be used as a thin film on silicon because their relatively low resistivity prevents the attainment of the requisite high fields without the loss of one of the carrier types required for the electron-hole recombination.

1.4 An outline of the thesis

The general form of a display device based on silicon has been described in the previous section. Having decided on this line of approach the immediate research programme must be concerned with the materials to be used to give the optimum light output. The necessity for high field d.c. electroluminescence

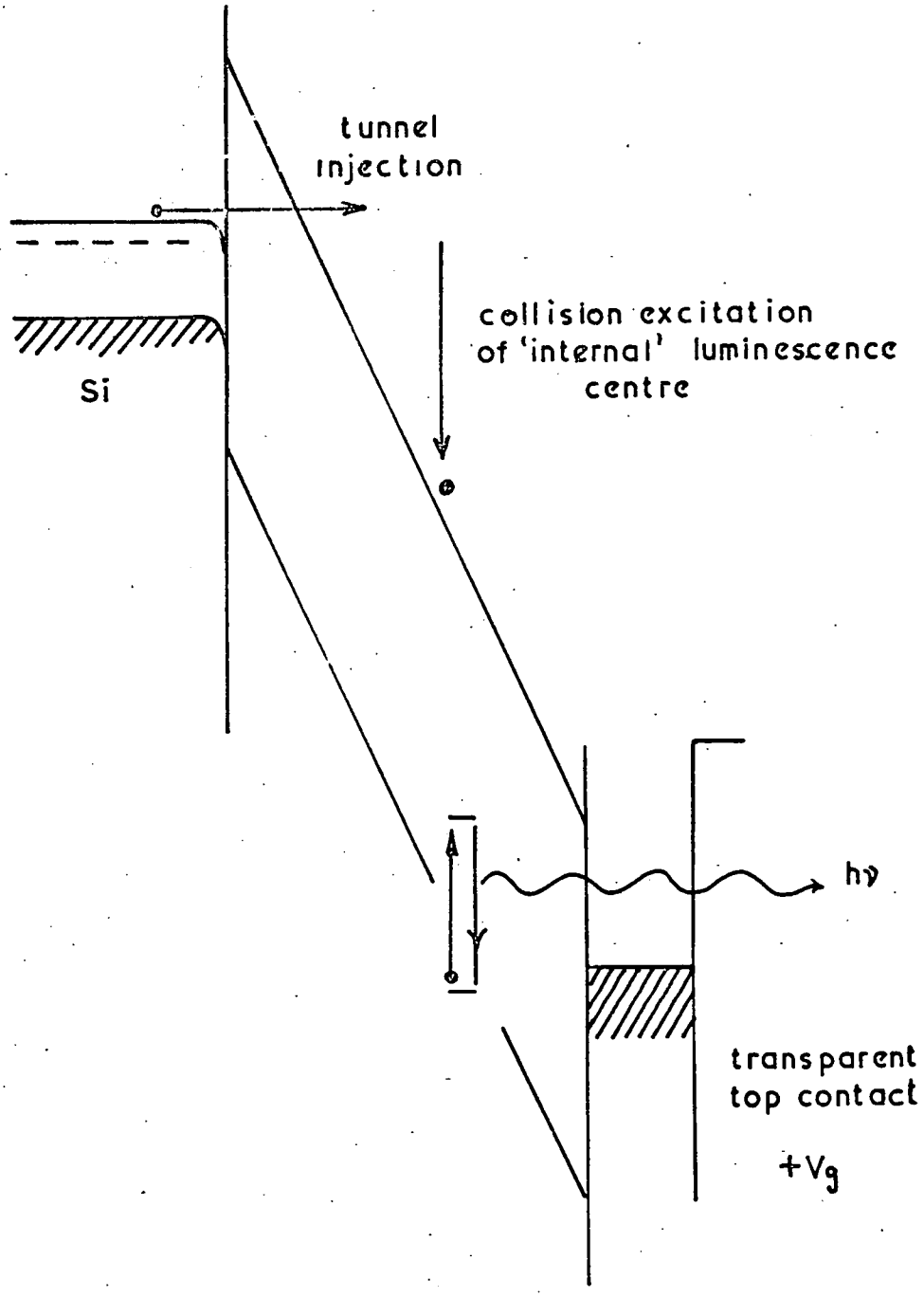


Fig 1-2 Schematic energy diagram of a possible thin film electroluminescent device based on silicon

led naturally to the work on the production of ZnS films on silicon described in Chapter 3. It was hoped that the films would be capable of sustaining high fields and would be electroluminescent if they could be produced as single crystal layers on the silicon. The prime reason for choosing ZnS initially was that it is a well-proven phosphor which is capable of attaining very high luminescent efficiencies in cathodoluminescence. It also had a crystal structure that is potentially compatible with silicon for epitaxial growth.

Unfortunately the work on the deposition of ZnS on silicon was not a success mainly because of the difficulty of obtaining reproducible films but also because it was found to be difficult to activate the films once they had been deposited.

This part of the work also included a certain amount of silicon technology (Chap 2). This was thought to be necessary not only because silicon was being used as a substrate but also because it is anticipated that the silicon will later contain p-n junctions and the effect that the phosphor layers had on these will be important. A knowledge of the phosphor-silicon interface also necessitated some silicon work.

The ZnS work was terminated after about one year. In addition to the unsuccessful state of this work its termination was prompted by the discovery by the author of another phosphor which already appeared to have considerable advantages over other materials when used as a film on silicon. The phosphor, manganese

activated zinc orthosilicate (Zn_2SiO_4:Mn) is used extensively for cathode ray tube screens and it also occurs in nature as the mineral willemite.

The main advantages of willemite are in the method of forming a thin film of the phosphor on silicon and in its compatibility with the established silicon technology. The process used was a novel one involving the conversion of thermal oxides on silicon into willemite by the reaction with an evaporated film of the phosphor ZnF_2:Mn. (Chap 4). The films were brightly green cathodoluminescent and they compared favourably with commercial willemite cathode ray tube phosphors.

The structure of this new type of phosphor film was studied by several methods including X-ray diffraction , electron microscopy and infra-red absorption, and these are described in Chap 5.

The willemite films are relatively insulating at low field strengths and information on their electrical properties was obtained largely by measuring the capacitance-voltage (C-V) characteristics of metal-willemite-silicon (MWS) structures and comparing them with similar thermal oxide devices. It was thus possible to establish the nature of charge movements in the willemite and to estimate the number of 'active' surface states at the silicon-willemite interface. (Chap 6).

The current-voltage characteristics of the films are described in Chapter 7. The most significant result to come

from these measurements was that although the films are insulating (about 10^{14} ohms) for a negative bias on the metal up to fields of 2.10^{6} V cm^{-1} and for a low positive bias relatively large currents can be passed through them at high positive field strengths (about 15 V across 750 Å). This was attributed to tunnelling from the silicon due to field intensification caused by ionic space charge selectively building up in front of the silicon. Further confirmation of this mechanism was given by the C-V work.

The most important consequence of this discovery was that the relatively high currents are accompanied by weak green EL, which occurs only in these conditions, i.e. with a high positive field. This was discussed together with other luminescent properties of the films in Chap 8.

A study of the literature showed that EL had occasionally been observed previously in willemite but only in powder phosphors and usually with high a.c. voltages. (Sec 4.2) However it has not been investigated in any detail. As far as can be ascertained this is the first time that d.c.EL has been seen in thin films of willemite and at voltages as low as 15 V in this material. Although the EL was weak, being visible only in a darkened room, it is considered that this observation together with the novel method of production of these films represents a significant step in the realisation of a completely monolithic display device based on silicon.

The nature of the work described in this thesis is mainly exploratory with the unfortunate consequence that each particular phenomenon studied could not be examined in as great a depth as would have been liked in the time available. For example, no completely detailed mechanisms are proposed for either the I-V characteristics or the EL phenomenon itself. However, several aspects of the work are novel and it is hoped that some of it will form the basis for a more detailed examination of the physical processes involved and a more detailed viability study of a practical display system. Further work in this direction is suggested in Chapter 9 together with an overall discussion to link together the preceding chapters.

References for Chapter 1

1-1 G.K. Megla, Applied Optics 5, 945 (1966)

1-2 G. Heilmeier, Proc. I.E.E.E 56, 803 (1968)

1-3 D.L. Bitzer and H.G. Slottow, Proc. Full Joint Computer Conference (1966)

1-4 R.W. Lomax and J.G. Simmons, The Radio and Electron. Engineer 35, 265 (1968)

1-5 A. Vecht, N.J. Werring, R. Ellis and P.J.F. Smith Brit. J. Appl. Phys. (J. Phys.D) Series 2, 2, 953 (1969)

1-6 A.G. Fischer, S.S. Electron. 2, 232 (1960)

CHAPTER TWO

SILICON TECHNOLOGY.

2.1 Introduction

In order to produce thin luminescent films on silicon it was first necessary to master several aspects of modern silicon technology. Furthermore, as it was envisaged that finally these films would be deposited on to silicon which might contain both p-type and n-type regions forming part of integrated circuitry within the chip, it was necessary to produce simple diffused p-n junctions in some of the silicon substrates. The first part of this Chapter is concerned mostly with the aspects of silicon when it is used as a substrate and Sec. 2. 4 describes work on the production of diffused p-n junctions and the interactions of these junctions with surface films on the silicon.

2.2 Silicon preparation

Czochralski grown single crystal slices of silicon were obtained from Metallurgie Hoboken (Belgium) with one side of each slice alumina polished to a mirror finish and the other side lapped. Both - p-type and n-type silicon were obtained doped with boron and arsenic respectively to give resistivities in the 1-10 ohm-cm range. The orientation of the slices was (111) $\pm$ 1 degree.

The 1 inch diameter slices, which were about 100 microns thick, were scribed with a diamond and cleaved to produce chips which were usually about 1 cm by 0.5 cm for use as substrates. The silicon

slices, as bought, were claimed to be nominally dislocation free, and to check this the dislocation densities were measured both before and after cleaning, by using Sirtl etch (Ref 2-1) which was made up from a 33% by weight solution of CrO_3 in water and 49% HF mixed in the proportions of 1:2 by volume. This etch preferentially reveals dislocations on the (111) face of silicon after an immersion of about 2 min. A 15 sec immersion reveals small scratches and stacking faults on the silicon surface. The dislocation density of silicon slices obtained from Hoboken was usually less than 100 cm^{-2}. The preliminary work used silicon slices obtained from Ferranti Ltd. and these had dislocation densities of 40,000-10,000 cm^{-2}. If the slices cleaved badly arrays of additional dislocations could be generated in the chip. Clumsily used tweezers could also produce this effect , particularly in the poorer quality Ferranti silicon. An example of the induced dislocations is shown in Fig 2-1

The surface preparation of the silicon chips prior to deposition is obviously very important particularly as the silicon is required not only as mechanical support but also as a contact for injecting carriers into the thin film to produce electroluminescence. Ideally a clean, optically flat surface free of work-damage is required. Chemical polishing etches, consisting primarily of nitric acid and hydrofluoric acid, are usually used in the production of semiconductor devices, often with a moderator of acetic acid. For compositions relatively rich in nitric acid non-selective etching

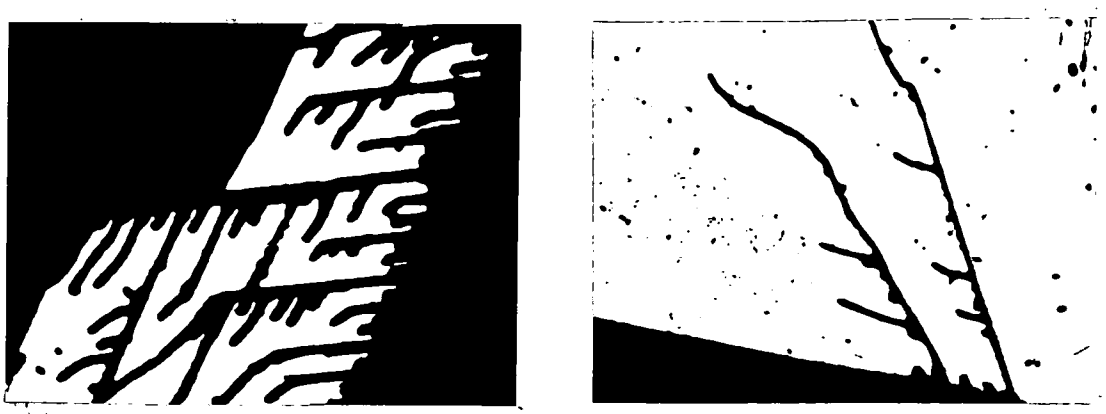


Fig 2-1 Dislocation arrays in maltreated silicon

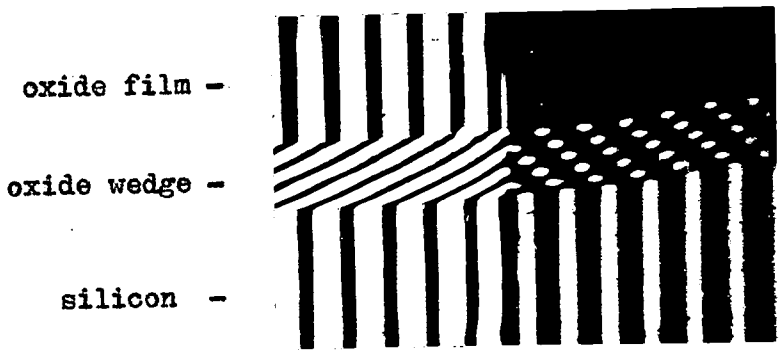


Fig 2-3 Interference fringes reflected from a partially metallised oxide step on silicon

occurs which removes the work damage and leaves the surface shiny but with a uniform rippling known as the 'orange peel' effect . Boiling 5% sodium hydroxide produces a bright shiny optically-flat surface with deep pits at dislocations.

Substrates on which ZnS was deposited in vacua (Sec 3.2.4) were prepared by degreasing in electronic grade trichloroethylene and methyl alcohol using Soxlets, washing in deionised water and then using one of these chemical polishes. Typically the acid etch that was used had a composition, by volume, of 1 part HF, 6 parts HNO_3 and 10 parts acetic acid. An immersion of about 2 min was used for this mixture, whereas for the alkali etch which was more vigorous and effervescent a shorter time was used, usually about 30 sec. No correlation was obtained however between the quality of the ZnS films and the silicon preparation, probably because of the general lack of reproducibility in the evaporations. This is described more fully in Sec 3.7.

Prior to the epitaxial deposition of silicon on silicon using the tetrachloride, a hydrogen chloride gas etch is usually used followed by a hydrogen etch, to produce very smooth damage-free surfaces. A vapour flow technique was initially envisaged as a possibility for the deposition of ZnS on silicon (Chap 3) and if this method had shown more promise (Sec 3.3) an adaptation of the gas phase etching technique would have been very convenient.

Electrolytic etching can also be used with hydrofluoric acid mixtures containing additives such as glycerol which increase the viscosity and produce the correct electropolishing action by the selective dissolution of high spots. This method has the attraction of producing surfaces that combine the optical flatness of a mechanical polish with the damage free finish of a chemical polish. However, it was not used because of the increased complexity of electropolishing and the necessity of making electrical contact to the silicon chip before the deposition of the luminescent film.

It has been shown that most wet chemical etches can introduce undesirable impurities on to the silicon surface even when electronic grade chemicals are used. Silicon surfaces also oxidise very rapidly on exposure to an oxidising gas to produce oxide layers about 50 Å thick. Even in a vacuum of 10^{-6} torr a silicon surface is oxidised in less than a second by the gases which are inevitably present in the system (Ref 2.2).

A further method of surface prepartion utilizes the thermal oxidation of the silicon surface itself. Mechanical damage at a polished surface is generally believed to extend to a depth of 3-4 times the size of the abrasive particles. The final mechanical polish usually uses 0.25 micron alumina so that damage may extend to about one micron into the surface . If the surface is then oxidised this silicon can be removed by incorporating it into the oxide. Approximately 0.4 of the oxide thickness is composed of silicon from

the underlying substrate, so that to remove all the damage an oxide layer 2-2.5 microns thick must be grown. The oxide film is then etched off with hydrofluoric acid to produce a damage-free surface which retains the flatness of the initial mechanical polish. A disadvantage of this method is the time required to grow the relatively thick oxide layers and the possible redistribution of impurities in the silicon which may take place due to different solubilities of impurities in the oxide and the silicon.

Thermal oxidation of silicon proceeds by the inward motion of the oxidising species through the oxide (Ref 2-3)(unlike the anodic oxidation of silicon which proceeds by the outward motion of silicon)(Ref 2-4). The silicon which produces the initial monolayers of oxide therefore remains at the surface of the oxide after oxidation. This means that any impurities which may have been introduced during the polishing or subsequent processing of the chip, such as abrasive particles, grow out of the silicon surface and leave an oxide-silicon interface which has never been in contact with the environment. Some impurities may however diffuse through the protecting oxide layer during the high temperature oxidation itself.

Thermal oxidation as a surface preparation was used in the preparation of planar p-n junctions (Sec 2.4). The oxide in this case also acted as a mask for impurity diffusion. Oxidation was used to prepare the surface of substrates used in the work described in Chapter 4. In this the oxide again served a double purpose, being used here as one of the reactants in the production of willemite films on silicon.

2.3 Oxidation of silicon

2.3.1 Apparatus and techniques

Oxides were grown by two processes depending on the thickness required. The apparatus used for thermal oxidation of silicon in dry oxygen is shown schematically in Fig 2-5b and the open tube steam oxidation system is shown in Fig 2-5a. The furnace was heated resistively by four Crusilite elements connected in series and equispaced around the outer alumina furnace tube. Oxidations were carried out in a silica tube, one inch in diameter, inside the furnace tube. Although the furnace was capable of running at 1400°C at full power this was seldom used because of the crystallisation of silica that occurs above 1200°C. Temperatures were controlled by using a Eurotherm Type PID/SCR-10 Controller with a Pt/Pt13% Rh thermocouple placed between the alumina tube and the silica.

Oxygen was supplied from a standard cylinder and dried by passing over trays of P_2O_5. A needle valve was used to control the gas flow which was measured with a Rotameter. Polypropylene connecting tubing was generally used in preference to vinyl which not only allows diffusion of ambient gas molecules through its walls but also contains undesirable impurities. A submicron filter is usually considered necessary in the gas flow immediately before the gas enters the reactor tube to prevent dust particles contaminating the oxides, and in later work a sintered glass filter was incorporated into the system. A more common source of contamination however, was small particles of cristabolite which gradually formed on the silica ware after long usage.

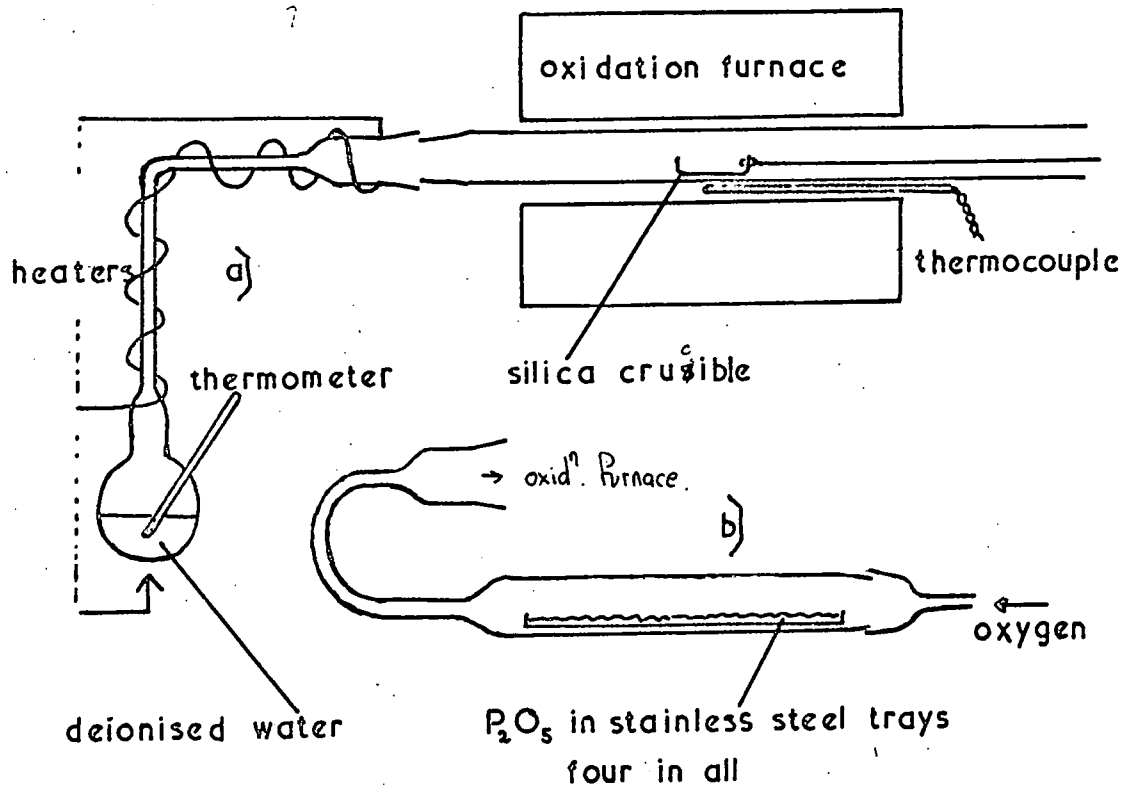


Fig 2-3 Apparatus for the oxidation of silicon by a) steam and b) dry oxygen

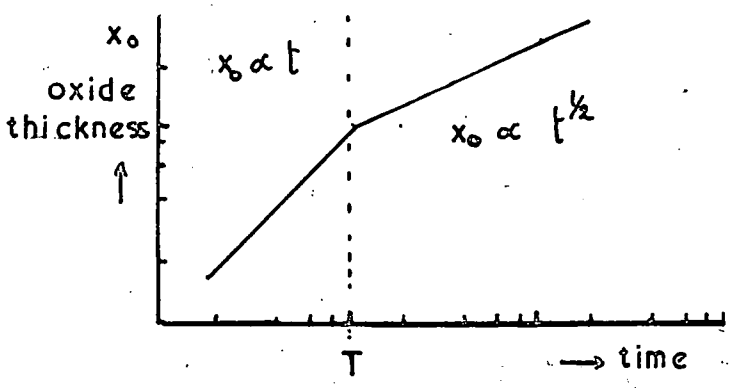


Fig 2-4 Idealised oxide growth-rate curve

Steam for oxidation was produced from deionised water in a flask which was connected to the reactor by glass tubing which was heated to prevent condensation. After degreasing in electronic grade trichloroethylene the polished silicon chips were inserted into the hot furnace in a silica boat using a silica push-rod.

The oxidation kinetics of silicon are now well understood. (Ref 2 - 5). The duration of oxidation relative to a 'time constant' T, determines whether the growth rate is linear or parabolic Fig 1 - 4. The time sonstant incorporates most of the oxidation parameters such as temperature and concentration of the oxidant. When the reaction time is less than T the rate of oxide formation is linear and depends on the rate of the reaction at the silicon/oxide interface. For times greater than T the rate is parabolic and is determined by the transport of the oxidising species through the growing oxide layer. The speed of the reaction and the diffusivity increase exponentially with temperature and therefore both the linear and the parabolic rate constants increase in the same way.

In this work thin oxides, about 1000 Å thick, were generally produced using dry oxygen typically for 1 hour at 1100°C. These are the conditions for the parabolic growth so that the much longer times required to produce substantially thicker oxide would be impractical. Increasing the oxidation temperature was also considered impractical due to the deterioration of the silica ware. In order to produce thicker oxides the steam oxidation therefore was used.

Water vapour is 500 times more soluble in silicon oxide than oxygen and consequently, even though it has a slightly lower diffusivity, this means that the flux of oxidant molecules through the oxide is considerably increased. In the parabolic region of growth this considerably increases the rate of formation of oxide. Steam oxidations typically produced layers 1 micron thick after 2 hours at 1100°C. provided the temperature of the water in the steam generator was sufficient to produce a pressure of one atmosphere of water vapour in the vicinity of the silicon.

2.3.2 Physical Assessment of Oxide layers

The thickness of the oxide films was measured using a Watson 16 mm interference objective attached to a Beck Epimax microscope. Interference occurs between the monochromatic light reflected from the specimen and from a reference plane in the objective ~~recombines~~. The fringes are displaced if there is a step on the surface and knowing the wavelength of the incident light the height of the step is simply calculated.

In the present work the step was produced by etching away the oxide so that its thickness could be measured. Masking to produce a step was done using Apiezon 'W' which was dissolved in trichloroethylene and painted over the desired oxide area. Ideally a step is required which is wide enough to allow the displacement of the fringes to be counted easily. Such a step can be produced by using 48% HF to etch the oxide because this etch undercuts the mask sufficiently to produce a graded step. Part of the surface was

metallised with aluminium to give a reflecting layer. Fig 2 - 3 shows the fringe pattern reflected from a partially metallised oxide step. In the unmetallised region displacement of the fringes is due to an optical path difference between light reflected from the bare silicon and from the silicon underneath the oxide. The direction of displacement of the fringes in this case is opposite to that with the metallised step.

If it is assumed that a complete 180° phase change occurs at the silicon/oxide interface then, knowing the refractive index of the silica film (1.46 for a thermal oxide) (Ref 2 - 6), the thickness of the oxide film can be calculated from fringe displacements across an unmetallised step. Weak reflections also occur from the surface of the oxide. These can be seen in Fig 2 - 3 as faint continuations of the fringes across the metallised/unmetallised boundary at the step. Metallisation was not necessary when these fringes were prominent. Incidentally the continuous nature of these fringes indicates that any phase change that may accompany reflection at the oxide and silicon surfaces is not important. A disadvantage of using an interference microscope for measuring film thickness by this method is that fringes are formed from only a few reflected light beams, ~~because of the small aperture of the objective lens.~~ This produces the rather broad fringes shown in Fig 2 - 3, unlike the finer fringes which can be obtained using the more tedious multiple beam interferometry. However, fringe displacements were usually estimated by eye to one tenth of the fringe spacing (about 250 Å using sodium light)

and by photographing the interference patterns this was reduced to about 100Å.

The interference colours of oxides can also be used as a guide to their thickness, and colour charts have been produced for doing this (Ref 2 -7). A uniform hue under times 100 magnification usually means that the oxide has no large discontinuities. However, small pin-holes and flaws, not usually discernable by direct observation can be enlarged by chlorine gas etching (Ref 2 - 8). To do this the oxidised silicon chips were heated to 900°C, in the furnace shown in Fig 1 - 5 using a modified silica furnace tube which allowed a small flow of chlorine gas from a Lectern bottle to be introduced into the main ~~argon~~ gas flow which was argon in this case. The chlorine gas was bubbled through NaOH solution at the exit vent. Any pore or microcrack in the protective oxide allows chlorine to penetrate to the silicon beneath, where it reacts to produce a visible mark on the surface. Typical pinhole densities obtained for oxides of 1000 Å thickness were 100 cm^{-2}. The thicker oxides used for masking against diffusion had very few pin-holes or none at all.

Chlorine etching was also used to detach oxide films from their silicon substrates for examination by transmission electron microsopy. An electron micrograph and a transmission diffractograph of a detached film are given in Fig 5-3a showing the uniformity of the films and their amorphous nature. This part of the work is described in more detail in Sec 5.3.2.

Chapter 4 describes the conversion of oxide into luminescent willemite. Further measurements of the oxide properties made for comparison with willemite, will be described in Chap.5,6 & 7.

2.4 Impurity diffusion in silicon

2.4.1 Apparatus and techniques

Planar p-n junctions were produced by the diffusion of phosphorus into p-type silicon using a solid source of 99.9999% P_2O_5. The apparatus was built in the department and is shown in Fig 2-6a. The diffusion furnace is similar to that used for oxidation with the addition of a separate source furnace, the temperature of which could be controlled to produce the temperature profile shown in Fig 2-6b. A monotonic increase in temperature from the source to the diffusion furnace is required to prevent condensation of the P_2O_5 vapour at the front of the furnace. Quartz wool is used to prevent P_2O_5 dust being picked up in the carrier gas and contaminating the silicon surface. The P_2O_5 is vapourised in the source furnace and is carried to the silicon by a suitable carrier gas where it reacts with the silicon surface to produce a glassy phosphorus-rich oxide layer on the silicon. Predepositions of this type were carried out with the silicon at 900°C and the source temperature about 200°C. The P_2O_5 was then removed from the source furnace and the temperature of the diffusion region increased to about 1100°C for the main diffusion of the phosphorus into the silicon. Predepositions and drive-in diffusions were usually carried out in an oxygen atmosphere to enhance the formation of an oxide on the silicon so as to prevent out-diffusion of phosphorus during the high temperature processing. The gas was dried by passing

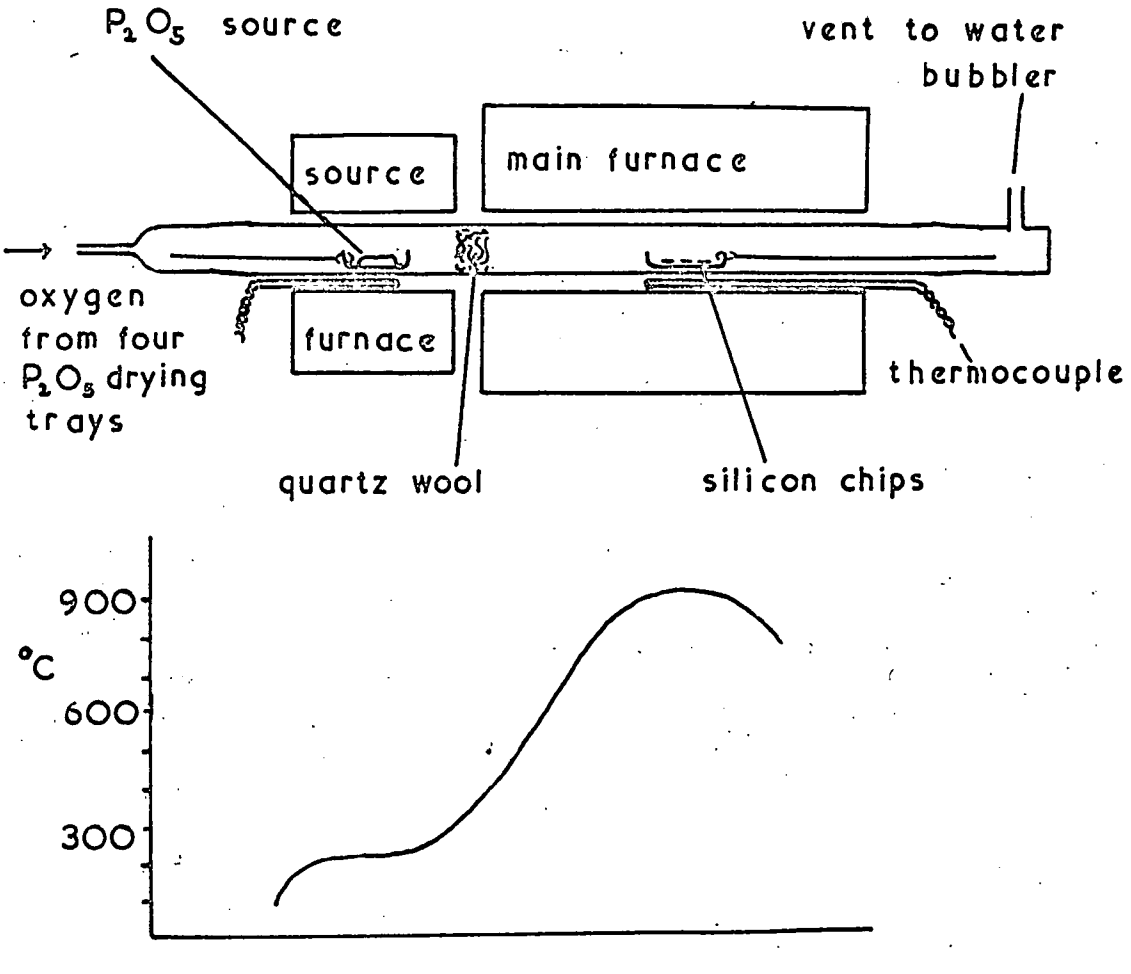


Fig 2-6 Diffusion furnace and its temperature profile at predeposion.

it over trays of P_2O_5 upstream of the furnace.

The final surface concentration of phosphorus on the silicon depends on the vapour pressure of the P_2O_5 as controlled by the source temperature. However the vapour pressure of the P_2O_5 is considerably affected by the presence of water vapour which it absorbs very quickly. Special precautions were therefore taken to prevent water pick-up in the transfer of the P_2O_5 from the sealed ampoules to the furnace.

The disadvantage of using P_2O_5 as a source for phosphorus diffusion is that it is very difficult to obtain low surface concentrations reproducibly and the surface concentrations are usually greater than 10^{20} cm^{-3} after diffusion. Modern diffusion systems therefore use liquid $POCl_3$ or phosphine gas which allow a greater control of the impurity distribution. However, it was decided, at least initially, to use the solid P_2O_5 source because of its less toxic nature and its convenience.

Oxide masking was used to control the area of the diffused regions. A thickness of about 0.5 microns is required to prevent penetration of the diffusing phosphorus for a typical diffusion of one hour at 1100°C (Ref 2-9). Photoresist techniques were used to define the oxide masks, although for simple geometries it was often simpler to use Apiezon 'W' dissolved in trichloroethylene. The oxides were etched with a standard HF etch buffered with ammonium difluoride to reduce undercutting of the masks. The etch was made by

dissolving 8 parts by weight of NH_4FHF into 15 parts of water and then mixing 4 parts by volume of this solution with 1 part by volume of 48% HF.

2.4.2 Assessment of diffused layers

The junction depths were measured by bevelling the sample, delineating the p-n regions with a selective electrochemical plating solution and then using interferometry. The bevelling was approximately 5° and was produced using a specially angled jig and successive polishes of 3 micron, 1 micron and 1/4 micron diamond paste. The electroplating solution consisted of 20 gm of copper sulphate dissolved in 100 gm of water to which 1 ml of 48% HF had been added (Ref 2 - 10). Plating occurs first on the n-type region of the bevel. When the junction was sufficiently delineated the solution was flushed away. The interferometer described in sec 2.3.2 was used and the depth of the junction measured by counting the fringes down the bevel along an extension of a fringe on the unbevelled region of the chip.

The surface concentrations of diffused layers in silicon were measured using the four-point probe method (ref 2 - 11) The probe apparatus was similar to that of Kennedy (ref 2 - 12). Probes were made of 0.008 inch tungsten wire electrolytically pointed (ref 2 - 13) and equispaced 0.010 inch apart. The pressure on the probes could be adjusted using a threaded weight and the specimen was held in a small vacuum chuck. Current, I, through the outer probes produces a voltage drop V across the inner probes. A potentiometric

measurement of V ensures that contact effects are eliminated. This was done using a Pye Portable potentiometer with an external Tinsley galvonometer having a sensitivity of 2.4 mm/nanamp. The values of V and I together with a knowledge of the junction depth can then be used to obtain the surface concentration of the impurity (ref 2 - 14, 2 - 15).

In the present work typical surface concentrations for a diffused n-region 4 micron deep were about $10^{20} cm^{-2}$ but this could be reduced to less than 10^{18} by oxidising and stripping.

The resistivity of initial silicon slices was also measured using the four point probe method in the range 0.1 - 10 ohm-cm and found to be as expected.

2.4.3 Contacts

Permanent electrical contact was made to the silicon devices by bonding on thin aluminium or gold wire. For p-type silicon 0.002 inch thick aluminium 1% silicon alloy wire was ultrasonically bonded to the bare silicon surface using a Kerry's 20 watt Fine Wire Bonder. To ensure a good ohmic contact a brief bake in air at a temperature below the silicon/aluminium eutectic of 577°C was required. Aluminium was also used to make contact to heavily doped n-type regions. Although Al^{+++} is an acceptor ion it is thought that an effective ohmic contact is formed due to the preferential solubility of Al^{+++} over P^{---} in silicon producing a tunnelling $p^+ - n^+$ junction.

Ohmic contacts to moderately doped n-type silicon

are usually made by using gold wire doped with about $\frac{1}{2}\%$ antimony. However, pure gold wire 0.002 inch thick when thermocompression bonded to a bare silicon surface could be made ohmic by electroforming. The wire was first threaded through fine capillary tubing and a small length melted in a gas-oxygen flame to produce an annealed ball of gold at the end of the wire. The silicon was heated to just below the gold/silicon eutectic at 370°C and the ball pressed on to a clean oxide-free surface with sufficient force to produce about 50% deformation of the ball. The bonds were simply electroformed by passing a current through them which was sufficiently large to produce some localised heating and diffusion at the gold/silicon interface.

2.5 Surface properties of p-n junctions

2.5.1 Introduction

The production of simple planar p-n junctions in silicon was undertaken because it is envisaged that in devices electroluminescent films will be produced on silicon which would contain p and n-type regions. The solid state circuitry produced within the substrate might be used either for addressing the display device or for storage. It may also be possible to use the high field regions associated with the p-n junctions themselves to produce the electroluminescence. Because of this it was thought necessary to consider the effect of dielectric films on the properties of p-n junctions beneath them. Section 2.5.2 gives a brief review of this topic.

2.5.2 Space charge regions in the silicon surface

The surface properties of p-n junctions can be explained in terms of the formation of space charge regions in the silicon surface adjacent to the p-n junction. The theory of these space charge layers will be described in more detail in sec 6 . 2 where their capacitance in MOS structures is used to give information about dielectric layers of oxide or willemite on silicon. Physically, the space charge region is due to the repulsion of majority carriers in the silicon by surface fields. On a p-type substrate, depletion, and finally inversion, of the surface is achieved by the application of an increasingly positive potential. Where a bulk junction comes to the surface the field-induced junction due to the inversion layer is in parallel with the true metallurgical junction and consequently both the reverse leakage current (I_r) and the breakdown voltage (V_b) can be dominated by the surface effects.

i) Leakage current

Fig 2 - 7 is a schematic diagram showing the effects that a field induced junction has on the leakage current of a n^+p diode. A negative surface potential (fig 2 - 7a) prevents the formation of any space charge regions in the surface of the p region and the reverse current is governed solely by the characteristics of the true metallurgical junction. For silicon at room temperature this means that I_r is constant depending mainly on the number of generation-recombination centres in the depletion region.

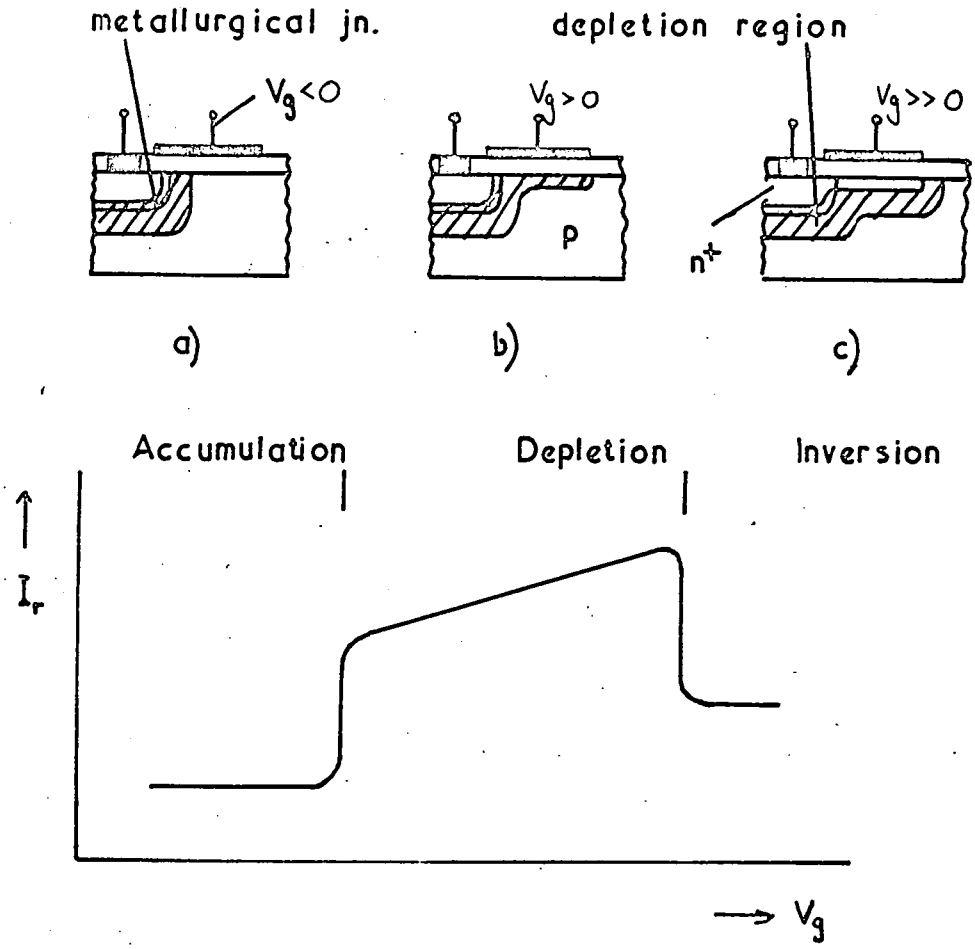


Fig 2-7 The effect of variation of surface potential on the reverse current of a n^+p diode.

With a positive surface potential (fig 1 - 7b) a depletion layer forms that includes the defect-laden surface and this not only increases the number of centres available to generate I_r but also results in I_r being dependent on the surface potential. Increasing the surface potential further however (fig 1 - 7c) causes the depletion layer to leave the surface, as the inversion region forms, and this decreases the concentration of defects in the depletion layer which produces a corresponding decrease in I_r.

ii) Junction breakdown

Breakdown of a p-n junction usually occurs by an avalanche multiplication process in the depletion regions and is therefore determined ideally by the maximum field which exists there. This maximum field is dependant on the resistivities of the p and n regions and the gradient of the impurity concentration. It is also dependant on the degree of field intensification that occurs at the corners of the junction. Space charge regions in the silicon surface adjacent to the junction alter the field distribution around a planar junction and thus effect V_b. However if field induced junctions exist at the surface they have their own breakdown voltage which is usually less than that of the metallurgical junction. Breakdown often occurs at defects in the depletion region. This is more likely with induced junctions because of the proximity of the surface. At breakdown, current flows along the inversion layer or 'channel' to the diffused regions and saturates at a value depending

on the conductance of the inversion layer. On increasing reverse bias after saturation has been reached the true breakdown at the metallurgical junction is usually reached. If, however, several different breakdowns occur the reverse characteristics are usually stepped due to the channel current saturating after each event. Metal ions can also lead to a similar 'softening' of the reverse characteristics. In this case it is thought that local breakdown occurs at metal precipitates within the depletion layer (Ref 2-17). These metal impurities can usually be removed and the characteristics 'hardened' by a process similar to the predeposition of phosphorus.

2.5.3 Surface films on p-n junctions

The thermal oxidation of silicon is limited by the rate of diffusion of the oxidising species through the growing oxide and consequently excess silicon must exist in the oxide near the silicon surface. This produces an effective fixed positive charge (Q_{ss}) in the oxide. The positive surface potential produces depletion regions in p-type silicon beneath. Annealing the silicon in oxygen reduces Q_{ss} by allowing oxygen to diffuse through the oxide and react with the excess silicon. The leakage current of the junction and its breakdown voltage can thus be controlled by careful annealing processes which tailor the surface potential to produce the required characteristics.

Current-voltage measurements were made on finished diodes using a Tetronix Curve Tracer Type 575. The characteristics of a

typical n^+p diode produced by the phosphorus diffusion process described in Sec 2.4.1 are shown in Fig 2-8a. The junction area was typically about 0.05 cm^2. Similar diodes to these , measured with the passivating 'phosphorus glass' still on the silicon surface, were usually 'hard' probably due to the gettering action of the phosphorus treatment. Removal of the phosphorus glass produced a very unstable characteristic (Fig 2-8b) and on reoxidising the surface the diode was still soft due to the formation of space charge regions (Fig 2-8c). The effect of annealing in oxygen at 500°C for 10min is shown in Fig 2-8d, the decrease in I_r being due to the removal of Q_{ss} at the silicon surface.

Hydrogen is extremely mobile in silica and produces an effective increase in Q_{ss} probably due to the formation of OH^- ions from O^{--} in the silica glass (Ref 2-18). Even a brief exposure to hydrogen has an extremely deleterious effect on the diode characteristics as shown in Fig 2-8e. The subsequent removal of Q_{ss} by further annealing in oxygen to regain the original 'hard' characteristics is shown in Fig 2-8f.

The formation of willemite films over p-n junctions and their effect on the diode characteristics is discussed in Sec 8.2.

2.5.4 The gate-controlled diode

A gate -controlled diode structure was made to illustrate the effect of surface potential on the reverse characteristics. This was constructed by evaporating an aluminium ring electrode on the

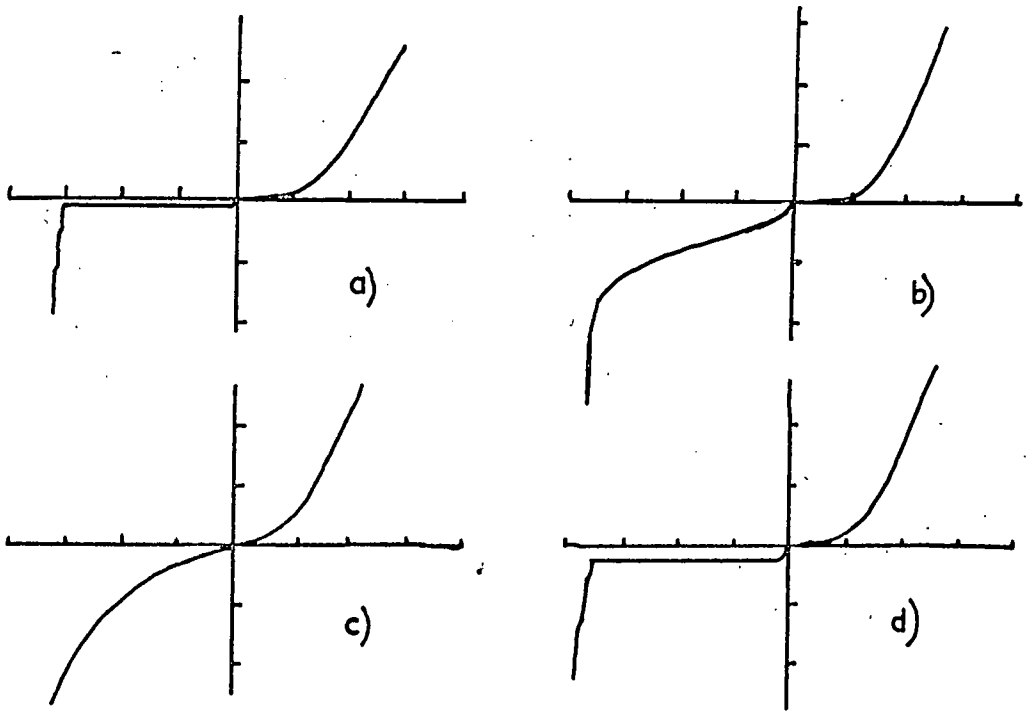
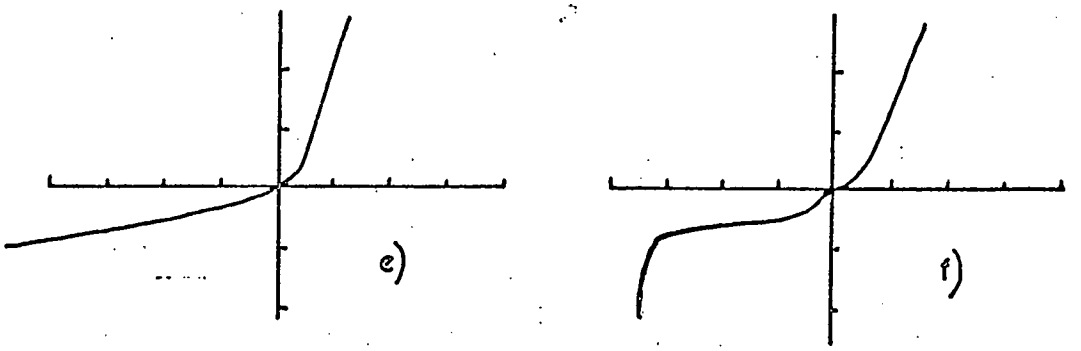


Fig 2-8 The I-V characteristics of a diffused n^+p diode
forward 0·1 mamp 0·2 V / div., rev. 0·1 mamp 20V / div.
a) untreated after diffusion b) after removal of the phosphorus glass c) after reoxidising at 1100°C in O_2. d) after a further bake in O_2 at 500°C for 10 min.



e) after 2 min in H_2 at 500°C. 1 mamp, 1 V/div and rev. 1 mamp, 10V/div.
f) after a further 10 min in O_2 at 500°C. 1 mamp, 1 V/div : 1 mamp, 20V/div.

oxide above the p-n boundary to form the gate (Fig 2-10). The gate voltage V_g could be varied with respect to the p-region and the reverse characteristics were photographed for several gate voltages (Fig 2-11). Negative values of V_g counteract the positive space charge trapped in the oxide and give decreased values of I_r whereas positive V_g increases I_r due to the formation of field induced junctions. On increasing the positive potential on the gate it should have been possible to observe inversion of the surface and the corresponding drop in I_r as the depletion region leaves the silicon surface. The maximum possible gate voltage was however limited to the breakdown voltage of the oxide, which was about 25 volts in this case.

2.6 Conclusions

The following has been achieved :-

i) cutting and etching the silicon to produce chips suitable as substrates.

ii) controlled oxidation and oxide thickness measurements.

iii) contact fabrication.

iv) n-type diffusion to give the expected diode characteristics.

All this agreed with expectations and therefore demonstrated an adequate control of the technology. The use of the above techniques on a routine basis for the study of luminescent films on silicon was now possible.

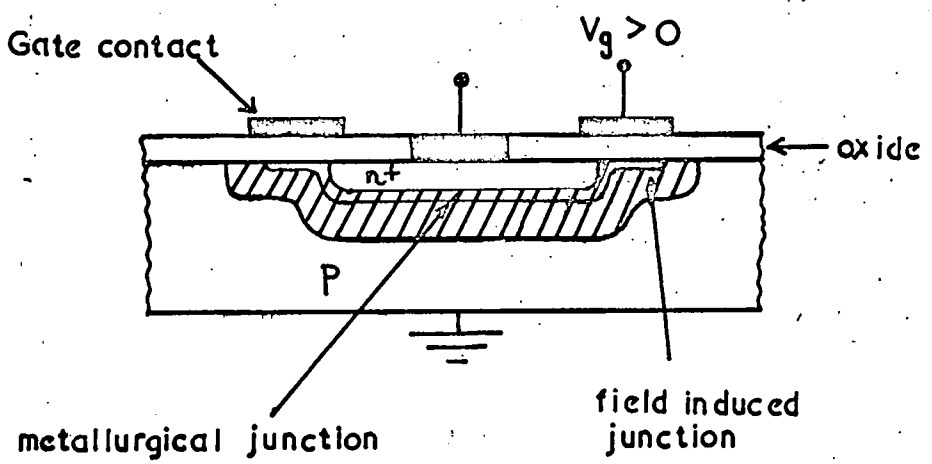


Fig 2-10 Gate-controlled diode structure

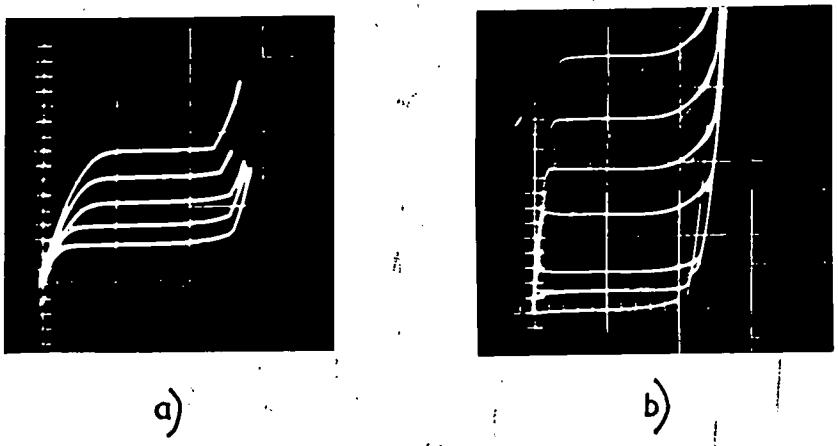


Fig 2-11 Reverse characteristics of a gate-controlled diode. Increasing increments of reverse current correspond to gate voltages 0, +6, +12, +18, +24 V for a) 1 mamp/vert div. 20 V/horiz. div. and -2·5, -5, -7·5, -10, -15, -20 V for b) 0·1 mamp/vert.div.

References for Chapter Two

2-1 E. Sirtl and A. Adler, Z. Metallk. 52, 529 (1961)

2-2 F.G. Allen, J. Eisinger, H.D. Hagstrum and J.J. Law, J. Appl. Phys. 30, 1563 (1959)

2-3 J.R. Ligenza and W.G. Spitzer, Phys. Chem. Solids, 14, 131 (1960)

2-4 E.A. Benjamini, E.F. Duffek, C.A. Mylroie and F. Schulenburg, Electrochemical Society Meeting, New York (1963)

2-5 B.E. Deal and A.S. Grove, J. Appl. Phys. 36, 3770 (1965)

2-6 B. E. Deal, J. Electrochem. Soc. 110, 527-33 (1963)

2-7 W.A.Pliskin and E.E. Conrad, IBM J. Res. Develop. 8, 43-51 (1964)

2-8 S. Wing, R.E. Morrison and J.E. Sandor, J. Electrochem. Soc 109, 221-6 (1962)

2-9 C.T. Sah, H. Sello and D.A. Tremere, J. Phys. Chem. Solids 11, 288 (1959)

2-10 D.R. Turner, J. Electrochem. Soc. 106, 701-5 (1959)

2-11 L.B. Valdes, Proc. IRE. 42, 420 (1954)

2-12 J.K. Kennedy, Rev. Sci. Instrum. 33, 773-5 (1962)

2-13 H.C. Torrey and C.A. Whitmer, Crystal Rectifiers p318 New York (1948)

2-14 F.M. Smits, Bell System Tech. J. 37, 711 (1958)

2-15 J.C. Irvin, Bell System Tech. J. 41, 387 (1962)

2-16 A. Goetzberger and W. Schockley, J. Appl. Phys. 31, 1821 (1960)

2-17 K.H. Zaininger and G. Warfield, Proc. LEEE. 52, 972-3 (1964)

CHAPTER THREE

ZINC SULPHIDE FILMS ON SILICON

3.1 Introduction

This chapter describes the deposition and properties of ZnS films on silicon. The ultimate aim of this work was the production of an electroluminescent film for applications to display devices. The devices envisaged would contain their own addressing and possibly storage capabilities in the form of integrated circuitry in the silicon substrate with the light output being controlled by the potential of the underlying silicon. The general considerations of electroluminescence relevant to thin films are discussed in Sec 1.23.

These considerations show that high-field carrier injection at the Si-ZnS interface and at the top electrode should produce conditions which were favourable to the d.c excitation of the luminescent centres in the film. A phosphor film was therefore required that could sustain very high fields and the use of epitaxial film of ZnS on silicon seemed to be a reasonable first objective. The interplanar distances of silicon and cubic beta-ZnS agree to within 0.5% and the atomic arrangement in the basal plane of hexagonal alpha ZnS also fits closely with that of the (111) plane in silicon. It therefore seemed likely that epitaxy of ZnS on single crystals of silicon would be possible.

The first part of this chapter describes the experiments which were carried out and the apparatus used in the

preliminary work designed to assess the feasibility of producing the required display device using ZnS phosphor films. The results of this work and the assessment of the films that were produced are followed in Sec 3.7 by a discussion of the problems involved.

3.2 Evaporated films of ZnS on silicon

3.2.1 The evaporator system

The evaporator used in this work was built in the conventional Department to a design. It consisted of a 12 inch bell-jar evacuated by an Edwards Type E03 oil diffusion pump used in conjunction with a liquid nitrogen trap. Pressures were measured using an Edwards Pirani gauge for the backing-pump pressure range and an Edwards ionisation gauge for lower pressures. The system was capable of reaching an ultimate pressure of about 10^{-6} torr.

The electrical power units for the evaporator comprised two 30amp a .c. supplies, a 3kV 30mamp H T supply for ion bombardment cleaning and sputtering, and a 3 A electron gun filament supply capable of operating at -2kV from earth. Ionisation gauge degassing facilities were also incorporated into the system.

Up to four temperatures could be measured in the bell-jar using Chromel-Alumel thermocouples via an 8-way instrumentation lead-through. A Pye low resistance multi-way switch was used to indicate each temperature on either a Honeywell Temperature Indicator (Type05-655) or a potentiometer as required. The Honeywell indicator could also be used as an on-off temperature controller for a resistance heater. A rotary lead-through enabled a movable shutter to be used within the bell-jar.

3.2.2 Evaporation sources

ZnS sublimes at about 1050°C in vacuum. The sublimation is usually accompanied by a large amount of decrepitation. This is due not only to the ZnS vapour being produced from within the bulk of the charge but also because commercial luminescent grade ZnS can contain volatile impurities, usually chlorides, up to 0.1% by weight. Luminescent grade ZnS powder was obtained from Levy West & Co. Ltd. for the evaporations. This was further purified for some evaporations using the flow system technique described in Sec 3.3. and the small crystals so produced were used as the evaporation charge.

Initially a pepperpot source was used for evaporating ZnS. This consisted of a standard Edwards molybdenum boat with a perforated lid of molybdenum foil fitted over the centre dimple. The temperature of the source was measured by a Chromel-Alumel thermocouple spot-welded to the underside of the dimple. These sources were simple to construct and produced uniform films with relatively little loss of heat to the surroundings. However, ZnS reacts slowly with all refractory metals at high temperatures so that these molybdenum sources had a very limited life and they also probably produced contaminated films. To prevent this a quartz crucible source was developed.

The crucible consisted basically of two concentric silica tubes (Fig 3-1a). The inner tube was sealed to contain the thermocouple and it was used as a support for the crucible. A quartz

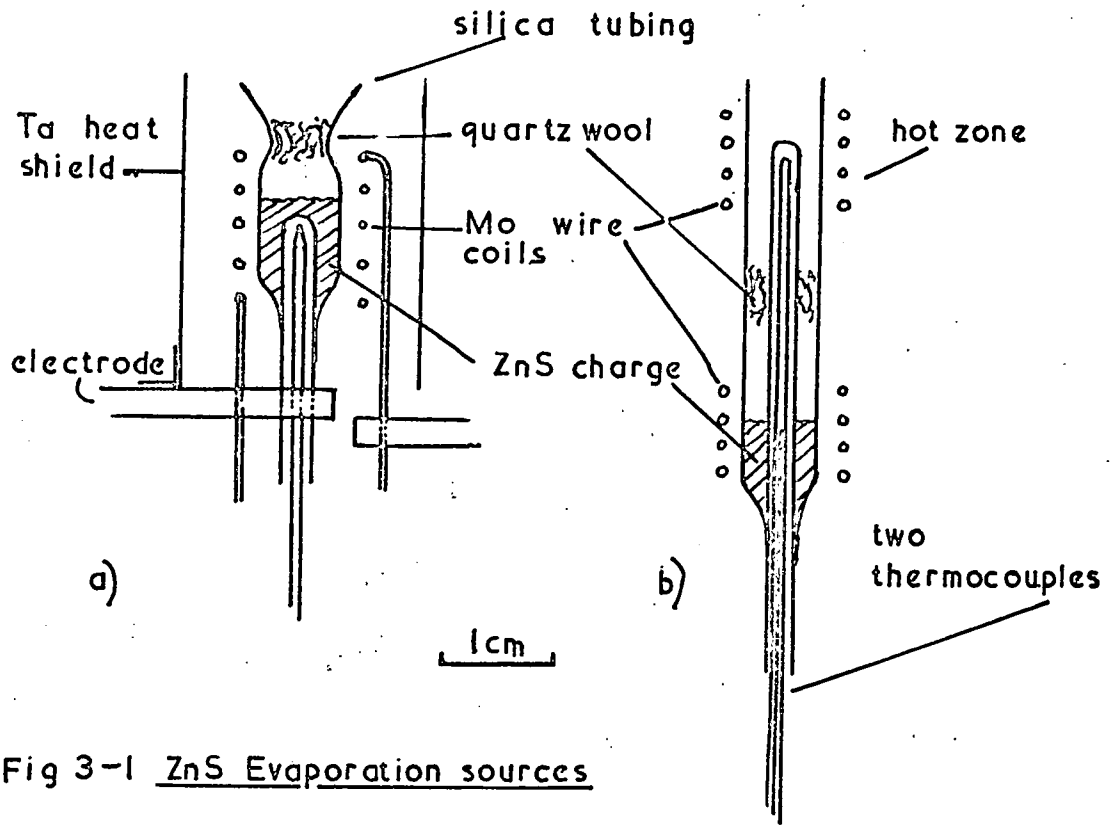


Fig 3-1 ZnS Evaporation sources

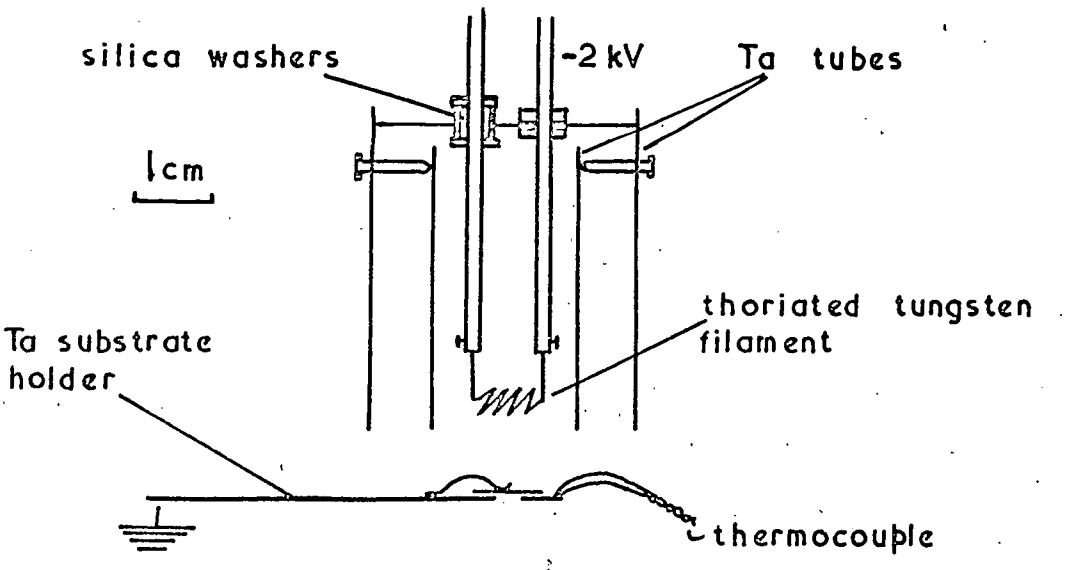


Fig 3-2 Electron beam substrate heater

wool plug was inserted into the belled mouth of the outer tube and prevented excessive spitting of the charge. The ZnS was sublimed using the radiant heat from a coil of molybdenum wire surrounding the crucible. Tantalum tubing was used as a heat shield around the molybdenum coil.

A later version of the quartz source was slightly different. The tube was longer and the upper part of the crucible was heated by a separate molybdenum coil to provide a second hot-zone the temperature of which was measured by a second thermocouple inserted up the centre tube as shown in Fig 3-16. This crucible was used in an adaptation of the 'hot-wall bell-jar method' of evaporation (Ref 3-1) in which the hot wall of the extended silica tube prevented excessive condensation of the ZnS, or of one of its dissociation products, on to the colder parts of the surroundings. This technique was used in an effort to improve the reproducibility of the evaporated films and to improve their stoichiometry.

3.2.3 Substrate heaters

In order to grow crystalline films and to get good adhesion it is necessary to condense them on to a heated substrate. One effect of heating the substrate is to change its state of oxidation. Silicon forms an oxide on its surface very rapidly even in a vacuum of 10^{-6} torr (Ref 3-2) and it was hoped that this could be thermally etched away by heating to about 1100°C before cooling the silicon in the presence of ZnS vapour to form the phosphor film. To obtain these relatively high temperatures without contamination from a heater

strip an electron-beam heater was developed based on the design of Wales (Ref 3-4) and this is shown schematically in Fig 3-2 The electron gun filament operated at a negative potential with respect to earth. It was made of thoriated tungsten wire shaped in the form of a flat zig-zag to produce a wide beam of electrons. The electron beam was focussed on to the earthed tantalum substrate holder 2 cm beneath the filament by a tantalum tube surrounding the filament and at the same potential as it. To produce the maximum emission for a particular filament temperature the filament was 'flashed' before use. The gun was operated in the temperature limited region to give a typical emission current of 10mamps at 2 kV and this produced substrate temperatures up to 1100°C.

The substrate temperature was measured by a Chromel-Alumel thermocouple spot-welded to the tantalum substrate holder next to the silicon sample. The silicon was clipped to the top surface of the holder which then acted as a mask for the evaporations from below. It is impossible to use Chromel-Alumel thermocouples in contact with silicon at temperatures above about 800°C due to the chemical reaction.

A resistive heater was used for heating the substrate to lower temperatures where contamination was not so likely. It consisted of a 0.004 inch thick molybdenum strip which was directly heated. The silicon substrates were clipped over a hole in a smaller piece of molybdenum which acted as an evaporation mask and also carried the thermocouples. This holder was then fitted over a hole

in the heater strip, through which the evaporation took place. Substrate temperatures of up to 700°C could be obtained with this heater.

3.2.4 Experimental

The evaporation experiments consisted mainly of determining the effect of substrate temperature on the quality of the evaporated films. Other parameters which effect film growth are background pressure, rate of evaporation, the condition of the substrate and source material. In investigating the effect of substrate temperature the other variables were kept as constant as possible by using a standard substrate treatment and by keeping the rate of evaporation constant. Ideally a film thickness monitor was required but, as this was not available, it was decided to use the source current to control the evaporation rate and to use a constant quantity of source material for each experiment. It was realised at the time that this was not a reliable way of controlling the evaporation rate due to uncontrollable properties of the evaporant such as the amount of charge actually in contact with the crucible and the undefined evaporating area. However, for the initial work at least, it was decided that the simple method of control was probably sufficient.

Silicon substrates were etched as described in Sec 2.2. Electronic grade ZnS powder obtained from Levy-West Ltd, was used as the source for most of the evaporations. The alternative charge for some of the evaporations was the finely ground flow crystals (Sec 3.3.) Extensive degassing of the source and substrate was carried out prior

to the evaporation by heating the source slowly to just below its sublimation point and the substrate to a temperature higher than that to be used during the evaporation. Degassing was continued until the pressure had dropped to below 10^{-5} torr. During degassing a movable shutter was interposed in the 2.5 cm gap between the substrate and the source. The evaporated areas were typically about 5 mm^{*2} on silicon substrates which were usually about 0.5 cm by 1 cm. A typical evaporation lasting 15 min with a source temperature of 1050°C produced film thicknesses ranging from 2000 Å for substrate temperatures of 400°C to one micron for substrates at about 100°C. The deposition was terminated by interposing the shutter between the substrate and the source and the substrate was then cooled slowly to room temperature in about 15 min. Evaporations were in general non-reproducible and the films were often discoloured and non-uniform. General conclusions can however be drawn from the results and these are discussed in Sec 3.7. together with the possible reasons for the lack of reproducibility.

3.3 Vapour flow deposition of ZnS on to silicon

Single crystals of ZnS can be grown by sublimation using a flow technique (Ref 3-5). An inert carrier gas passed over subliming ZnS powder carries the vapour to cooler parts of the system where crystal growth takes place. This method was used for obtaining flow crystals for use in later experiments. It was also decided to use this method to try to grow films of ZnS on silicon.

The furnace used was similar to the oxidation furnace described in Sec 2.3.1 with the addition of silica liners inside the furnace tube to enable the crystalline growths to be removed easily. Argon was used as the carrier gas and the outlet fumes were bubbled through hydrochloric acid and then water at the exit vent.

Several flow runs were first carried out to determine the position of the crystalline growth along the furnace tube for a fixed source temperature and gas flow. The results are shown in Fig 3-4. The ZnS charge was positioned slightly upstream of the point of maximum temperature in the furnace to assist crystal formation. Crystalline growths appeared first at about 1100^{o}C and consisted of small needles and platelets with the occasional large crystal up to 40 mm^{2} in area. Moving downstream the crystallite size decreased and the crystals became increasingly yellow and finally dark grey indicating non-stoichemetric excesses of sulphur and zinc respectively.

Growth on silicon was investigated by placing several freshly etched silicon chips in small silica holders into the flow furnace equispaced along the crystal growth region (Fig 3-4).

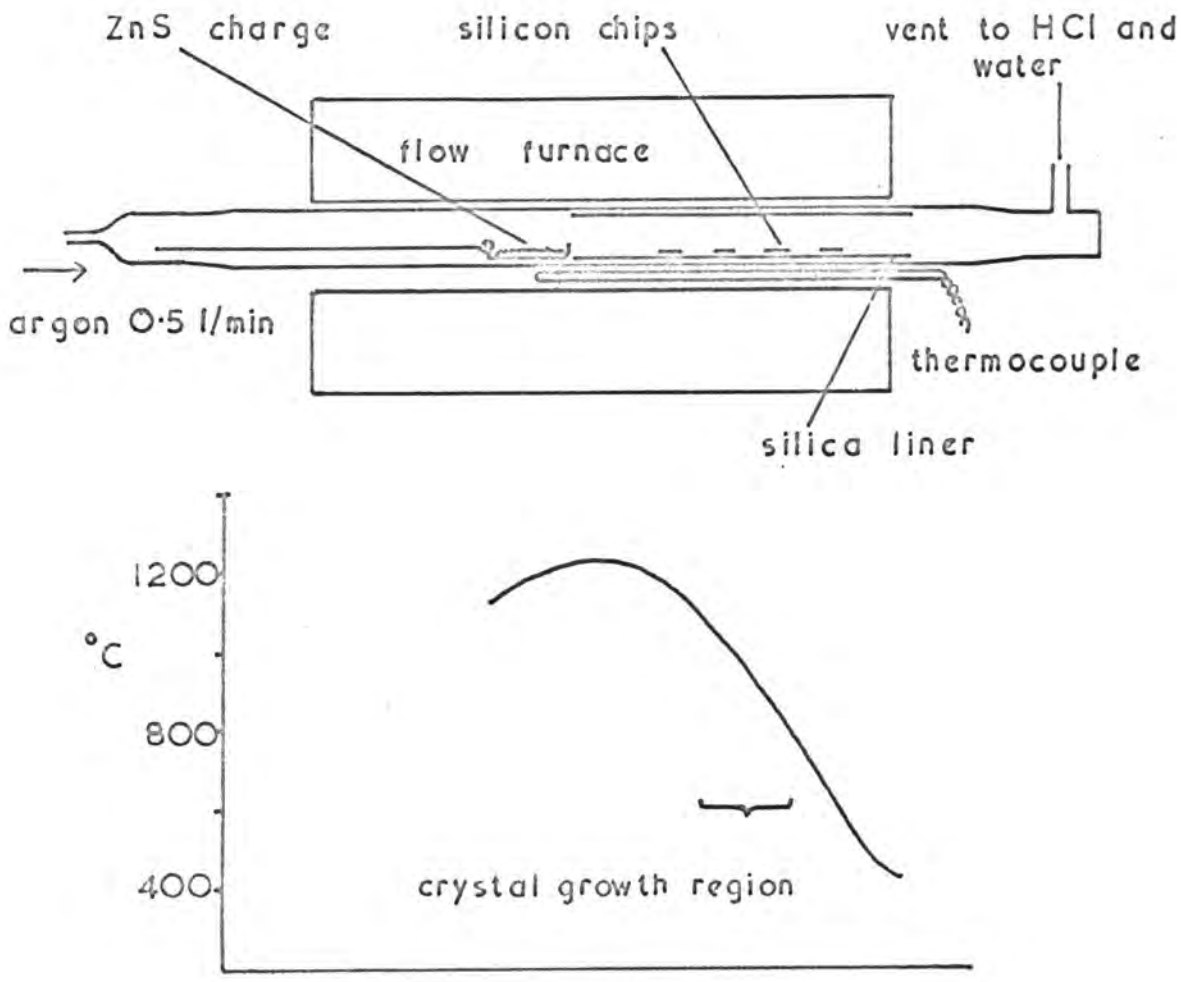


Fig 3-4 ZnS flow furnace and temperature profile

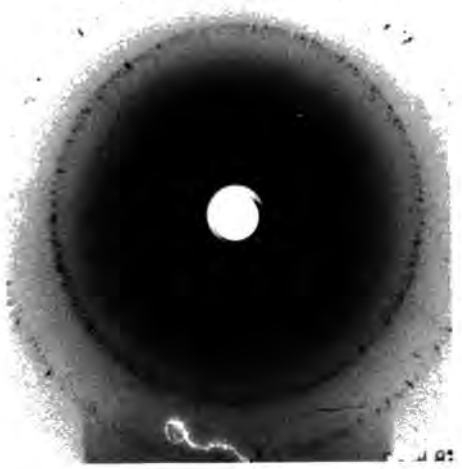


Fig 3-5 Back-reflection X-ray photograph of ZnS film on silicon formed in growth region above.

A typical flow run lasted 12 hrs with a source temperature of 1170°C and gas flow of $0.5 l\,min^{-1}$. The results of this work are described in Sec 3.6.2.

3.4 Chemical deposition of ZnS on to silicon

The third method of growing ZnS on silicon that was tried was homogeneous precipitation. This involves the formation of the sulphide from a metal salt solution at a controlled rate governed by the dissociation rate of a sulphur containing compound in the solution. It has been used for the growth of epitaxial films of PbS on germanium. In this the sulphide is produced by precipitation from an aqueous alkaline solution of a lead salt (Ref 3-6). Homogeneous precipitation has also been used in the final purification process for CdS. It was decided to see if films of ZnS could be deposited on to silicon using an adaptation of this process.

A solution of 20ml ~~of 20ml~~ of 0.1 M zinc nitrate and 40 ml of 0.5 M sodium hydroxide was made up. If this solution is sufficiently alkaline the Zn^{++} ions do not precipitate as the hydroxide but form the soluble zincate ion. In order to maintain a clear solution the zinc nitrate was therefore added slowly to the alkali particularly for the last few mls which were added very slowly with much stirring. The solution was then added to a second solution consisting of 20 mls of N/1 thiourea and the mixture was heated to about 80°C. A freshly etched chip of silicon was suspended from a P.T.F.E. clip face downwards in the liquid. After about half an hour a fine white powder of ZnS had deposited on the surface of the silicon.

These deposits were however very non-uniform and had poor adhesion to the substrate, probably due to the oxidation of the silicon in the water. Also the ZnS seemed to have no preference to forming on surfaces unlike PbS which produces mirror-like layers on any surface in contact with the precipitating solution.

Because of the lack of success this work was discontinued. However, more recently Nagao and Watanabe (Ref 3-7) have published a method for the deposition of CdS using homogeneous precipitation from a mixutre of $Cd(NO_3)_2$, NaOH, NH_4NO_3 and thiourea, the ammonium salt apparently being necessary to produce the soluble cadmium tetramine ions. The CdS films produced had very good adhesion to their glass substrates and it seems possible that this method might also be adopted for the production of ZnS films. However, because of the necessity for using a silicon substrate in the present work and the likelyhood of it oxidizing in any aqueous solution it was decided not to attempt this.

3.5 Post-evaporation treatment

In order to obtain ZnS films which have desirable luminescent properties it is necessary to incorporate the required electrically and optically active centres into the lattice. This process is called activation and it usually necessitates an increase in crystallinity of the film to locate the centres in the correct lattice sites. ZnS powder phosphors always condense in an unactivated form when they are evaporated, although on heated substrates they are often highly crystalline. This is due to the extreme difficulty of

obtaining the required control over the evaporation and condensation rates of the host, the activator and the co-activator simultaneously. Some post-evaporation treatment is therefore necessary to obtain highly luminescent films. The methods used to date are listed below (Ref 3-12).

i) Co-evaporation of the activator and the host lattice and subsequent heat treatment.

ii) Gilles and Van Cakenberghe's method (Ref 3-8) which involves the evaporation of the activator, usually a metal, on to the host film. In a subsequent heat treatment the activator diffuses into the film and also induces crystallisation.

iii) Embedding techniques which utilise the solid state diffusion of the activator at high temperatures from a phosphor powder in contact with the film.

iv) The organometallic technique (Ref 3-9) which relies on diffusion of the activator from a hot oil bath containing the activator in the form of organometallic ions.

Elsewhere ZnS films evaporated on to glass sustrates have been made luminescent by all the methods described above. However, in this work the necessity for using silicon substrates which may contain electrically active regions restricts the activation processes which can be used. In order to appreciate the problem more fully some of the methods for activating evaporated films on silicon substrates

were attempted using the Mn^{++} ion as the activator. Manganese is a particularly suitable activator for d.c. EL in the II-VI compounds because the luminescent transitions are internal to the ion itself and are thus relatively independent of the transport properties of the surrounding host lattice. Also manganese ions are divalent like the host metal ~~ions~~ so that charge compensation effects are eliminated. In theory, manganese ions should therefore be easily incorporated into the host lattice.

The simplest method was to activate the ZnS before evaporation. Electronic grade ZnS powder was activated with 1% by weight of Mn by forming a slurry with manganous chloride solution, drying it and then baking at 1000°C. The resulting phosphors luminesced with the typical orange colour under U.V. 3650 Å light and also cathodoluminesced in a 2kV discharge. This phosphor powder was used as the source for evaporation on to substrates at about 400°C. The micron thick films produced were not luminescent. After annealing these films in vacuo at 700°C (Ref3-12) for 30 min they remained non-luminescent and a similar anneal in argon at 700°C was no better and it usually caused the films to start peeling from their substrates.

The embedding technique was more successful. ZnS:Mn phosphor powder was placed on top of a deposited ZnS film which was then inserted into an argon flow at 700°C. The film produced by this method showed weak orange luminescence under both U.V. and electron excitation although their adhesion to the substrate was again poor.

A third technique involved the evaporation of a thin layer of manganese on to the ZnS film as in the method of Gilles and Van Cakenberghe.

The final film however was not luminescent and it also partly detached itself from the substrate during the heat treatment.

Electroluminescence experiments were not conducted on the above films due to their non-uniformity and lack of reproducibility in their preparation.

3.6 Results

3.6.1 Evaporated ZnS films

In general the evaporated films were non-reproducible . Evaporations carried out under apparently identical conditions often showed wide variations in the quality of the films. General results can however be noted.

i) Films deposited on to substrates with temperatures less than 100°C were usually transparent and amorphous as shown by their broad diffuse X-ray diffraction rings.

ii) Evaporations with substrate temperatures in the range 100-450°C produced crystalline films. The X-ray and electron diffraction analysis of these films is described in Sec 3.6.3.

iii) With substrate temperatures greater than about 600°C very poor non-uniform films were produced. This was probably due to re-evaporation and the increased contribution of substrate contamination. Above about 800°C ZnS vapour appears to react with silicon.

iv) Post-evaporation heat treatment of ZnS films described above usually resulted in the film becoming detached from its substrate.

v) Due to the lack of reproducibility of the evaporated films in general, it was impossible to ascertain whether the source material (i.e. Levy-West powder or flow crystals) or changing the source crucible and using the hot-wall bell-jar method affected their quality.

3.6.2. Vapour-deposited films.

The films of ZnS deposited on to silicon using the flow system were also non-uniform. The following observations were however made.

i) ZnS crystal growth occurs in the temperature range 1100-800°C using the system described in Sec 3.3. Silicon substrates in this region were covered with polycrystalline deposits of mainly hexagonal alpha Zns and a Laue back reflection X-ray diffractograph of one of these films is shown in Fig 3 - 5. From this it seems that the crystal habit favoured under these circumstances is incompatable with the formation of single crystal layers on silicon.

ii) At temperatures higher than those in the growth region ZnS vapour appears to react with silicon. It seems probable that prior to the polycrystalline growth of ZnS on silicon, described above, this reaction also takes place.

3.6.3. Physical assessment of films.

The ZnS films were examined under the microscope and their thickness was measured interforometrically as described in Sec 2.3.2. ZnS vapour reacts with silicon at temperatures greater

than about 800°C and under favourable conditions it etches the silicon surface to produce triangular pits on the (111) face probably by the formation of the volatile reaction product SiS. The triangular areas shown in Fig 3-6 were a few microns in size and at first they were confused with the initial nucleations of epitaxial growth. To resolve this problem of whether they were pits or deposits films were metallised and examined under a magnification of x 425 with an interference objective made by Vicker's Ltd. The direction of displacement of the fringes, clarified by the use of white light fringes, showed conclusively that these areas were in fact pits which incidentally were rotated through 180°C with respect to the etch pits produced by the NaOH etch.

The crystal structure of the films had to be examined using a glancing angle X-ray diffraction technique. Since the scattering of X-rays by thin (0.5 micron) films was insufficient for the standard Laue back reflection technique. The silicon substrate was mounted on a goniometer head with the ZnS film aligned to make a 5° angle with the incident X-ray beam. The forward diffracted X-rays were recorded on a plate film positioned about 3 cm behind the specimen. The area of the X-ray beam was about 1mm ^{2}. Due to the difficulty of measuring the film-specimen distance a calibration layer was sometimes deposited on top of the evaporated film. This could conveniently be produced by painting 'Silverdag' over the film.

ZnS evaporated on to substrates at about 300°C usually showed the diffraction rings of the low temperature beta-Zns. The rings

10μ

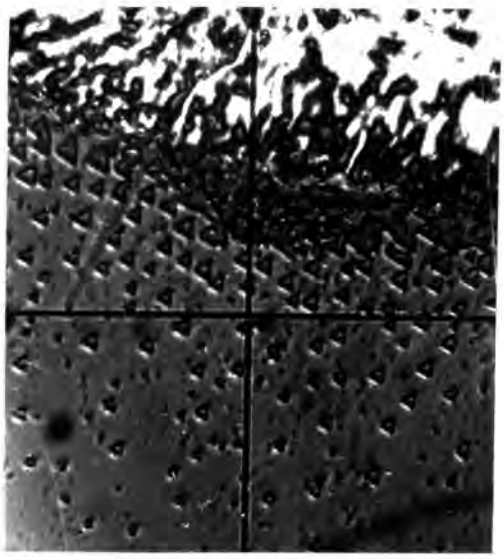


Fig 3-6 Two examples of the triangular etch pits produced by the action of 'ZnS' vapour on the (111) surface of a silicon substrate at 800°C. An evaporation mask has prevented the silicon in the lower parts of the photograph receiving the direct vapour beam but diffusion of the vapour under the edge of the mask has presumably occurred to produce the pits.

were unbroken and showed no evidence of arcing due to a preferred orientation in the film. When the film was deposited on to substrates at higher temperatures (about 400°C) the films condensed with the (111) axis of beta-ZnS as a preferred orientation normal to the substrate. Fig 3 - 8 is a diffractograph of one of these films taken with a glancing angle of 5° showing arcing of the rings. The angle subtended by the arcs at the centre of the ring is a measure of the deviation of the preferred axis about the normal and this is about ± 5° in this case. The single crystal spots are from the underlying silicon substrate. Fig 3-7 shows a Laue back reflection X-ray diffractograph of the same film which could be taken because it was particularly thick. It shows the randomness of the crystallites in the plane of the substrate. Diffraction rings due to the hexagonal alpha-ZnS, the high temperature modification, were also faintly visible in the glancing angle photographs.

A few of the crystalline films were also examined using the JEM 120 electron microscope. This microscope was acquired by the Department towards the end of ~~this~~ the ZnS work and consequently no detailed investigation was undertaken. Films about one micron thick were scraped from their substrates with a razor blade and then thinned by floating them on dilute HCl acid. A transmission electron diffractograph of such a ZnS film deposited on to a silicon sustrate at 200°C is shown in Fig 3-9. The electron beam illuminates about 10^{-5} mm^2 and consequently the diffraction rings from the beta ZnS are more spotty than in the X-ray case. The hexagonal alpha spots of a larger single

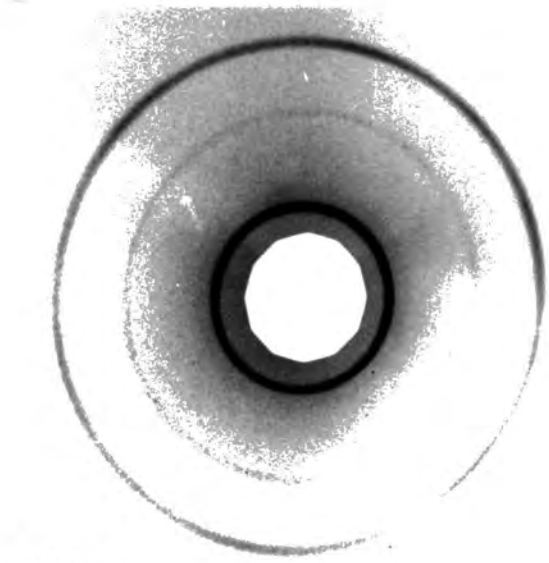


Fig 3-7 Laue X-ray back-reflection photograph of an evaporated ZnS film

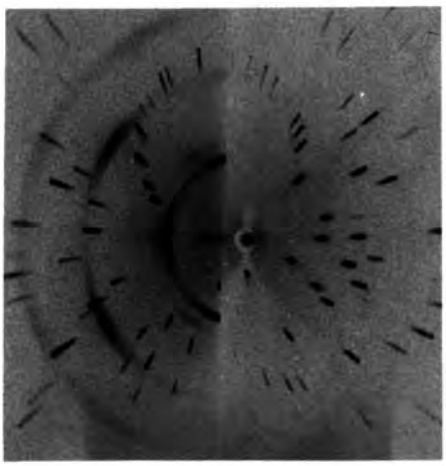


Fig 3-8 Glancing angle X-ray photograph of an evaporated ZnS film on silicon

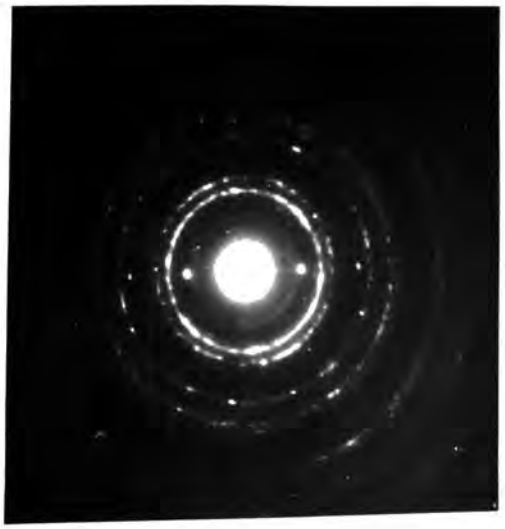


Fig 3-9 Transmission electron diffraction pattern from an evaporated film composed mainly of polycrystalline beta-ZnS

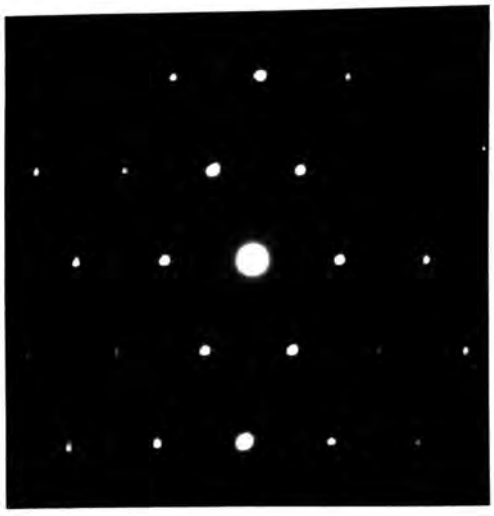


Fig 3-10 Transmission electron diffraction pattern from a small crystallite of alpha-ZnS in an evaporated film

crystallite of ZnS are also visible on this photograph and the diffractograph of a perfect hexagonal crystallite of Zns is shown in Fig 103-10. This however was not typical of the film as a whole which was mainly polycrystalline beta-Zns.

3.7 Discussion

The work described in this chapter occupied approximately one year, and though it was mainly of a preliminary nature several considerations had become apparant by the end of the year, which led to the termination of the ZnS work for the possible production of a solid state display based on silicon. These considerations are discussed below with emphasis mainly on the evaporated films. The reasons for discontinuing the chemical and vapour depositions are given at the end of their respective sections 3.3 and 3.4 though several of these considerations apply equally to all forms of deposition.

a) Reproducibility

The main problem encountered during the work on evaporated films was their lack of reproducibility. In order to produce uniform films much more precise control of all the evaporation and condensation parameters appears to be necessary. Several later workers (Ref 3-10) have also reported similar difficulties in the evaporation of reproducible ZnS films. Oil vapour from the diffusion pump, resistance heaters, the evaporant charge and the poor vacuum itself are all sources of uncontrollable contamination in the system used. If this work had been extended then the following features would have been incorporated into an improved system.

i) An oil-free U.H.V. system.

ii) Electron beam heating of the source and the substrate.

iii) Pre-evaporation purification of the evaporant charge and control of the evaporation and condensation rates.

iv) A method for cleaning or etching the silicon in high vacuum. Jones et al. (Ref 3-11) using a carefully designed U.H.V. system have more recently obtained epitaxial ZnS films up to 5 microns thick on silicon. A thermal etching method for cleaning the silicon in situ was adopted which involved the evaporation of the surface oxide and some of the silicon itself. Epitaxial growth occurred at 620°after rapidly cooling the substrate from a temperature of 920°in the presence of ZnS vapour.

Although the production of epitaxial films of ZnS on silicon using improved methods therefore seems possible it appears that this would involve the use of apparatus not currently available in the Department. Also for eventually producing devices the substrate preparation before evaporation must not be so rigorous as to prevent it being applied to silicon containing p-n junctions. The main reason for trying to produce single crystal ZnS film however was to make them electroluminescent under a high field and in this context the following considerations are also relevant.

b) Activation Problems.

Thin films of ZnS always require post-evaporation heat treatments to make them electroluminescent (Ref 3-12). The work done

on this was not extensive, due to the problems of non-reproducibility, and it was only relevant to polycrystalline films and not the single crystal films which might finally have been produced. However, the following detrimental aspect of activation may be noted.

i) Activation involves the controlled diffusion of certain impurities in-to the ZnS lattice. These impurities must not be allowed to diffuse into either the underlying silicon or the adjacent silicon which may contain electrically active regions. This restriction, though perhaps not impossible to overcome, could offer considerable technological difficulties.

ii) The coefficients of expansion of ZnS and Si differ slightly. For example along the (111) direction they are $7.10^{-6}/^{o}C$ and $2.5\ 10^{-6}/^{o}C$ respectively. (Ref 3-13). It seems that in post-evaporation treatments with films about a micron thick this results in the detachment of the film from its substrate. Jones et al. (Ref 3-11) in fact used the array of cracks appearing in their films after deposition at $620^{o}C$ as evidence of epitaxy. The mismatch may be a fundamental limitation to heat treatment after evaporation.

c) Excitation Problems

In order to produce electroluminescence the luminescent centres within the film must be excited and for d.c. operation this involves the injection of carriers from the silicon. Although the exact nature of interface states at heterojunctions is not fully understood it is probable that contamination and uncontrollable oxide layers at the interface would reduce the efficiency of the field emitting contact as envisaged for ZnS on Si. The following points seem

relevant to the problem of excitation of electroluminescence in films of ZnS on silicon.

i) Silicon surfaces rapidly oxidise and this oxide must be removed prior to deposition in order to achieve an efficient contact as well as promoting single crystal formation.

ii) ZnS dissociates on sublimation (Ref 3-14) and the vapour appears to react with silicon at temperatures greater than 800°C. If the reaction products are solids at these temperatures the interface will be contaminated with the consequent detrimental effects on the contact properties. On the other hand if the products are volatile it would seem to be feasible that, with careful control of temperature, the ZnS itself could also be used as an etchant for its own substrate in a similar way to the use of $SiCl_4$ and HCl mixture in the epitaxial growth of silicon on silicon.

iii) Single crystals of ZnS are usually densely twinned and contain many defects and it is probable that similar effects would occur in epitaxial films deposited on silicon. These defects would be undesirable from the point of view of sustaining a high field across the films and obtaining injection at the silicon or top electrode particularly if activator ions concentrated in the defects as often occurs in other materials.

3.8. Conclusion

In this chapter the preparation and properties of ZnS thin films deposited on silicon have been described. Three methods of preparation were tried but only evaporation seemed to be feasible and even this was in general non-reproducible. The general features of films evaporated on to substrates at different temperatures are described and the possible modifications necessary for reproducibility are discussed. It also seemed to be difficult to incorporate the required activators to obtain any form of luminescence.

The preparation of bright d.c. electroluminescent ZnS films on glass has now been achieved elsewhere (Ref 3-15) and also techniques for growing single crystal ZnS films on silicon have also been developed elsewhere (Ref 3-11). However, bringing these two processes together to produce electroluminescent ZnS on silicon introduces several new problems which have been discussed in Sec 3-7, The processes which appear to be necessary to produce electroluminescent films on silicon seem to be incompatible with the silicon technology necessary for the display device originally envisaged.

The difficulties encountered in the preliminary work on the deposition of ZnS films on silicon and those which could be anticipated in the production of a display device led to the decision to terminate the ZnS thin film work. This decision was fortified by the fact that electroluminescence of thin ZnS films were being extensively studied in other laboratories and, since the start of the project, considerable effort had been directed elsewhere towards the production

of ZnS epitaxial films on silicon. Moreover the author became aware at about this time of another luminescent material which seemed to show considerable advantages over ZnS for display devices based on silicon. The new material willemite, had the added advantage that it has scarcely been studied elsewhere in thin film form. The rest of this thesis is concerned with these films and the next chapter describes their production on silicon.

References for Chapter 3

3-1 R.R.Addiss, Tenth National Symposium on Vacuum Technology Transcs, New York (1963) p 354.

3-2 F.G. Allen, J. Eisinger, H.D. Hagstrum and J.T. Law, J. Appl. Phys. 30, 1563 (1959).

3-3 A.P. Hale, Vacuum, 13, 93-100 (1963).

3-4 J. Wales and A.J. Jeal, J. Sci. Instruments, 41, 521 (1964).

3-5 A. Kremheller, Sylvania Techn. 8, 11 (1955).

3-6 J.L. Davis and M.K. Norr, J. Appl. Phys. 37, 1670 (1966).

3-7 M. Nagao and S. Watanabe, Jap. J. Appl. Phys. 7, 684-5 (1968).

3-8 J.M. Gilles and J. Van Cakenberghe, Nature, 182, 862 (1958).

3-9 A. Vecht and A. Apling, Phys. Stat. Sol. 3, 1238 (1963).

3-10 B.A. Unvala, J.M. Woodcock and D.B. Holt, Brit. J. Appl. Phys. (J. Phys. D) Series 2, 1, 11 (1968).

3-11 P.L. Jones, C.N.W. Litting, D.E. Mason and V.A. Williams, Brit. J. Appl. Phys. (J. Phys. D) Series 2, 1, 283 (1968).

3-12 The Physics of Thin Films, Vol 3,, New York (1966) Chap 13, A. Vecht.

3-13 American Institute of Physics Handbook, New York (1957).

3-14 A.D. Pogorelyi, J. Phys. Chem. (USSR), 22, 733-45 (1948).

3-15 A. Vecht, N. J. Werring, R. Ellisand, P.J.F. Smith, Brit. J. Appl. Phys. (J. Phys. D) Series 2, 2, 953 (1969).

CHAPTER FOUR

WILLEMITE FILMS ON SILICON

4.1 Introduction

4.1.1 General

Willemite is a naturally occurring luminescent mineral composed of manganese activated zinc orthosilicate. As a photoluminescent phosphor excited mainly by 2537 Å ultraviolet radiation it was used in one of the first fluorescent lamps, although its photoluminescent properties are very little used now. The cathodoluminescent properties are however widely used for cathode ray tube screens. The light output is usually green, the broad structureless emission band centred about 5250 Å being typical of the Mn^{++} ion in the rhombohedral crystal structure of alpha-Zn_2SiO_4, and it is practically independent of the excitation method used. Weak luminescence has also been observed in the yellow and red which is associated with Mn^{++} in a pseudo-rhombohedral structure of willemite called beta-Zn_2SiO_4 and to an amorphous phase respectively. (Ref 4-1). A self-activated phosphor Zn_2SiO_4 :Si can also be made which luminesces a pale violet with electron and X-ray excitation (Ref 4-5)

4.1.2 Electroluminescence of Willemite

The photoluminescent and cathodoluminescent properties of willemite phosphors were extensively studied in the early days of phosphor technology. Various attempts have also been made to obtain electroluminesence (subsequently referred to as EL) from Zn_2SiO_4:Mn powders but the output has always been quite low. Destriau in 1947

gave the first report of EL from silicates. (Ref 4-2). Bramley and Rosenthal (Ref 4-3) embedded the powder in a low melting point glass and obtained EL with a.c. fields of the order of 5.10^{54} volt/cm. The light emission occurred near the electrodes and was attributed to field emission across the electrode-phosphor interface.

Ueta (Ref 4-4) suspended Zn_2SiO_4:Mn in polystyrene and obtained green EL with 500 volts at 50 Hz applied across a 20 micron thick powder layer with 40 microns of castor oil to form an insulating film between the phosphor and the electrode. Similar results were obtained by Luyckx and Stokkink (Ref 4-6) who studied the brightness waves of the light output and observed high frequency oscillations at their peaks indicating the approach of breakdown. The potential required to produce similar light output from a ZnS powder phosphor was only a tenth of that required for Zn_2SiO_4: Mn. Mizushima (Ref 4-7) in a short note, mentions the observations of both light and electron emission from a zinc silicate phosphor cold cathode.

A patent by Rulon and Butler (Ref 4-8) describes a method for achieving EL in Zn_2SiO_4:Mn. $SnCl_4$ was sprayed over the phosphor powder and in a subsequent heat treatment conducting spots on the phosphor particles of tin oxide were formed to produce conditions suitable for EL. The brightness achieved however was still low being 0.015 ft-L with an applied voltage of 600 volts at 60Hz.

Nicol and Kazan (Ref 4-9) observed EL in conventional cathode ray tubes. A direct voltage was applied between the

aluminium backing of the phosphor screen and a transparent coating on the outside of the tube. EL was observed when the screen was heated to allow sufficient current flow. A brightness of 1.5 ft-L was obtained at a current density of 350 μamp/cm^2 produced with 200 volts. This effect was observed both when the tube was evacuated and when the vacuum had been destroyed.

An explanation of the above effect was suggested by Bowell and Bate (Ref 4-10) who also obtained EL with zinc ~~die from~~ silicate in which the particles were first coated with an ionisable salt such as potassium silicate, a material which is used as a binder in the production of cathode ray tube screens. The emission was localised at the anode for d.c. operation but the output decayed due to polarisation effects. Most, but not all photoluminescent materials could be made to EL in this manner.

Kolomoitsev et al. (Ref 4-11) obtained similar effects to Bowell and Bates when they coated Zn_2SiO_4: Mn phosphors with Na_2SiO_3 but they found that the light emission disappeared in vacuum. They concluded that the motion of the ions in the powder matrix allows the build-up of a sufficiently large potential barrier at one of the electrodes to produce a corona discharge within the powder at relatively low d.c. voltages. This also explained the decrease in light output that occurred when the frequency was increased with a.c. excitation. At high frequencies there was thought to be insufficient time between cycles to allow for the redistribution of the ions in the powder.

Lehmann (Ref 4-12) showed that many photoluminescent but non EL phosphors, such as a Zn_2SiO_4: Mn powder, could be made EL by first mixing the phosphor with sharp conducting particles, usually of metal. A slurry of the mixture with castor oil was then inserted between a conducting glass plate and a metal electrode covered with a thin insulating film. When a.c. at a frequency of about 20 kHz was applied across these cells light output was observed at the contacts between the phosphor particles and the conducting particles. Light was also observed when the conducting particles were of copper sulphide. ~~Ultra-violet~~ Ultra-violet excitation from gas discharge was discounted because some phosphors showed different colours for U.V. photoluminescence and for contact EL.

The most recent reference to EL in Zn_2SiO_4: Mn appears to be a paper by Jones (Ref 4-13) who constructed three basic types of EL cells using Zn_2SiO4:Mn, amongst other cathode ray tube phosphors. The Type I cells which were similar to Lehmann's but without the conducting particles in the phosphor powder exhibited a.c. EL although they glowed only weakly if at all in vacuum. Their Type 11 cell incorporated a layer of the semiconductor cadmium pyroantimonate sandwiched next to the phosphor layer and this showed a.c. EL both in air and in vacuum. In the Type 111 cells he removed the insulating film from between the electrodes which resulted in EL with direct voltages applied in vacuum. A comparison of these cells with a commercial ZnS EL panel showed that equal light outputs could be obtained for the same currents although 20 times more voltage was

required for the Zn_2SiO_4:Mn cells. Jones concludes that the EL was due to the bombardment of the phosphor by electrons either at the air-phosphor interface or at a semiconductor (in this case cadmium antimonate) to phosphor interface. Although the provision of the latter contacts considerably reduced the threshold voltage required for EL he does not discuss the exact nature of this semiconductor/phosphor interface.

It appears that the EL of Zn_2SiO_4:Mn powders requires much more intense fields than for the more-usual sulphide type phosphors. Many of the earlier observations of EL in willemite phosphors can probably be attributed to cathodoluminescence or ionoluminescence due to corona discharges within the powder matrix which could be enhanced by special phosphor treatments. The observations of EL from within a highly evacuated cathode ray tube by Nicoll and Kazan however, means that this is perhaps not always the case. Considering also the work of Lehmann and Jones it does not seem impossible that some high field injection mechanism which will excite EL may also occur at a heterojunction with zinc silicate. It appears that both a metal and a semiconductor can fulfil this role.

EL studies of Zn_2SiO_4: Mn have so far been devoted entirely to powders. The high fields which appear to be necessary to obtain the required injection phenomena in willemite can however be obtained most conveniently by the use of thin films; although thin films of willemite have been produced by other people noøne appears to have considered their electroluminescent properties.

The production of thin uniform films which can also sustain the high fields necessary for EL is described in Sec 4.2. The next section describes the methods of production of willemite phosphors which have been used in the past and more specifically those which have been applied to the preparation of thin films.

4.1.3 Willemite thin films

Since Zn_2SiO_4 melts at 1785°C commercial phosphors are usually prepared by a solid state reaction between a zinc compound and silica. The zinc compound is often ZnO though many variations exist. In the 'hydrofluoric acid' process (Ref 4-15) more intimate mixing of the solid state reactants is achieved by utilizing the effervescent action of the volatile SiF_4 which is produced when hydrofluoric acid, ZnO, and silica are mixed. Zn_2SiO_4 is formed when the products of the reaction are heated to 1000°C. The 'carbonate process' relies on the co-precipitation of the carbonate and silica in colloidal form from a zinc nitrate and ammonium carbonate solution. Again a high temperature bake is required to produce the silicate. The manganese activator required for luminescence is usually added in the same way as the zinc compound in these processes.

Zinc silicate phosphors are used mainly for cathode ray tube screens. Granular screens produced from a powder have the disadvantages of very poor thermal contact with the glass substrate and of specular scattering of the light output from the cathodoluminescent spot which results in a loss of resolution. The light output as the electron beam moves across the screen is also very noisy.

A transparent thin film screen should have none of these disadvantages and it would also give the dubious advantage of bilateral viewing. It was mainly for these reasons that thin willemite films have been investigated in a number of laboratories and several techniques developed for their growth.

The earliest mention of willemite films is in the work of Feldman and O'Hara (Ref 4-16) who evaporated Zn_2SiO_4: Mn from a tantalum boat at 1300°C on to silica glass substrates. The films condensed in a darkened condition due to the reduction that occurs during evaporation and they required firing at 1100°C for 30 minutes in oxygen in order to make them green luminescent. Various heat treatments were described in a later paper (Ref 4-17) in which Feldman and O'Hara reported the formation of red and yellow luminescent films. One particular heat treatment involved the diffusion of boron from a borosilicate glass substrate and demonstrated the strong reaction that can occur between these films and their glass substrates.

A reaction technique involving a glass substrate and a metal fluoride for the production of transparent luminescent films was patented by Rottgardt (Ref 4-18). In this, a manganese activated zinc fluoride phosphor was evaporated on to the inside of a glass cathode ray tube envelope which was then baked between 300°C and 650°C. The resulting luminescent film was the product of a reaction between the silicate glass and the fluoride. Koller and Coghill (Ref 4-19) used a variation of this technique to produce silicate phosphor films by evaporating ZnF_2:Mn on to a high silica-containing glass substrate

which had been heated to about 600°C.

A hot spray method was utilized by Kirk and Schulman (Ref 4-20) which also involved the reaction with the substrate surface. A methanol solution of $ZnCl_2$ containing the requisite amount of $MnCl_2$ was sprayed on to Pyrex substrates at 500-1000°C to produce green cathodoluminescent films of Zn_2SiO_4:Mn. The recrystallisation of a glass substrate surface which results in the embedding of willemite in a vitreous matrix to produce luminescent films is the subject of a patent by Veres (Ref 4-21).

Substrate reaction techniques for the preparation of thin luminescent films offer several advantages over the more usual methods of production and these are enumerated more specifically in Sec. 4.4. The work described in this thesis is concerned with the preparation and properties of luminescent films on silicon with a view to making a fully integrated display device and the development of a substrate reaction technique for the production of willemite phosphor films on a silicon rather than a glass substrate is described in the next section.

4.2 The preparation of willemite films on silicon

The substrate reaction methods described in Sec 4.1.3 have been used successfully by several workers to produce luminescent willemite films on borosilicate and silica glasses and it was decided to try to develop similar techniques for silicon substrates.The use of the naturally occurring oxide on silicon in the formation of a phosphor film forms the main original idea in this thesis and is the reason for

the choice of willemite as a luminescent material highly compatible with silicon technology. Many other advantages became apparent later and these are listed in Sec 4.4.

The final process for the formation of willemite on silicon consisted basically of the oxidation of the silicon surface, the deposition of a layer of ZnF_2:Mn and a subsequent heat treatment. The details of this method and the different techniques that were tried before the final process was adopted are described below.

The oxidation and preparation of the silicon sustrates is described in Sec 2.2. Oxide thicknesses of about 1000 Å or less were generally used because thin films were necessary so that high fields could be ~~purchased~~ produced with relatively low voltages.

Electronic grade ZnF_2 both unactivated and activated with 1% by weight of manganese fluoride was obtained from Levy West and Co. Ltd. It was decided to use ZnF_2:Mn because unlike other compounds that could be used (e.g. $ZnCl_2$) both ZnF_2 and MnF_2 have very nearly the same vapour pressure and melting points (872°C and 856°C respectively) so that the phosphor can be evaporated and condensed without the loss of the activator. The requirement for the by-products of the substrate reaction (SiF_4 in this case) to be volatile is also fulfilled with ZnF_2:Mn. The volatility of SiF_4 and consequent effervescent action was in fact the basis of the ' hydrofluoric acid ' process used for producing willemite powder as described in Sec 4.1.3.

Some of the early work used unactivated ZnF_2 to which manganese had been added. This activation process was carried out by

making a slurry of ZnF_2 powder with a measured solution of $MnCl_2$, drying it and then baking in argon at a temperature just below the melting point of the fluoride. Ideally this should be done in a platinum crucible as ZnF_2 reacts with silica. In fact, the use of a quartz boat results in zinc silicate forming on the surfaces of all the nearby glassware as ZnF_2 sublimes slightly at these temperatures.

Early experiments on activation involving this baking process led to attempts at making willemite by a reaction technique from a gas flow. A flow method which could be used without a vacuum system should also have an advantage from the silicon technology point of view. Oxidised silicon chips were placed slightly downstream of some subliming ZnF_2:Mn at 800°C in a tubular furnace with an argon gas flow. Silicate films formed on the silicon giving green cathodluminescence. However, the films were frosty and non-uniform and excess ZnF_2: Mn seemed to prevent the formation of luminescent films. Masking these films would be difficult although an adaptation of the standard oxide masking technique used in silicon technology might have been possible.

In order to have more control of the amount of ZnF_2:Mn deposited on to the hot silicon substrates it was decided to use a vacuum evaporation method instead of the gas flow. The substrates were heated in the bell jar system described in Sec 3.2. using a molybdenum strip heater and the phosphor was evaporated from a pepperpot boat which is also described in Sec 3 2. Molton ZnF_2 reacts

with molybdenum and ideally a platinum crucible would have been used but it was found that the most reproducible evaporations were performed by subliming ZnF_2 powder at about 825°C which is about 50°C below its melting point and this could be done using a molybdenum boat. Over 50 evaporations carried out with one molybdenum pepperpot led to no apparent corrosion.

It was essential to thoroughly outgas the source prior to evaporation at a temperature lower than about 450°C. The manganese fluoride activator reacts with oxygen at temperatures greater than this and the high melting point oxides produced remain in the crucible as a black deposit after the evaporation. The fluorides are also decomposed by any water vapour in the system. The evaporations were carried out in the 10^{-6} torr pressure range.

As the evaporating ZnF_2 molecules reach the substrate they react with the hot oxidised silicon to produce zinc silicate. The films produced by this process, however, were only weakly green cathodoluminescent and again they appeared to be frosty and non-uniform. The reaction of the silica with the impinging ZnF_2 molecules is determined to some extent by the rate of arrival of the molecules so that too fast a rate of evaporation resulted in a film which showed the yellow luminescence of unreacted ZnF_2: Mn on the surface and the weak green luminescence of a thin willemite layer beneath. The ZnF_2: Mn could however be washed away with ammonium hydroxide leaving the silicate layer intact.

In the next technique tried the dependence of the substrate

reaction on the rate of evaporation was removed by evaporating the ZnF_2:Mn on to unheated oxidised substrates. These were then removed from the vacuum system and baked at 1100°C in oxygen to give the reaction that forms willemite. The approximate thickness of fluoride required to react with the oxide film was estimated using the known density values and assuming that the simple reaction during the baking was :-

$$ZnF_2:Mn + SiO_2 = Zn_2SiO_4:Mn + SiF_4$$

The proporations of ZnF_2:Mn and silica required were about 1.5:1 by volume respectively so that a typical evaporation at 825°C produced fluoride films that were about 1000Å thick. When converted to willemite these films gave bright green cathodoluminescence showing for the first time that the phosphor could be produced satisfactorily on silicon substrates.

The baking procedure required to produce the brightest and most uniform luminescent films involved inserting the ZnF_2:Mn film into the furnace at 1100°C so that the fluoride melted before any appreciable reaction with the oxidised silicon surface occurred. Gas ambients containing oxygen appeared to be necessary during the 5-10 min bake because probably the oxygen allowed the formation of the completely oxidised silicate necessary for bright luminescence. Feldman and O'Hara (Ref 4-16) also found that films of Zn_2SiO_4:Mn evaporated from the powder phosphor were only luminescent after a 30min bake in oxygen at 1100°C.

In spite of the reaction equation above there seemed to be a wide tolerance in the relative thicknesses of ZnF_2 and silica films which could be used successfully to produce green luminescent films. ZnF_2:Mn films which were of greater than about 5 times the thickness estimated from the above equation produced willemite films which were frosty and only weakly luminescent. Greater proporations of ZnF_2 resulted in dark non-luminescent films.

If only a relatively small amount of ZnF_2 was evaporated onto the oxidised silicon surface, films were obtained that cathodoluminesced yellow. This observation was in agreement with previous workers who obtained yellow luminescent zinc silicate powders when excess silica was used in their production.

Bright green cathodoluminescent films could also be produced when up to 1000Å thickness of ZnF_2:Mn was evaporated on to silicon substrates which had not been thermally oxidised, though they probably still had a thin oxide coating. In this case gas used in the reaction bake probably provided most of the oxygen required for the formation of Zn_2SiO_4.

The maximum thickness of the oxide that could be used was governed by the depth of penetration of the reaction. This is discussed in the next chapter together with a structural assessment of the willemite formed by the substrate reaction process.

The process described above, involving the evaporation of the ZnF_2:Mn on to unheated oxidised silicon substrates followed by a bake in oxygen at about 1100°C was capable of producing uniform

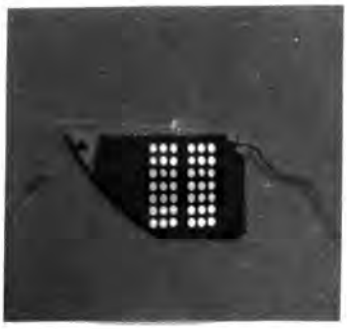
green cathodoluminescent films with brightness comparable to that of commercial willemite powders. Over 80 willemite films were prepared altogether and of these all those made by this process within the conditions outlined above (about 50) showed the same green cathodoluminescence. The films were always uniform apart from a few films that had one or two small specks of contamination probably produced by dust particles picked up during the transfer from the evaporator to the furnace.

4.3 Device preparation

The phosphor films made by the final process described in Sec 4.2 usually had dimensions of about 0.5 x 0.2 cm on silicon substrates about 1 cm by 0.5 cm. Metal top contacts were evaporated on to the films after baking usually in the form of an array of 21 dots each about 10^{-3} cm^2 in area. Gold, aluminium and indium were used as contact metals. The evaporation masks for the contacts were made by standard photolithographic techniques. Contacts were also evaporated on to the neighbouring oxide surface so that both MOS and metal-phosphor-silicon structures were available for electrical measurements, the MOS devices being used for comparison purposes. A photograph of a typical willemite and a schematic diagram is shown in Fig 4-1.

Ohmic electrical contacts were made to the silicon by first etching the oxide from selected areas of the device and then bonding wires to the silicon surface using the techniques described in Sec 2.4.3.

Transparent tin oxide contacts were used on a few devices.



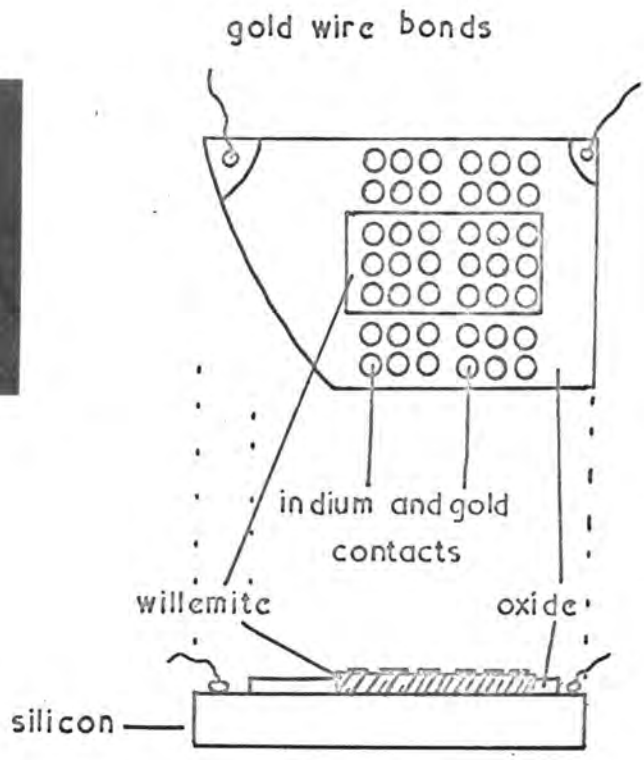


Fig 4-1 A typical willemite device

These were prepared by exposing the surface, heated to about 300°C, to a stream of $SnCl_4$ in argon. Masking of these contacts was difficult as no etch for tin oxide could be found that did not also attack the silicate film underneath. Strips of tin oxide 0.010 inch wide were finally produced across the willemite by first evaporating aluminium on areas where the tin oxide was not required and after the deposition etching it away with NaOH solution to take the overlying tin oxide with it.

The devices were mounted on perspex holders and placed under a microscope so that the contacts could be probed in turn. This was done by using two micromanipulators made by Research Instruments Ltd. which were fitted with gold probes. The electrical measurements carried out on these devices are described in Chapters 6 and 7.

4.4 Discussion

The willemite films, prepared by the substrate reaction technique described above, offer considerable advantages over other possible materials for the eventual production of an electroluminescent display device based on silicon. The immediate advantages are listed briefly below although the whole subject will be discussed in more detail in Chapter 9.

1) The method of production of willemite films is simple and highly compatible with modern silicon device technology in which controlled oxidation is extensively used.

ii) The interface between the silicon and the phosphor should be clean and free from contamination. The oxidation of silicon proceeds by the inward motion of the oxidising species so that the oxide-silicon surface is never exposed to the ambient although diffusion of impurities can still occur during the high temperature processing.

iii) The use of a uniform silicon oxide film which may be grown by well established methods for conversion to the phosphor considerably reduces the chance of obtaining pinholes.

iv) The films are electrically insulated from the adjacent silicon surface by the surrounding oxide layer and junctions in the silicon remain passivated.

v) The process appears to be capable of a high degree of reproducibility which is probably at least partially due to the fact that the process parametes do not seem to be critical.

vi) The unique feature of the phosphor ZnF_2:Mn (used as the starting material) enables it to be evaporated and condensed without loss of luminescent efficiency so that the manganese activator can be easily incorporated into the willemite film during its preparation.

vii) Evaporation masks can be used to closely define the areas of willemite produced.

viii) By using the reaction of the substrate to produce the phosphor films excellent adhesion is assured.

References for Chapter Four

4-1 H.G. Peiffer and G.R. Fonda, J. Electrochem. Soc. 99, 140 (1952)

4-2 G. Destriau, Phil. Mag. 7, pp 700, 774, 800, 885 (1947)

4-3 A. Bramley and J.E. Rosenthal, Phys. Rev. 87, 1125 (1952)

4-4 M. Ueta, J. Phys. Soc. Jap. 8, 429 (1953)

4-5 H.W. Leverenz and F. Seitz, J. Appl. Phys. 10, 479-93 (1939)

4-6 A. Luyckx and A.J. Stokkink, Brit. J. Appl. Phys. Suppl 4, 59 (1955).

4-7 Y. Mizushima, J. Phys. Soc. Jap. 14, 979 (1959)

4-8 R.M. Rulon and K.H. Butler, U.S. Patent 2,844,540 (1958)

4-9 F.H. Nicol and B. Kazan, Proc. I.R.E. 43, 1012 (1955)

4-10 J.N. Bowtell and H.C. Bate, Proc. I.R.E. 44, 697 (1956)

4-11 F.I. Kolomoitsev, Opt. i. Spek. 12, 127 (1961)

4-12 W. Lehmann, J. Electrochem. Soc. 104, 45 (1957)

4-13 S. Jones, J. Electrochem. Soc. 111, 307 (1964)

4-14 P. Lenard, F. Schmidt and R. Tomaschek, Handbuch Experimentalphysik Vol 23, Leipzig (1928)

4-15 H. Leverenz, U.S. Patent 2,210,087 (1941)

4-16 C. Feldman and O'Hara.M. J. Opt. Soc Amer. 47, 300 (1957)

4-17 C. Feldman and M. O'Hara, J. Opt. Soc. Amer. 48, 816 (1958)

4-18 J. Rottgardt, U.S. Patent 2,876,129 (1959)

4-19 L.R. K$_0$ller and H.D. Coghill, J. Electrochem. Soc. 107, 973 (1960)

4-20 R.D. Kirk and J.H. Schulman, J. Electrochem. Soc. 108, 455 (1961)

4-21 F. Veres French Patent 1,406,464 (C1 C 03c) (1965)

CHAPTER FIVE

The Structural Assessment of Willemite Films on Silicon

5.1 The Crystal Structure of Willemite

The first detailed X-ray examination of zinc orthosilicate, willemite, was undertaken by Bragg and Zachariasen (Ref 5-1) in 1930. The structure was shown to be based on a rhombohedral space lattice with a unit cell of dimensions a = 8.76 Å and $\alpha = 107^{\circ}\ 45'$, where 'a' is the length of the sides of the rhombohedron and α the obtuse angle between any two of them. Each unit cell was composed of 42 atoms comprising six Zn_2SiO_4 groups.

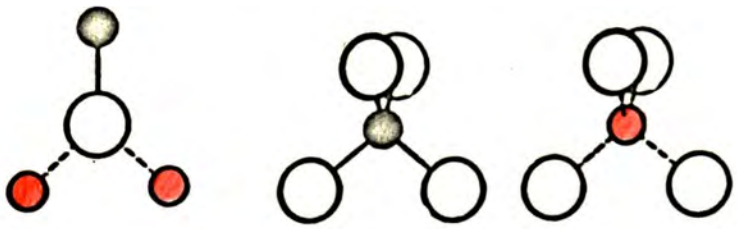
Silicates are classified, in general, by the Si–O arrangement which they display. The basic building block consists of a silicon atom surrounded by four oxygen atoms at the corners of a regular tetrahedron. The Si–O distance is about 1.6 Å and the O–O distance 2.6 Å. The various types of Si–O arrangements that occur are due to the tendancy of these tetrahedral (SiO_4) groups to polymerise to form chains, networks and framework structures.

Bragg, in his early classification of the silicates, included willemite in the group which was characterised by the existence of separate (SiO_4) groups (i.e. tetrahedra with oxygen atoms linked only to the one central silicon atom). Zoltai (Ref 5-2) in 1960 however attempted to reclassify silicate structures by introducing an additional criterion based on the relative sizes of the cation and the anion and he included Zn_2SiO_4 in the class characterised by an unterminated tetrahedral framework structure.

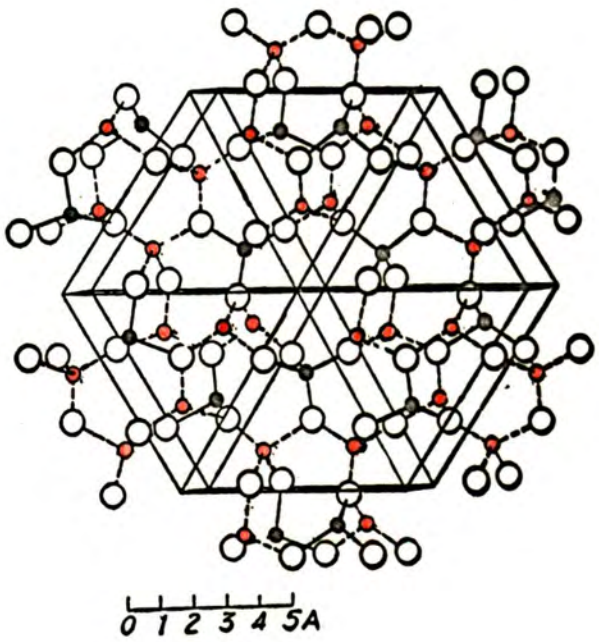
It is difficult to get a clear picture of the atomic arrangement in the willemite structure because of the large unit cell and the rhombohedral space lattice on which it is based. Fig 5-1 is reproduced from Ref 5-5 and is an attempt to illustrate some aspects of the structure. The structure is basically formed by the interlinking of SiO_4 tetrahedra with ZnO_4 tetrahedra around a trigonal axis in such a way that each oxygen atom of the SiO_4 group also forms part of two neighbouring tetrahedra around Zn atoms. Each Si atom and each Zn atom is then surrounded by four O atoms and each oxygen atom is linked by two Zn atoms and one Si atom at the corners of an equilateral triangle (Fig 5-1a)

The formation of the strongly bound SiO_4 group in zinc orthosilicate is due to the sp^3 hybridisation of the Si atoms. The mixed covalent-ionic bond that occurs between the Si atom and the four surrounding O atoms results in a SiO_4 radical which has a net charge of -4. The bonding of the Zn^{++} ion to this radical appears to be mainly ionic (Ref 5-3) each $(SiO_4)^{-4}$ radical being balanced by two Zn^{++} ions.

Zinc silicate phosphors are made by incorporating manganese ions into the crystal lattice. The Mn^{++} ion enters the lattice as an isomorphous substitution for the Zn atom. Solid solutions up to 45 mol % of Mn_2SiO_4 have been reported by Kroger (Ref 5-4), the only structural change being a small expansion of the lattice due to the slightly larger size of the Mn^{++} ion. These mixed cation silicates



a)



b)

Fig 5-1 The crystal structure of willemite – αZn_2SiO_4

O – oxygen atom ● – Zn atom ● – Si atom

a) The local atomic configuration of each atom

b) The atomic structure looking down the trigonal axis

in fact occur naturally as the mineral troustite ($ZnMnSiO_4$). For concentrations greater than 45% structural changes appear in the phosphor and X-ray analysis shows that crystals of the formula $MnSiO_3$ are formed. These luminesce red, as does the pure $MnSiO_3$.

Zinc silicate can also crystallise in a trigonal form that resembles a distorted tridymite or cristobalite structure (Ref 5-6). This alltropic form is called β-Zn_2SiO_4 and its preparation suggests that it is only an intermediate stage in the formation of the usual α Zn_2SiO_4 to which it reverts on heating above 800°C.

5.2 Reaction Depth Studies

In order to determine the depth of the substrate reaction, the silicate film was dissolved preferentially in concentrated hydrochloric acid. This left both the unreacted silica film and the underlying silicon substrate intact. The thickness of the silica film which had been converted to willemite was then measured using the ~~multiple-beam~~ interferometric technique described in Sec 2.3.2.

The maximum thickness of oxide that could be converted to willemite by the reaction process described in Sec 4.2. (i.e. the reaction between an evaporated film of ZnF_2 and its underlying oxide) was about 1000 Å. The use of oxides thicker than this resulted in a willemite-oxide-silicon sandwich structure as shown by the presence of oxide after etching off the willemite. The thickness of willemite films produced on oxides thinner than 1000Å was determined by the oxide thickness, any excess ZnF_2 presumably being lost by evaporation.

The distribution of the luminescent Mn^{++} ions throughout the depth of the phosphor film was investigated by bevelling the film and then observing cathodoluminescence down the bevel. The phosphor film and substrate were first 'potted' in a metallurgical mounting compound supplied by North Hill Plastics Ltd., and then lapped to the required angle, usually about 1^{o}, using 600 grit, 3, 1, and 0.25 micron diamond paste in succession followed by a final polish with gamma alumina powder. The mounting compound was then removed with hot trichloroethylene. The mounting procedure prior to lapping was necessary to prevent excessive rounding of the bevel in the thin phosphor film.

The bevelled device was mounted inside a small 'Spectrosil' U.V. absorption cell with the phosphor film close to one of the optical flats of the cylindrical cell. A rotary pump was connected to the mouth of the cell and a needle valve in the vacuum line controlled the ultimate pressure in the cell. The film was made to luminesce by using a glow discharge excited from outside the cell with a Tesla coil. Maximum brightness was achieved by localising the discharge with a small metal ring placed outside the cell above the phosphor film. The luminescent bevel was then examined with a microscope focussed through the centre of the metal ring on to the specimen beneath.

The light output from the phosphor film was photographed through the microscope using no external illumination using a film, Ilford HPS, which was relatively insensitive to the light from the discharge. The extent of the luminescence down the bevel was

determined by comparing these micrographs with interferographs of the same part of the bevel after it had been metallised. These photographs, together with one of the bevel under external illumination and a schematic diagram of the device are shown in Fig 5-2.

The phosphor film shown in these micrographs was prepared from a 1000 Å thick oxide film and it can be seen that within the limits of resolution of the interferometric technique (about ± 200 Å) the whole length of the bevel is luminescent. However, this does not prove conclusively that all the oxide has been converted to a phosphor because smearing of the luminescent centres down the bevel may have occurred during the lapping process. Scattering of the luminescence within the film may also be important. However, the existence of a relatively sharp contrast at the interface between the luminescent bevel and the non-luminescent oxide and silicon bevels seem to indicate that both smearing and light scattering effects are negligible. It was concluded from these experiments that as far as could be determined the willemite was uniform through a thickness of up to 1000 Å.

5.3 Electron microscopic examination of the Films

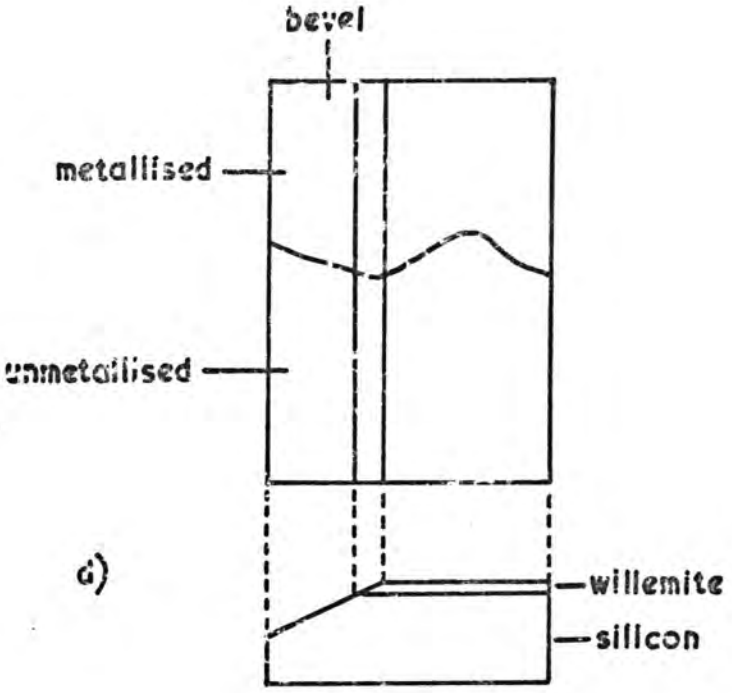
5.3.1 The Microscope

The electron microscope which was purchased on an S.R.C. Contract partly for this work was a J.E.M. 120 made by Japan Electron Optic Laboratories Co. Ltd. Transmission electron micrographs and selected-area diffraction patterns were obtained in the standard manner. A high resolution diffraction stage could also be

a)

b)

c) 5μ



d)

Fig 5-2 A bevelled willemite film

a) A micrograph of the light emitting bevel and film

b) Sodium light interference fringes from the same area as a)

c) A micrograph of the bevel under reflected illumination

d) A schematic diagram of the device

incorporated into the instrument and this enabled reflection diffraction patterns to be obtained. The microscope was normally operated with an electron accelerating potential of 120 kV. All the micrographs were recorded on Ilford N50 photographic plates.

5.3.2 Transmission Studies.

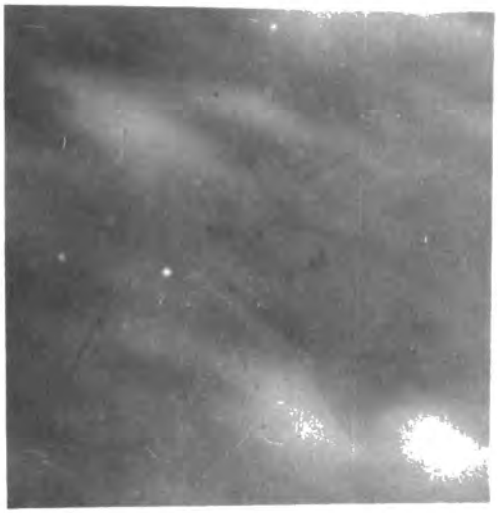
Since electrons interact strongly with solids the electron microscope specimen has to be very thin in order to obtain appreciable transmission. Samples must be of the order of 1000Å thick for 120 kV electrons. Many standard techniques exist for thinning bulk specimens (Ref 5-7) but most of these were unsuitable for removing thin (1000 Å) silicate films from silicon substrates. For example, a common technique for thinning Si is jet-etching. All the etches which were tried however dissolved the silicate much more rapidly than the silicon even if the etch only touched the film in the final stages of the thinning procedure.

Oxide films can be removed from silicon by using chlorine gas as a preferential etch (Sec 2.3.2) and this technique was also used for removing the willemite. To do this the underside of the silicon was first stripped of its oxide. The silicon together with its surface film was then inserted into a furnace at 900°C. After about one hour in a gas flow of argon containing a small amount of chlorine only the willemite and oxide films remained and these were picked up from their crucibles by using the electron microscope grids covered in 'Sellotape' adhesive dissolved in chloroform.

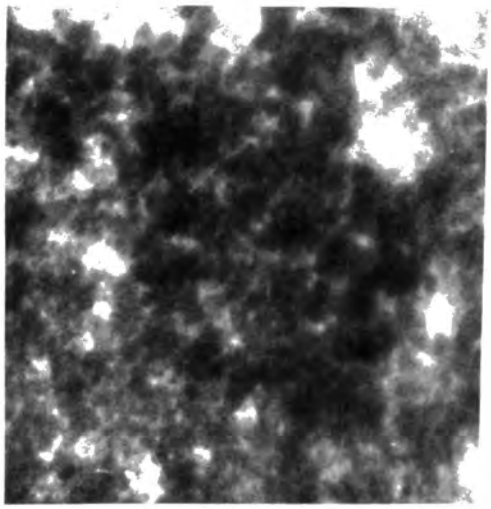
The selected area diffraction patterns of both the oxide

and the willemite films which had been subjected to the chlorine treatment were however identical (Fig 5-3c) consisting of the diffuse halos characteristic of an amorphous material. Fig 5-3a shows a micrograph of the oxide. This has a uniform appearance the only features being due to thickness contrast produced by wrinkles in the film. The films are grown at 1100°C and differential contraction between the oxide and the silicon occurs on cooling, so that when the substrate is removed the strained film wrinkles. A micrograph of the converted oxide is shown in Fig 5-3b. The granular appearance of this is probably due to the reaction that occurs during the conversion process.

The willemite films were not cathodoluminescent after their chlorine gas treatment. It appears that either the chlorine acts as a 'killer centre' for the luminescence when the willemite films become amorphous or that the gas reacts with the willemite and converts it to a non-luminescent amorphous material, probably returning it to the oxide in this case. The oxide film appeared to be unaffected by the chlorine treatment and it could be converted to willemite after it had been removed from its substrate. The converted film luminesced green but it was difficult to determine its uniformity because of the wrinkled nature of the fragile surface. These films were unstable in the electron beam and this produced a twinkling effect in the diffraction patterns which were obtained. The reaction process on the very thin oxides probably introduced considerable strain into the films and the annealing effect of the electron beam would account for this instability.



a) 0·1μ



b) 0·1μ

Fig 5-3 Transmission electron micrographs a) of Cl_2 detached oxide and b) of converted oxide after Cl_2 treatment

Fig 5-3c) Selected area transmission diffraction patterns from A) and b) above

The electron micrographs of oxide films converted to willemite after removal from the silicon (Fig 5-4a and b) showed roundish islands often in an otherwise uniform film. These were probably small crystallites of converted oxide which were surrounded by unreacted oxide in regions where there was insufficient ZnF_2. Diffraction patterns of the uniform areas showed them to be amorphous whereas the diffractographs of the islands (Fig 5-4c) were characteristic of a crystalline material. The apparently random nature of the multitude of diffraction spots was probably at least partly due to the instability of the films. The selected area of diffraction was about one micron square. The presence of diffraction spots rather than rings indicates that only a few crystallites were being illuminated by the electron beam. A mean crystallite dimension would then be about 0.1- 0.5 microns which is consistent with each island of Fig 5-4 being a single crystallite.

5.3.3. Reflection Diffraction Studies.

A reflection technique offers the advantage that thin surface films can be examined without removing them from their substrates. The diffraction patterns from single crystal surfaces are however very difficult to interpret. Polycrystalline arc patterns present difficulties in analysis because of the inevitable shadowing of the innermost arcs that occurs and the uncertainty in the position of the central undiffracted spot. The exact distance between the photographic plate and that part of the specimen from which diffraction occurs is also difficult to determine.

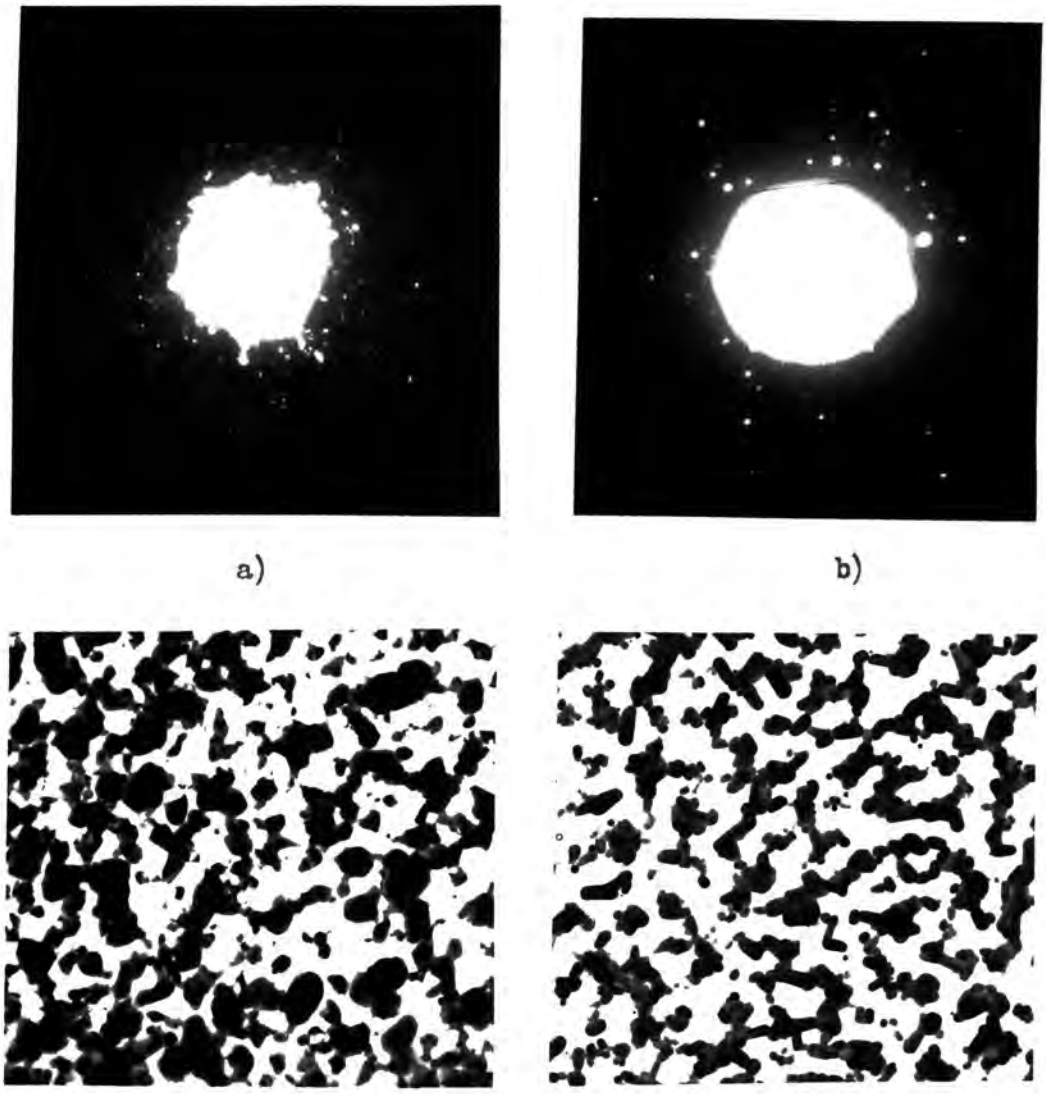


Fig 5-4 Transmission electron diffraction patterns and corresponding electron micrographs of two willemite films prepared from oxide layers which had first been detached from the silicon

For reflection microscopy the silicon specimens were stuck to the reflection stage with 'Aquadag' to prevent excessive charging of the specimen. Nevertheless charging sometimes occurred for specimens with oxide coating greater than about 0.5 micron which could not therefore be studied. This could have been overcome with a charge neutralizer for the microscope but this was not available at the time.

The reflection stage was capable of movement in two directions perpendicular to the electron beam and it could also be tilted about one of these directions and rotated about an axis normal to the specimen surface. In this way the specimen was positioned at a few degrees glancing angle to the electron beam so that the forward diffracted electrons were focussed on to the fluorescent screen or a photographic plate as required.

The reflection diffraction pattern from a bare (111) surface of silicon is shown in Fig 5-6. The Kikuchi lines produced are representative of a perfect single crystal. For all the thermally oxidised silicon surfaces which were examined a diffuse halo diffraction pattern (Fig 5-7) characteristic of an amorphous structure was obtained. The willemite films produced diffraction ring patterns which will be described later. These are characteristic of a polycrystalline material.

The area of the specimen from which reflection difraction occurs is of the order of 1 mm^2 due to the grazing incidence of the beam. This is extremely large compared with selected area transmission

Fig 5-6 Reflection electron diffraction pattern from the (111) surface of silicon

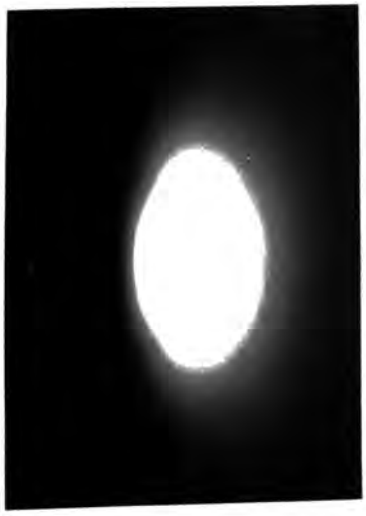


Fig 5-7 Reflection electron diffraction pattern from the oxidised surface of a silicon substrate

diffraction where specimen areas as small as one micron square can be defined by the use of apertures. A specimen which produces a spotty diffraction pattern in transmission (characteristic of diffractions from only a few crystallites) would produce broad rings in reflection because as many as 10^6 times more crystallites would be illuminated by the beam.

The thickness of film which is examined in reflection is undefined because of the uncertainty involved in assessing the depth of penetration of the electrons normal to the surface. Some idea however can be obtained from the fact that 200 Å thick silicate films often showed faint Kikuchi lines from the underlying silicon in reflection when examined.

Because of the difficulties (described in Sec 5.4.2) with films removed from silicon the reflection diffraction technique was the primarily method used for the identification and analysis of the willemite films. The ring pattern is analysed by a combination of Bragg's law and a simple geometrical construction which for small angles leads to the standard formula for the interplanar distance

$$d = \frac{n\lambda L}{r} \qquad 5\text{-}1$$

where L is the distance between the photographic plate and the diffracting part of the specimen, λ is the wavelength of the electrons (0.0334 Å for energies of 120 keV), r is the radius of the diffraction ring and n is an integer.

Diffraction patterns of several specimens are shown in Fig 5.8 ~~and 9~~ together with microdensitometer traces corresponding

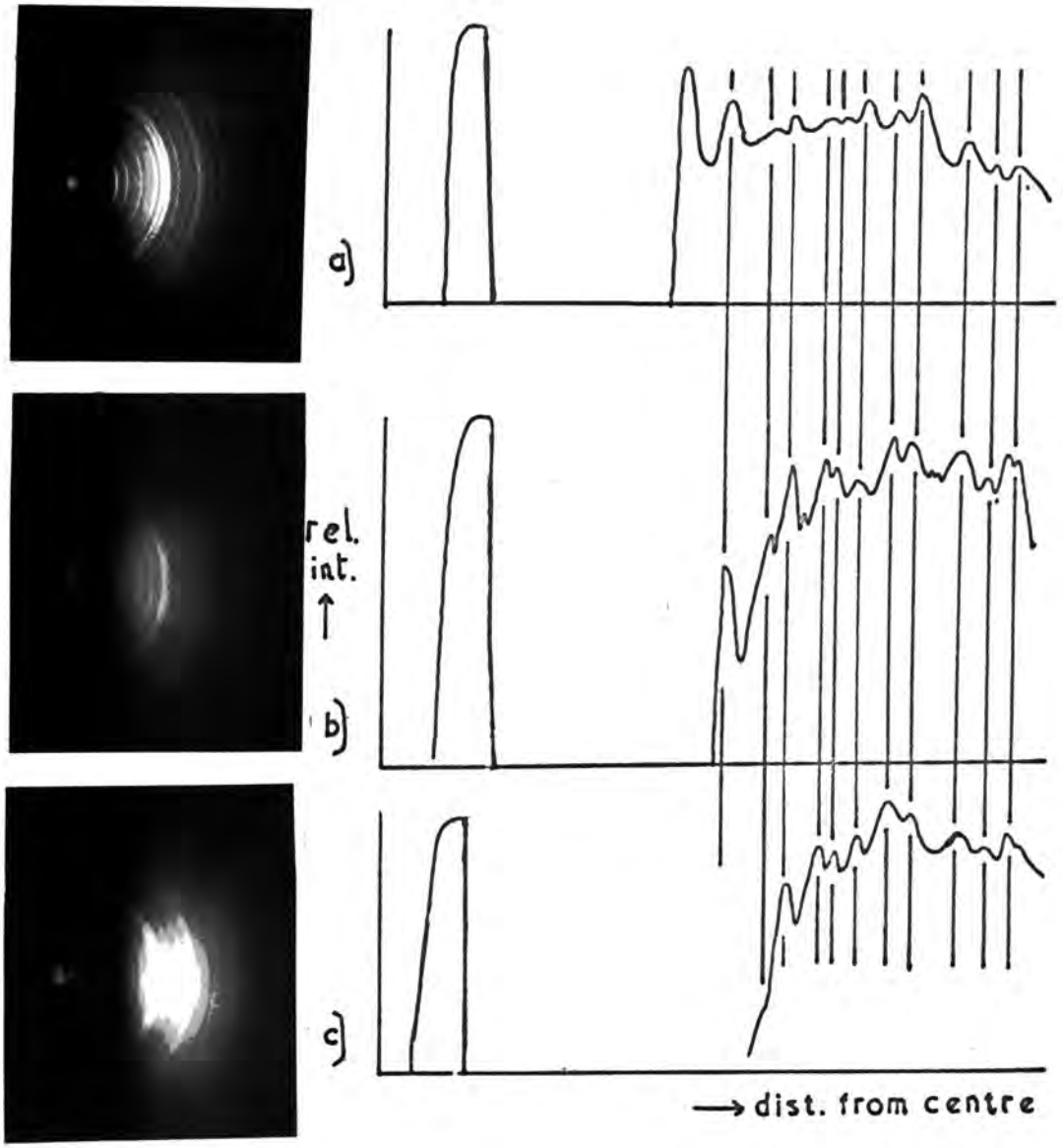


Fig 5-8 Reflection electron diffraction patterns of willemite films and the corresponding microdensitometer traces. The specimen number, silicon type, oxide thickness and firing conditions are given below for each film.

a) EW 118, n-Si, 1200 Å, 15 min at 1100 °C in O_2.
b) EW 107, p-Si, 1000 Å, 5 min at 1100°C in O_2. no Mn in ZnF_2.
c) EW 104, p-Si, 800 Å, 5 min at 1100°C in O_2.

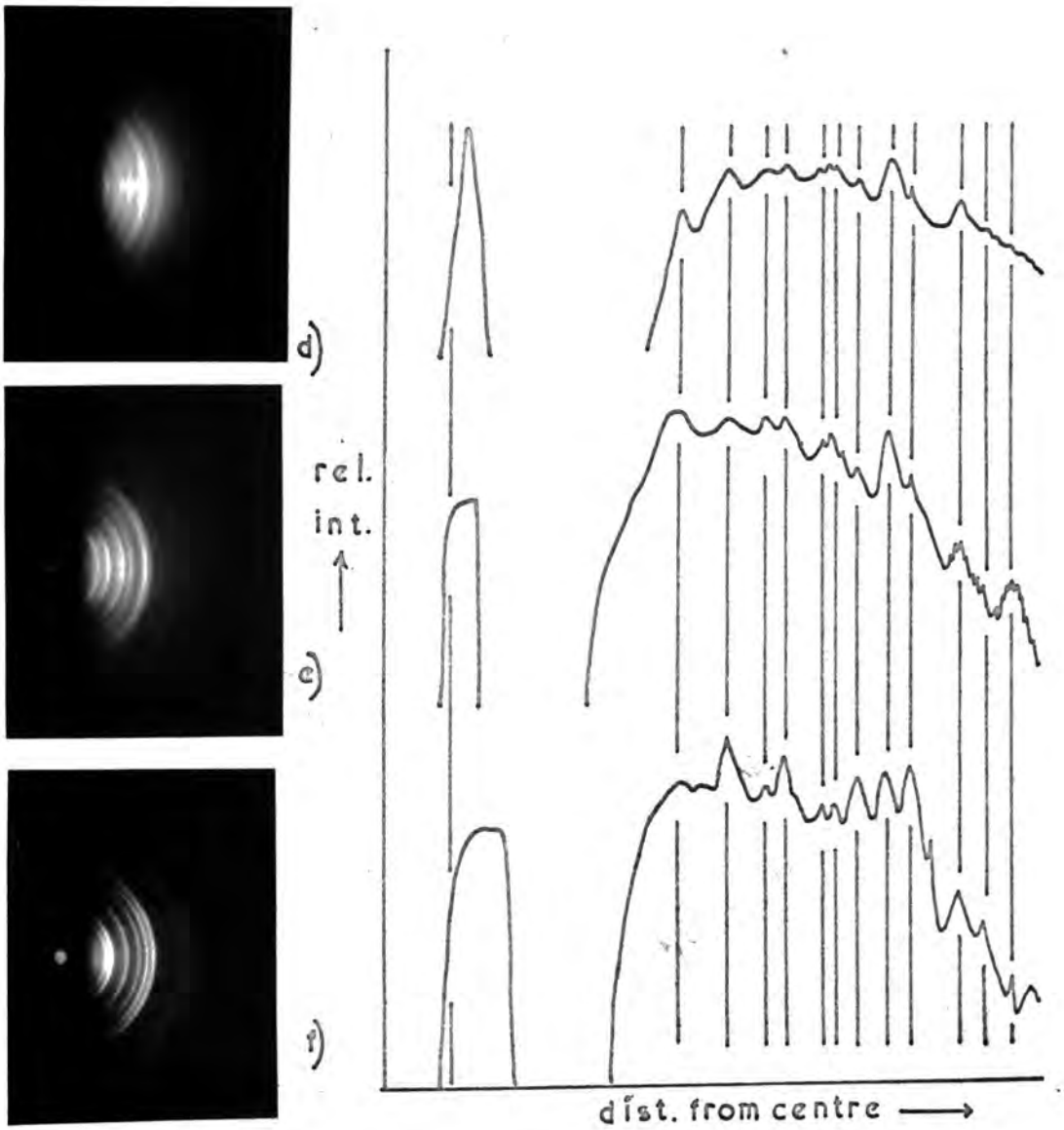


Fig 5-8 cont.)

d) EW76, p-Si, 1000 Å, 1000 °C in O_2.

e) EW103, p-Si, no oxide, 1100°C in O_2.

f) EW119, n-Si, 1000 Å, 900°C in air.

to a scan scross the plate through the central spot. The overall shape of each densitometer trace changes due to variations in the background density of the plates because of different amounts of inelastic scattering from the specimen surface and the varying degrees of plate exposure. Some of the rings were also slightly spotty. In spite of this the agreement between the relative positions of the peaks, corresponding to diffraction rings, is very good.

Due to the uncertainty in the location of the central spot of each diffractograph, the positions(s) of the peaks in all the traces were measured relative to a common point, an unknown distance (x) from the exact centre of the rings. After allowing for the scaling factor between the microdensitometer chart and the plate (1.5 inches ≡ 1 mm/ respectively in this case) the radius of the rings is given by

$$r = s + x \qquad 5\text{-}2$$

and rearranging Eqn 5-1

$$s = \frac{n\lambda L}{d} - x \qquad 5\text{-}3$$

A graph of s against the reciprocal of the d-values obtained from the A.S.T.M. X-ray index card 8-492 for α Zn_2SiO_4 is shown in Fig 5-10. The d-values are shown on the abscissa with the relative intensities as obtained from the X-ray index, expressed on a linear scale 1-100. The straight line relationship obtained indicates that the positions of the diffraction rings are consistent with diffraction from α-Zn_2SiO_4. The gradient of the straight

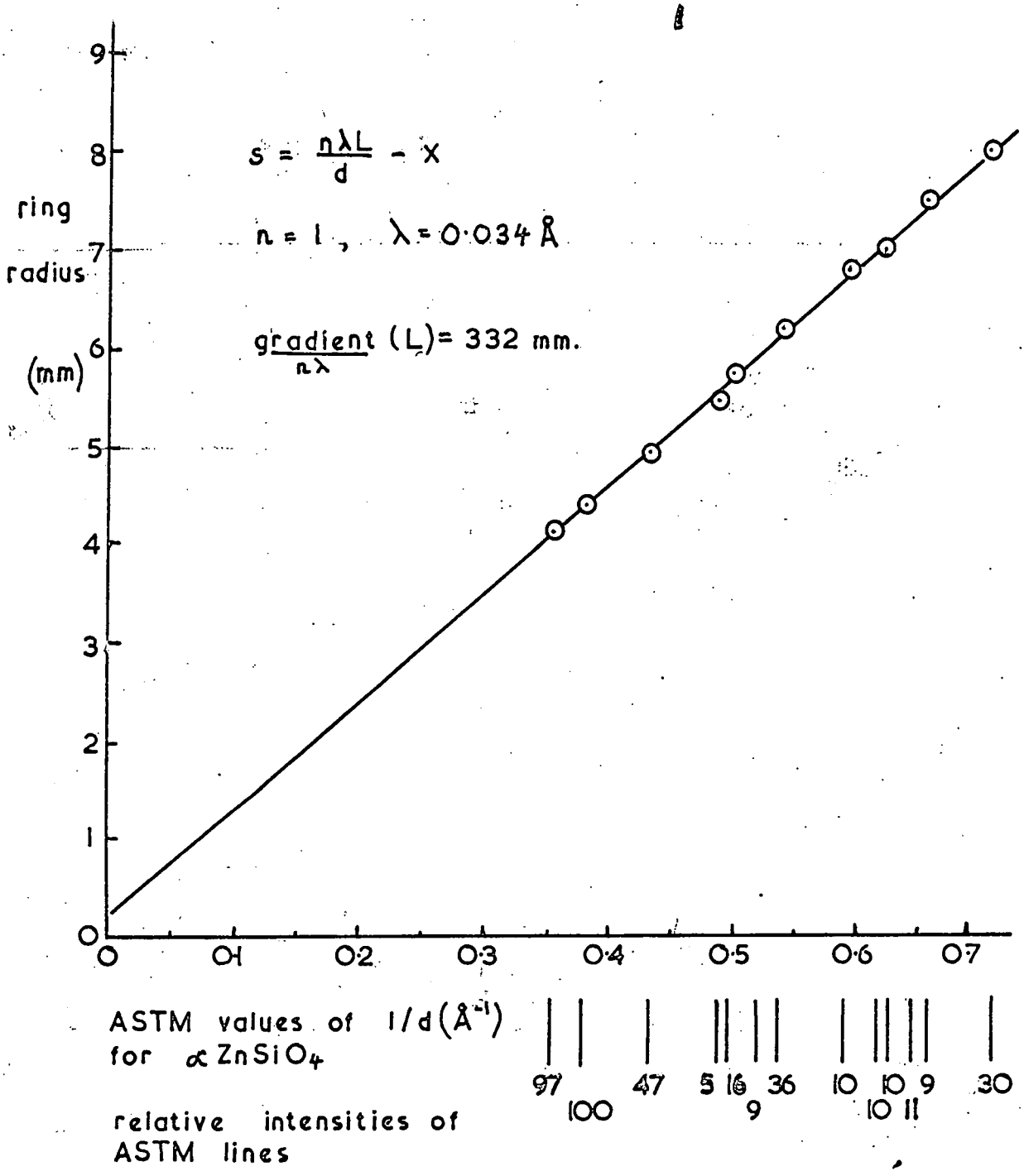


Fig 5-10 Graph of $1/d$ for αZn_2SiO_4 obtained from the ASTM index against the ring radii measured from the electron diffraction reflection pattern of a typical willemite film.

line ($n \lambda L$) was measured and for λ =0.0334 Å and n = 1 gives a value for the camera length, L, of 332 mm. The intercept on the ordinate represents x the difference between the estimated and true position of the centre spot. A calibration diffraction, using an evaporated gold film and adopting the same analysis procedure as described above, gave a value of 336 mm. for the camera length with n = 1, in good agreement with that obtained above. This was also consistent with JEOL information. Several willemite films which had been prepared under slightly different conditions were examined by reflection diffraction and the following general considerations about their structures can be noted.

i) Films which had been fired at temperatures between 900°C and 1100°C always showed the same general diffraction pattern. Fig 5-8 ~~and 9~~ shows a representative number of these.

ii) Films fired below the melting point of ZnF_2 produced a much more granular looking diffraction pattern which was probably due to the larger but more irregular crystallite growth likely in a solid state reaction.

iii) Films prepared on oxide-free silicon surfaces showed diffraction patterns which were indistinguishable from those of i) above.

iv) Manganese-free silicate films were non-luminescent but showed diffraction patterns similar to i) above.

v) No structural effects could be distinguished between films which had been produced on B-doped p-type and As doped n-type silicon.

vi) Films which had been grown on oxides thicker than about 0.5μ (and these were usually produced by steam oxidation) showed much more diffuse patterns, probably due to excess silica in the films and the consequent increase in the amorphous characteristics of the diffraction pattern.

vii) A few phosphor films were produced which gave yellow cathodoluminescence (see Sec 8.1). These were prepared by using very thin ZnF_2: Mn films in the reaction process. The diffraction patterns from these films were very diffuse and characteristic of an amorphous material.

viii) Apart from those mentioned in vii) all the films which were examined gave green cathodoluminescence in the electron beam.

5.4 X-Ray Diffraction Studies

Although the electron diffraction work described in Sec 5.3 showed that the reaction films had the general features of willemite, it could not be used to determine the interface spacing d, with very high accuracy. The minimum spacing Δd, between two d-values d_1 and d_2 (corresponding to ring radii r_1 and r_2) which can be resolved by diffraction in a perfect electron microscope can be estimated from Eqn 5-1 which gives

$$\Delta d = d_1 - d_2 = \lambda L \left(\frac{1}{r_1} - \frac{1}{r_2} \right) = \frac{\lambda L \Delta r}{r_1 r_2} \qquad 5\text{-}4$$

For the J.E.M. 120 high resolution diffraction stage and 120 kV electrons λL = 11 Å-mm and an average value of r was about 10 mm. The accuracy with which the radius of the rings could be measured, Δr ,

was about 0.1 mm. Substituting these values in the above equation gives a value for Δd of about 0.05 Å which is at least an order of magnitude greater than that obtainable with conventional X-ray diffraction. Consequently in order to absolutely identify a material such as willemite (which has many closely spaced d-values) an X-ray diffraction experiment is required.

The willemite films produced on silicon substrates were too thin to produce any measurable diffraction effects with X-rays and therefore a powder technique was adopted. A mixture of 99.9999% SiO_2 (obtained from Koch Light Ltd.) and ZnF_2:Mn in the appropriate molar proportions was finely ground together and baked in exactly the same manner as in the preparation of willemite films on silicon. X-ray powder photographs were taken, using the standard Debye-Sherrer technique, both before and after firing the mixture and these are shown in fig.5-11 together with a reference powder photograph of some commercial willemite powder phosphor (Type P1 obtained from Derby Luminescents Ltd). The diffraction lines from the fired powder agreed exactly with those of the reference and further confirmation that these lines were due to willemite was carried out by the standard X-ray diffraction analysis techniques. The fired powder mixture was green cathodoluminescent like the commercial phosphor

5.5 Observations with the scanning electron microscope and the microprobe

5.5.1 Introduction

Scanning electron instruments employ a finely focussed beam of electrons (about 1-2 microns in diameter) which is scanned in a

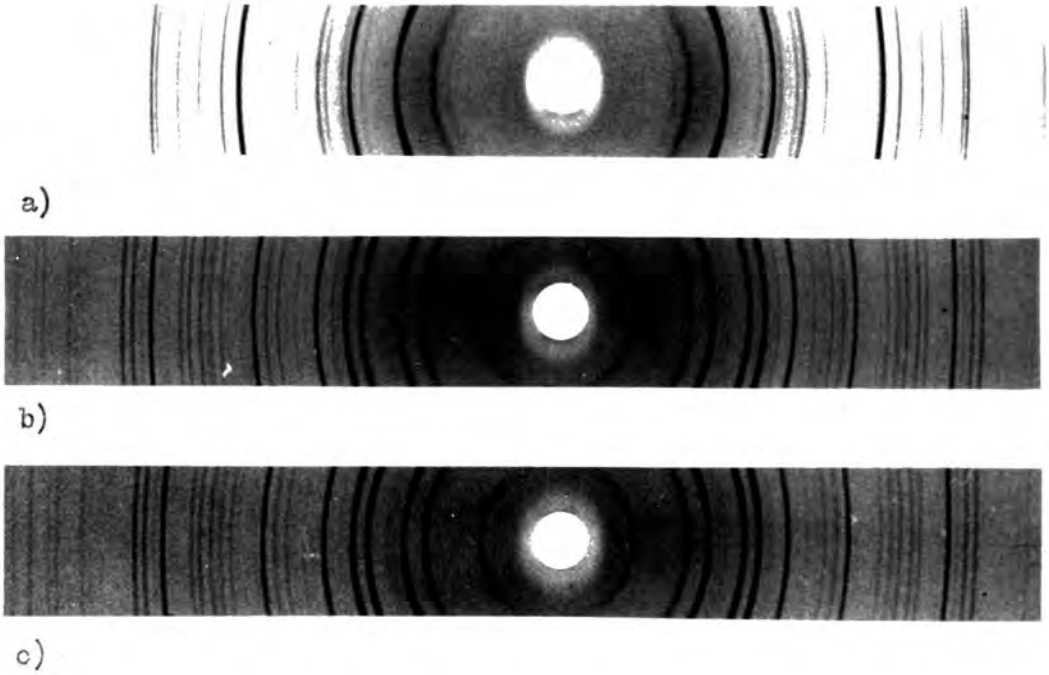


Fig 5-11 X-ray powder photographs of :-

a) a mixture of ZnF_2:Mn and SiO_2 in the proportions of 2:1 by weight

b) the same mixture after a bake at 1100°C in oxygen for 10 min

c) a commercial willemite phosphor powder (P1)

raster across the surface of the specimen. The scan generator which controls this raster also generates a raster on an oscilloscope, in sychronism with the primary beam. The various signals which are produced by the interaction of the primary electrons with the specimen are then used to intensity modulate the oscilloscope spot, so that pictures of the spatial distribution of these signals throughout the specimen are obtained. This information is usually photographed on Polaroid film. Line scans of the primary beam along the X-direction can also be obtained. In this case the deflection of the oscilloscope spot in the Y-direction is controlled by the intensity of the beam-induced signal from the sample. A chart output is possible with this method.

The modes of operation of these instruments, depending on the various types of interaction of the primary beam with the specimen, are as follows :-

i) Secondary and reflected primary electrons can be used to obtain information about the surface of the specimen. The emitted electrons are collected and accelerated into a scintillator which emits light. This light is then fed by a light pipe to a photomultiplier and after amplification this signal is used to intensity modulate the oscilloscope spot. The mechanisms of electron emission enable contrast in the micrograph to be obtained from topographical features and from the composition of the surface. The potential of the surface can also effect the intensity of electron emission.

ii) Fluorescent X-rays, excited by the primary beam and character-

istic of the elements within the specimen, can be used to determine the chemical composition of the specimen and its variation throughout the specimen. A Bragg spectrometer is used to select the wavelength which is required and the intensity of these X-rays is then used to control either the brightness of the oscilloscope spot or the ordinate of the chart output. Instruments designed primarily for this mode of operation are called electron probe microanalysers.

iii) The recombination radiation induced by the primary beam can be used in the study of luminescent materials. A photomultiplier can be used to measure the intensity of the light output as the electrons scan the specimen. Alternatively, the electron beam can be left unscanned and unfocussed. The cathodoluminescence is then photographed through the specimen-viewing microscope. This method is not possible in instruments with concentric electron and optic axes.

iv) Beam induced currents provide information about charge carrier mechanisms and those properties of the specimen that either enhance or impede the charge collection. This mode is often used in the study of semi-conductor devices where the depletion layers of p-n junctions provide suitable contrast effects.

The properties of willemite films on silicon corresponding to the first three modes of operation described above have been studied. In addition, both the cathodoluminescent and fluorescent X-ray modes have been used in conjunction with interferometry to assess the thickness of the phosphor films.

A scanning electron instrument was not available in the

Department, and therefore a prolonged study of the above features could not be undertaken. Three instruments at various other places were used and these are described briefly below :-

a) The Stereoscan (Cambridge Instruments Ltd.) is used primarily to display reflected and secondary electron images of specimen surfaces. The instrument at the Research Laboratories of J. Jobling Ltd. was used.

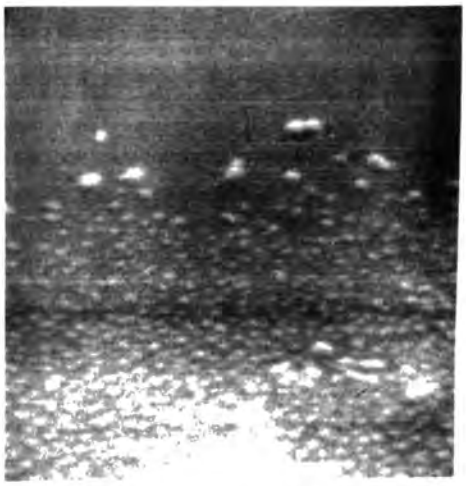
b) The Geoscan (Cambridge Instruments Ltd.) is an electron probe microanalyser which enables both X-ray and electron images to be obtained. The observation of cathodoluminescence was however, not possible with this instrument which belonged to the Geology Department, University of Durham.

c) The JXA-3A (Japan Electron Optic Laboratories Ltd.) is a microprobe similar to the Geoscan but with the added facility for the observation of cathodoluminescence. This instrument was used at the School of Engineering Science at Bangor.

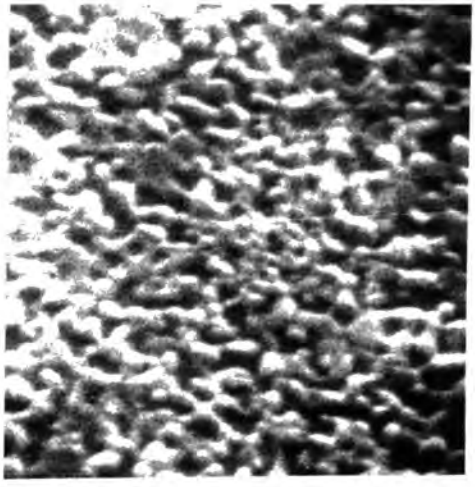
5.5.2 Results.

Fig 5-12 shows the reflected electron images of two luminescent willemite films obtained with the Stereoscan using electron energies of 20 kV. One of the films EW102 was produced from oxide 1000 Å thick whereas the other EW100 was produced on the bare silicon surface. ~~The reactive melt process employed in the preparation of the films is evident from the micrographs~~. The globule formation of the silicate on the bare silicon and at the edges of EW102 are presumably due to the scarcity of one of the constituents of the reaction in these regions.

a)



b)



c)

Fig 5-12 S.E.M. scattered electron images of willemite films

a) Specimen EW102 produced from 1000 Å thick oxide

b) A close-up of the edge of EW102

c) Specimen EW100 formed on a bare silicon surface.

Most of EW102 appeared uniform in the electron image. The 'crater' in the lower part of the micrograph was not typical of the film as a whole and was probably due to some contamination of the oxide surface prior to film formation. Globule formation can also be seen around this defect. Lack of contrast between the oxide (the upper regions of Fig 5-12 a and b) and the phosphor was probably due to the thinness of the films and also to the fact that their secondary emission coefficients are probably comparable (both have a similar structure and are insulators).

The cathodoluminescent image shown in Fig 8-3 was obtained with the JXA-3A using a beam current of 3.10^{-6} amp at 25 kV with the beam unfocussed and unscanned. The photograph was taken with a colour negative film (Kodacolor X) and printed to give a colour reproduction which was slightly blue compared with visual observations. The cathodoluminescent 'islands' are at the oxide-phosphor interface and correspond to the globules in the electron image described above. The green cathodoluminescence of the films was generally uniform with the occasional non-luminescent spots presumably due to some form of contamination. A blue luminescent spot is visible in the micrograph which is probably due to a luminescent contaminant. This luminescence and that of the films generally is discussed more fully in Sec 8.2.

The depth of the phosphor films was assessed using the Geoscan electron microprobe and a technique similar to the optical one described in Sec 5-2. Instead of measuring the extent of the cathodoluminescence down a bevel fluorescent X-rays characteristic of

the Zn Kα radiation were excited by a line scan across the bevel in the microprobe. Film EW115 (prepared from 1000 Å thick oxide) was bevelled to 1° and the length of the bevel, as measured by the X-rays, was compared with the length measured by the interference technique described in Sec 2.3. Within the limits of the interference technique (± 200 Å) both lengths were the same and this showed that all the oxide had been converted to a compound containing Zn atoms, presumably Zn_2SiO_4. The resolution of this technique is greater than with cathodoluminescence because of the fineness of the electron probe (about 1 micron). Light scattering effects within the film are also eliminated. Any smearing of the silicate down the bevel however, would still obscure the true depth of the phosphor film.

A cathodoluminescent image of the bevel was also obtained with the JXA-3A. A comparison of the length of the luminescent bevel with the X-ray scan confirmed the assessment of the phosphor depth given above.

The X-ray line scan also provided information about the spatial distribution of the zinc silicate in the film. Large variations in the intensity of the fluorescent X-rays appear to occur as the electron beam scans the film ($30\ \lambda\ \text{min}^{-1}$) and these could be interpreted as variations in the distribution of Zn_2SiO_4 within the film. The signal from the small number of Zn atoms in the thin film was however relatively weak and most of the variations on the trace are comparable with the relative noise level on the Zn deficient part of the trace (i.e. the bare silicon.). Variations larger than this

could be due to damage in the surface of the film introduced by the bevelling procedure.

5.6 Infra-red absorption studies

Thin films of infra-red active materials on silicon can conveniently be examined by transmission infra-red spectroscopy because of the relative transparency of the silicon substrates to infra-red light. The SiO_4 tetrahedra found in most glasses, silicates and silicon oxides retain their identity independently of the rest of the structure and absorb strongly in the infra-red due to the polar nature of the Si-O bond. The constraining nature of the structure surrounding these radicals and the degree of polymerisation of the tetrahedra produce absorption spectra which are characteristic of the material. The identification of zinc orthosilicate films on silicon was therefore attempted using this method.

A Grubb-Parsons double beam infra-red spectrometer in the Chemistry Department at Durham was used for these measurements. Specimens consisted of oxide and willemite films about 1500 Å thick grown on silicon substrates 200 microns thick with resistivities of about 2-3 ohm-cm. The absorption of these substrates was appreciable and to obtain a workable differential between the two beams and to reduce substrate effects an unoxidised silicon chip was inserted in one of the beams. Interference effects within the silicon slices produced oscillations in the output of the spectrometer and these have been averaged out in the spectra shown.

Fig 5-14 shows the infra-red absorption spectra of a

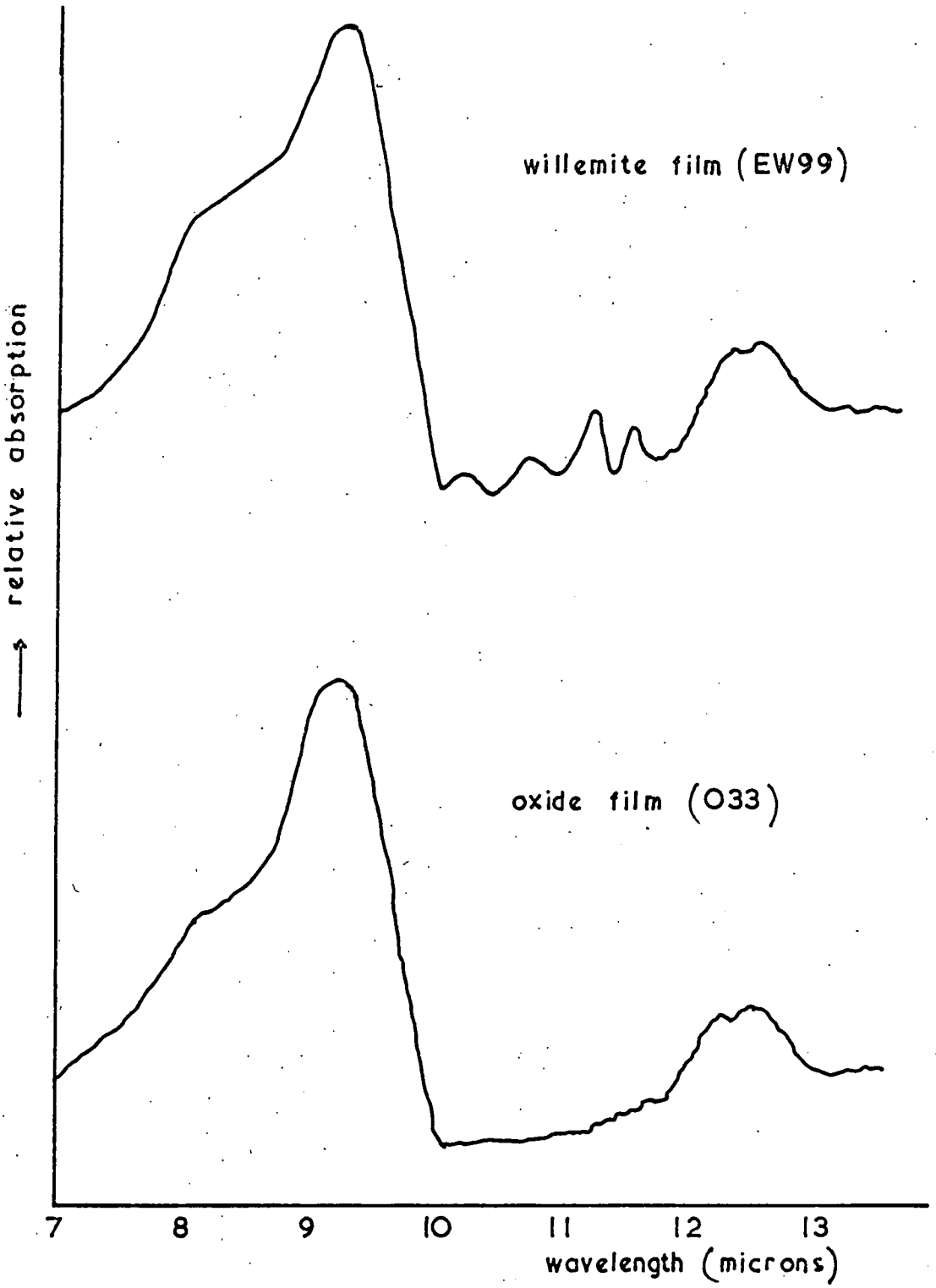


Fig 5-14 Infra red absorption spectra of a typical willemite film and the oxide layer from which it was made.

thermal oxide both before and after conversion. Unfortunately, the converted oxide still retained an oxide coating on one of its sides. The strong absorption characteristic of the Si-O vibrations for thermal oxides (9.2 microns and 12.5 microns) (Ref 5-8) are therefore visible in both spectra. In a molecule containing N atoms there are 3N - 6 independent degress of freedom but, for the regular tetrahedral configuration of the SiO_4 radical (N = 5), of the 9 possible degrees of freedom two are triply degenerate and one doubly degenerate, producing four separate vibrations (Ref 5-9). These have been identified for the mineral willemite most recently by Saskena (Ref 5-10) at 10.2 , 10.7 , 11.1 and 11.5 microns. These values agree very closely with those obtained from the spectra of the willemite films on silicon. An absorption band for willemite at about 17μ has also been identified as being due to vibrations of the ZnO_4 tetrahedra in the lattice (Ref 5-11). This could not however be resolved, due to the insensitivity of the infra-red spectrometer at these longer wavelengths.

5.7 Conclusions

The product of the reaction between silica and ZnF_2 has been conclusively identified by the X-ray powder technique as α-$Zn_2\ SiO_4$. Confirmation that willemite is also produced in the slightly different conditions of thin film formation has been achieved by using reflection electron diffraction and infra-red absorption. The films are polycrystalline consisting of a randomly oriented array of small crystallites with a mean dimension of about 0.1 - 0.5 microns.

A substrate reaction technique is capable of producing films which are thin and which have very good adhesion with their substrates which of necessity are chemically similar to the films themselves. These properties, although advantageous from the device point of view, increase the difficulties in assessing the structure of the films because neither X-ray diffraction nor the conventional electron diffraction techniques can be used.

Reflection electron diffraction was used as a non-destructive method of identification. The necessity for the examination of large areas with this method and the reduction in resolution due to the small crystallite size make absolute identification difficult particularly for complex substances with many closely spaced d-values. Infra-red spectroscopy however does not rely on the crystalline nature of the specimen but looks at the characteristic vibrational modes of the molecules involved. It is particularly convenient for the examination of thin films on silicon substrates because the transparency of high resistance silicon enables the simpler transmission technique to be used.

The microscopic mechanisms involved in the substrate reaction are not very clear but it seems that the reaction must be accompanied by the evolution of the gas SiF_4. Whether this seeps through from the deepest parts of the nascent phosphor film or whether gas evolution occurs on the surface and the silicate is formed in some metastable and perhaps molten form such as β-Zn_2SiO_4 before it finally crystallises into the small crystallites, are obviously very

difficult things to assess. However, it seems reasonable to assume that the evolution of this gas contributes to the formation of such small randomly oriented phosphor crystallites.

Depth studies have shown that this reaction can occur to depths of up to 1000 Å and that the Mn^{++} ion in this part of the film is situated in a crystalline field suitable for luminescence. The transmission electron diffraction studies indicate that crystallites of the silicate can exist surrounded by a matrix of amorphous silica. The similarities which exist between the structure of silica glass (i.e. a random array of interlinked SiO_4 tetrahedra) and the silicate itself do not preclude the possibility of some form of silicate-oxide interface bonding. The resolution of both the fluorescent X-ray and the cathodoluminescent microprobe techniques which were used however, was insufficient to distinguish between these small silicate crystallites in the phosphor film.

References for Chapter Five

5-1 W.L. Bragg and W.H. Zachariasen, Zeit. f. Krist. 8, 518 (1930)

5-2 T. Zoltai, Am. Mineralogist, 45, 960-73 (1960)

5-3 D.P. Grigor'ev, Fundamentals of the Constitution of Minerals, Isreal Program for Scientific Transl., Jerusalem (1964)

5-4 F.A. Kroger, Physica, 6, 764 (1939)

5-5 W.L. Bragg, Atomic Structures of Minerals, London (1937)

5-6 H.F.W. Taylor, Am. Mineralogist, 49, 932-44 (1962)

5-7 P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Paschley, M.J.Whelan Electron Microscopy of Thin Crystals, London (1965)

5-8 W.A. Pliskin, Proc. I.E.E.E. 52, 1468 (1964)

5-9 G. Hertzberg, The Infra-red Spectroscopy and Raman Spectroscopy p 100 London (1938)

5-10 B.D. Saskena, Trans. of Faraday Soc. 57, 250 (1961)

5-11 P. Tarte, Spectrochima Acta, 18, 471 (1962)

CHAPTER SIX

THE CAPACITANCE-VOLTAGE CHARACTERISTICS OF WILLEMITE DEVICES

6.1 Introduction

The capacitance-voltage (C-V) characteristics of metal-insulator-silicon (MIS) structures can be used to assess the quality of the insulator and the semiconductor-insulator interface. The essence of the technique lies in the comparison of experimental observations with idealised characteristics which are calculated on the basis of a simple model involving space charge regions in the surface of the semiconductor (Sec 6.2).

Deviations from this simple theory arise, however, when the insulator and the interfaces are not perfect. For example, the insulator may contain charges of one kind or another and charge transport may also occur across the interfaces. The C-V curves are modified by these effects and this enables certain properties of the insulator-substrate structure to be deduced (Sec 6.3).

Willemite films on silicon are sufficiently insulating (the current-voltage characteristics are discussed in Chap 7) to enable them to be used as the dielectric in MIS devices and the C-V characteristics obtained with these devices are discussed in Section 6.5. MOS structures made with unconverted oxide adjacent to the willemite films were also measured for comparison purposes. The properties of the silicon oxide-silicon system are so well-known that they were used to establish the reliability of the measurement technique which is described in Section 6.4.

6.2 The Ideal MIS Capacitor and its C-V Characteristics

6.2.1 Introduction

The theory of the ideal MIS capacitor, based on the formation of surface space charge regions in the semiconductor is now well established. The effects of a distribution of charge at a semiconductor surface were first considered by Shockley in 1949 (Ref 6-1). The mathematical relationships cannot be evaluated explicitly and later works (Ref 6-2 and 3) extended the treatment by using numerical methods.

The development of silicon technology and particularly of methods for growing controlled surface oxides on silicon led to the production of MOS structures. It was soon after these were proposed as voltage variable capacitors (Ref 6-4 and 5) that they were first used in the study of the silicon oxide-silicon interface (Ref 6-6,7 and 10). More recently the theory of the MIS capacitor has been developed by Grove et al. (Ref 6-9) to bring out the physical significance of the mathematical results.

The theory is briefly presented here in order to establish the physical model and to develop the formulae and notation which will be used in the interpretation of the experimental C-V results (Sec 6-5). The theory is based on a p-type semiconductor but it can be extended to n-type material by the appropriate changes in signs and polarities of the parameters involved.

6.2.2 The physical model

The energy band diagram of an ideal MIS structure with a p-type semiconductor is shown in Fig 6-1, for three different bias conditions. The diagram incorporates the following assumptions:-

i) The work functions of the metal electrode and the semiconductor are equal.

ii) The insulator and its interface are perfect and free from space charges.

iii) There is equilibrium in the semiconductor so that its Fermi level (E_F) remains constant constant regardless of the applied voltage V_g.

iv) The charge layers are of uniform density through their thickness.

The surface potential, ϕ_s, is represented by the bending of the bands from the flat-band condition (at which $\phi_s = 0$) and is measured positive downwards. The position of the Fermi level in the bulk of the semiconductor, ϕ_f, is measured with respect to E_i, the intrinsic Fermi energy and is positive in this case. All capacitances are related to unit area.

With a negative voltage applied to the gate, holes accumulate at the surface (Fig 1a) and under these conditions the capacitance of the device is solely that of the oxide, C_o.

For a slight positive bias (Fig 1b) holes are repelled from the surface leaving behind a negative space charge of uncompensated acceptor ions. As the concentration of electrons (n_p) in this

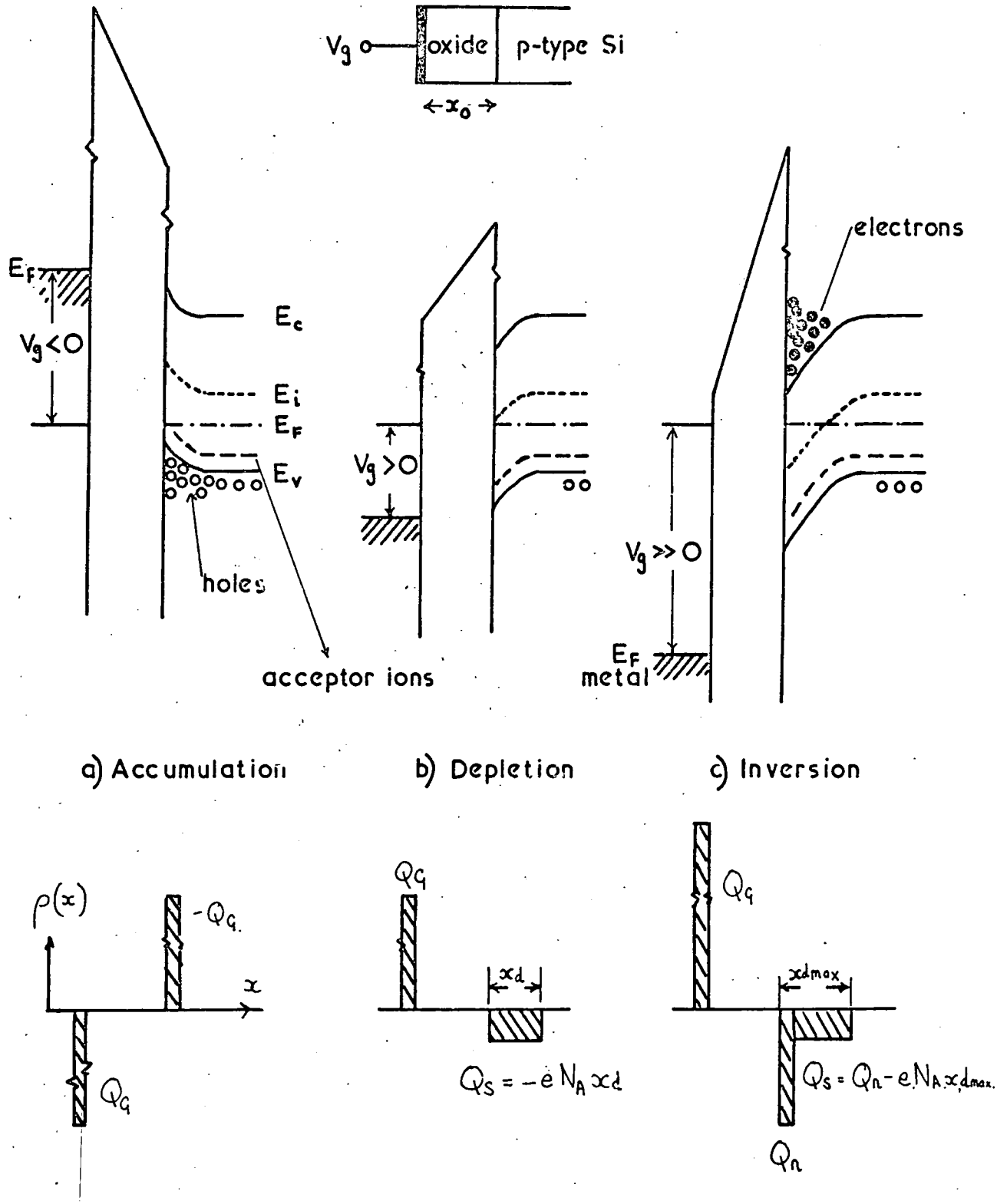


Fig 6-1 Energy band and charge distribution of ideal MOS structure, for various bias conditions

region is so small that it can be neglected the surface is depleted of carriers and the charge per unit area in the space charge region is

$$Q_s = - e.N_a.x_d \qquad 6\text{-}1$$

where N_a is the density of acceptor ions, e the electronic charge and x_d the width of the depletion region.

If larger positive voltages are applied to the gate (Fig 1c) the bending of the bands increases until, as the conduction band approaches the Fermi level, the surface inverts to n-type, i.e. the concentration of electrons at the surface (n_p) is greater than n_i, the intrinsic carrier concentration. With a small amount of inversion the space charge in the semiconductor is still dominated by the exposed acceptor ions but, with increasing voltage more and more electrons occupy the conduction band and a point is reached when $n_p = N_a$. The intrinsic Fermi level is then as far below E_F as, in equilibrium, it was above and

$$\varphi_s > 2\varphi_F \qquad 6\text{-}2$$

For higher applied voltages the additional space charge induced in the semiconductor is accommodated largely by the electrons in the narrow inversion region and not by any appreciable increase in the space charge of the depletion layer. This is because of the very rapid rise of n_p with energy due to the Boltzmann relation. The width of the depletion region therefore reaches a maximum x_{dmax},

and the space charge in the semiconductor.

$$Q_s = Q_n + e.N_a.x_{dmax} \qquad 6\text{-}3$$

where Q_n is the charge per unit area of electrons in the inversion region.

The total capacitance, C, of the MIS structure in the depletion and inversion regions is the capacitance due to the space charge region C_s in series with that of the insulator, i.e.

$$\frac{1}{C} = \frac{1}{C_o} + \frac{1}{C_s} \qquad 6\text{-}4$$

where

$$C_s = \frac{dQ_s}{d\phi_s} \qquad 6\text{-}5$$

and

$$C_o = \frac{K_o.\epsilon_o}{x_o} = \frac{Q_s}{V_o} \qquad 6\text{-}6$$

ϵ_o is the permittivity of free space , K_o the dielectric constant of the oxide x_o is the oxide thickness and V_o is that part of V_g which appears across the oxide so that

$$V_g = V_o + \phi_s \qquad 6\text{-}7$$

6.2.3 The net surface space charge

The total space charge in the semiconductor is

$$Q_s = \int_0^\infty \rho(x)\,dx = \int_{u_s}^0 \frac{\rho(u)\,du}{du/dx} \qquad 6\text{-}8$$

where $\rho(x)$ is the charge distribution in a direction normal to the surface and u is the reduced electrostatic potential in the semiconductor (e.g. $u_s = \phi_s e/kT$ where k is Boltzmann's constant and T the absolute temperature).

Poisson's equation gives

$$d^2u/dx^2 = \frac{\rho(u)}{K_s \epsilon_o} \qquad 6\text{–}9$$

where K_s is the dielectric constant of the semiconductor. The Boltzmann distributions of electrons and holes are respectively

$$n_p = n_i \exp(u - u_F) \qquad 6\text{–}10$$

$$p_p = n_i \exp(u_F - u) \qquad 6\text{–}11$$

In the bulk of the semiconductor (where u = 0) charge neutrality exists

$$\rho(0) = (p_{p_o} - n_{p_o} - N_a).e = 0 \qquad 6\text{–}12$$

and with Eqn 6–10, 6–11 ~~and 6–13~~ this gives

$$N_a = 2 n_i \sinh(u_F) \qquad 6\text{–}13$$

Assuming that $p_{p_o} > n_{p_o}$ then

$$N_a = 2 n_i \cosh(u_F) \qquad 6\text{–}14$$

In general, from Eqn 6–10, 6–11 and 6–13

$$\rho(u) = 2e n_i \left(\sinh(u_F - u) - \sinh(u_F) \right) \qquad 6\text{–}15$$

Integrating Poisson's equation once provides du/dx which can than be substituted in Eqns 6–8 with 6–15 to give after integration

$$Q_s = -2e n_i L_D F(u, u_F) \qquad 6\text{–}16$$

where the intrinsic Debye length,

$$L_D = \left(\frac{kT}{e}\right)^{1/2} \left[\frac{K_s \epsilon_o}{2 e n_i}\right]^{1/2} \qquad 6\text{-}17$$

and

$$F(u, u_F) = \left[2\left[\cosh(u-u_F) - \cosh(u_F) + u.\sinh(u_F)\right]\right]^{1/2} \qquad 6\text{-}18$$

The part of Q_s which is due to electrons in the inversion region can also be calculated in a similar manner to give

$$Q_n = \int_0^{x_i} n(x)\,dx = \int_{u_s}^{u_F} \frac{n(u)\,du}{du/dx} \qquad 6\text{-}19$$

where x_i is the depth of the inversion region inside the surface as defined by the intrinsic boundary where $u = u_F$.

The capacitance of the ideal MIS device in the depletion and inversion regions can be obtained exactly by differentiating the charge relations, Eqns 6-16 and 6-17 together with Eqns 6-5 and 6-4. The process is complicated due to the explicit nature of the functions involved and numerical methods are necessary. Fortunately, graphical representations, obtained by computer, are available (Ref6-8).

6.2.4 The depletion approximation

A good approximation to the exact calculation of the capacitance is possible for the range of applied voltages in which the surface is depleted. In the depletion region Eqn 6-1 provides a realistic value for Q_s which can be used with Poisson's equation to give a value for the surface potential

$$\phi_s = \frac{e N_a x_d^2}{2 K_s \epsilon_o} \qquad 6\text{-}20$$

A combination of this with Eqns 6-4, 6-5, 6-6 and 6-7, after some

manipulation gives

$$\frac{C}{C_o} = \left[1 + \frac{2 K_o^2 \epsilon_o V_g}{e N_a K_s x_o^2} \right]^{-1/2} \qquad 6\text{-}21$$

This equation shows that the capacitance of the MIS structure will fall with a positive voltage on the top contact increasing from zero. This fall will stop as soon as the surface starts to become inverted, beyond which the depletion layer width will remain constant. The theoretical capacitance variation is shown in Fig 6-2.

6.2.5 Variation of capacitance with frequency

The capacitance of the MIS structure varies with frequency when it is biased into the inversion region. The computer calculation of capacitance in the inversion region, described above, assumes that all the charges contributing to the space charge take part in the measurement process and that they are capable of moving at the same frequency as the a.c. measurement signal. At high frequencies this is only possible for the majority carriers which have dielectric relaxation times of about 10^{-9} sec. The ability of the minority carriers to follow the measurement signal is governed by their generation-recombination rates and by their mobility as they move in and out of the inversion region passing from the carrier rich bulk of the semiconductor through the surrounding depletion layer. These processes are too slow for the minority carriers to contribute to the a.c. capacitance at high frequencies. The minority carriers do, however, limit the width of the depletion layer to a maximum,

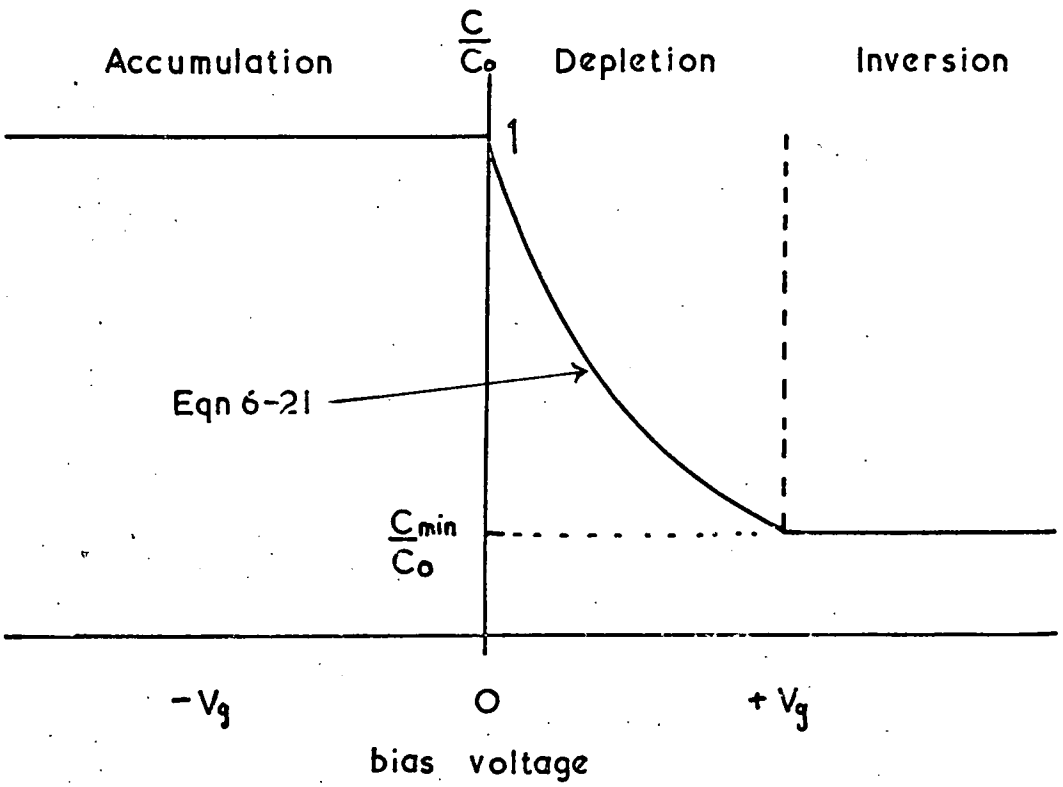


Fig 6-2 Idealised C-V characteristics of an MOS structure on p-type silicon

x_{dmax}, as has been explained above. This means that the capacitance of the MIS device, after the onset of strong inversion, levels out to a minimum (C_{min}) which is independent of V_g.

Using Eqns 6-5 and 6-20, the high frequency capacitance per unit area of the space charge layer in the inversion bias region is

$$(C_s)_{min} = \frac{K_s \epsilon_o}{x_{dmax}} \qquad 6\text{-}22$$

Substituting for x_{dmax} from Eqn 6-20 and applying Eqn 6-2 as the onset condition for inversion

$$(C_s)_{min} = \frac{1}{2}\left[\frac{K_s \epsilon_o e N_a}{(kT/e)\ln(N_a/n_i)}\right]^{1/2} \qquad 6\text{-}23$$

where

$$\ln(N_a/n_i) = e.\phi_F/kT$$

from Eqn 6-14 and 6-13. This can then be combined with Eqn 6-4 to give the minimum capacitance of the device

$$C_{min} = \frac{C_o.(C_s)_{min}}{(C_o + (C_s)_{min})} \qquad 6\text{-}24$$

At low frequencies, usually less than about 100 Hz, the minority carriers can follow the measurement signal. With increasing positive bias more and more electrons then fill the inversion layer and C_s therefore increases again so that the total capacitance of the device (the series combination of C_s and C_o) increases to C_o.

The low frequency C-V characteristics are also observed at higher frequencies than normal when the minority carrier generation-recombination rates are increased either by light or heat. Fig 6-3 shows such an effect in an experimental C-V plot, obtained by the author, for an MOS device at 159Hz. The low frequency type characteristics were obtained when the structure was illuminated but reverted to the high frequency type in the dark.

6.2.6 The flat-band capacitance

The flat-band capacitance, C_{FB}, is the capacitance of the device when no potential appears across the space charge region. The energy bands are then flat and $\phi_s = 0$. Further information on the space charge properties of the insulator and the interface can be obtained by measuring the voltage (V_{FB}) required to bring about flat-band conditions. In the ideal case discussed above $V_{FB} = 0$ However, in practice non-ideal characteristics are usually observed and V_{FB} rarely equals zero.

6.3 Non-ideal C-V characteristics

6.3.1 The flat-band capacitance

It is convenient to discuss here a practical method of calculating C_{FB} from an experimental C-V curve (Sec 6-5) based on the measurement of the maximum capacitance (C_o) and the minimum high frequency capacitance (C_{min}). This follows the treatment by Lehovec (Ref 6-11).

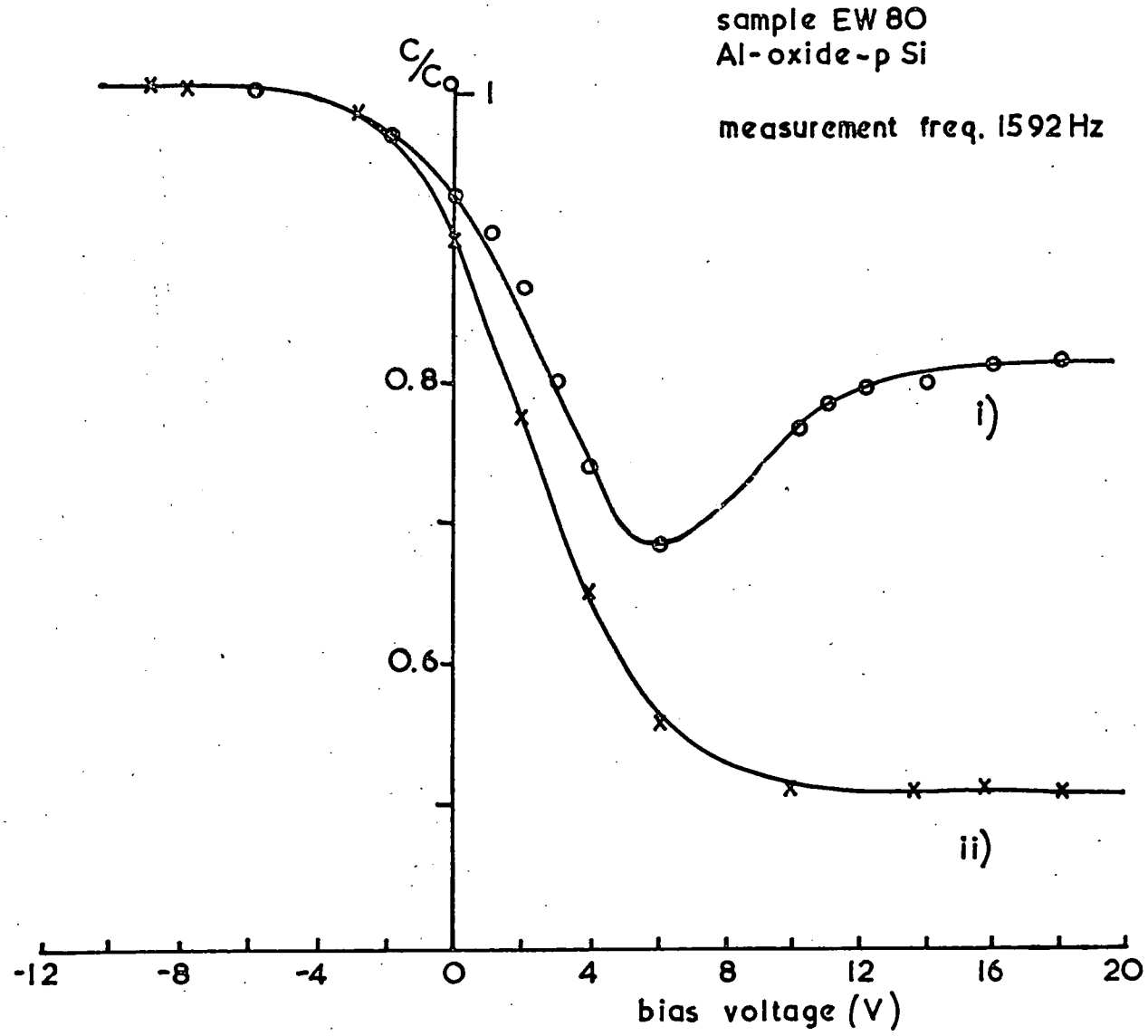


Fig 6-3 C V curves for an MOS device i) illuminated
ii) in dark

In the flat-band region, the surface is not depleted and the total surface charge in the semiconductor must be considered by means of Eqn 6-16. If $p_p \gg n_p$ in the bulk (and this is usually the case for a p-type semi-conductor at room temperature) then Eqn 6-13 and 6-14 hold and substituting these Eqn 6-16 yields

$$Q_s = (kT/e)^{1/2} (2 N_a . e K_s \epsilon_o)^{1/2} (u + e^{-u} - 1)^{1/2} \qquad 6\text{-}25$$

The capacitance of the space charge region in the vicinity of the flat-band condition is given by Eqn 6-5 which, in terms of reduced potentials gives

$$(C_s)_{FB} = (e/kT) \left. dQ_s/du \right|_{u \to 0} \qquad 6\text{-}26$$

Differentiation of Q_s, in the limit as u tends to zero, can be accomplished by using the series expansion of the exponentials ($u \ll 0$). Substituting this in Eqn 6-26 above gives

$$(C_s)_{FB} = (N_a . e K_s \epsilon_o)^{1/2} (e/kT)^{1/2} \qquad 6\text{-}27$$

and from Eqn 6-23 and 6-24 this becomes

$$(C_s)_{FB} = 2 (\ln(N_a/n_i))^{1/2} \frac{C_o . C_{min}}{(C_o - C_{min})} \qquad 6\text{-}28$$

The capacitance of the device at flat-band is obtained by inserting this equation into Eqn 6-4. The resulting expression can be written in the form

$$\frac{(C_o - C_{FB})}{(C_o - C_{min})} = \frac{C_o}{C_o + C_{min}\left[2(\ln(N_a/n_i))^{1/2} - 1\right]} \qquad 6\text{-}29$$

In this expression $\left[2 \left(\ln (N_a / n_i) \right)^{1/2} - 1 \right]$

is relatively insensitive to the value of N_a, ranging from 6.3 for 1 Ω-cm and p-type silicon at room temperature, to 5.8 for 10 Ω-cm material (taking $n_i = 1.45 \cdot 10^{10}$ cm^{-3} at room temperature). The above expression for C_{FB} can then be approximated to within an accuracy of a few per cent to

$$\frac{C_o - C_{FB}}{C_o - C_{min}} \simeq \frac{C_o}{C_o + 6 C_{min}} \qquad 6\text{–}30$$

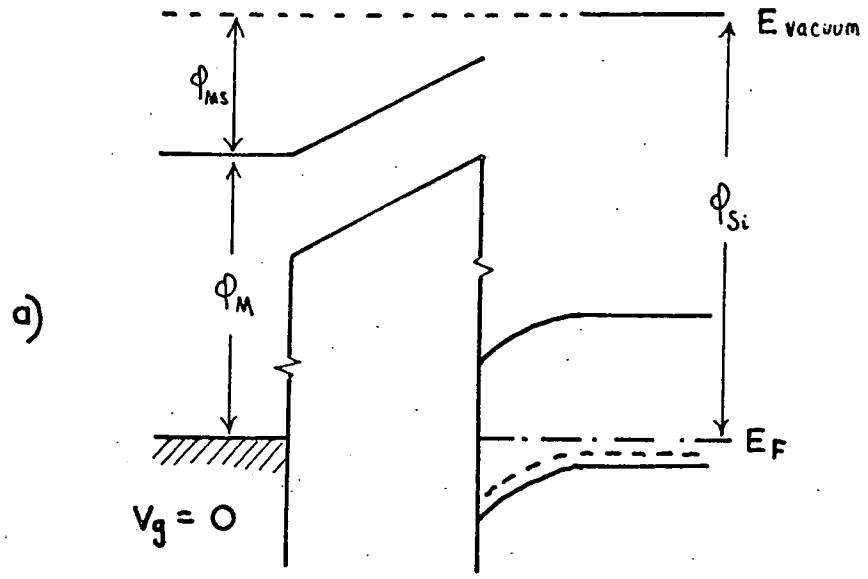
This provides a very convenient way of measuring C_{FB} and V_{FB} which does not depend on an accurate knowledge of N_a, so that effects such as the redistribution of impurities during oxidation, can be neglected.

~~6.3 Non-ideal C-V characteristics~~

6.3.2 Work function differences and charges in the insulator

As discussed in Sec 6–2 space charge can be induced within the surface of a semiconductor by the application of a suitable bias to the gate of a MIS structure. However, such space charges can also be produced by charges inside the insulator or by a work function difference ϕ_{MS} between the gate electrode and the semiconductor. The effect of these adds to that of V_g and it results in a shift of the C-V characteristics along the voltage axis. This shift is represented by the value of the flat-band voltage V_{FB} ($V_{FB} = 0$ for the ideal case above).

Fig 6–4 shows the energy bands of an MIS structure



for flat band condition $V_g = V_{FB} = \phi_{Si} - \phi_M = \phi_{MS}$

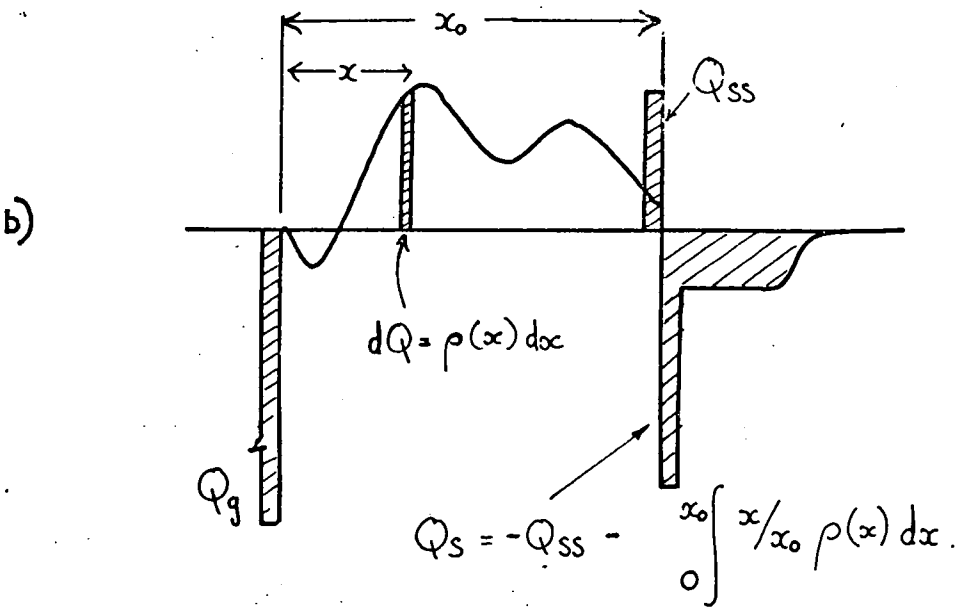


Fig 6-4 The energy bands and charge distribution of an MIS structure a) the effect of work function difference ϕ_{MS} b) the effect of a fixed charge Q_{SS} and an arbitrary one $\rho(x)\,dx$.

with the effect of ϕ_{MS} included together with the effects of an interface charge $+Q_{SS}$ and an arbitrary space charge $\rho(x)$ distributed throughout the insulator. It can be shown by the use of Gauss' Law from the superposition of elemental charges of the arbitrary distribution that

$$V_{FB} = \phi_{MS} - Q_{SS}/C_o - {}^{1}/_{C_o} \int_0^{x_o} ({}^{x}/_{x_o})\, \rho(x)\, dx \qquad 6\text{-}31$$

A positive space charge therefore produces a negative shift in V_{FB} and vice versa.

The exact nature of the space charge in the insulator and the effect of charge movement on V_{FB} will be discussed together, with the experimental results in Sec 6.5. A value for V_{FB} can be obtained from an experimental C-V curve by using Eqn 6-30 to calculate C_{FB}, from C_{min} and C_o, and reading V_{FB} off the voltage scale. Eqn 6-31 can then be used with C_o to obtain a measure of the effective charge $(Q_{SS})_{eff}$ at the silicon-insulator interface.

$$(Q_{ss})_{eff} = (V_{FB} - \phi_{MS})\, C_o \quad C\,m^{-2} \qquad 6\text{-}32$$

Values of ϕ_{MS} for various metal/silicon combinations are given in Table 6-1.

6.3.3 Surface states

The charges considered above have been assumed to be constrained within the insulator. We will now consider the effects of interfacial surface states or traps that are capable of interchanging charge with the semiconductor. Since the charge in these states is governed by their proximity to the Fermi level it depends greatly

Metal	φ_M	φ_{MS} eV			
		p-type		n-type	
	eV	10 ohm-cm	1 ohm-cm	0.1 ohm-cm	1 ohm-cm
Gold	4.8	-0.08	-0.18	+0.6	+0.5
Indium	4.0	-0.88	-0.98	-0.2	-0.3
Aluminium	4.2	-0.68	-0.78	0	-0.1

Table 6-1 Various metal / silicon work function differences (The work functions have been obtained from Ref 6-28 and 6-29 respectively and φ_{MS} calculated from Eqn 6-38)

on the amount of band bending that occurs at the surface.

If the capacitance of the structure is measured at a sufficiently high frequency the surface states cannot respond to the measurement signal so that they can be considered as fixed charges whose only effect is to alter the amount of space charge in the depletion layer. As the energy bands bend downwards with increasing bias the amount of charge in the surface states varies and this produces shift in the C-V curve which itself varies with applied bias. The result is to distort the shape of the characteristics in a way which is governed by the distribution of surface states within that energy range swept by the Fermi level at the surface. (From the onset of depletion to strong inversion this is about $2\phi_F$). For a continuum of states in this range a continual change in the slope of the characteristics is therefore observed.

By measuring the slope of the C-V curve at V_{FB} it is possible to obtain a value for the effective density of surface states (D $cm^{-2}eV^{-1}$). The procedure for this is outlined briefly below, following an extension of the theory in Sec 6.2.6.

Assuming that the surface states are of the acceptor type (i.e. negative when occupied by an electron and neutral otherwise) and that the surface potential increases from the flat band position to a value ϕ then the increase in the interface charge is $-De\phi_e$ An extension of Eqns 6-7 and 6-6 to include this then gives

$$V_g = Q_s/C_o - \frac{D.ekT.u}{C_o} + \frac{ukT}{e} + \text{const.} \qquad 6\text{-}33$$

where the constant includes any interfacial charge at flat band and all other contributions to V_g not dependent on u.

Differentiating this yields

$$\frac{dV_g}{du} = \left[\frac{dQ_s}{du} - D.ekT \right] . \frac{1}{C_o} + \frac{kT}{e} \qquad 6\text{-}34$$

Differentiating Eqn 6-4 with respect to V_g and using Eqn 6-34 to change the variable from dV_g to du gives

$$\frac{dC}{dV_g} = \frac{(C_o - C)^2}{C_o^2} . \frac{dC_s}{dV_g} = \frac{(C_o - C)^2}{C_o^2} . \left[\frac{dQ_s}{du} - DekT - \frac{C_o kT}{e} \right]^{-1} \qquad 6\text{-}35$$

This can then be evaluated in the limits where $u \to 0$ (i.e. flat band) using Eqn 6-26 and 6-27, and substituting C_{FB} for C_s using Eqn 6-4 to give

$$D = \frac{(C_o - C_{FB}) . C_{FB}}{3\ dC/dV_g \big|_{V_{FB}} . ekT} - \frac{C_o^2}{(C_o - C_{FB})} . \frac{1}{e^2} . eV^{-1}cm^{-2} \qquad 6\text{-}36$$

where C_{FB} can be obtained from the experimental C-V curves using Eqn 6-30 and $(dC/dV_g)V_{FB}$ is the slope of these curves at the point (C_{FB}, V_{FB}). The value of D obtained, together with the effective density of charge at the interface $(Q_s)_{eff}$ can then be used to characterise the quality of the interface and of the insulator.

Apart from the fast surface states described above, traps can occur within the insulator which are sufficiently close to the interface to allow carrier exchange between the silicon (or the metal electrode) and the insulator, either by tunnelling (Ref 6-21) or by a hopping mechanism (Ref 6-15). These exchanges take place relatively slowly and do not effect the slope of the C-V curve but

can lead to drift of V_{FB} with time at a particular temperature. Also, due to the different probabilities for emptying and filling the traps and the different tunnelling range of the electrons and holes, hysterisis effects can be produced in a C-V curve, as the bias increases and then decreases.

6.3.3 Minority carrier extraction

Sec 6.2.2 described the limiting effect of the formation of an inversion layer on the width of the depletion region and the consequent levelling off of the C-V curve to C_{min}, at high frequencies. However, if the minority carriers in the inversion layer cannot accumulate at the surface because of conditions favouring injection into the insulator, then the capacitance goes on decreasing at a rate governed by the Schottky depletion capacitance given in Eqn 6-23. A plateau region at C_{min} is also possible if these conditions occur only after the formation of the inversion layer.

With increasing bias the depletion region extends even further into the semiconductor so that an increasing proportion of the applied voltage appears across it and a point is eventually reached where the current through the device is governed not by the insulator or its interface properties but by the generation of minority carriers in the semiconductor. This provides a limit to the current capable of flowing through the device. These effects have been observed in pyrolytically deposited films of Si_3N_4, SiO_2 and Al_2O_3 on silicon (Ref 6-12).

In the above treatment, current flow was assumed to be restricted to the semi-conductor minority carriers because the Fermi level is closer to the conduction band (for p-type) than the valence band for depleting voltages. If current flows through the insulator by mechanisms involving the injection of majority carriers from the silicon however, no change can be expected in the C-V curves provided that the measurement apparatus is capable of measuring only pure capacitance and that the currents are not excessive.

6.4 Experimental

C-V measurements have been made on a large number of willemite films produced on silicon substrates as described in Chap 4. Circular dot contacts (2.10^{-3} cm^2 in area) were evaporated on to both the willemite and the adjacent unconverted oxide (Sec 4-3) and electrical contact to the silicon was made as described in Sec 2.4.3. Aluminium, gold and indium dot contacts were used. The resultant MIS structures were then mounted under a microscope and the dot contacts were probed in turn with a gold probe held in a micromanipulator (Sec 4-3).

The C-V characteristics were usually obtained at room temperature in the dark, and were automatically recorded on an X-Y plotter. Preliminary measurements were, however, made point by point using a Wayne-Kerr B.221 bridge operating at 1592 Hz with the applied bias adjusted step by step manually. Fig 6-3 shows a typical C-V plot obtained in this manner. This method was very tedious and if the number of measurements was not sufficiently high significant data

could be overlooked. Also, due to the time required to balance the bridge at each point it was very difficult to measure any drift that may have occurred in the C-V characteristics. For these reasons it was decided to use an automatic C-V plotter.

Several designs of automatic C-V plotters have been published which use either phase sensitive detection (Ref 6-13) or automatic bridge balancing techniques. However, because the electronic equipment for these was not available in the Department a beat frequency method was developed (Ref 6-14). Fig 6-5 shows a schematic diagram of the system. A continuously variable bias was applied to the MIS structure using a Feedback waveform generator Type TWG300 and the device was included as a voltage variable capacitor in the LC circuit of a stable Franklin oscillator designed to oscillate at a frequency, f = 350 kHz. The voltage across the device was kept as small as possible (about 200mV). With a total capacitance, C_{osc}, in the resonant LC circuit of the oscillator.

$$f^2 = (\text{const.}) / C_{osc.}$$

On differentiating this gives

$$df = (\text{constant}) / (f\, C^{3/2})\, dC$$

so that provided $df \ll f$ and $dC \ll C_{osc}$, df is proportional to dC, the change in the capacitance of the device. The small changes in oscillator frequency (df) were detected by beating the oscillator signal with the signal from the beat frequency oscillator (B.F.O) of an Eddystone 830/7 communications receiver to give an audio signal the frequency of which was proportional to the capacitance of the device

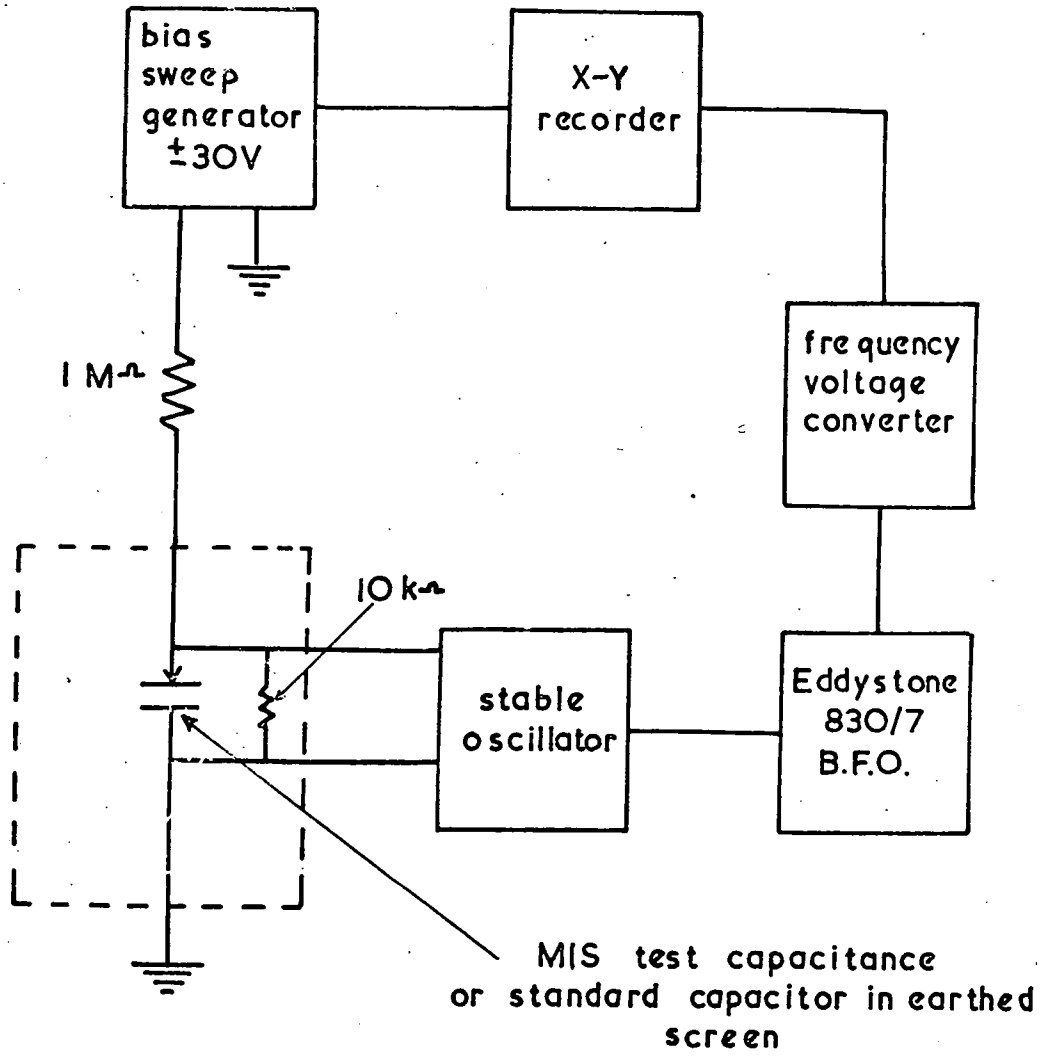


Fig 6-5 Block diagram of the system used for the automatic plotting of C-V characteristics

A linear frequency-to-voltage converter was then used to display the capacitance as the Y variable on a Bryan's X-Y recorder (Model 21001). The bias voltage on the device was simultaneously applied to the X-input of the recorder. The sweep rate was typically about 50 sec/cycle. Calibration of the C-V plots was achieved by using standard capacitances in place of the device. The sample holder capacitance was eliminated by slightly adjusting the B.F.O. frequency. A 10 kohm resistor was placed in parallel with the device to ensure that any variations in sample resistance did not effect the oscillator frequency, although this would only be a second order effect.

The C-V characteristics of several devices were measured after bias-temperature (BT) treatments in order to ascertain the properties of charge redistribution within the willemite. The devices were heated to temperatures of up to 150°C whilst still in contact with the probe, by mounting them on a silica plate which was arranged on a small stainless steel support so that it could be heated from beneath by a coil of Kanthol resistance wire. The temperature was measured by a Chromel-Alumel thermocouple in contact with the silica plate and adjacent to the silicon. The complete heating stage was enclosed in a metal box through which a steady stream of argon was maintained. After the heat treatment (about 5-10 min) the device was cooled quickly to room temperature with the bias still applied and the C-V characteristics were then measured. This process is usually referred to as the BT test in the literature and this abbreviation will be adopted here.

6.5 Experimental results and discussion

6.5.1 General observations on oxide structures

A typical C-V obtained curve for an aluminium/thermal oxide/p-type silicon structure is shown in Fig 6-9. The theoretical curve for a similar device but with a perfect oxide and interface (Ref 6-8) is also included in the figure. The negative shift of the experimental curve with respect to the ideal case indicates the presence of 3.10^{11} positive charges reflected at the interface (Sec 6.3.2). The two curves have the same gradients (about 6 pF/V) and this shows that there are very few fast states at the interface. Substituting this gradient into Eqn 6-36 with the experimental values for C_o and C_{FB} gives $D = 10^{12}$ $eV^{-1}cm^{-2}$ which represents approximately the lowest detection limit for this technique.

The results of BT measurements on the oxide devices are summarised in Fig 6-7. It can be seen that the application of a positive bias, at temperatures of about 150°C, shifts V_{FB} in the negative voltage direction. The initial characteristics can be recovered by applying a negative bias. With continuing negative bias shifts of V_{FB} in the positive direction were usually small and the final value of V_{FB} was always negative, corresponding approximately to the above number of interface charges. These BT drift effects are typical of positive ion migration in the oxide which is usually attributed to sodium or hydrogen ion contamination of the oxide (Ref 6-15).

Although the oxides were prepared in a controlled manner (Chap 2) no special precautions were taken to eliminate

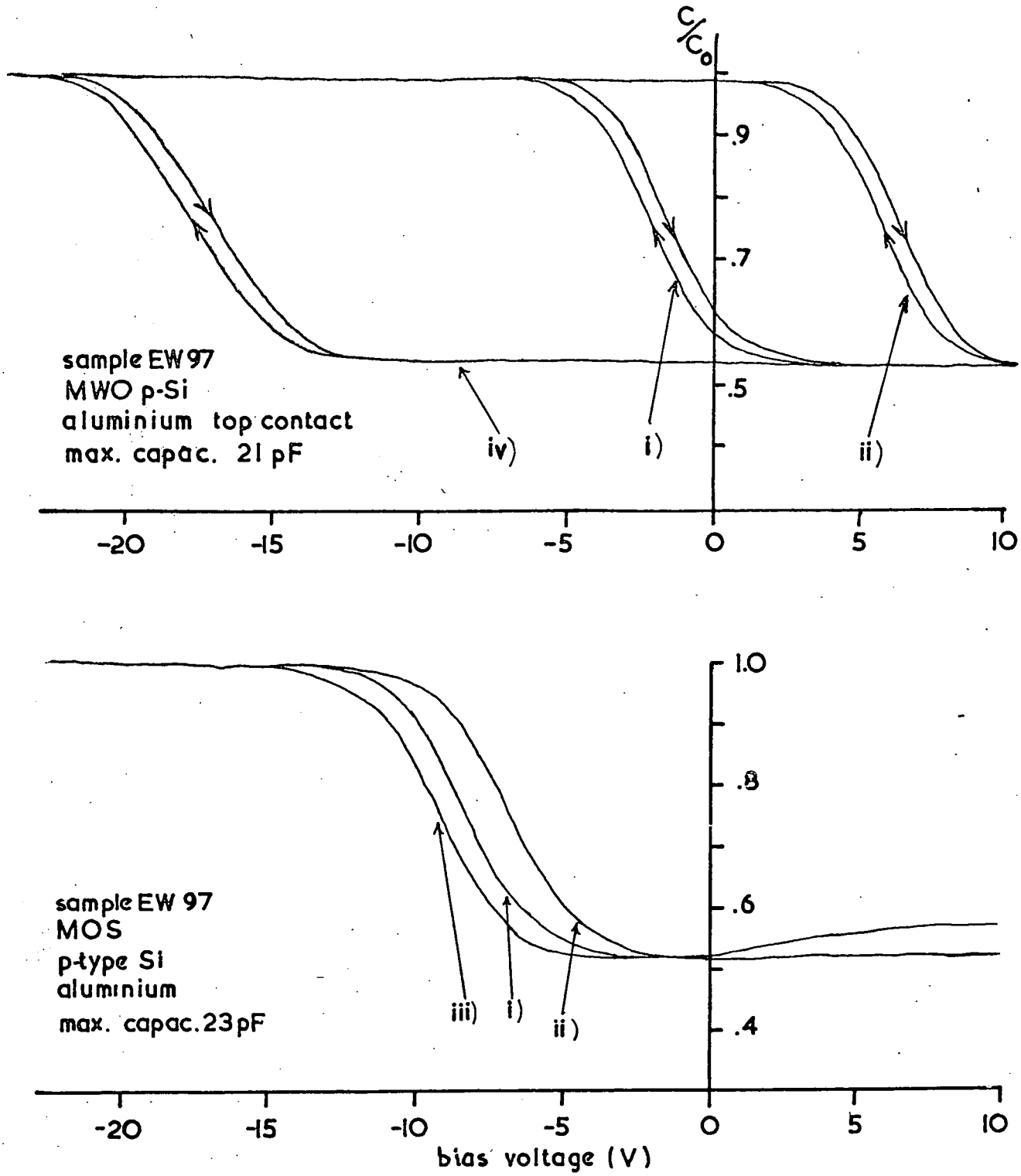


Fig 6-7 (bottom) and Fig 6-12 (top) Shifts in C-V curves of MOS and MWOS devices after various BT treatments.

i) original curve and after 5 min 165°C zero bias ii) 5 min 165°C −20 V
iii) 5 min 165°C +20V iv) 2 min 165°C +20 V.

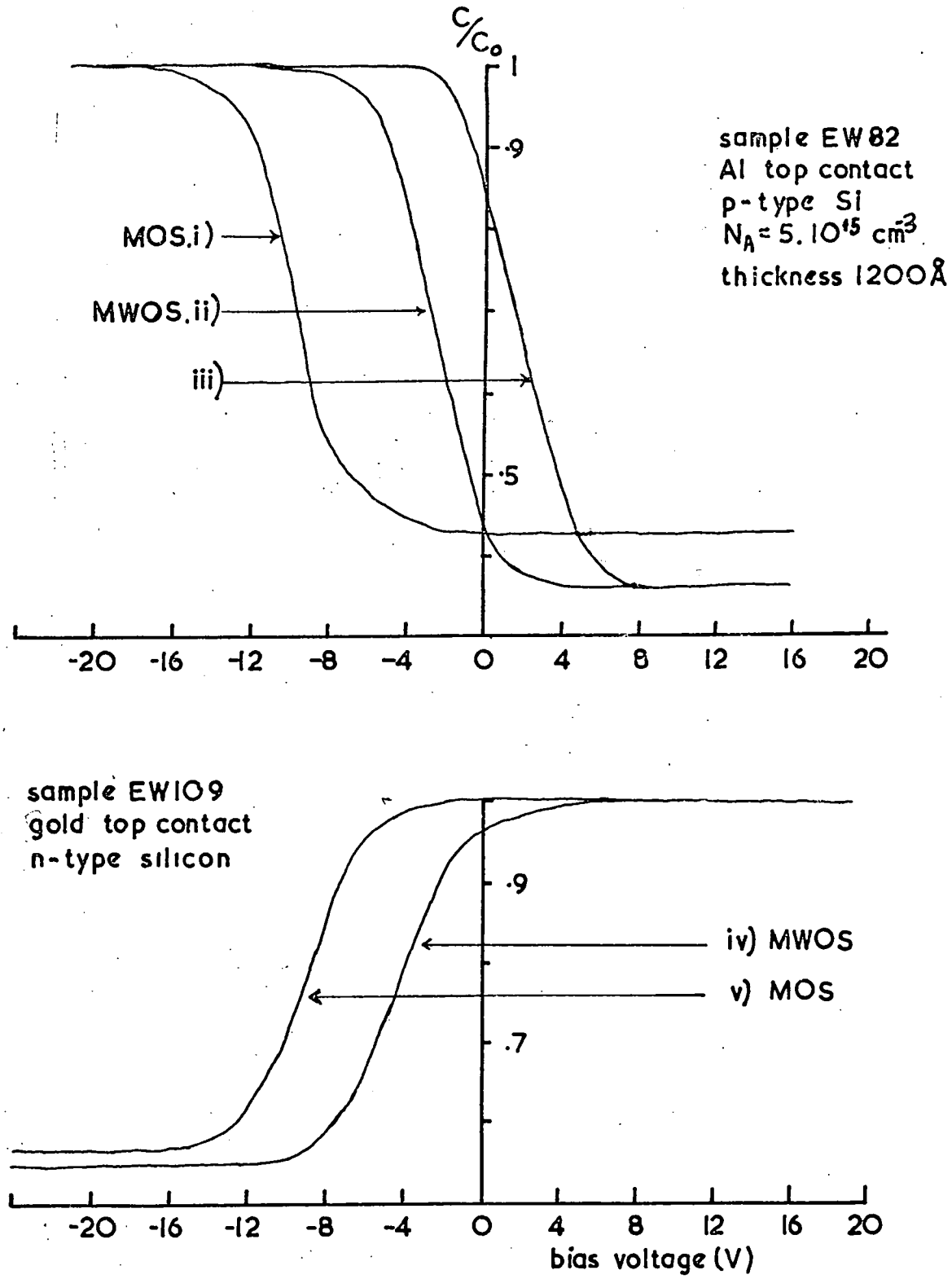


Fig 6-9 C-V curves ii) for MWOS made from oxide with C-V curve (top) shown in i), iii) theoretical curve for ideal MOS (Ref 6-8)

Fig 6-8 C-V curves for v) MOS device adjacent to MWOS device (bottom) with curve shown in iv)

sodium contamination which can occur extensively when oxides are grown in silica glassware using resistively heated furnaces (Ref 6-16). Traces of hydrocarbon solvents and water can also produce hydrogen ion contamination of the oxide (Ref 6-15). It is therefore reasonable to assume that the thermal oxides grown here were exposed to varying and uncontrollable amounts of this sort of contamination and this was confirmed by the varying extents of drift that occurred for the several devices that were examined.

The distribution of the contaminant ions within the oxide was also variable and this led to varying amounts of positive V_{FB} shift from the initial C-V curve. For example if the contamination was introduced just prior to the evaporation of the contacts the positive ions would initially reside at the metal-oxide interface and a negativo bias would not shift V_{FB} (Ref 6-17). On the other hand, contamination that is distributed throughout the oxide can be swept to the metal-oxide interface by a negative bias and produce a small positive shift in V_{FB}.

A negative value of V_{FB} after prolonged negative BT treatment indicates a fixed positive charge near the silicon-oxide interface. This is generally attributed to excess silicon in this region and its extent depends only on the final heat treatment or oxidation conditions of the device, for a particular silicon orientation (Ref 6-18). In the oxides considered here the final heat treatment was the bake required for the conversion of the adjacent oxide into willemite. In spite of this additional heat treatment the value of fixed surface charge given above compares favourably with

values given in the literature for similar oxides.

The tendency of the capacitance to increase on the inversion side ($+V_G$) of the C-V curve as shown in Fig 6-7 is thought to be due to the accummulation of positive charges on the oxide surface around the metal electrode. This effectively increases the capacitance of the space charge in the underlying silicon by increasing its surface area so that the series capacitance of the oxide and the silicon tends to increase to a maximum of C_o. This effect was more noticeable at higher temperatures when presumably the increase in surface conduction allowed the charge to accumulate over a larger surface area (Ref 6-6).

The general results described above for oxides grown on p-type silicon were also observed with n-type silicon. As expected, the C-V curves (Fig 6-8) were a mirror image of the p-type characteristics but retained a negative shift in V_{FB} due to the fixed positive charge near the silicon. BT tests showed qualitatively similar effects which could again be attributed to positive ion contamination.

The C-V measurements were usually carried out in the dark. At the measurement frequency used (350kHz), the ambient lighting in the laboratory had no effect on the characteristics and even a focussed microscope lamp only increased C_{min} slightly. This is consistent with an increase of n_i produced by the light (Eqn 6-23).

The thickness of the oxides ranged from 500-1500 Å as measured by an interference method (Sec 2.3.2) and was in agreement with oxide capacitance measurements in the accumulation region

(when no space charge capacitance is present) using the known value of 3.4 for the dielectric constant of thermal oxide (Ref 6-19) and the measured contact area. The largest error was in measuring the exact area because the contact was sometimes scratched by the gold probe.

The conclusion reached on the C-V measurements on silicon oxide was that they were all in perfect agreement with findings reported in the literature and with the theory for films containing a positive ion impurity which was to be expected in the growth conditions used. This agreement gave confidence in the new measurements made on the adjacent willemite films.

6.5.2 Willemite MIS devices

The C-V measurements made on the willemite films could be divided into two groups depending on the thickness of the reacted layer. This was determined by preferential etching and interferometry after the C-V measurements had been completed (Sec 5.2) It was found that penetration of the oxide by the reaction was generally about 500 Å producing willemite films which had a total thickness of about 750 Å approximately 250 Å of which was raised above the surrounding oxide surface. The dielectric constant of willemite does not appear to have been measured in the past and it was difficult to calculate from the capacitance measurements because of the limited resolution of the interference technique ($\pm$ 250 Å) for the thickness. The approximate value of 6 $\pm$ 2 was as good as could be estimated. The willemite devices were categorised into the two general types :-

(a) those made from thick (1500 Å) oxides which resulted in a metal-willemite-oxide-silicon structure (MWOS device) and

(b) those which were made from thin (500 Å) oxides so that no oxide (or at least only a very thin film) was present at the silicon interface (MWS device).

a) MWOS structures

Fig 6-8 shows the C-V curve for an MWOS device made by converting about 500 Å of the 1200 Å thick oxide into willemite (curve iv). The unconverted oxide adjacent to the willemite has been used to form the MOS device with the characteristics shown in curve v). It can be seen that the conversion of the outer layer of the oxide into willemite has shifted V_{FB} in a positive direction. All the willemite devices that were measured on both p-type and n-type silicon showed a similar shift so that V_{FB} was often about zero as in the case above.

It is thought likely that this positive shift in V_{FB} is due to the formation of a negative space charge in the oxide next to the silicon and that this neutralizes the fixed positive charge normally found in thermal oxides. This negative space charge is probably due to the diffusion of Zn^{2+} ions in the willemite moving away from its interface with the oxide during the high temperature bake (1100°C) required in its preparation. The evidence for this hypothesis is discussed below.

Using values obtained from Ref 5-3 it is found that the Zn^{2+} ion has a radius of 0.38 Å which is comparable with the

radius of the network-forming the Si^{4+} ion (0.39Å) in the oxide (the radius of the network-bridging O^{2-} ion is 1.36 Å). The zinc ion can therefore replace the silicon substitutionally to produce an effective negative space charge. This is in agreement with Chang and Tsao (Ref 6-20)who have recently investigated the C-V properties of diffused zinc(from an evaporated metal film) in thermal silicon oxide.

The willemite used for the MWS structures was luminescent and contained about 1% by weight of manganese. However, the Mn^{2+} ion occupies a similar place in the willemite lattice as the Zn^{2+} ion (Sec 5.1) and it also has the same valency. It was therefore considered that it would behave in exactly the same way as the zinc ion as far as the C-V measurements are concerned. For this reason none of the effects observed are thought to be influenced directly by the manganese impurity.

The range of values of V_{FB} that were observed for different MWOS devices were probably due to the different degrees of zinc diffusion as related to the quality and thickness of the oxide through which the Zn^{2+} ions diffuse. However, the fact that V_{FB} was often zero and never positive seems to indicate that the region of disorder in the oxide (i.e. the region of fixed positive space charge due to the excess silicon)allows the Zn^{2+} ions to be easily accommodated in the network with an effective negative charge. When the disorder has been removed and the space charge neutralised, it is more difficult for the Zn^{2+} ions to be accommodated with a negative charge so that a positive V_{FB} corresponding to a net negative

space charge in the oxide was never observed.

The slopes of curves 1 and 2 in Fig 6-9 are very similar and this indicates that no fast surface states have been introduced at the silicon-oxide interface by the fabrication of the MWOS device. However a few anomalous C-V curves were obtained such as that shown in Fig 6-10. In this it can be seen that there is a considerable difference between the slopes of the MWOS device and the MOS device, although V_{FB} is still about zero for the willemite device. The effective density of fast surface states for this MWOS device was $1.0 \cdot 10^{13}$ $cm^{-2}eV^{-1}$ as calculated from Eqn 6-36 whereas the slope of the MOS curve corresponded to a density of less than $2. \ 10^{12}$ cm^{-2} eV^{-1}. An extra $8.0 \cdot 10^{12}$ cm^{-2} eV^{-1} states were therefore added at the interface in this device.

An explanation of this could be that, although the oxide seems incapable of supporting an overall negative space charge due to the zinc ions some of the ions diffuse right through these particular oxides and act as fast acceptor traps at the silicon-oxide interface. Zn^{2+} is an acceptor impurity in silicon so that any further diffusion of the zinc into the silicon itself should produce an increase in the conductivity of the p-type silicon or a decrease for the n-type silicon. This effect was never observed with the anomalous devices, probably because the bake was relatively short (usually about 5 min) but also because it was very difficult to ascertain how much of C_{min} was due to the insulator capacitance (in this case the willemite and the oxide) and how much was due to

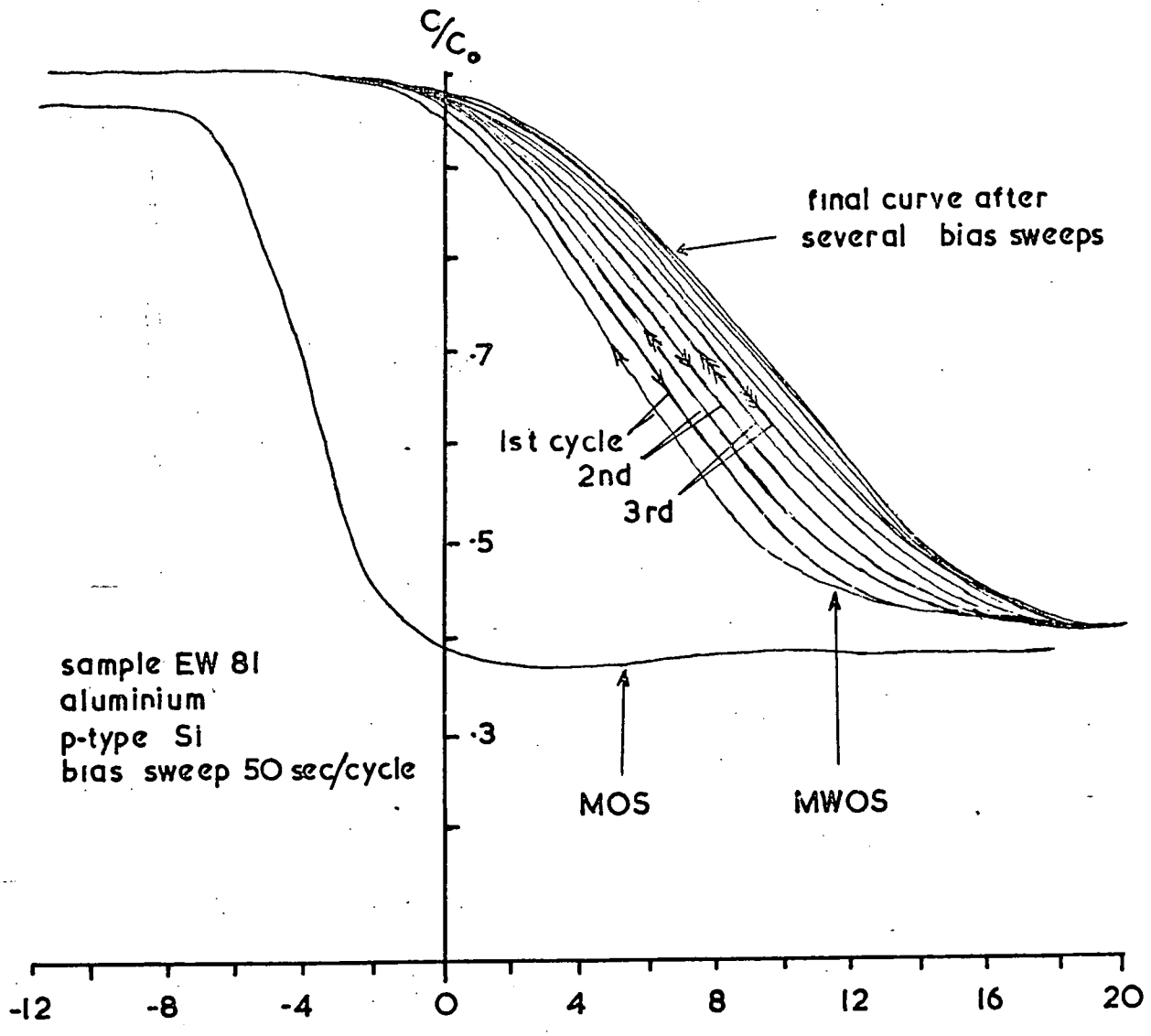


Fig 6-10 Anomalous C V curve showing deterioration of characteristic with increasing number of bias sweeps

the space charge capacitance in the silicon.

Support for this explanation of the anomalous C-V curves was also given by further anomalous characteristics shown in Fig 6-10. In this graph several sweeps of bias have been recorded with curve 1 corresponding to the characteristics of the virgin device. With every positive excursion of the bias more and more fast states seem to be created as indicated by the continuing decrease in the slope of the curve. This is consistent with Zn^{2+} ions being swept to the interface where they are trapped to form fast surface states. The continual drift of the C-V characteristics with increasing number of bias sweeps was only observed with some of the anomalous devices presumably because a particularly open structure or poor quality oxide is required to allow the ions to drift through it comparitively rapidly.

All the C-V characteristics of the willemite devices that were measured showed at least some hysteresis with applied bias. The hysteresis was always in such a sense as to make V_{FB} more negative after the maximum positive excursion of applied bias. This is opposite to that which would be obtained if successive carrier trapping and retrapping occurred at the silicon-oxide interface (Ref 6-21) and it was attributed to ion migration in the willemite. The possibilities of polarisation or carrier trapping at the metal-willemite interface are discussed later.

Typical C-V curves obtained after BT experiments are summarised in Fig 6-12. They show a similar effect to the

room temperature hysteresis, namely that a positive bias, at temperatures of about 150°C, produced a negative shift in V_{FB} and a negative BT treatment produced a positive shift in V_{FB} The latter was even sufficient to make V_{FB} positive and it could be attributed to mobile positive ions (probably Zn^{2+}) moving in a relatively immobile negative matrix so that with $-V_g$ applied the positive ions migrate to the top contact and leave behind a fixed negative space charge due to the larger SiO_4^{4-} radical. The positive ions reflect most of their charge into the adjacent metal and do not affect V_{FB} whereas at least part of the negative space charge is reflected in the silicon to produce a positive value of V_{FB}. Similar affects to this have been observed in lead silicate glasses by Snow and Dumesnil (Ref 6-22). It is also possible for the positive ions or ~~some~~ at least some of them to discharge at the metal and the results of the I-V work showed that this was probably the case (Chap 7). With the C-V experiments alone it is not possible to determine whether the ions discharge or not at the metal because the effects of the space charge closest to the silicon always predominate. However, the C-V experiments do show conclusively that positive ions do not discharge to the silicon when they pile-up with positive bias and this is in agreement with the conduction model proposed in Chap7.

The fact that the BT curves in Fig 6-12 show a much larger negative than positive shift was probably partly due to an additional positive ion component from sodium ion contamination and partly to the fact that the mobile positive ions can move close to

the silicon surface and affect V_{FB} more than the distributed charge of the immobile negative ions (Eqn 6-31).

During the BT experiments it was assumed that no ionic transfer occurred between the willemite and the oxide and that any zinc which had diffused in to the oxide during the bake remained there. This is in agreement with the fact that after zero bias temperature treatment the curve relaxed to its initial position showing that there was no change in the number of zinc ions in the oxide. It should also be noted that a positive B.T. treatment produces a negative V_{FB} shift due to the accumulation of positive ions at the willemite-oxide interface, whereas zinc ions actually inside the oxide have an effective negative charge that produces the opposite shift in V_{FB}, i.e. in the positive direction.

Hysteresis and BT shifts which make V_{FB} more negative after a positive bias excursion and vice versa can also be attributed, in theory, to two other effects besides the migration of positive ions described above. These are (1) carrier exchange at the metal-insulator interface and (2) polarisation of the insulator. The reasons why these effects are not considered to be significant here are briefly discussed below.

(1) It is usually assumed that interchange of charge at a metal-insulator interface involves short range penetration of the insulator because the carriers either tunnel into trap states within the insulator (Ref 6-21) or the potential wells of the trap overlap with the electrode so that carrier hopping can occur (Ref 6-15).

It therefore seems reasonable to assume that the space charge regions produced by these processes lie close to the metal electrode and that they reflect very little of their charge into the silicon so that only small shifts in V_{FB} occur ($x \ll x_o$ in Eqn 6-31). This was not observed in practice.

(2) Polarisation phenomena in insulators are due to one or more of the following mechanisms :- (a) the displacement of the electronic charge surrounding the molecule (b) the displacement of the ions within the molecule (c) the alignment of dipolar molecules in the applied electric field or (d) to interfacial polarisation - a process that involves the separation of opposite charges inside some macroscopic conducting phase within the insulator.

The first two of these mechanisms contribute to the high frequency dielectric constant of the insulator. They are too fast to contribute to any hysteresis or BT effects and they are also independent of temperature. The alignment of dipolar molecules within a crystalline lattice is unlikely and again the correct temperature dependence for this is not observed. (As the thermal energy of the molecules increases the amount of alignment should decrease which is contrary to what was observed.)

Interfacial polarisation due to a non-uniform dielectric seems to be the only likely polarisation mechanism for the hysterisis and BT effects in the willemite films because in this case the effect increases with temperature as the mobility of the charges increases. This form of polarisation has been observed by Snow and Deal (Ref 6-23)

in phosphosilicate glasses on silicon. Their shifts in V_{FB} were however small and always less than the polarising voltage unlike the shifts shown in Fig 6-12 and they were attributed to about 6% by volume of conducting material in the glass. It is possible that interfacial polarisation may be occurring, at least partly, in the willemite films but it probably occurs by ion migration within some macroscopic phase in the crystal such as a small crystallite and as such it is practically indistinguishable from ion migration across the film.

b) MWS structures

A typical C-V curve for a thin MWS device is shown in Fig 6-13. In this device the willemite film was made from oxide that was about 500 Å thick and it was shown by preferential etching of the willemite that no detectable oxide (i.e. < 100Å) remained at the silicon interface after the conversion process. The C-V characteristics of the MOS device made from unconverted oxide adjacent to the willemite are also given in Fig 6-13.

The voltage hysteresis of MWS devices can be seen to be relatively greater than for the thicker MWOS devices. This is because the ions that give rise to the hysteresis are now much closer to the silicon so that they reflect more of their charge in the semiconductor and also the structures are much thinner which considerably increases the factor x/x_o in Eqn 6-31 relating V_{FB} with the position of charges in the insulator.

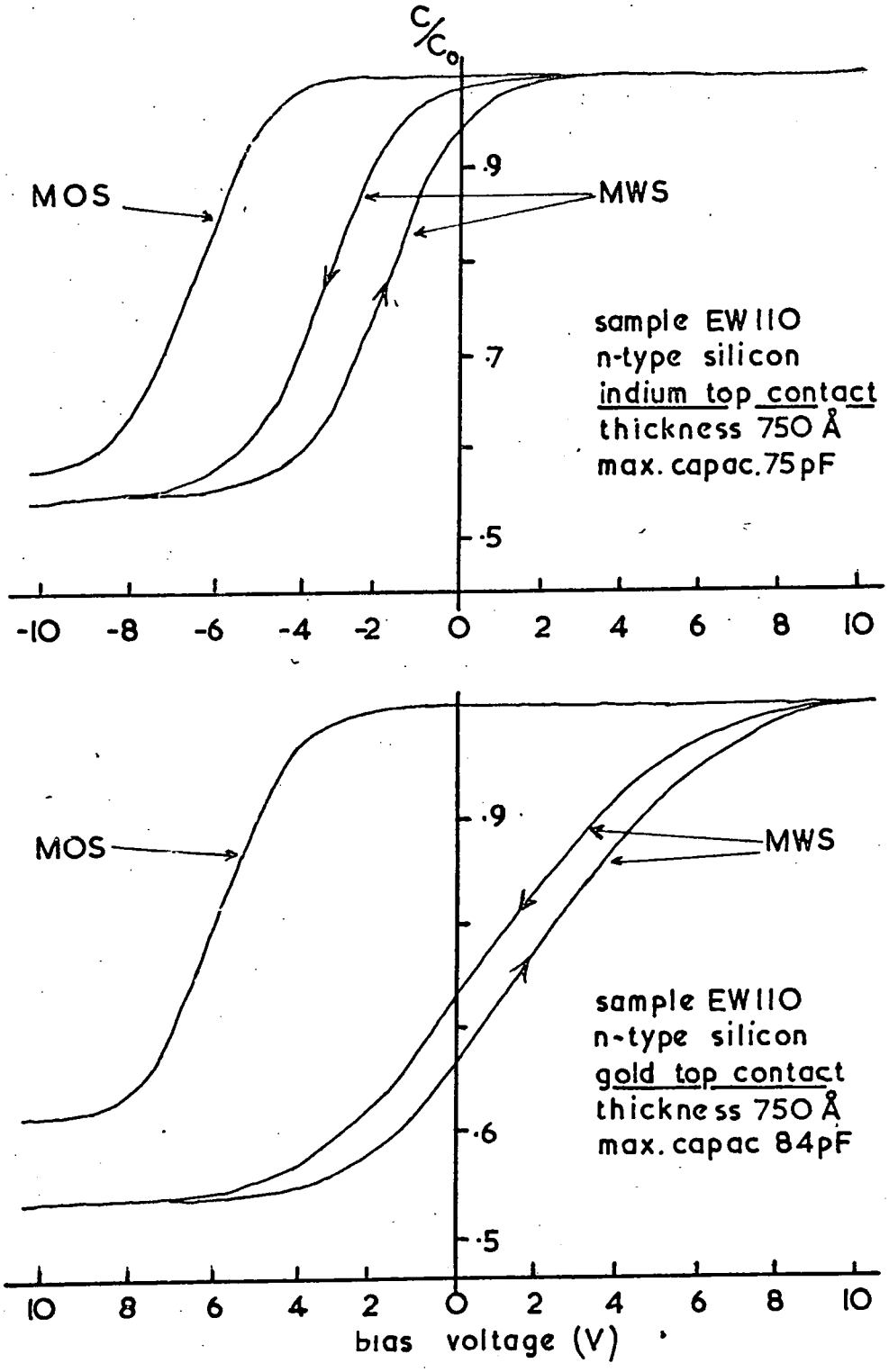


Fig 6-13 and Fig 6-14 (bottom) C-V curve for two MWS devices (top) and adjacent MOS structures.

BT experiments on MWS devices were qualitatively similar to the MWOS ones although the shifts in V_{FB} were again relatively larger than for the MWOS devices after similar BT treatment. This is probably for the same reasons as those given above for the increased hysteresis effects. In the MWS devices the position of the initial C-V curve was however found to depend on the type of metal that was used for the top contact and this effect will now be discussed.

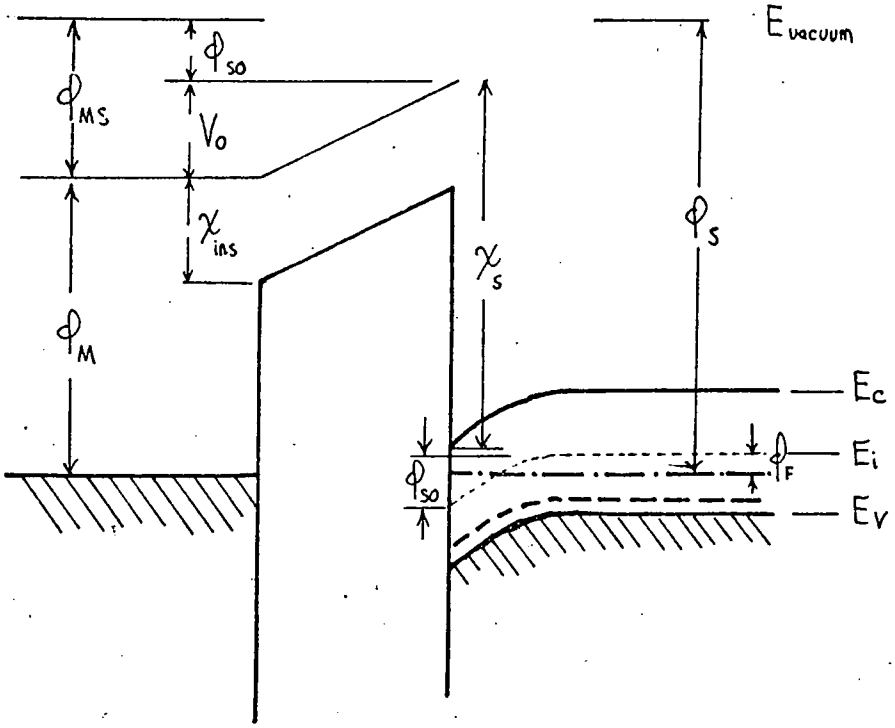
Although indium, gold and aluminium were used as top contacts for the MWOS experiments with both n and p-type silicon no significant differences were observed between them (i.e. within the limits of reproducibility). However, for the thinner MWS structures which were all made of n-type silicon, a considerable difference was observed in the initial C-V curves for devices with gold and indium top contacts. These characteristics shown in Figs 6-13 and 6-14 were obtained prior to any BT treatment. The different values for V_{FB} in these two curves can be attributed to the same positive ion drift as discussed earlier for the BT case but this time the ions must drift at room temperature in the electric field set up by the difference in work functions (ϕ_{MS}) between the silicon and the metal. This drift presumably occurs in the time between the contacts being applied and the C-V curves being measured which was usually a few days. Ion drift in the contact potential field would be expected to be much greater in the thin MWS devices

than in the comparitively thick MWOS case, as observed.

Fig 6-15 shows the energy levels of an MIS device with zero applied voltage. The values of ϕ_{MS} for the three metals and for silicon of various resistivities have been calculated using Eqns 6-13 and 6-14 and these are listed in Table 6-1. This table shows that only a gold contact with n-type silicon produces a negative voltage on the metal (i.e. positive ϕ_{MS}). This is in agreement with the positive shift in V_{FB} that was observed for this combination only. This shift would correspond to the accumulation of the Zn^{2+} ions at the metal and the formation of a negative space charge of $(SiO_4)^{4-}$ radicals near the silicon surface. If the Zn^{2+} ions discharge to the metal however (and there is evidence for this in the I-V work) this could not be detected in the C-V shifts. The existence of an electric field due solely to ϕ_{MS} and of sufficient magnitude to cause ion migration, is further substantiated by the increase in capacitance on the inversion side of the characteristics in Fig 6-14 due to the accumulation of negative ions on the surface of the device. This effect is also shown in Fig 6-7 for a MOS device but here it occurs only during BT treatment.

Negative values for V_{FB} were observed for all the devices that were made with the other metal/silicon combinations shown in Table 6-1, and this could be attributed to ion drift in the reverse direction due to the electric fields set up by the negative values of ϕ_{MS}.

Although the polarity of V_{FB}, as indicated by the



$$\phi_{MS} = \phi_M - \phi_s = \phi_M - (\chi_s + E_g/2 + \phi_F) = V_o + \phi_{so}$$

Sustituting for ϕ_F from Eqn 6-13

$$\phi_{MS} = \phi_M - (\chi_s + E_g/2 + \frac{kT}{e} \ln_e \frac{N_A}{2n_i})$$

Taking χ_s = 4.05 eV (Ref 6-27), E_g = 1.1 eV and $n_i = 1.45 . 10^{10} cm^{-3}$ at 27°C then

$$\phi_{MS} = \phi_M - 4.6 + 0.026 \ln_e (N_A / 2.9\ 10^{10}) \qquad 6\text{-}38$$

Fig 6-15 The contact potential difference of an ideal MIS device at zero bias

sign of ϕ_{MS}, was maintained from device to device, its exact value usually varied. This seems reasonable because the precise history of each device prior to measurement was usually different and this would allow different degrees of ion migration to occur before V_{FB} was actually measured.

The rate of zero-bias ion migration also depends on the value of ϕ_{MS} and the magnitude of the resultant electric field in the willemite. In the thicker MWOS devices, where the fields were smaller and the amount of ion migration correspondingly less, this would account for the fact that no zero bias effects were observed. Another reason for this was probably that charge movement in the willemite has a smaller effect on V_{FB} in the thicker films which also have a relatively thick oxide layer between the silicon and the willemite.

When no ion drift effects occur at zero bias the only effect of different contact metals is to shift the C-V curve by an amount equal to ϕ_{MS} (Eqn 6-31). This can be seen in Fig 6-14 where the characteristics of the MOS device made from the unconverted oxide of device EW110 are shown for both gold and indium electrodes. The shift in V_{FB} between these two curves represents the difference in ϕ_{MS} for the two devices (assuming that other charges are the same for both oxides) and equals about 1.0 volt in good agreement with the theoretical value of 0.8V obtained from Table 6-1.

The formation of a negative space charge in the insulator near the silicon appears to favour the introduction of fast

surface states at the interface. This can be seen in Figs 6-14 and 6-13 where the C-V curve of the device with the indium top contact has a V_{FB} = 2.0 V (corresponding to an effective density of 5.10^{11} cm^{-2} positive charges at the interfaces) and a slope which is similar to that of the unconverted oxide, whereas the device with a gold electrode has a C-V curve with V_{FB} = +5 V ($1.4.\ 10^{12}$ cm^{-2} negative charges at the interface) and a much smaller gradient corresponding to an effective density of fast surface states of $1.7\ .\ 10^{12}$ eV^{-1} cm^{-2} as calculated from Eqn 6-36. This is in agreement with the C-V curves of the anomalous MWOS devices shown in Fig 6-10 where a similar density of fast surface states was attributed to a negative space charge at the interface due to zinc ions in the oxide.

It has been shown by several authors (Ref 6-23) that electroforming effects can occur in thin dielectric films and these can be attributed to the field injection of metal ions from the contacts into the dielectric. These effects seem to occur mainly with gold electrodes and apparently devices can only be electroformed in a vacuum of less than about 10^{-3} torr. In spite of the fact that the C-V measurements described above were all carried out at atmospheric pressure, the difference between the C-V curves of gold and indium contacted MWS devices could possibly have been due to an electroforming effect viz. gold ion injection into the willemite under the influence of the electric field set up by ϕ_{MS}.

Cagnina and Snow (Ref 6-25) have investigated gold diffusion in gold/thermal-oxide/silicon structures using the C-V technique and found no evidence for either gold injection with fields of up to about 10^6 V cm^{-1} (at atmospheric pressure) or for gold diffusion into the oxide for temperatures up to about 250°C. At temperatures higher than this they found that gold entered the oxides with an effective positive charge in a similar way to sodium ions (the ionic radii of gold and sodium are 1.37 Å and 0.98Å respectively and either can therefore replace the network bridging oxygen ion which has a radius of 1.36 Å).

The effective charge of the gold ion in willemite is not known, but it is likely to be positive as it is in the oxide and in most other oxygen containing dielectrics. If this were so, the positive space charge produced by the gold ions in the willemite would produce a negative shift in V_{FB} which is opposite to what was observed. Also in order to field inject positive gold ions ϕ_{MS} would have to be negative whereas for the gold-n-type silicon combination it is in fact positive. For these reasons it seems unlikely that injection of ions of the contact metal can account for the observed contact dependance in the MWS devices.

Carrier transfer mechanisms at the metal-willemite interface could also account for the contact effects that occur with MWS devices. It was considered above, earlier in this Section, that any interchange of carriers at this interface would probably involve space charge formation close to the metal electrode and this would

have only a small effect on the C-V curves. Even if the measurements were sensitive to charge injection from the metal this could not account for the positive shift in V_{FB} for gold electrodes (with n-type silicon) and the negative shift for indium electrodes, because electrons would have to be injected by the gold and holes by the indium (both under the influence of ϕ_{MS}). From a consideration of the work functions of these two metals it would however seem more reasonable if the opposite occurred and in fact gold is generally considered to be a blocking electrode compared to indium. Also if charge injection occurred at the metal contact then the hysteresis effects would also probably be due to this and the amount of hysteresis would ~~then~~ be dependent on the shift of V_{FB} which would then be an indication of the ease with which charge occurs at the interface. However, from Figs 6-12 and 6-14 it can be seen that the hysteresis is approximately the same for both the indium curve with $V_{FB} = -2.0$ V and for the gold curve with $V_{FB} = 5$ V, so that once again charge injection seems to be eliminated as an important process.

From the above considerations it is considered that neither electroforming effects nor carrier injection can account for the contact effects obtained in the C-V curves of the MWS devices, and that charge movement in the contact potential field is much more probable.

6.6 Conclusion

The C-V characteristics of willemite films on silicon have been studied and compared with those of MOS devices made from unconverted oxide surrounding the willemite. The MOS devices have standard C-V characteristics that can be explained in terms of the theory presented in Sec 6.2, with the addition of two space charge components in the oxide namely, a mobile positive charge and a fixed positive charge near the silicon. Other workers have identified similar charges as being due to sodium or hydrogen ion contamination of the oxide and excess silicon at the silicon-oxide interface respectively, and both these are likely with the fabrication processes used here. It appears that the willemite conversion process introduces no further contamination into the surrounding oxide.

The C-V characteristics of willemite films depended on the depth of oxide that was used up in the conversion process, so that either metal-willemite-oxide-silicon (MWOS) or metal-willemite-silicon (MWS) structures resulted. The thicker MWOS devices displayed characteristics that were consistent with a reduction in the amount of fixed positive charge in the oxide sandwich. This has been attributed to zinc diffusing from the willemite into the underlying oxide during the high temperature conversion process. The silicon-oxide interface however, usually remains free from fast surface states, as indicated by the slope of the C-V curve. Voltage hysteresis and bias-temperature experiments on these devices indicated that ion migration

occurs in the willemite probably in the form of Zn^{2+} ions moving in a matrix of $(SiO_4)^{4-}$ radicals. The effects of both polarisation and carrier exchange at the interfaces have been discounted.

The C-V curves of the thinner MWS devices were more sensitive to charge movement in the willemite and this enabled a contact effect to be seen in these devices that was not observed with the MWOS structures. MWS devices with top contacts of different metals displayed characteristics having varying shifts in both directions along the voltage axis. This has been attributed to ion migration occuring under the influence of the zero bias field set up by the difference in work functions between the silicon and the metal. Possible alternative explanations in terms of either electroforming effects of carrier exchange at the metal-willemite interface have been shown to be inconsistent with the experimental results. It has been possible to produce MWS devices with no extra fast surface states at the willemite-silicon interface, as shown by the slope of the CV curve, in comparison with the characteristics of the unconverted oxide.

The theory shows that the hypothesis of ionic movement in the contact potential field in the MWS devices could be confirmed by further experiments using p-type substrates, and it was unfortunate that there was no time available for the further work required. Table 1 shows that the contact potential fields with gold and indium contacts are in the same direction with p-type silicon

so that the shifts of V_{FB} should also be in the same direction unlike the case with the n-type substrates.

Minority carrier injection from the silicon into the willemite produces a characteristic change in shape of the C-V curve, although this was not observed in any of the devices that were examined. It seems reasonable to assume that current mechanisms through the willemite will therefore involve majority carrier effects. The current-voltage characteristics of willemite films will be considered in the next chapter.

References for Chapter Six

6-1 W. Shockley, Bell Syst. Tech. J. 28, 435 (1949)

6-2 R.H. Kingston and S.F. Neustadter, J. Appl. Phys. 26,718 (1955)

6-3 C.E. Young, J. Appl. Phys. 32, 329 (1961)

6-4 W.G. Pfann and C.G.B. Garret, Proc. I.R.E. 47, 2011 (1959)

6-5 J.L. Moll, Wescon Convention Record, Part 3, 32 (1959)

6-6 L.M. Terman, Solid State Electron. 5, 285 (1962)

6-7 K. Lehovec and A. Slobodskoy, Solid State Electron. 7, 59 (1964)

6-8 A. Goetzberger, Bell Syst. Tech. J. 45, 1097 (1966)

6-9 A.S. Grove, B.E. Deal, E.H. Snow and C.T. Sah, Solid State Electron. 8, 145-63 (1965)

6-10 R. Linder, Bell Syst. Tech. J. 41, 803 (1962)

6-11 K. Lehovec, Solid State Electron. 11, 135-7 (1968)

6-12 F.H. Hielscher and H.M. Preier, Solid State Electron. 12, 527-38 (1969)

6-13 J. Shewchun and A.Waxman, Rev. Sci. Instruments. 37, 1195-1201 (1966)

6-14 P.G. Martin, M.Sc. Thesis, Durham (1970) unpublished.

6-15 S.R. Hofstein, Solid State Electron. 10, 657 (1967)

6-16 E. Yon, W.H. Ko and A.B. Kuper I.E.E.E. Trans. Electron Devices ED-13, 276 (1966)

~~6-17 E.H. Snow, M. Sklar, A.S. Grove and E.H. Snow, J. Electrochem 114,266 (1977)~~

6-18 B.E. Deal, M. Sklar, A.S. Grove and E.H. Snow, J. Electrochem Soc. 114, 266 (1967)

6-17 E.H. Snow, A.S.Grove, B.E. Deal, C.T. Sah, J. Appl. Phys. 36, 1664-73 (1965)

6-19 B.E. Deal, J. Electrochem Soc. 110, 527-33 (1963)

6-20 Chun-Yen Chang and Kuey-Yeau Tsao, Solid State Electron. 12, 411-15 (1969)

6-21 F.P. Heiman and G. Warfield, I.E.E.E. Trans. ED-12, 167-78 (1965)

6-22 E.H. Snow and M.E. Dumesnil, J. Appl. Phys. 37, 2123-31 (1966)

6-23 E.H. Snow and B.E. Deal, J. Electrochem Soc 113, 263 (1966)

6-24 J.G. Simmons and R.R. Verderber, Proc. Roy. Soc. A301, 77-102 (1967)

6-25 S.F. Cagnina and E.H. Snow, J. Electrochem Soc. 114, 1165 (1967)

6-26 S.R. Hofstein and G. Warfield, Solid State Electron, 8, 321 (1965)

6-27 F.G. Allen and G.W. Gobel, Phys.Rev. 127, 150 (1962)

6-28 H.B. Michaelson, J. Appl. Phys. 21, 536 (1950)

6-29 J.C Reverie, Proc. Phys. Soc. Lond. B70, 676 (1957)

CHAPTER SEVEN

MEASUREMENTS OF CONDUCTION CURRENTS IN WILLEMITE DEVICES

7.1 Introduction

This chapter describes the conduction measurements that have been made for willemite films on silicon. The conduction mechanisms that are theoretically possible for a perfect metal-insulator-metal (MIM) structure are discussed briefly in Sec 7.2. The quantitative application of these theories to experimental results is notoriously difficult and this is particularly the case with willemite because of the ionic movement that has been shown to occur within the films by the C-V work described in Chapter 6 and the distortion of the internal electric field that this produces. The structural assessment of the films (Chapter 5) has shown that they have a fine polycrystalline texture and this, together with the rather unknown nature of the silicon-willemite interface due to the substrate reactive preparation (Chapter 4), leads to further departures from an ideal structure. Complications also arise because one of the electrodes is silicon which is capable (in its non-degenerate form)of sustaining space charges at the surface with the result that the field penetrates the electrode to a certain extent. This has been discussed extensively in Chap 6 with respect to the MIS structure and effects that this may have on the conduction mechanisms are discussed in Sec 7.2.4

Electroluminescence (EL) has been observed from willemite films under certain conditions with very high d.c. fields and this is discussed in Chapter 8 with other luminescent properties

of the films and the simple theories of dielectric breakdown which are partly relevant.

Current measurements have not been extensive, largely because of lack of time, and it is not possible to be certain about the conduction mechanism. In order to elucidate the exact mechanism in thin films it is usually necessary to measure I-V curves with both the temperature and the film thickness as a parameter. Although this has not been possible in this case, certain interesting polarity results have emerged from the simple I-V measurements and these are discussed with the results in Sec 7.4.

Barrier heights at the electrode interfaces can be investigated by using the photoemission of carriers from the electrodes into the insulator. This type of measurement was attempted for a willemite film but with inconclusive results although intrinsic photoconductivity of the willemite was found as described in Sec 7.4.2

7.2. Conduction Mechanisms in Insulating Films.

7.2.1 General considerations

According to the band theory of solids, insulators are characterised by a full valence band separated from an empty conduction band by a forbidden gap of a few electron volts. For significant electronic conduction to take place additional carriers must therefore be introduced into the insulator. The resulting current is then controlled by the region of highest resistance which is either at the electrodes or in the bulk of the material. This leads to theories of injection limited currents and bulk limited

currents. Various mechanisms are possible within this classification and these are illustrated schematically in Fig 7-1. The theory will be only briefly discussed here as several reviews have appeared on the subject (Ref 7-1, 7-2).

7.2.2 Injection-limited conduction

Thermal emission of electrons from the contact electrode to a conduction level in the insulator takes place when electrons moving in the required direction have sufficient thermal energy to surmount the interface barrier φ . Once injected into the conduction band of the dielectric as shown in Fig 7-1a they are swept away to the anode under the action of the applied field. The barrier height is reduced by the ~~image force on the electrons leaving the electrode~~ applied field and this introduces the field dependant term into Richardson's thermionic emission equation to give the Schottky relation for current density.

$$J = A^* T^2 \exp\left[\frac{-e\left(\varphi - \frac{eE^{1/2}}{K\epsilon_0}\right)}{kT}\right] \qquad 7\text{-}1$$

Where T is the absolute temperature, A^* the effective Richardson coefficient, E the electric field strength and K the dielectric constant of the insulator.

Field emission occurs with large applied fields which produce a potential barrier sufficiently thin to allow electrons to tunnel through it into the conduction band of the insulator as shown in Fig 7-1b. For very thin insulator films (less than about 100 Å) electrons can also tunnel completely through the insulator

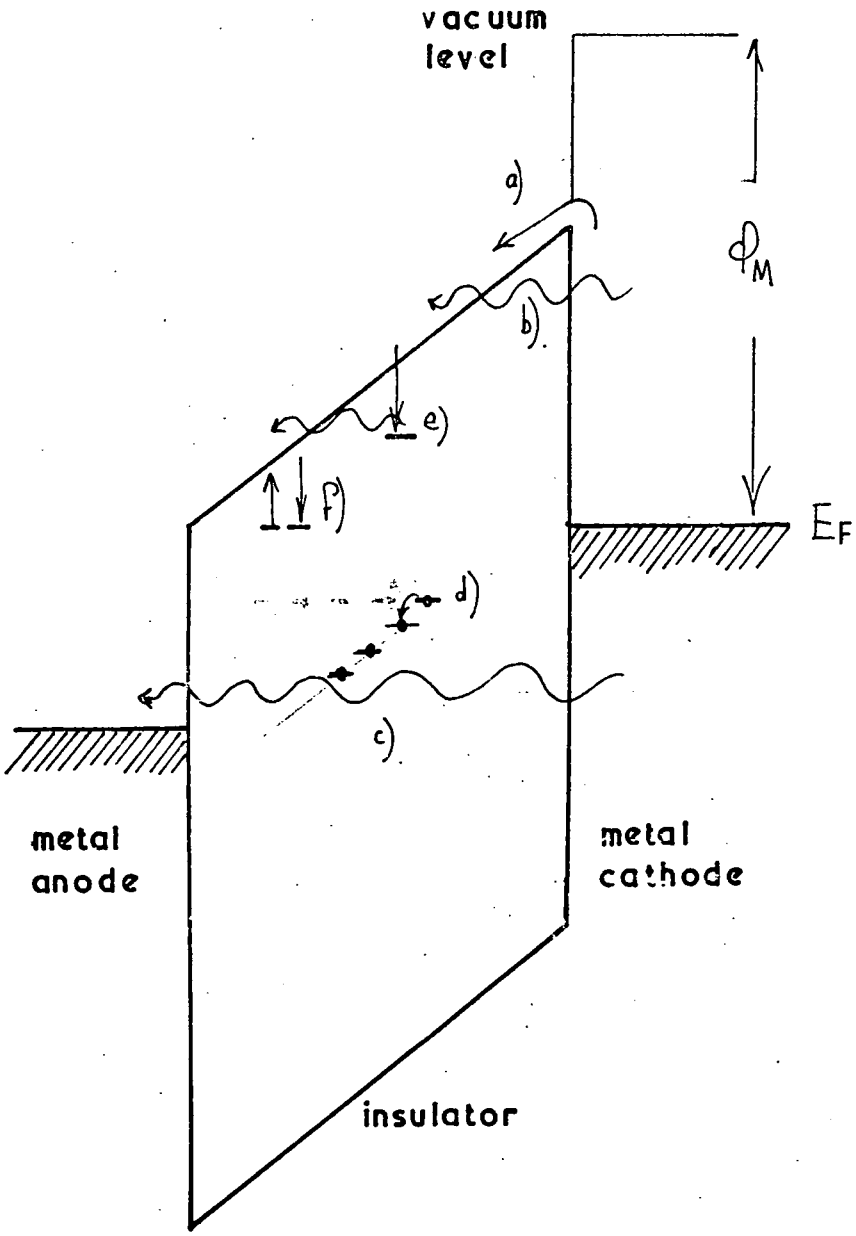


Fig 7-1 Schematic energy diagram of an ideal MIM structure showing several possible conduction mechanisms a) Schottky emission b) field emission c) 'direct' tunnelling from cathode to anode. d) impurity 'hopping' conduction e) Poole-Frenkel mechanism via traps f) which can dominate space-charge effects.

to reach the electrode as 'hot electrons'. Current densities for quantum mechanical tunnelling mechanisms can be obtained by integration from the expression

$$J = \int e N(T,E)\, D(E)\, dE \qquad 7\text{-}2$$

where N(T.E) is the number of electrons which are incident on the barrier per second per unit area having energy within the range E to E + dE and D(E) is the tunnelling probability. The integration is taken from the effective constant potential inside the source electrode to infinity. This equation can be worked out to various degrees of sophistication and the simplest gives the Fowler-Nordheim equation.

$$J = \frac{e^2 E^2}{8\pi h \phi} \exp\left[- 4 \frac{(2m)^{1/2}}{3he E} \right] \qquad 7\text{-}3$$

Other treatments (Ref 7-3) take into account the image force barrier lowering, the effective mass of the electron in the insulator and a term which takes into account the small temperature dependence of N(T.E). All of them however, retain the same general functional behaviour as the Fowler-Nordheim equation.

At high temperatures and comparitively low fields Schottky emission usually predominates, while at high fields tunnel emission takes over. However, between these two there is an intermediate region where thermally assisted tunnelling occurs. The complete treatment for all three regions is given in Refs (7-4,7-5)

7.2.3 Bulk-limited conduction

So far only those current mechanisms governed primarily by the properties of the electrode have been considered. However, when injection into the insulator is not the rate limiting process bulk effects predominate.

When the injecting electrode can supply carriers at a rate which is greater than the rate at which they flow away through the insulator space charge build-up occurs either in the conduction band or at trapping centres in the insulator. The ability of a contact to inject carriers into a semi-insulator is usually characterised by the work function (ϕ_M) of the contact material, and the electron affinity χ, the band gap (E_g) and the work function ϕ_s of the semi-insulator. (Ref 7-6). In theory ohmic and blocking contacts result for p-type material when

$$\phi_M \geq E_g + \chi \qquad \text{and} \qquad \phi_s > \phi_M$$

respectively. For n-type material these contacts result when

$$\phi_M \geq \chi \qquad \text{and} \qquad \phi_s < \phi_M$$

respectively. However, the extension of these criteria to large band gap solids is of doubtful validity and also there is the difficulty of determining the values of ϕ_m and χ which actually occur at the interface, so that in practise it is usually difficult to decide theoretically whether a contact will be ohmic or blocking.

Nevertheless if a low impedance contact is established several I-V relations operate depending on the space charge conditions, the trap density and the carrier trapping kinetics within the

insulator. These have been reviewed by Wright (Ref 7-7). For a simple single carrier model the I-V curve with increasing applied bias is initially ohmic followed by a classical space charge region, with I proporation to V^2, until a trap filled limit is reached when the current rises sharply to a trap free space charge region.

Double injection, i.e. of electrons at the cathode and holes at the anode, can also occur. In this case the transport and trapping kinetics of the electrons and holes then determine the type of I-V characteristics which are obtained and a wide variety of I-V curves is found.

Another bulk-limited conduction mechanism is due to electrons that are trapped at centres in the insulator themselves acting as a source of carriers. The process by which this occurs is called the Poole-Frenkel or 'internal' Schottky effect. Similar considerations apply as for the emission of electrons from the electrodes but with slight differences because the number of electrons available for emission is reduced and because the ~~image force~~ lowering of the potential barrier surrounding the trap is twice as large in this case due to the immobility of the positive charge. The theoretical current density for thermally assisted emission of trapped electrons into the conduction band of the insulator is given by

$$J = \text{const. } E \exp\left[-\frac{e\left(\phi_b - \left(\frac{eE}{\pi\epsilon_0 K}\right)^{1/2}\right)}{kT} \right] \qquad 7\text{-}4a$$

where ϕ_b is the height of the potential barrier around the trap.

If the concentration of the impurity centres is large so that a certain amount of overlap occurs between the wave functions of the centres then the electrons can tunnel or hop between the centres instead of being excited into the conduction band as in the Poole-Frenkel effect. This process is called impurity conduction. Several quantitative treatments exist for this situation and one of these has been given by Simmons. (Ref 7-8) and applied to MIM structures of Al_2O_3 and SiO. The field dependence of this mechanism is given by

$$J = \text{const.} \sinh\left[\text{const.}\, V\right] \qquad 7\text{-}4b$$

The impurity centres described above are immobile and act as"stepping stones" for electrons passing through the insulator. However, if the impurity centres are ionised and their mobility is sufficiently large then ionic conduction occurs. The ions move by way of defect sites (or vacancies) in a manner analagous to solid state diffusion and the ionic current density has a similar functional form which for large fields (10^5 V/cm) can be expressed by

$$J = J_o \exp\left[-\frac{\phi_d}{kT} - \frac{eEl}{kT}\right] \qquad 7\text{-}5$$

where ϕ_d is the height of the potential barrier between defect sites 'l' is the ir spacing and J_o is a constant.

Although the ionic current given by 7-5 is determined by the properties of the bulk, electrode processes can also influence ion currents in a way that depends on the nature of the charge transfer at the electrode. If electron transfer occurs readily at

the electrodes then 7-5 is obeyed and solid state electrolysis occurs. Failure of the ions to discharge results in the build-up of ionic space charge at the electrodes and a gradual decrease in the current due to polarisation effects. Injection of ions of the electrode metal is also possible as an electrode transfer process (Ref 7-9).

The factors that determine whether or not an ion discharges at an electrode include not only the work function of the electrode and the electrode potential of the discharging ions but also the physical nature of the electrode surface and in particular its ability to absorb the ion on its surface (electrocatalysis). In electrolytic solutions the potential required, in excess of the theoretical value (as obtained from the electromotive series) to electrodeposit ions at the electrode is termed the overvoltage, a term which must be determined empirically. In sec 7.4.3 it is concluded that Zn^{++} ions moving in the willemite films discharge at the metal (i.e. In, Au on Al) electrode of the MIS device whilst remaining ionised in front of the silicon electrode. In view of the above factors this will be shown to be a reasonable deduction.

7.2.4 Conclusions

The following conduction mechanisms in thin insulating films have been discussed

i) Schottky emission —— $\ln (J/T^2) \propto E^{1/2}/T$

ii) Field emission —— $\ln (J/E^2) \propto 1/E$

iii) Space charge limited currents a) Single carrier injection

b) Double carrier injection

iv) Poole-Frenkel conduction — $\ln(J/E) \propto E^{1/2}$

v) Ionic conduction — $\ln(J) \propto E/T$

In spite of the fact that all these processes have different field dependencies it is very difficult to distinguish between them on this basis, particularly as over a small range of V most will produce straight lines on a log plot and also some of the constants in the relations are very difficult to measure (e.g. the emitting area and the effective work function) (Ref 7-13).

A further complication is that although the possible conduction mechanisms discussed above usually occur in thin insulator films in isolation they can be inter-related in certain conditions. Several examples of this have been reported by other workers. For example electroforming effects in thin films of SiO and the subsequent voltage controlled negative resistance phenomena (Ref 7-8) have been ascribed to ion injection at the anode followed by impurity conduction with a 'memory' effect attributed to space charge build-up in certain voltage ranges. Similarly silicon nitride films on silicon have been shown (Ref 7-10) to exhibit three conduction mechanisms for particular voltage or temperature ranges depending on whether trapped electrons are field emitted to the conduction band (with or without thermal assistance) or whether they simply tunnel between the trap centres (Ref 7-10) Also thermal silicon oxide MIS structures have been shown to exhibit Fowler-Nordheim emission from either of the electrodes (Ref 7-11) whilst

double sandwiches of both Si_3N_4 and SiO_2 (Ref 7-12) give rise to further charge memory effects in addition to the different conduction mechanisms of the constituent layers.

The above two examples of conduction processes in thin films are particularly relevant because they used silicon as one of the electrodes compared with the MIM structures described in the theoretical discussion. Space charge effects in the surface of the silicon of an MIS structure have been discussed in the previous chapter. Electrode controlled conduction in thin films is predominately concerned with the injection of electrons at the contacts and it can be shown that for a positive bias on the metal contact electrode of a MIS structure both p-type and n-type silicon provide a degenerate supply of electrons. For the p-type silicon this occurs only after the inversion of the surface. The voltage drop across the silicon space charge in this case can, however, never be greater than the energy gap because of the effective limit to the width of the depletion layer in the surface (Sec 6.22 Eqn 6-2 and 6-20). It follows that for a moderate bias (greater than about 10 volts) only a small part of the voltage is dropped in the depletion layer and the field in the thin film is approximately equal to the applied voltage divided by the film thickness.

7.3 Experimental Procedure

The willemite devices used for the conduction measurements were the same as those used for the C-V work and their preparation is described in Sec 4.3 (Fig 4.1). Gold, indium and aluminium top

contacts were used. The currents in the willemite devices were usually less then 10^{-10}amp for most of the bias range. These were measured with a Model A-33C Vibron vibrating-reed electrometer which had built-in measuring resistors of up to 10^{12} ohm to give full scale deflection for currents of 10^{-14} amp to 10^{-6}amp. The devices were mounted on a $\frac{1}{2}$ inch thick block of Perspex which was fixed on to the movable stage of a microscope so that the top contacts could be selected in turn using a gold probe held in a micromanipulator and the whole arrangement was enclosed in an earthed metal box to prevent pick-up. A schematic diagram of the apparatus is given in Fig 7-2.

Apart from the samples with dot contacts (Fig 4.1) on which most of the experiments were performed, a guard ring contact arrangement was used on one device to see if there was any leakage across the surface of the willemite. With the guard ring at the same potential as the dot inside the ring, leakage current cannot flow. However, for all the contacts on this device the guard ring potential did not affect the current. This showed that leakage currents across the surface were negligible probably due to the thinness of the dielectric compared with the long paths for any surface currents.

Photocurrents measurements were carried out by suspending the device in the monochromatic light beam of an Optica spectrophotometer Type CF4 which used a hydrogen lamp source. The device was mounted normal to the horizontal light beam and positioned

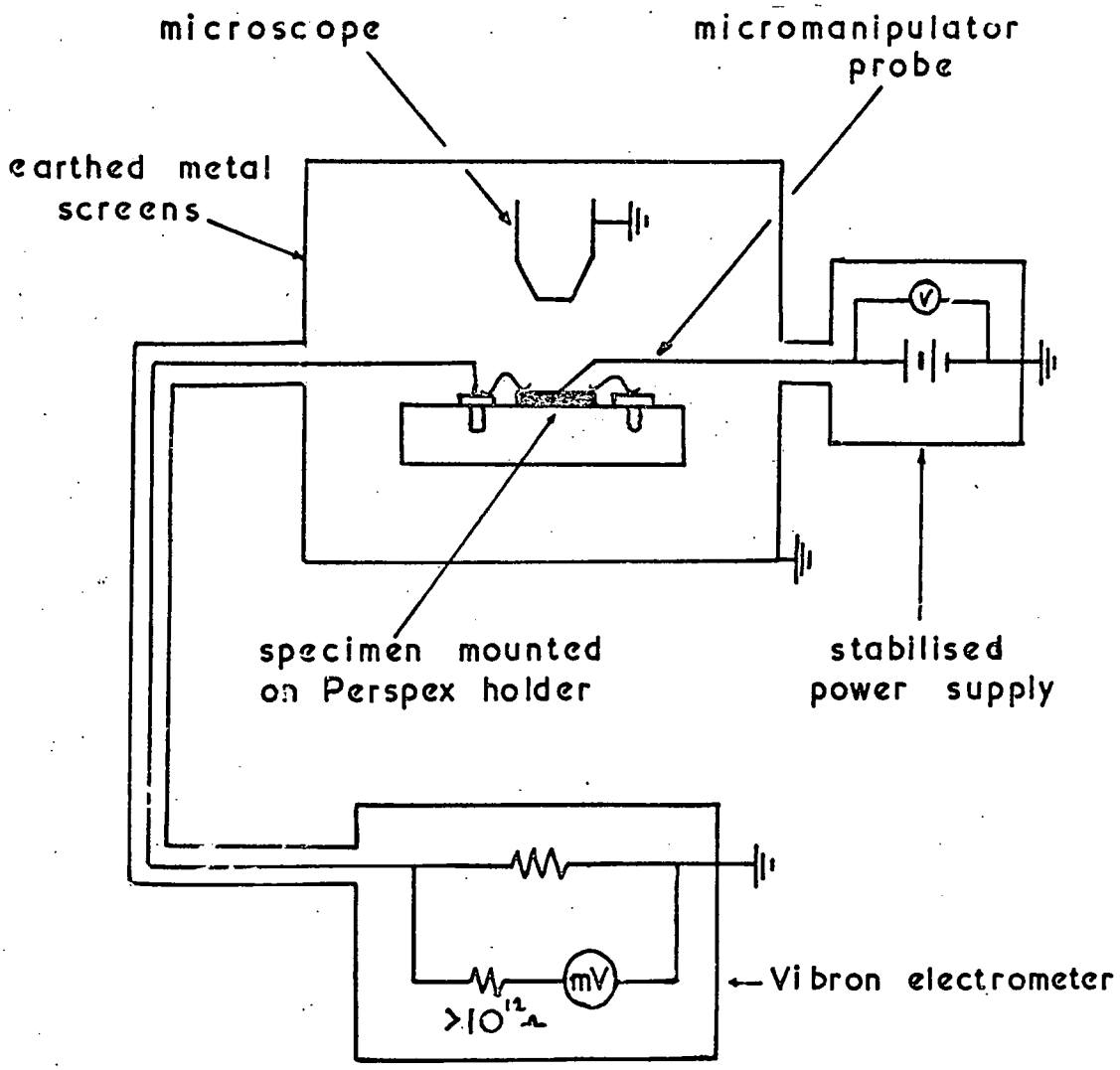


Fig 7-2 A schematic diagram of the apparatus used for the current-voltage measurements

using a X-Y microscope stage in a vertical plane. A microprobe was also attached to the X-Y movement so that the device could be contacted before being positioned in the light beam.

For the photocurrent experiment transparent gold top contacts (about 10^{-3} cm^2 in area) were evaporated on to the willemite in the standard manner. The degree of transparency was assessed by a simultaneous evaporation on to a silica substrate next to the device. Any absorption of light in the gold could then be accounted for by measuring the absorption spectrum of the gold on transparent silica.

The wavelength of the monochromatic light from the Optica was automatically scanned at a rate of 65 Å min^{-1} starting at 1800 Å and the photocurrents were measured with the Vibron as described above. This was repeated for a few set bias voltages of both positive and negative polarity.

7.4. Results and Discussion

7.4.1 Experimental I-V curves

Current measurements were carried out on both MWOS and MWS devices. The MWOS devices had films thicker than about 1000 Å with a detectable oxide film at the willemite-silicon interface and the MWS type had thinner films in which all the oxide had been converted to willemite. Both p-type and n-type silicon sustrates were used for the MWOS samples but the later MWS ones were restricted to n-type silicon due to the lack of time. It was found that the substrate type and resistivity (in the range 1-10 ohm cm) did not

affect the I-V curves of the MWOS devices. This was probably due to the formation of degenerate inversion layers at the silicon surfaces which provided contacts that were effectively metallic. This is described in more detail later. Similar lack of substrate dependence is also expected for the MWS devices.

The typical I-V characteristics for a MWO/p-Si device (EW73) and a MW/n-Si device (EW110) are shown in Figs 7-3 7-4 and 7-5 for both negative and positive voltages applied to the top contact. (All voltages refer to the polarity of the top contact with respect to the silicon). Several top contacts were tested on each device and they all showed very similar characteristics. Both types of devices show a broadly similar I-V behaviour with larger currents flowing for positive bias after a certain critical field strength has been reached, although the increase was greater for the MWS devices and the critical field about one half that of the MWOS devices.

The experimental uncertainty on a particular I-V plot is primarily due to the varying current. Two distinct types of variation were evident corresponding to different regions of the I-V curves. These were (i) polarisation effects (i.e. the current decaying slowly with time) which occurred throughout the negative bias range and in low positive bias regions and (ii) spiky and relatively high frequency variations in the current for the high positive field region. In the polarisation region a rough estimate of the error on the current measurements was made by noting the

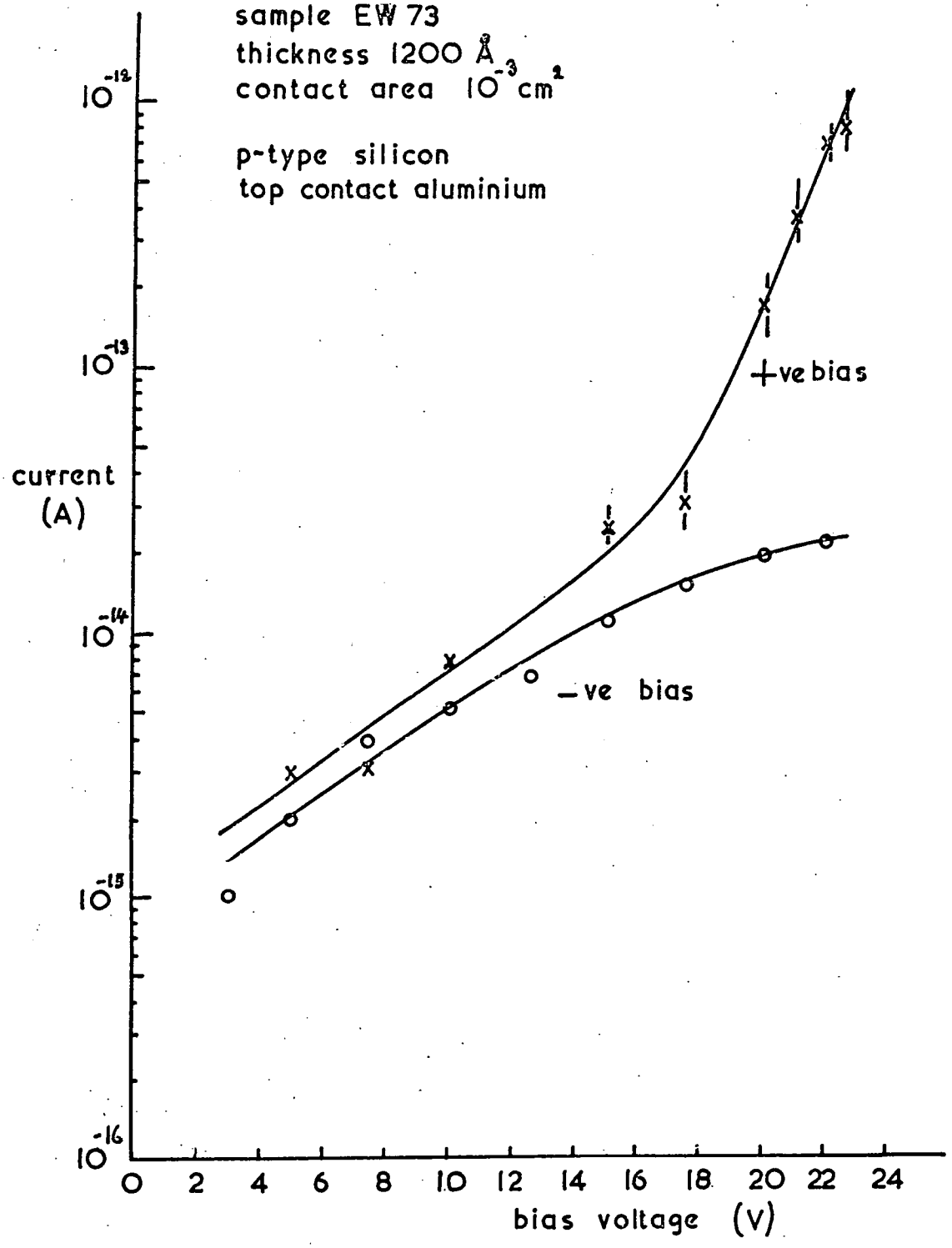


Fig 7-3 Current-voltage characteristics of an MWOS device

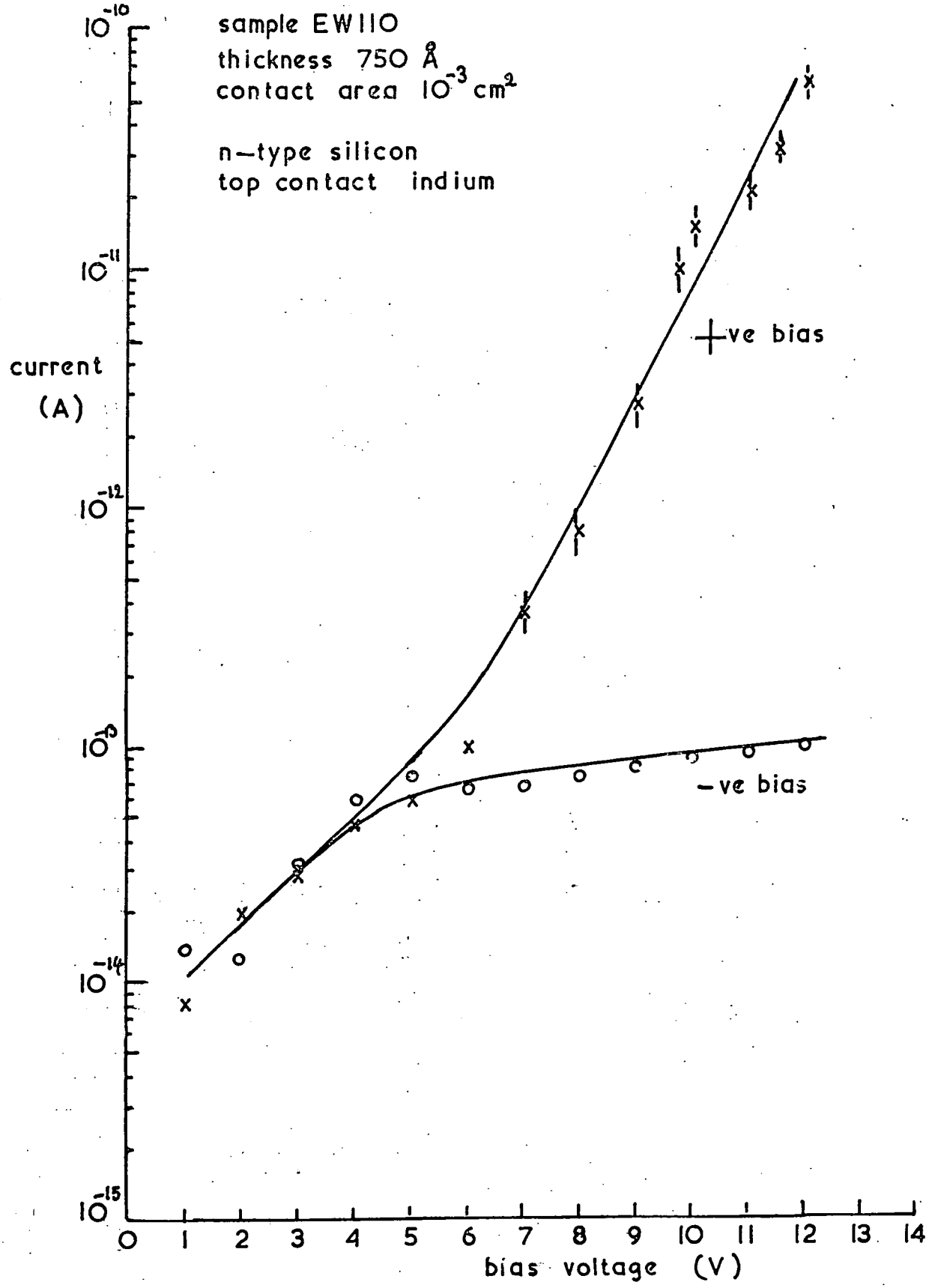


Fig 7-4 The I-V characteristics of an MWS device

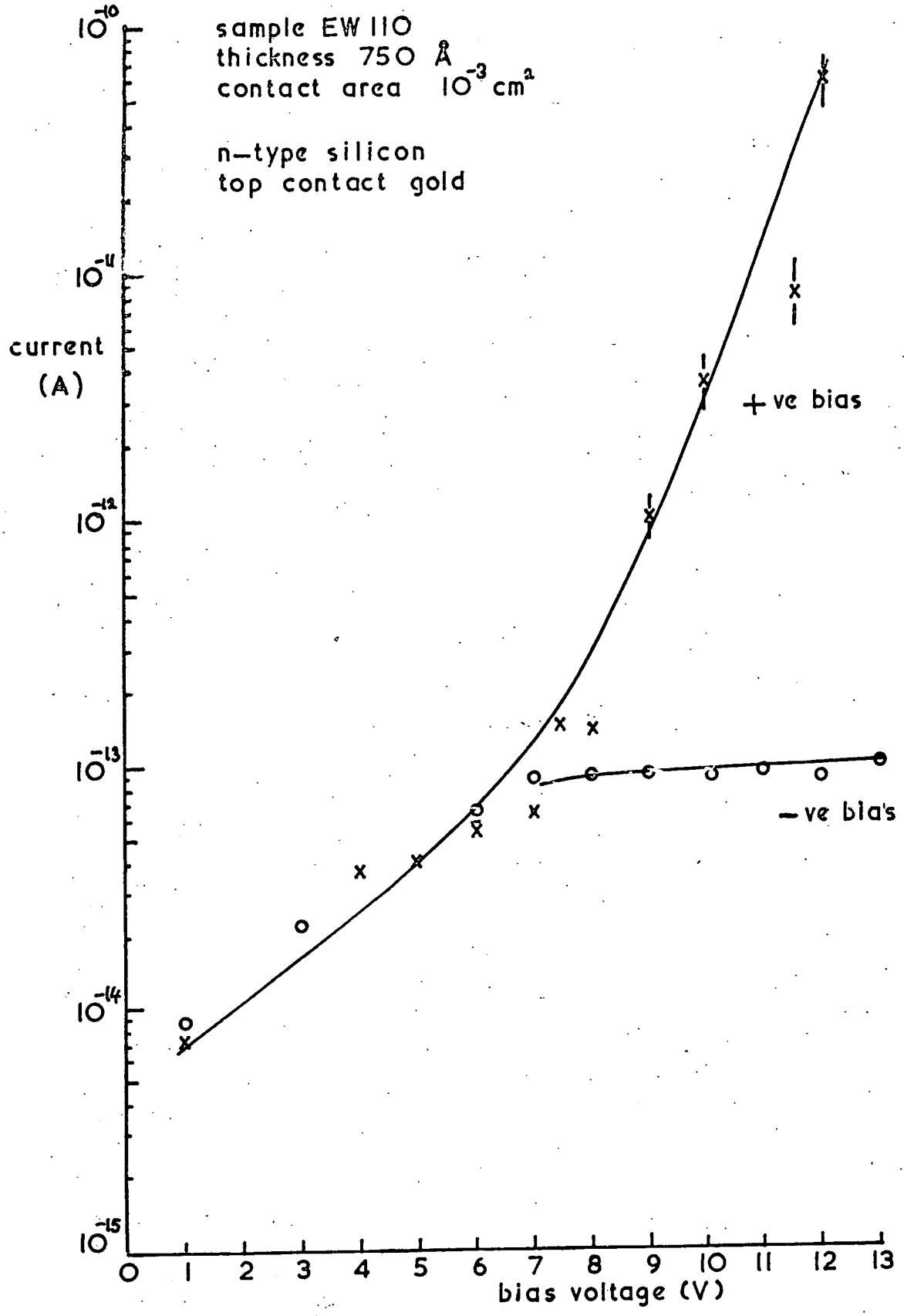


Fig 7-5 The I-V characteristics of an MWS device

current values 5 min and 10 min after the start of each measurement. Experimental errors in the high positive bias region were estimated from the recorder trace over a period of about 10 min. These are indicated by a vertical line in the I-V curves. Aluminium, gold and indium top contacts were used in the experiments but no difference in the general shape of the I-V characteristics was observed. This can be seen from Figs 7-3,7-4 and 7-5.

For values of positive bias greater than the maximum shown in the above graphs further increases in current occurred, but these are not included because they were very erratic and difficult to measure so that only a few readings were possible. At first these even higher currents were confused with breakdown of the device but it later proved possible to reproduce the readings on the same device when the experiments were repeated. A typical value was about 100 μamp at 15 V for the device used in Fig 7-4 and this provided an approximate order of magnitude agreement with an ~~straight line~~ extension of the positive bias curve shown in the graph. Because of this and because these much higher currents occurred only for positive bias it seemed likely that the conduction mechanism was the same throughout the whole high positive bias region.

Dot contacts on the unconverted thermal oxide next to the willemite film (Fig 4.1) were also used for current measurements. The oxides always showed polarisation effects for both negative and positive bias throughout the full voltage range, and the I-V curves

were essentially symmetrical and similar to the negatively biased willemite devices. Only a few I-V measurements were made on the MOS devices as these structures have been the subject of extensive study elsewhere (Ref 7-11). Their prime use here was as a rough comparison with similar measurements made on the willemite.

The most important observation from the I-V measurements is the rapid increase in the high field current in the willemite which occurs for a positive bias only and which is not found in silicon oxide. This increase will later be seen to be important for possible EL application (Sec 8.3).

7.4.2 Photoconduction measurements.

Measurements of photoconduction were initiated to see if carriers in the contact material could be photoexcited over the potential barrier at either electrode and into the willemite. The relationship between the resultant photocurrents and the wavelength of the incident light could then be expected to yield values for the barrier heights at the interfaces. This has been done successfully by Williams using SiO_2 (Ref 7-14) and by Goodman (Ref 7-15) with Si_3N_4. It was expected that a knowledge of the barrier heights obtained by this method, and particularly their variation with different contact materials, would help to elucidate the conduction mechanisms in willemite films. In practice this effect was swamped by the bulk photoconduction of the willemite which, as will be seen, incidentally produced information that helped to confirm the conduction model proposed in Sec 7.4.2.

Fig 7-6a shows the photoresponse of a 1000Å thick willemite film (EW 97) with 20V applied. This curve was identical for both positive and negative bias voltages. In these measurements the wavelength of the light was automatically scanned at a rate of 65 Å / min. Fig 7-6b shows the response curve for both bias polarities plotted together after a correction factor has been applied to take account of the spectral response of the hydrogen lamp.

Polarisation of the currents occurred when the wavelength scan was stopped at any part of the exciting region. Negative currents flowed in the measuring circuit when either the voltage was reduced to zero or the spectrophotometer slits were closed. These phenomena were consistent with the redistribution of space charge in the willemite films.

The space charge in the film could have been produced by photo-injection from the contact of electrons which were subsequently trapped in the willemite but this would have meant that the barriers at both electrodes would have to be exactly the same height to account for the symetrical photoresponse. It was more reasonable to attribute the photocurrents to intrinsic photoconduction within the willemite and this was substantiated by two further points. Firstly, Kröger in 1939 (Ref 7-16) found the absorption edge in a willemite phosphor to be 2200Å which is very similar to the response shown in Fig 7-6 (2250 Å), and secondly photocurrents were not observed in the thermal oxide test points next to the willemite as they should have been according to Williams (the initial photoresponse was at 2900 Å in his experiment on SiO_2)

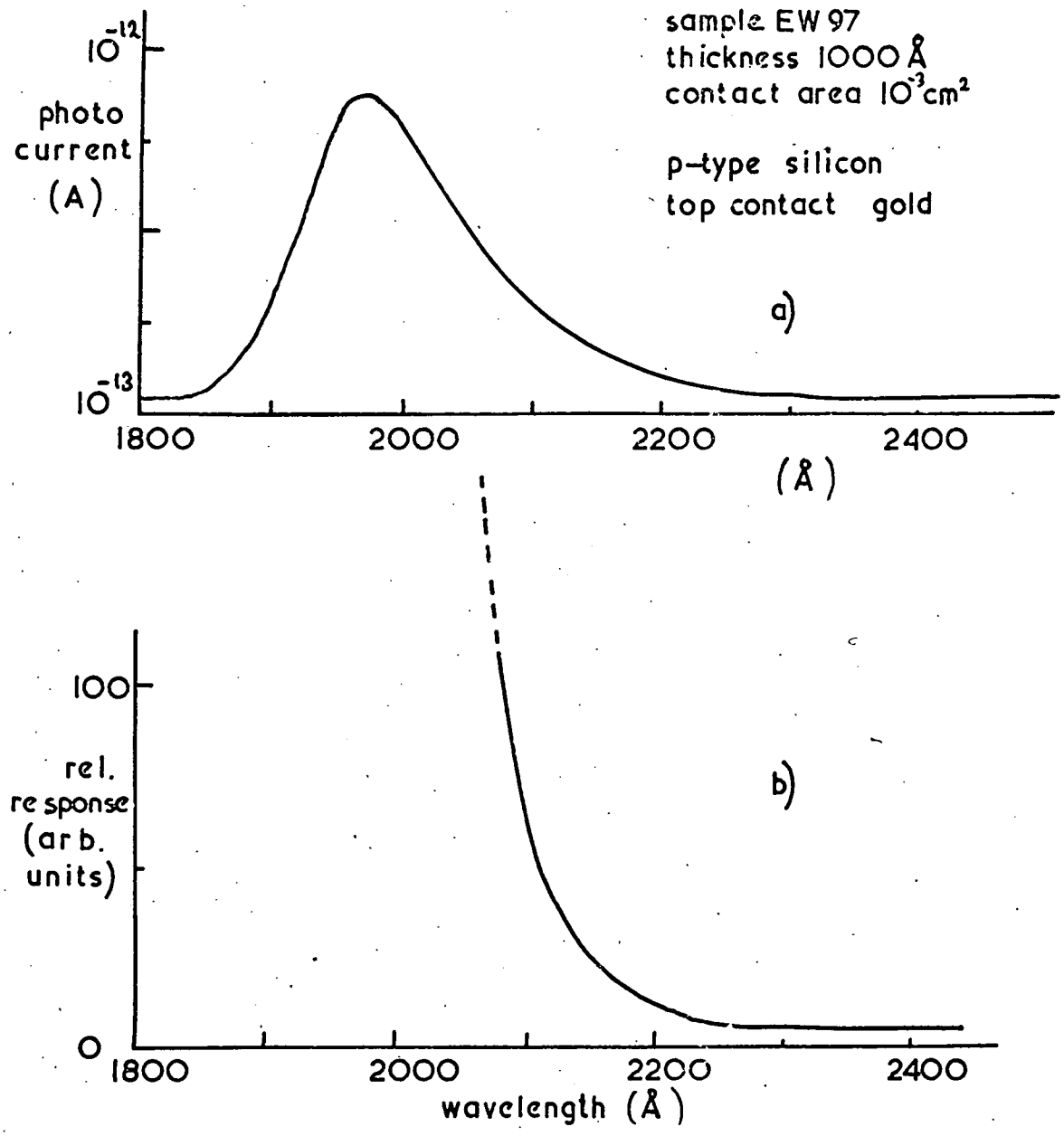


Fig 7-6 a) Photocurrent response for both +20V and −20V bias

b) Response corrected for the spectral distribution of the hydrogen lamp (detector insensitive < ~2050 Å)

This lack of photocurrents in the SiO_2 also ruled out the possibility that photoemission from the contact into the air might account for the photocurrent because the photoemission from gold contacts on the willemite and on the oxide would then be the same. Photovoltaic effects were also discounted as the source of the observed currents because any voltage induced in the device as a result of the light would then be divided between the high resistance of the device and the relatively low measuring resistance in the Vibron circuit so that only a very small part would be recorded.

It was therefore concluded that the monochromatic ultraviolet light that was used in these experiments was too weak to produce any measurable photoexcitation from the contacts and that the photoeffects were due to intrinsic photo-conduction (Ref 7-17) in the willemite. An estimate of the optical band gap of the willemite films obtained from the cut-off of the photocurrent was 5.5 ± 0.1 eV. which was in very good agreement with previous work#.

Transient photo-effects were also observed when the slits of the Optica were first opened at any wavelength from 3200 Å to about one micron, but these were attributed to the creation of electron-hole pairs in the silicon substrate and their subsequent recombination. This effect could easily be distinguished from that described above.

The fact that the high-field photocurrent was independent of polarity whereas the dark current was considerably greater

with a positive bias is confirmation of the conduction model proposed in the next section.

7.4.3 A model for the conduction process

A model for the conduction of willemite films on silicon must explain the observations described in Secs 7.4.1, 7.4.2 and 6.6. These are summarised briefly below :-

i) Relatively large and noisy currents flow across the device when high positive voltages are applied to the top electrode.

ii) Throughout the negative bias region and for a low positive bias the currents are comparatively small and they show polarisation effects.

iii) These characteristics are, in general, unaffected by the metal of the top contact and also the conductivity type of the silicon. It is of particular interest that both indium and gold contacts with MW n-Si devices produce a similar I-V curve because indium has a smaller work function than the electron affinity of silicon and vice versa for gold. The thermal properties of the top contacts do however effect the breakdown properties of MIS structures (i.e. their ability for self-healing) and this is discussed in Chap 8.

iv) The photocurrents produced by irradiating the willemite device under bias with ultra-violet light are about the same for both polarities and show no increase in the high positive bias region.

v) The C-V work described in Chap 6 showed that ion drift was occurring in the willemite films under bias. An estimate of the ionic conduction can be obtained from the C-V curves by assuming that in each sweep of the applied voltage all the charge movement is reflected in the hysteresis of the C-V curve (Sec 6.5.2). In Fig 6-13 the hysteresis voltage is 2 volt and the capacitance about 80 pF so that with a sweep time of 25 sec this gives a value for ionic current of about 5.10^{-12} amp, which is in approximate agreement with the currents observed in similar devices with negative and low positive voltages applied.

When the I-V results were plotted in accordance with the simple conduction relationships described in Sec 7.2 it was found that straight lines were obtained for most of the mechanisms in the negative and low positive bias ranges and similarly in the high positive field region. The inability to distinguish a single mechanism in this way was mainly due to the small voltage ranges involved and to the experimental error on the results. Nevertheless not one of the simple conduction mechanisms could explain all the facts described above for all the voltage range and because of this a model was necessary that incorporated a combination of processes.

Consideration of these points led to the following simple model for the conduction process which appears to fit all the facts. It is proposed that in the negative bias region the very small currents are mainly ionic and due to the transport of Zn^{2+} ions in the willemite films via defect sites. Under the influence

of the negative field these ions move towards the cathode where at least some of them will discharge because the cathode is metal. Ionic conduction produces some solid state electrolysis in the willemite but this need not affect the film greatly because of the small currents involved.

With positive bias the ions move towards the silicon electrode where, in contrast to the metal electrode, it is proposed they do not discharge but instead form a positive space charge that builds up in front of the silicon. The size and distribution of this space charge is governed by the rate of arrival of ions at the interface and the opposite flux of ions due to their electrostatic self-repulsion and self-diffusion. For low positive fields however, the space charge is small and currents are due to the ionic conduction itself.

Polarisation effects in the negative and low positive bias region are probably due to the ionic nature of the currents. In addition, some electrons may be thermally emitted from the electrodes and contribute to polarisation effects by becoming trapped in the willemite. However, the fact that no work function dependence was noted shows that currents due to this mechanism are probably negligible.

With a high positive bias, space charge enhances the electric field at the cathode (Fig 7-7) until this is sufficiently large to allow electrons from the silicon to tunnel through it. The current then increases rapidly with voltage in accordance with the

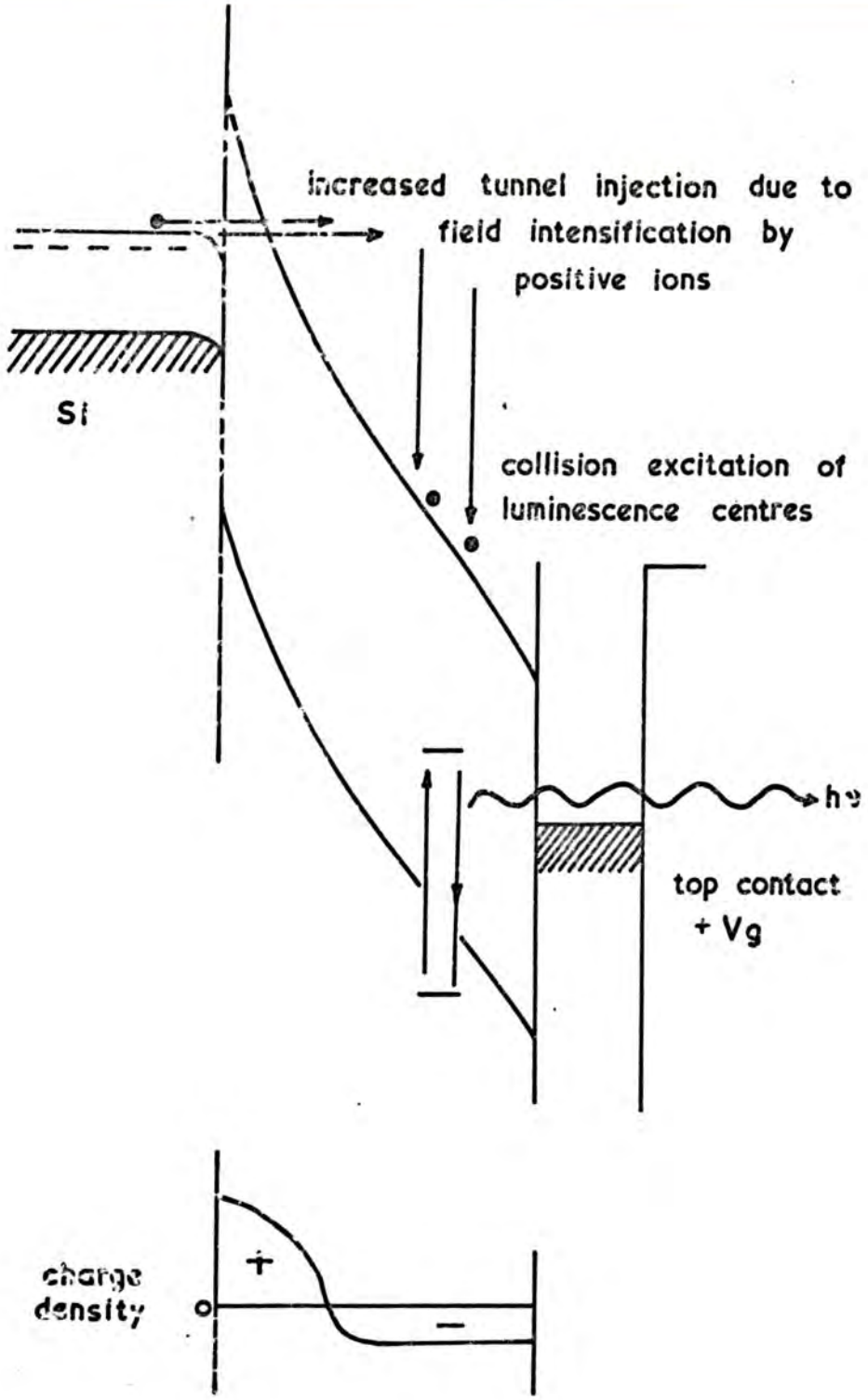


Fig 7-7 Schematic energy diagram and charge distribution for the proposed mechanism for increased current in MWS devices with positive applied bias.

Fowler-Nordheim relation (Sec 7-2). However, any attempt to interpret the constants in this equation in terms of barrier heights and effective masses will be very difficult. In general with high field conduction it is not easy to attach any significance to these values because of uncertainties involved in defining the emitting area and because of 'patch effects' but in this case it is particularly difficult because of the field intensification that is thought to occur so that the actual field value is not easily estimated.

In the tunneling process it is proposed that the electrons passing through the positive space charge do not combine with the oppositely charged ionic space charge because the time taken for them to tunnel is much smaller than the recombination time of an electron with a positive ion. Neutralization of the space charge probably does occur however when electron-hole pairs are generated in the willemite as a result of the ultra-violet light. This accounts for the photocurrents being about the same for both bias polarities (Fig 7-6) whereas the dark current is considerably greater at high fields just for the positive/positive where the build-up of space charge allows tunnel emission to occur.

The fact that the positive ions do not discharge to the silicon can be explained by assuming that a thin oxide layer exists at the interface. This layer would be greater than about 100 Å thick for the MWOS devices and less than this or even only a monolayer in the MWS device. The ions then crowd up against the

oxide and are unable to discharge. They therefore enhance the electric field which reaches a critical strength when electrons can tunnel through the oxide layer into the conduction band of the willemite, in much the same way as they do in a hot electron triode (Ref 7-2 p 109.) ~~N.P.~~ An alternative explanation of the lack of ionic discharge at the silicon as opposed to the metal interface is that the ions can be thought of as having energy levels in the forbidden gap of willemite and that these ~~could~~ might correspond to energies in the forbidden gap of the silicon. This would mean that an ion drifting up to silicon interface would have a very small chance of finding an electron in the silicon to neutralise it. However, the C-V curves have shown the presence of surface states at the interface and these could effect the neutralisation process. In view of this, the first explanation given above seems more probable although both processes could contribute to the effective insulation of the ions from the silicon.

The possibility of the enhancement of the electric field at a contact by mobile positive charges has often been used in the explanation of dielectric breakdown, e.g. in NaCl by Seitz in 1949, and more recently (Ref 7-19) in thermal oxide films on silicon which have been found to show different breakdown properties depending on whether the silicon is positive or negative. With a positive bias on the top electrode the breakdown of silicon oxide obeys Peek's law which states that the breakdown voltage is proportional to $t^{-\frac{1}{2}}$ where t is the time that elapses after the initial application

of the bias. On the other hand for negative bias the breakdown is abrupt and occurs at a constant voltage independent of time. These observations could only be explained in terms of positive ions drifting under the influence of the applied field but discharging only at the metal. The ions in this case are presumably provided by some form of contamination (e.g. Na^+) but the above observations of Peek's Law would then lend support to the model proposed above for the conduction mechanisms in thin willemite films.

7.5. Conclusion.

The I-V characteristics of willemite films on silicon have been investigated and a model is proposed for the conduction mechanism. This involves mainly ionic conduction for the whole of the negative voltage range and for low positive bias values. The mobile positive ions are thought to discharge only when the cathode is the metal. For the positive polarity, when the cathode is silicon the ions produce field enhancement at the interface which results in quantum mechanical tunnelling of electrons from the silicon. Relatively high currents then flow with high positive applied fields only. This observation and its explanation are of great significance for the possible use of willemite in display devices.

An important aspect of these experiments has been left out of the above discussion, and this is the existence of green d.c. electroluminescence that accompanies the large currents in the high positive field region. This will now be discussed in the following chapter together with other luminescent properties of the willemite films.

References for Chapter Seven

7-1 R.M. Hill, Thin Solid Films, 1, 39-68 (1968)

7-2 D.R. Lamb, Electrical Conduction Mechanisms in Thin Films London (1967)

7-3 R.H. Good Jr. and E.W. Muller, Handbuch der Physik, Vol XXI p 176 Berlin (1956)

7-4 E.L. Murphy and R.H. Good, Phys. Rev. 102, 1464-1473 (1956)

7-5 W. Tantrapom, Solid State Electron. 7, 81-91 (1964)

7-6 A.J. Dekker Solid State Physics, p 358 London (1957)

7-7 G.T. Wright, Solid State Electron. 2, 165-89 (1961)

7-8 J.G. Simmons, R.R. Verderber, Proc. Roy.Soc. A, 301, 77-102 (1967)

7-9 R.R. Verderber, J.G. Simmons and B. Eales, Phil. Mag. 16, 1049 (1967)

7-10 S.M. Sze, J. Appl. Phys. 38, 2951 (1967)

7-11 M. Lenzlinger and E.H. Snow, J. Appl. Phys. 40, 278 (1969)

7-12 D. Frohman-Bentchkowsky and M. Lenzlinger, J. Appl. Phys. 40, 3307 (1969)

7-13 M.J. Morant, Proc. Phys. Soc. 68, 513 (1955)

7-14 R. Williams, Phys. Rev. 140 A, 569 (1965)

7-15 A.M. Goodman, Appl. Phys. Letters 13, 275 (1968)

7-16 F.A. Kroger, Physica 6, 764 (1939)

7-17 R.H. Bube Photoconductivity of Solids, London (1960)

7-18 F. Seitz, Phys. Rev. 76, 1376 (1949)

7-19 M. Yamin, I.E.E.E. Trans. Electron Devices ED-12, 88-96 (1965)

CHAPTER · EIGHT

THE LUMINESCENT PROPERTIES OF WILLEMITE FILMS

8.1 Introduction

Zinc silicate is an insulator with an energy band gap of about 5.5 eV (Sec 7.4.2) and in its 'pure form' it exhibits only a very feeble pale violet luminescence. This luminescence is thought to be due to an excess of silicon in the lattice which results in the formation of the self-activated phosphor Zn_2SiO_4: Si (Ref 8-1). When zinc silicate is activated with manganese however a brightly green luminescent phosphor, Zn_2SiO_4 : Mn, is produced which also occurs in nature as the mineral willemite.

The manganese ion in willemite normally substitutes for the zinc ion in the lattice, where the local energy environment is such that the green luminescence occurs as opposed to the more usual yellow emission of the Mn^{2+} ion in other phosphors (e.g. ZnS:Mn). The configurational co-ordinate diagram has been derived by Klick and Schulman in 1952 (Ref 8-2) who attributed the emission transition to a spin reversal of one of the 3d shell electrons of the Mn^{2+} ion, the mode of vibration being a radial one for the four oxygen ions surrounding the manganese. (Fig 5-1 shows a diagram of the crystal structure.)

The luminescence spectrum of willemite is broad and featureless even at low temperatures (e.g. 20°K, Ref 8-3) which is characteristic of emission from localised centres. The peak emission occurs at about 5200 Å for 1% by weight of Mn, although it has been found to move slightly longer wavelengths as the percentage of

manganese is increased. This is due to small changes in the configurational energy of the activators caused by some overlapping of their potential wells.

In addition to the normal green emitting form of willemite both red and yellow luminescent phases exist. These were first noted in 1923 by Schleede and Gruhl (Ref 8-4) who found that when the green luminescent powder was fused and rapidly quenched a vitreous looking product was obtained which showed a red luminescence. Less rapid cooling resulted in a phosphor with a yellow luminescence. The formation of the red phase of willemite is thought to be due to the evaporation of ZnO from the melt leaving an excess of silica which on cooling forms crystals of the red luminescent phosphor SiO_2: Mn. The yellow phase, on the other hand, has been attributed to the formation of a pseudo-cristobalite form of willemite (β-Zn_2SiO_4) in which the manganese takes up a network modifying position in place of some of the silicon ions. The yellow phase is, however, unstable and it reverts to the green form above about 800° C. Red and yellow luminescent films have also been produced by Feldman and O'Hara who evaporated the green emitting powder on to glass substrates under various conditions and then subjected the films to certain heat treatment (Ref 8-5).

In spite of the two luminescent variations described above the green emitting phase is not only the most common form of willemite but it is by far the brightest and most efficient phosphor. The green emission may be excited either by ultra-violet light or by

electron bombardment, although the efficiencies may be different for each type of excitation, depending in an empirical way on the precise details of the preparation (Ref 8-1). Electroluminescence has also been observed from willemite and past worker's reports of this are discussed fully in Sec 4.2 . . The observation of d.c. EL from the willemite films used in this work is described in Sec 8.3. in which a possible mechanism is also discussed.

The photoluminescent excitation spectrum of willemite has a cut-off at about 2250 Å for a weakly activated phosphor. However, as the concentration of manganese increases, the absorption band of the manganese ions themselves appear on the long wavelength side of the cut-off. Further than this the nature of the energy transfer process that excites the manganese ion has been very little studied by modern workers. This is also partly true of the cathodoluminescent spectrum where excitation occurs by way of secondary electrons generated by the incident primaries. The emphasis on most of the willemite phosphor research in the past has been on the development of bright green, efficient cathode ray tube screens and once this had been achieved very little further work was done. Once the green emission of willemite had been perfected the phosphor researchers moved to new materials that were capable of efficient emission in other colours that were not possible with willemite.

8.2 Experimental observations of cathodoluminescence

The willemite films made as described in Sec 4.3 were brightly green cathodoluminescent. However, they showed no sign of

photoluminescence when irradiated with ultra-violet light from a high pressure quartz hydrogen lamp (2537 Å) either at room temperature or at liquid nitrogen temperature. One reason for this could have been that the precise structure and the chemical composition of the films were incompatible with the required photo-luminescent processes. Another possibility was that because the films were so thin (about 1000 Å) the ultra-violet light interacted only weakly with the luminescent centres in the film. (Some inter-action with the films must have occurred because intrinsic photo-conductivity was observed for wavelengths ~~greater~~ shorter than about 2200Å (Sec 7.4.2).

Commercial willemite powder obtained from Levy West Ltd (Type P1) behaved in a roughly similar way to the films, i.e. it was brightly green cathodoluminescent but only very feebly photoluminescent with ultra-violet light. This was probably because the commercial powder had been developed empirically to maximise the cathodoluminescent output and this also applied to the films themselves to a certain extent.

The emission spectrum of a cathodoluminescent willemite film is shown in Fig 8-1a. This curve was obtained using the Optica spectrometer (Type F4) with the phosphor excited at room temperature by means of a 2 kV glow discharge in a rough vacuum chamber. The curve is broad and featureless as expected (Ref 8-6), and the peak wavelength of 5195 Å agrees well with the value obtained on the same apparatus for the commercial P1 powder (5200 Å). However, very little more

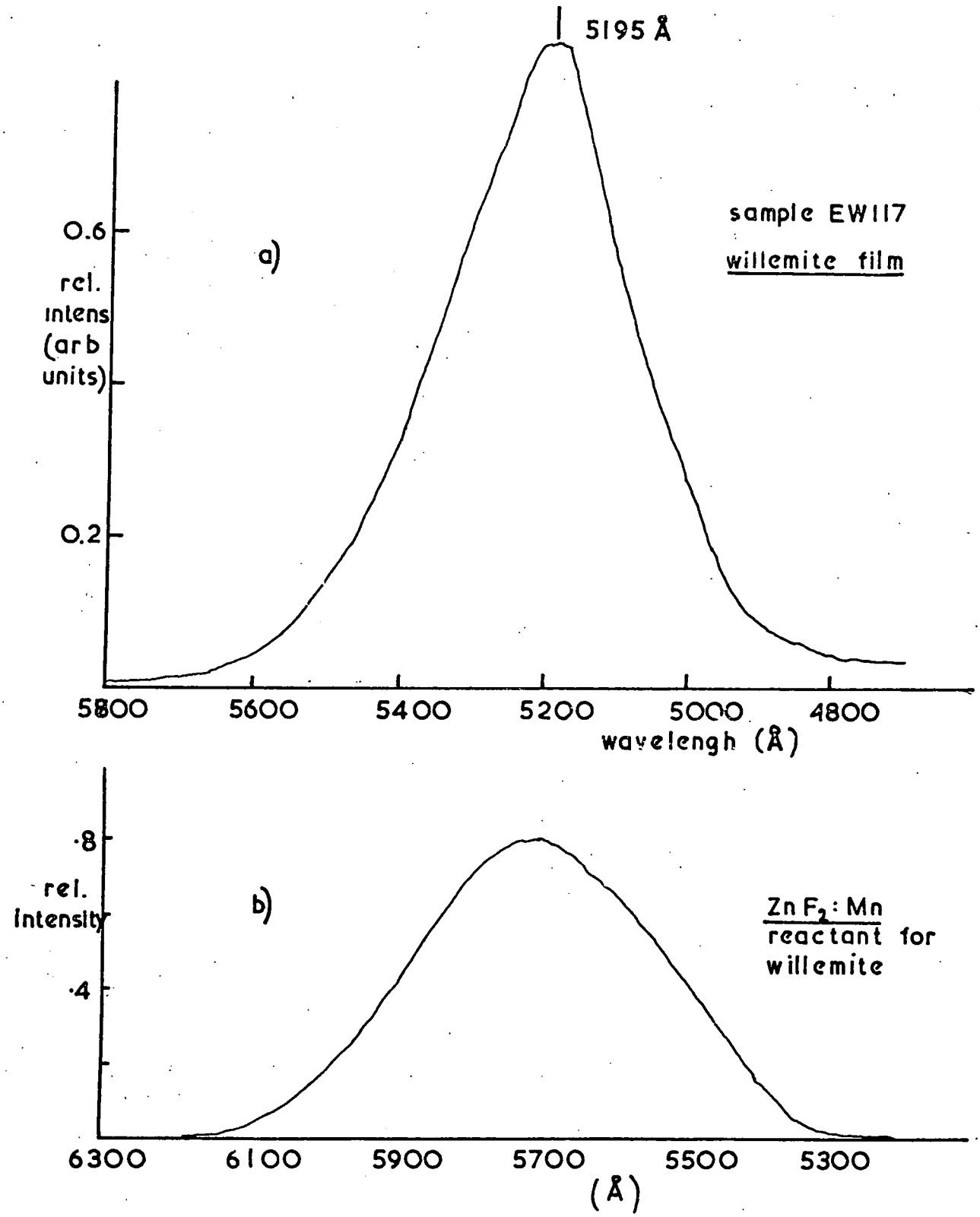
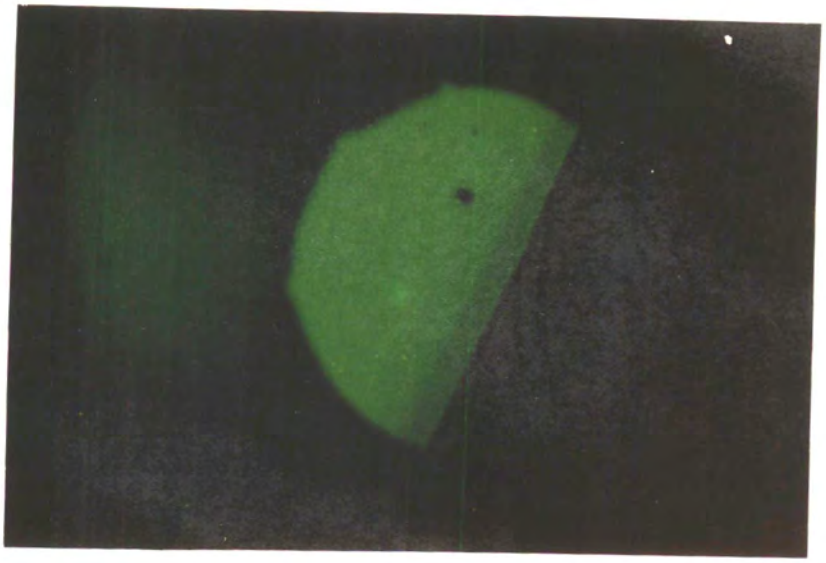


Fig 8-1 Cathodoluminescence spectra a) willemite film
b) ZnF_2 : Mn reactant.

useful information can be obtained from the emission spectrum using such a crude form of excitation and any further cathodoluminescent work would have had to use an exciting beam of electrons of known energy. Attempts at making a simple electron gun to enable this to be done were abandoned because of technical difficulties in the limited time available. The cathodoluminescent spectrum of the ZnF_2 : Mn powder from which the willemite films were made is shown in Fig 8-1b for comparison.

Green cathodoluminescence was also observed from the willemite films when they were examined with either the 100 kV electron microscope (Sec 5.3) or the 25 kV scanning electron microscope (Sec 5.5.) A colour photomicrograph (Kodacolor X print) of a willemite film in the JXA-3A microprobe (Sec 5.5.2) is shown in Fig 8-3. The purpose of the S.E.M. work was to assess the uniformity of the light output from the film. In this context the photograph is not typical of the film as a whole which gave a very uniform green light over most of its area. The apparent non-uniformities in this photograph were in fact chosen by the microscopist as something on which to focus. The blue spot is particularly interesting and it may be due to some sort of carbon contamination (probably from an organic solvent) reacting with the silicon to produce a blue luminescent form of silicon carbide. A different shade of green was observed at the edge of the film and this is thought to be due to contrast effects caused by the film being thinner at its edges. The blue spot and the different shades of green may however be due to



25μ

a)

2mm.

b)

Fig 8-3 a) The cathodoluminescent image of a willemite film in a S.E.M.

b) A simple cathodoluminescent display in willemite

different intensities of the same green colour which may affect the photographic process.

A few films were also produced that showed a pale yellow cathodoluminescence. These were made by using only a thin layer of ZnF_2 in the reaction process and then baking at about $800^{o}C$ so that excess silica was present (Sec 4.2). The yellow luminescence changed slowly to green for films baked above this temperature. These films also showed crystalline properties when examined by electron reflection diffration (Sec 5.3) although the exact determination of the crystal structure was impossible because of the diffuse nature of the diffraction rings. In view of these facts it seemed reasonable to attribute the yellow luminescent phase to the formation of beta-Zn_2SiO_4: Mn in agreement with previous works (Sec 8.1).

Non-luminescent films were produced when the willemite was made on substrates which had been heavily doped with phosphorus. The diffusion of phosphorus into silicon for the preparation of planar p-n junctions is described in Sec 2.4. One effect of the solid source method of diffusing phosphorus is that the surface of the Si is very heavily doped probably up to the solubility limit of about 10^{20} atoms cm^{-3} so that the thermal oxide grown on top is also heavily doped (Sec. 2.3) When willemite was produced by reaction with the oxide on these planar junctions it was found that only part of the film, that over the p-type silicon, was cathodoluminescent. The willemite produced on the heavily doped n-type region was not

luminescent and this was attributed to the poisoning effect of the phosphorus on the activator centres . If the surface layer of the heavily doped n-type region was first removed by steam oxidation and stripping of the oxide the normal green luminescent films of willemite could be formed on the remaining Si.
The luminescent willemite could also be produced on the arsenic doped n-type silicon that was used in other experiments.

A cathodoluminescent alpha-numeric display is shown in Fig 8-3. This was produced by simply evaporating ZnF_2: Mn through a mask on to oxidised silicon and then baking it in the usual manner (Sec 4.2). It is shown here to demonstrate that, although a fusion reaction occurs in the bake, the precision of the original evaporation of ZnF_2:Mn is retained in the willemite. The micrograph of Fig 8-3 was taken using Kodacolor X film and excitation by means of a 2kV glow discharge. The final objective of the work described in this thesis was to obtain similar alpha-numeric displays but with controlled electroluminescence rather than cathodoluminescence. Needless to say, this objective has not been achieved although an important step towards it has been the observation of d.c. electroluminescence of the films as described in the next section.

8.3 D.c electroluminescence of willemite films

8.3.1 Experimental observations

For electroluminescent experiments the willemite devices were positioned under a microscope in a darkened room. Bias voltages were applied in the same way as for the I-V experiments (Sec 7.3) but with either a 10 kΩ or 100 kΩ protective resistor in series with the device. Electroluminescence (EL) occurred when high electric fields were applied and the light output was photographed through the microscope using ASA 3000 Polaroid film. Only MWSdevices on n-type silicon were used for the EL experiments because of the lack of time.

The light output with electrical excitation was extremely small. This was partly due to the small area of the test points as well as the limited brightness. Thus it was always necessary for the eye to be well dark-adapted and to use the microscope with about x100 magnification. The problem of measuring the light output or its spectrum was not tackled due to the lack of time.

Two types of light output were noted. The first, which will be referred to as Type I, occurred for both positive and negative applied voltages, and it consisted of bluish-white flashes which were accompanied by localised destruction of the top electrode. These are thought to be due to the electrical breakdown of the willemite. The other form of light output Type II,occurred only with a positive voltage applied to the metal electrode and consisted of a relatively uniform green glow thought to be EL which was sometimes accompanied by flashes of Type I light also. The thermal-oxide devices adjacent to

the willemite exhibited only Type I light for both positive and negative voltages.

The light output is usually only visible because destruction of the top electrode occurs during the process. For Type I light this involves vaporisation of the metal contact for a few microns around the discharge point. A smaller area of the insulator is also destroyed inside this region exposing the silicon substrate underneath. After the breakdown event no part of the metal is in contact with the silicon (i.e. the breakdown is self-healing) and this results in the full voltage being reapplied across the device as soon as the MIS capacitor recharges. Further breakdown events then take place if suitable sites exist or alternatively if the bias is increased slightly. A continuation of the above processes can lead to gross destruction of the contact area.

Self-healing breakdowns also occur when the device is biased with the metal positive though to a lesser extent. It is thought that self-healing breakdowns play no part in producing the Type II light output. The latter was characterised by a green glow which produced much more widespread damage than the Type I process. Obviously the light cannot be observed through an opaque metal electrode unless some destruction occurs and it therefore seemed possible that the EL process could have been occurring beneath a greater part of the electrode than just the damaged area. One particular device in fact showed a faint ring of EL around the dot electrode, as though the light was being scattered out from beneath the contact.

Ideally a transparent top contact was required and, to this end, tin oxide contacts were deposited as described in Sec 4.3. These were not very satisfactory as the deposition process was mechanically vigorous, particularly with the masking, so that only a few undamaged films with tin oxide contacts were achieved. However, these showed the same general characteristics as those with the metal contacts although more breakdowns, producing the Type I light, were visible for both polarities, probably because of the SnO_2 deposition techniques.

A further shortcoming of the SnO_2 electrodes was that their breakdown was not always self-healing, so that only the most perfect willemite films could be used. The virtue of contacts that did display self-healing breakdowns was that any flaws in the film which would effectively short circuit the device, could be effectively isolated by the self-healing event leaving the more perfect willemite film behind to withstand the high fields necessary for electroluminescence.

Aluminium, gold and indium were used as contact metals but it was found that they all produced qualitatively similar results. The only practical criteria for a satisfactory contact appeared to be that it should be thin enough to produce self-healing breakdowns. Gold evaporated films having a thickness of about 1000 Å were found to be the best in this respect.

A few simple I-V measurements were conducted in conjunction with the electroluminescent work. The Type I light output was usually accompanied by a noisy intermittent current which

was characteristic of breakdown and which usually occurred for a bias of about 15 V across a 750 Å film. The current during the Type II light output was much less noisy, and the intensity of the EL was found to be approximately proportional to the current as far as could be judged by eye. Typical current values for maximum brightness with 15 V across 750 Å was about 100 μamps. The current density was impossible to determine because the actual area through which the current was flowing was unknown. The life time of the devices displaying this sort of EL was probably only a few hours but again this was difficult to quantify because the failure was usually due to breakdown at the damaged contact and clearing the short by a self-healing breakdown often regenerated the EL.

Although the spectral distribution of the Type II EL was not measured, a rough check on this was obtained by using a Chance optical interference filter which had a pass band from 5300 Å to 4950 Å, When this was placed in front of the electroluminescent film there seemed very little difference in the intensity of the light as far as could be judged by eye. This was in agreement with the cathodoluminescent spectrum (Sec 8.2) as the transmission band of the filter passed about 80% of the cathodoluminescent spectrum of willemite.

8.3.2 Discussion

The light output from willemite films described in the previous Section only occurs with very high applied fields and information on dielectric breakdown is obviously of relevance. Some of the recent published work on breakdown of thin films is described in Appendix A which shows that self-healing as well as propagating breakdown processes are well known and that the light output from discharges has also been studied.

The general features of the Type I light emission observed for negative bias with willemite films and for both polarities with thermal oxide devices (Sec 8.3.1) were consistent with observations of electrical breakdown. The nature of the discharge point and its degree of destruction were also similar to the effects of breakdown in silicon oxide reported in the literature. To the eye these (P.T.O)

discharges appeared as flashes of bluish white light in contrast to the green glow of the Type II light emission which was observed only with willemite films subjected to positive bias and which could last several hours.

Propagating breakdowns can last for several minutes but in this case the track which the discharge leaves behind is easily recognisable. No evidence of this discharge phenomena was found either for the oxide or for the willemite devices and it was concluded that the external resistance of 100 kohm was too large in this case for the propogation of electric breakdowns.

The light emission that accompanies the discharges described above is probably better thought of as a gas discharge. In distinction to this, however, it is proposed that the Type II light was a true solid state phenomenon and further that the EL was a direct consequence of the large increase in current that was observed in the I-V experiments for the same positive bias range. The high currents arise as a result of field intensification in front of the silicon caused by mobile positive ions in the willemite and it seems likely that the 'hot' electrons passing through the willemite excite the luminescent Mn^{2+} centres which then relax to give out their characteristic emission spectrum. This light output was green and as far as could be judged has the same spectral distribution as the cathodoluminescence of the willemite films. Unfortunately from a practical point of view, it also seems probable that the same ion movement that produces the favourable conditions

for EL may also lead to the deterioration of the device with time.

Although this EL phenomenon is roughly similar to the cathodoluminescent processes it should be noted that the voltages involved here (10-20 V) are much less than the sparking potential for air so that there is no possibility of glow discharges occurring and exciting the phosphor by electron bombardment. Incidentally the fact that green luminescence was not observed near the gas discharges and arcs that were caused by breakdown of the dielectric for negative applied voltages showed that any cathodoluminescent contributions to the light output were too small to notice.

A full correlation of the results of the C-V, I-V and electroluminescent work together with the relevant structural aspects of the willemite films will be given in the final chapter of this thesis.

References for Chapter Eight

8-1 H.W. Leverenz, Introduction to the Luminescence of Solids London (1950)

8-2 C.C. Klick and Schulman J.H., J. Opt. Soc. Amer. 42,910 (1952)

8-3 C. Vlam, Physica 15, 609 (1949)

8-4 A. Schleede and A. Gruhl, Z. f. Electrochemie 29,411-412 (1923)

8-5 C. Feldman and M. O'Hara, J. Opt. Soc Amer. 48,816 (1958)

8-6 P. Goldberg, Luminescence of Inorganic Solids, New York (1966)

8-7 R.M.Burger and R.P. Donovan (Ed.) Fundamentals of Silicon Integrated Device Technology p 109, London (1967)

8-8 J.W. Allen, Physics Exhibition Handbook p81 I.P.P.S. (1970)

CHAPTER NINE

CONCLUSIONS

9.1 Summary and discussion

The work described in this thesis was directed towards the eventual production of a monolithic display device based on silicon. Such a device, using integrated circuitry within the silicon, would be capable of overcoming the interconnection problems associated with the addressing and memory functions of modern display devices. In the proposed device the light output from an electroluminescent film would be controlled by the potential of the underlying silicon. Electroluminescence in such a structure relies on the high field injection of carriers from the silicon into the wider gap phosphor material (Chap 1).

The use of silicon as a substrate for this type of device was considered absolutely essential because of the need for complex solid state circuitry. The silicon is required to be both an injecting contact and a substrate material compatible with the phosphor technology. It is a good substrate because almost perfect single crystals of silicon are readily available commercially and the simpler aspects of the technology are well publicised. Very little silicon technology had been attempted in the Department prior to this work and one of the first requirements was to master some of the relevant techniques. Controlled cutting, etching and oxidation of commercially polished silicon wafers has been mastered and diffused planar p-n junctions, MOS structures and gate controlled diode

structures have been made (Chap 2). All the results on silicon agreed with the established practice and theory of the technology.

The first phosphor material tried was zinc sulphide. This was chosen because it is pre-eminent for the type of high field EL that is to be used. D.c. electroluminescent thin films and powders have been made by several workers elsewhere and it seemed likely that the sort of mechanism proposed (tunnel injection from conducting inclusions of copper sulphide and collision excitation of manganese centres) might be extended to injection from silicon substrates provided the electric field was high enough for tunnelling. To withstand the high field, single crystal epitaxial films of ZnS on the silicon were considered to be desirable. The good match of the silicon and ZnS lattice spacings was an additional point in favour of trying ZnS.

Three methods of deposition were tried, chemical deposition from aqueous solution, vapour deposition from a carrier gas flow, and vacuum evaporation. In all three the main difficulty seemed to be due to poor compatibility with silicon. Chemical deposition from a solution of zinc salts in thiourea produced powdery films with poor adhesion which was attributed to oxidation of the silicon in the water. In order to produce epitaxial films heated substrates are probably required although it was found that ZnS chemically attacks silicon at temperatures greater than about 800°C in both the vapour deposition and the evaporation technique. At lower temperatures polycrystalline films were formed. For the flow technique the films

consisted of randomly oriented crystallites mainly of α-ZnS, whereas the evaporated films were mainly β-ZnS oriented with the (111) axis perpendicular to the substrate surface (the (111) plane of silicon). This was in agreement with other workers who have shown that the preponderance of the two crystal types of ZnS depends mainly on the exact deposition parameters during crystal growth. Because the evaporated films grew at least partly oriented, it seemed that this technique had the most promise although the films produced were never reproducible and were often of poor quality. This was attributed to uncontrollable contamination of the substrates during evaporation which could have been overcome in any further work by the use of a much cleaner evaporation system then the one used. In addition considerable difficulties were experienced in trying to activate the polycrystalline films with Mn.

For a number of reasons discussed in Sec 3.7 the work on ZnS films was terminated. The ZnS experiments were not very extensive (e.g. the growth rate of the films was never fully investigated) and although unsuccessful in depositing epitaxial films it was stopped largely because of the discovery by the author of another phosphor, willemite which seemed to be admirably suited for deposition on silicon.

In contrast to ZnS, the compatibility of willemite with silicon was excellent. EL from willemite had been observed in the past by several works (Sec 4.1.2) but the output had always been low. Furthermore the EL of willemite thin films had not been studied, nor had films of willemite been deposited on silicon so

that there was plenty of scope for original work. Also the deposition technique which was developed was entirely novel and was subsequently the subject of a Patent Application.

The formation of willemite on silicon uses a substrate reaction technique which involves the oxidation of the silicon surface followed by the evaporation of a thin layer of the phosphor ZnF_2: Mn. Heat treatment results in the ZnF_2:Mn reacting with the silicon oxide to form the willemite (Zn_2SiO_4:Mn) layers which are brightly green cathodoluminescent. A full description of the deposition technique is given in Chap 4.

The main advantage of this technique was its compatibility with modern silicon MOS technology and, in contrast to the ZnS work, the reproducibility was very good. Over 80 films were produced under identical conditions with no significant change in either their physical nature or cathodoluminescence. Control of the silicon-phosphor interface was also considerably better than with ZnS because it is never in contact with the ambient during the formation of the phosphor. Incorporation of a suitable luminescent centre (Mn^{2+} in this case) is also very conveniently carried out by co-evaporation, because both ZnF_2 and its activator MnF_2 have very similar vapour pressures.

The technique of converting the highly insulating oxide into a phosphor eliminates pinholes which are filled with unreacted oxide in this technique. Also the willemite film as a whole is embedded in a surrounding oxide so that edge effects are reduced.

The adhesion of the film is excellent. The surface oxides on the silicon, that had seemed to be a problem with some of the ZnS work, were now being used for conversion to the phosphor itself. From a long term viewpoint, this type of deposition also offered the capability for very small displays with the resolution limited mainly by the evaporation masks.

In view of these considerable advantages there was no doubt that a continued study of these films was necessary. The first requirement was to find out more about their structure (Chap 5). The thickness was governed mainly by the depth of reaction and this was determined both by bevelling and by preferential etching. It was found that up to 1000 Å of oxide could take part in the reaction to give approximately this thickness of willemite which was ideal for a low voltage, high field device.

As the films were so thin conventional X-ray analysis using Laue back reflection was not possible. X-rays were used, however, with the powder technique to identify α-Zn_2SiO_4:Mn as the reaction product of ZnF_2:Mn and SiO_2 powders when fired under similar conditions to the films (Sec 5.4). The identity and crystal structure of the films themselves was determined by electron diffraction. Due to the difficulty of removing the films from the silicon, reflection diffraction was mainly used (Sec 5.3.3) although some films that had been made from thin oxide films detached from the silicon prior to conversion to willemite were examined with transmission diffraction (Sec 5.3.2).

The structural studies showed that the films consisted of microcrystallites with average dimensions of about 1000 Å which were embedded in an amorphous material, probably the unreacted silicon oxide. A similar type of structure is found with other glass-ceramic structures where crystalline glass phases are formed within a glassy matrix. A scanning electron microscope used to study the films (Sec 5.5) gave electron images and cathodoluminescent images that were very uniform but the resolution of the instrument was unfortunately too low to distinguish between individual crystallites of the films.

The results of the structure determinations were a little disappointing because it had been hoped that the films would be amorphous to increase their breakdown strength. (In fact the first transmission electron diffraction studies using Cl_2-detached films did show an amorphous structure but this was later proved to be due to the hot chlorine gas which oxidised the willemite to silica). However, in spite of the apparently inhomogeneous nature of the films, later experiments indicated that they have a breakdown strength at least as high as that of the surrounding oxide.

Electrical measurements were made on devices of the conventional MIS type (Sec 4.3) using several top contact metals and silicon of both conductivity types. Capacitance-voltage (C-V) and current-voltage (I-V) measurements formed the major part of this work (Chap 6 and 7) but some photoconduction experiments were also carried out. (Sec 7.4.2). The most important result was the

observation of green d.c. EL from the willemite films. Unfortunately this was discovered towards the end of the work so that only a brief study was possible.

A simple model was eventually proposed to explain all the rather confusing observations. The I-V experiments showed that currents in the willemite films were polarity dependant. Relatively large currents accompanied by a faint green d.c.EL flowed for positive bias on the top contact whereas only very small currents, comparable with those observed in the surrounding oxide, flowed for negative bias. This asymmetry is usually associated with different barrier heights at the contacts but in this case the I-V curves seemed to be independent of the contact materials that were used.

The C-V experiments, however, showed that the position of the flat band voltage was different for gold and indium top contacts. Possible explanations of this, for example differences in the injection mechanisms at the metal contacts or differences in the metal work functions, could not explain the direction or the amount of the shift.

In an attempt to investigate contacts further, some photoemission experiments were attempted after the manner of Williams (Sec 7.4.2). The observed photocurrents could not be attributed to carriers photoemitted from the silicon and intrinsic photoconduction seemed to be more reasonable. The approximate value of energy gap obtained from the photocurrent cut-off was 5.5 eV and this was in excellent agreement with previous workers who had used a single

crystal of willemite. Although this was encouraging a puzzling feature was the fact that the photocurrents seemed to be independent of bias polarity unlike the dark current measurements.

The clue to these apparently conflicting results was provided by further C-V work. Shifts in the C-V curves after bias-temperature treatment and hysterisis effects showed that there was a considerable amount of ion movement in the films even at room temperature. The analysis of these effects indicated that positive ions were moving in a fixed negative space charge and, under the influence of a positive bias they were piling up at the silicon interface. With the opposite polarity the ions drifted towards the metal contact where they discharged. The nature of the charges could not be positively identified but it seems reasonable to assume that the positive charge is that of zinc ions and the fixed negative charge the SiO_4^{4-} radicals, in view of the crystal structure of willemite (Sec 5.1) and some of the C-V results.

The reason for the ions discharging at the metal and not at the silicon electrode is not completely understood, but a possible explanation could be the existence of very thin layers of oxide at the silicon-willemite interface. The effect of ions failing to discharge at the silicon interface is that field intensification due to the build-up of space charge occurrs only in front of the silicon and only with positive bias. It is suggested that this enables tunnel emission of electrons to occur from the silicon to the willemite, increasing the high field current for this polarity. The luminescence

then results from the excitation of Mn^{2+} centres in the willemite by 'hot' electrons moving through the films. A schematic diagram of the band diagram and charge distribution for this polarity is shown in Fig 7-7. With a negative bias, no field intensification occurs and consequently only very small currents flow, dominated mainly by the ionic component. On increasing the applied voltage further breakdown occurs similarly to the silicon oxide observations of Klein et al (Appendix A). The lack of any large tunnelling currents for either polarity when the films were illuminated with ultra-violet light was attributed to the neutralisation of the ionic space by carriers generated by the radiation.

The shifts in the C-V curves for different contact metals were also explained in terms of ion drift. With ionic movement the position of the resultant space charge, and therefore the shape of the C-V curve, is determined by the built-in electric field caused by contact potential difference across the MIS structure. Using published values for the work functions and electron affinity it was possible to show that the direction of the contact potential field was opposite for gold and indium contacts with n-type silicon and this agreed with the C-V shifts that were obtained.

It was also possible to explain most of the other results in terms of the motion of zinc ions in a fixed negative space charge as discussed in Sec 6.5. The shape of the C-V curves also provided information about the silicon -willemite interface showing that the conversion of oxide into willemite introduced no further

'active' states at the interface.

The C-V technique proved to be a very powerful tool for the study of conduction mechanisms in thin films particularly as space charges usually figure prominently in these. The overwhelming advantage that this technique confers on silicon as a substrate was not realised early in the work. The interpretation of C-V measurements on insulating films on silicon can be carried out to a high degree of sophistication because of the extensive investigation of silicon oxide. This means that as long as silicon can be tolerated as one of the electrodes it is the ideal substrate for the investigation of conduction processes in thin films of all types. Its use with other electroluminesence materials could solve some of the problems in the general understanding of this phenomenon.

The luminescent properties of the films are discussed in Chap 8. The cathodoluminescent spectrum was broad and featureless peaking at 5195 Å, in agreement with previous workers' results on powdered willemite. Unfortunately the EL was dim and visible only in a darkened room and it occurred typically with + 15 V bias and 100 μamp current with a light emitting contact area of about 10^{-3} cm^2. As a suitable photomultiplier was not available the spectral distribution of the EL was judged by eye using an interference filter and found to be very similar to that of the cathodoluminescence. The light output gradually decreased with time although a typical device usually lasted several hours before its EL could not be seen by eye.

Due to the lack of time there were only a few EL experiments and no attempt was made to increase the light output or investigate the decay process. However, as far as can be ascertained, this is the first time that d.c. EL has been observed in willemite. It is also the first time that EL of any type has been seen in willemite thin films.

It has been shown in Chap 1 that in order to obtain efficient injection from silicon into a wider gap phosphor, high field mechanisms are necessary. In most phosphors that exhibit this form of EL some selective field intensification occurs to enable tunnelling to take place. For example in ZnS powder phosphors it is thought that copper sulphide inclusions produce the requisite injection conditions. In some recent work (Ref 8-8) on ZnSe it has been shown that Schottky barriers can produce the same effect. Thin insulating layers have also been used to give tunnel injection and in willemite itself some early work used sharp-edged metal particles to produce the high fields. With the willemite films considered here however it has been shown that field intensification is produced by the selective build-up of ions in front of the silicon. This gives rise to tunnelling injection which is followed by excitation of Mn centres in the willemite to produce EL probably in a similar way to the other manganese activated phosphors. Although it is the ionic movement that enables the thin film willemite device to give EL output it may at the same time present a fundamental limitation to the life of a practical display device because of some type of solid state electrolysis. However, even the present ZnS

d.c. EL phosphors, which do not rely on ionic movement for injection, at the moment only have lifetimes of the order of 1000 hours so that willemite might eventually be competitive. Alternatively it may be possible to control the ion movement or to utilise some other form of field intensification in willemite for the realisation of a practical display.

Nevertheless, it is considered that willemite films on silicon merit further study particularly in view of their compatibility with silicon technology and the ease and reproducibility of the deposition process. In addition it is thought that the observation of d.c. EL from such films is a significant step in the production of a completely monolithic display device based on silicon.

9.2 Suggestions for further work.

Any future work on willemite films on silicon should be directed towards producing much brighter EL. A more detailed and quantitative study of the postulated model is required and its correlation with variations in physical structure of the films (e.g. thickness, MWOS structures and manganese content) would also be desirable. Light output measurements, combined with further I-V experiments, are probably the first requirement of any detailed study. Pulse measurements would be more informative than d.c. particularly over a range of temperatures to obtain more details of the conduction mechanisms, also the ion motion could be studied in more detail by extending the C-V work to include full a.c. impedance measurements and by incorporating a charge-integrating device in the bias circuit.

The easiest confirmation of the proposed model would probably be obtained by using some different contact combinations in the I-V and EL experiments. Thin films structures with both contacts of the same material should not show any polarity dependence. In particular, a silicon-willemite-silicon structure should show high currents and EL for both bias polarities, and this would probably be the easiest structure to make with the existing techniques, whereas MWM structures should show neither, provided the nature of both contacts can be guaranteed to be similar. The use of p-type silicon for MWS structures (Sec 6.5.2) in the C-V work would further establish the proposed theory of ion drift in the built-in field of the contact potential difference. An important aspect of any future work as far as a practical display system is concerned should also be the production of transparent top contacts possibly based on tin oxide.

The nature of the work described in this thesis has been mainly exploratory and as a consequence the various aspects could not be studied to as great a depth as would have been liked in the time available. Nevertheless, many aspects of the work are novel and it is hoped that some of it will form the basis for a more detailed examination of the processes involved and a more detailed viability study of a practical display system .

Appendix A

Dielectric Breakdown of Thin Films

Dielectric breakdown processes are classically divided into two basic types. The first of these is termed intrinsic breakdown and is electronic in nature depending on the presence of electrons in the conduction band of the dielectric. These are accelerated by the field and lose energy primarily by ionising the atoms of the solid to produce further electrons.

The second type of dielectric breakdown is called 'thermal' breakdown. In this, Joule heating occurs at some localised weak spot in the dielectric. When the loss of heat to the surroundings is less than the heat produced thermal runaway occurs and this results finally in breakdown. The main differences between the two types is that intrinsic breakdown occurs quickly (in about 10^{-8} sec) and it is only slightly affected by the ambient temperature whereas thermal breakdown tends to be much slower and more temperature dependent.

There are also other mechanisms for dielectric breakdown but these usually involve either electrochemical or chemical deterioration of the dielectric with time as a result of which one of the basic mechanisms leads to breakdown. Discharges within voids in the dielectric can also produce dielectric breakdown but this mechanism is not strictly a property of the dielectric.

Most of the early work on breakdown of solids was concerned with comparitively thick specimens so that large voltages

were required to produce the breakdown field strength (10^5V/cm). However, with the onset of thin film technology not only were the voltages required much less but it was found that because the films had fewer defects than their bulk counterparts the effective breakdown strength was substantially increased (e.g. 10^7 V/cm for thermal silicon oxide Ref 8-7). Another aspect of thin film breakdown that has helped considerably in their study has been the ability of MIM structures to form self-healing breakdowns so allowing a single film to be tested many hundreds of times before becoming shortcircuited and before complete destruction finally occurs.

Some examples of dielectric breakdown work on thin films will now be given. Budenstein and Hayes (Ref A-1) have studied evaporated thin film Al-SiO-Al structures. They found that the breakdown was initiated at unidentified dark spots in the silicon oxide about 0.1 micron in diameter. Moreover, the process was a destructive one in which small areas of the MIM structure were physically and chemically altered. By identifying the polycrystalline silicon globules at the post-breakdown sites they concluded that breakdown is due to an electrochemical reation in which the Si-O-Si bond is ruptured by the local high field and crystalline silicon and gaseous oxygen result.

A different breakdown mechanism has been proposed by Klein et al who also investigated evaporated silicon oxide MIM structures with both d.c. voltages (Ref A-2) and voltage pulses

(Ref A-3) using specimens with self healing properties. For d.c. and pulses of long duration, thermal runaway occurred as a result of an excessive increase in Joule heating and breakdown was accompanied by the simultaneous evaporation of nearly the whole capacitor. However, at voltages just less than those required for this process a second, very different, form of breakdown occurred that was termed electric breakdown. This had the nature of a random event the probability of which increased rapidly with voltage so that thermal breakdown had very little time to develop at room temperature. Klein et al proposed that the electric breakdown was due to a succession of processes. Firstly, the breakdown occurred at the weakest point in the dielectric where it was triggered by a small random current pulse caused by electron avalanches and initiated by a single electron at the cathode. The pulse produced local heating and thermal instability which resulted in a hole being evaporated through the dielectric and the metal electrodes. After this the MIM capacitor discharges via an arc between the metal electrodes and this causes the bulk of the destruction.

Another type of breakdown called 'propagating' breakdown was also observed by Klein et al when the resistance of the external circuit was small (less than about 10 k ohm). A propagating breakdown was started by a single electric breakdown which left behind a ring of hot dielectric around the discharge point. Due to the increased conductivity in this region further resistance was small enough to allow a sufficient supply of energy before the dielectric

had time to cool. Each successive breakdown then generated the conditions necessary for the next and a propagating breakdown resulted.

Thermal, electric and propagating breakdowns were also observed in thermal oxides on silicon (Klein Ref A-4). However, although the general results were substantially similar to those described for evaporated silicon oxide the silicon substrate had a distinct effect on the extent and nature of the breakdown. Klein summarised his results in the following way :- The resistivity of the silicon substrate and its polarity greatly influence the breakdown events which takes place within nano- to microseconds. Breakdown is most rapid in degenerate samples and they increase in duration in this order : medium resistivity samples with an accumulation layer (i.e. positive bias for n-type silicon), high resistivity samples, and medium resistivity samples with an inversion layer. The amount of damage due to the breakdown decreases in the same order and it increases with voltage.

These results were explained using a model that invoked a detailed analysis of the resistance of the discharge path and also the thermal properties of the silicon as compared with the metal electrode used earlier. A good correlation was obtained between calculations of the discharge energy using this model and measurements based on both oscillographs of the breakdown events and also a physical examination of the destruction around the discharge points.

A brief description of the dielectric breakdown of thin films has been included in this Appendix because so called

electroluminescence occurs at the discharge. Klein measured the spectrum of the emitted light and he found that it contained the atomic emission lines of the metal top contact (Al in his case). An estimate of the temperature of the metal ions was also made using the relative intensities of two of the spectral lines and found to be between 4000°K and 4500°K. The amount of light given out depended on the amount of energy expended by the discharge and this was controlled by the polarity and resistance of the silicon as explained above. However, there are several distinct differences between this sort of light output and the true Type II EL obtained from willemite (Sec 8.3.1)

References for Appendix A

A-1 P.P. Budenstein and P.J. Hayes, J. Appl. Phys. 38, 2837 (1967)

A-2 N. KLein and H. Gafni, I.E.E.E. Trans. Electron Devices, ED-13 281 (1966)

A-3 N. Klein and E. Burstein, J. Appl. Phys. 40, 2728-40 (1969)

A-4 N. Klein, I.E.E.E. Trans. Electron Devices ED-13, 788-805 (1966)