The deposition of coatings on to polymer substrates by planar magnetron sputtering

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/11227](https://dspace.lboro.ac.uk/2134/11227)

Publisher: © Charles Arthur Bishop

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
This book was bound by

**Badminton Press**

18 Half Croft, Syston, Leicester, LE7 8LD

Telephone: Leicester (0533) 602918.
THE DEPOSITION OF COATINGS ONTO POLYMER SUBSTRATES
BY PLANAR MAGNETRON SPUTTERING.

BY

CHARLES ARTHUR BISHOP

A Doctoral Thesis submitted in partial fulfilment of
the requirements for the award of Doctor of Philosophy
of the Loughborough University of Technology.

MAY 1986

Supervisor : Dr. R. P. Howson B.Sc. Ph.D.
Department of Physics.

© by C. A. Bishop.
Certificate of originality.

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a higher degree.

Signed. ..................... Date. .....................
To my family and friends, 
with my thanks for their patience 
and belief.
Abstract.

A vacuum deposition system was built enabling flexible polymer sheet, wound on a roll-to-roll transport mechanism, to be coated by any or all of three planar magnetron sputtering sources. Using this machine a large variety of coatings were produced in long lengths onto heat sensitive substrates and with controlled stoichiometry. Within the system the coating was monitored soon after deposition which allowed fast response to changing film properties, a critical factor in the continuous production of high quality coatings.

An area of current interest has been the production of large area optical filters, the basis of which is the deposition of thin metal and thin metal oxide films. Of particular interest have been the 'heat mirror' type filters, the transparent conducting oxides and more recently the electrochromic devices. All these have been deposited at high rates onto flexible polyester substrates ostensibly at room temperature.

Analysis of the films has been carried out by a variety of techniques in order to establish the precise chemical composition and structure of the films. Once optimum conditions had been established for individual films multilayer filters were produced and their optical performance determined. In the case of the 'heat mirror' type filters their performance was compared to theoretical predictions.
Index.

Abstract  

Index  

Acknowledgements  

1). The aims of the research.  

2). Introduction.  

2.1). References.  

3). The d.c. glow discharge.  

3.1). The Townsend discharge.  

3.2). The breakdown region.  

3.3). The transition region.  

3.4). The normal glow discharge.  

3.5). The abnormal glow discharge.  

3.6). The arc.  

3.7). References.  

1  

2  

14  

20  

22  

27  

30  

31  

31  

32  

33
4). D.C. diode sputtering. 34

4.1). The emission of neutral particles. 34
4.2). Secondary electron emission. 37
4.3). Positive & negative secondary ion emission. 37
4.4). Reflected incident ions. 38
4.5). The desorption of gases. 38
4.6). The emission of photons and x-rays. 38
4.7). Surface modification. 39
4.8). Chemical effects. 40
4.9). Reactive d.c. sputtering. 41
4.10). Planar magnetron sputtering. 42
4.11). References. 50

5). System considerations. 54

5.1). Introduction. 54
5.2). The pumping system. 55
5.3). The planar magnetron sources. 61
5.4). The magnetron d.c. power supplies. 69
5.5). The gas supply and gauging system. 76
5.6). The roll-to-roll coater and film monitoring equipment. 80
5.7). References. 86

6). Experimental procedure. 87
7). Techniques for thin film analysis.

7.1). Optical microscopy. 91
7.2). Scanning electron microscopy (SEM). 93
7.3). Transmission electron microscopy (TEM). 94
7.4). Energy dispersive analysis of x-rays (EDAX). 95
7.5). SEM plus EDAX. 99
7.6). TEM plus EDAX. 99
7.7). Electron microprobe microanalysis. 100
7.8). X-ray diffraction. 100
7.9). Auger electron spectroscopy. 102
7.10). X-ray photoelectron spectroscopy. 103
7.11). Laser induced mass analysis. 105
7.12). Ellipsometry. 107
7.13). The scanning spectrophotometer. 112
7.14). The emissometer. 114
7.15). The thickness by the stylus technique. 115
7.16). The Hall effect measurements. 115
7.17). References. 118

8). The optical performance of dielectrics and thin metals. 121

8.1). The Lorentz model. 123
8.2). The Drude model. 132
8.3). The reflectance and transmittance of thin films. 136
8.4). The loss of performance due to scattering. 140
8.5). References. 143
9). The deposition of coatings, results and discussion. 146

9.1). Single layer transparent coatings. 146
9.2). The transparent conducting oxide coatings. 147
9.3). The transparent non-conducting oxide coatings. 166
9.4). The thin metal and oxide-metal-oxide filters. 179
9.5). The electrochromic coatings and devices. 201
9.6). References. 222

10). Conclusion. 231

11). Appendix. Publications list. 233
Acknowledgements.

My thanks go to the many people who have helped facilitate the completion of the project.

Firstly to my supervisor Dr. R.P.Howson to whom I shall always be grateful for the opportunity to try my hand at research and for his continued enthusiasm throughout the project.

To the rest of the research group present and past, Messers A.G.Spencer, R.W.Lewin, K.Suzuki and M.I.Ridge.

Inevitably the project would have ground to a halt without the backup available from the mechanical and electrical workshop staff. In particular Mr. E.M.Stenlake who has fought valiently to keep up with all the repairs and modifications desired.

My thanks also go to the members of the Physics department. In particular to Professor J.F.Raffle and to Dr. J.W.Sturgess who, as Head of Department, welcomed me into the department.

The project was funded by grants from the E.E.C., S.E.R.C. and I.C.I. to whom I am endebted.

Finally I thank all the people who have helped with suggestions and discussions both at home and wherever the work has taken me.
1. The aims of the research.

To enable coated polyester substrates to be used in large areas required a process capable of depositing coatings at ostensibly room temperature to avoid degradation of the substrate. The advent of magnetron sputtering produced the opportunity to do this. The increasing cost of energy led to the research into alternative energy sources and the development of materials for energy saving. One such development was the 'heat mirror' type coatings that allowed light through a coated window into a room and would reflect the resultant infra-red radiation thus keeping the heat in the room.

The research aim was to further understand the critical parameters relating to the room temperature deposition of coatings many of which are normally associated with an elevated substrate temperature during deposition or as a post deposition annealing process. To do this it was necessary to develop the planar magnetron sputtering technique to be able to reproducibly deposit both oxides, by a reactive process, and metals onto polymer substrates at room temperature. A further aim was that in developing a reliable technique other coatings or combination of coatings could then be deposited onto polymers suitable for production in large areas at low cost.
2. Introduction.

The first reference to what is now known as sputtering was in 1852 by Grove who observed that during experiments 'on the electrochemical polarity of gases' a dark deposit was formed around the platinum electrode (1). This was more clearly commented on by Gassiot in 1858 who was experimenting using a gas discharge and noted that the negative electrode was corroded as well as the glass surrounding the electrode being coated (2). It was also Gassiot who first passed comment that placing a magnet close to the plasma caused some distortion of the plasma. The potential to use cathodic disintegration as a technique for the deposition of materials was recognised by Plucker in 1858 (3). Although the possible use of the technique was recognised early it was not until 1877 that Wright demonstrated that a whole variety of materials could be deposited this way (4). Crookes in 1891 conducted a whole series of experiments in which he compared the 'volatilities' of different materials as deposited by 'electric evaporation'. He was regarding the heating of the cathode in the plasma as being the cause of the ejection of material rather than ion bombardment to be the cause of the ejection and also the cathode being heated. Crookes also used the emission lines from the plasma to characterise the glow discharge (5). It was not until 1902 that Goldstein demonstrated that the material was removed by positive ion bombardment (6). In several of these early experiments the area of interest was the plasma and the experiments were aimed at furthering the understanding of why the plasma could be divided up into striations. This type of experimentation continued for many years with parts of the plasma being named after prominent experimenters (7,8,9). The start of the modern theories of sputtering can be traced back to Stark in 1908 who first postulated the 'billiard ball'
momentum transfer idea as the explanation of how sputtering occurred (10). The earliest use of the word 'sputtering' appears to be in 1911 by Thomson (11) who was a scientific expert for the Encyclopaedia Britannica. A use for the disturbance that a magnet could produce on the plasma was not found until 1912 when Arnold used a magnet to control electrons in a mercury vapour discharge tube as part of an amplifier design (12). This was not as successful as other devices that were being developed and was dropped. The use of the magnet to alter the path taken by electrons was later developed by Hull, beginning in 1916 he designed an alternative to the triode which he called a 'magnetron' (13). This design incorporated the crossed electric and magnetic fields which gave the electrons their spiralling motion which is still seen in descriptions of modern magnetron sputtering source designs. During this period there were advances being made in the performance of vacuum pumps. Gaede initially patented a design in 1915 of a diffusion pump although it was Langmuir in 1916 who produced the design of condensation pump that bears a remarkable similarity to products still on the market today (14,15). It was also Gaede who developed the rotary pump in 1907 which also looks similar to its modern counterparts (16). These developments took the base pressure limit from 0.13 Pa (10E-3T) to 1.3x10^-6 Pa (10^-8 T). A further step in the theory of sputtering was taken in 1923 by Kingdom and Langmuir who described the removal of atoms by a two step process that involved an incident ion producing a depression and a second incident ion bouncing off the bottom of the depression and then colliding with the atom at the rim on its way away from the surface taking the atom with it (17). The research here being stimulated by the desire to improve the lifetime of tungsten filaments. Other theories were also put forward as to how the
cathodic disintegration occurred, the 'hot spot' or 'local boiling' theory was the most popular. The theory was that the energy from the ion bombardment led to very localised heating resulting in the boiling and evaporation of a tiny spot of material. The main proponent of this theory was von Hippel (18). It was proved by Baum in 1929 that the sputtered particles left the cathode uncharged (19). This was confirmed by Ditchburn in 1933 who sputtered cadmium with a view to comparing results with earlier results of evaporated cadmium. In his paper (20) he also details how the bombardment of growing films can affect the deposition rate due to re-sputtering. This paper also detailed the basis of sputter or plasma cleaning. In the same year Overbeck was carrying out reactive sputtering to produce tin oxide (21). He saw similar effects to those seen earlier by Wright (4) of colours in transparent films and was able to name it as being caused by interference effects following work done by Edwards (22). Although these films were transparent there is no evidence that they were found to be conducting. Ditchburn also used 'filament maintained' discharges to enable him to work at lower pressures than normally used in self-maintained discharges, this is probably the earliest use of triode sputtering. Penning approached the same problem of low pressure discharges differently, by using a magnetic field the electrons spiralled around their drift path considerably extending the distance they travelled and thus were able to undergo many more ionising collisions. This design was reported in 1936 (23) and later patented (24). In work published in 1940 Penning and Moubis described characteristics of sputtering that are still relevant today. Whilst trying to deposit aluminium films they recorded the different voltages associated with an oxidised target compared to that of a metal target. They
also referred to the 'sudden change in the colour of the discharge' as the oxide surface was sputtered away and the aluminium metal started to be sputtered (25). In modern processes this is still seen and remains very difficult to control. The problem of oxidation had been noted much earlier (26) with regard to the lower sputtering rate of the oxide and it led to the attempt to use higher purity gases although as was later pointed out if the sputtering current is too low it could take days to remove all the native oxide from the target surface (27). The deposition of transparent conducting oxides in a commercial process was patented in 1942 by a major glass manufacturer (28). The main part of the patent referred to the spray pyrolysis of the coating but they did refer to sputtering or evaporation as alternative ways of depositing the coatings. The need for the coatings was for aircraft windscreens where a current was passed through the coating which acted as a heating element and prevented the windscreen from icing up. It was also suggested that it could be used equally as well in car windows, a use that is only recently being seriously considered. Townes persevered with the 'hot spot' theory of sputtering and tried to quantify the pressure dependance, the number of ions bombarding the target surface and the number of atoms released per incident ion (29). The design of sputtering sources was furthered by Holland who utilised the fact that if the anode encroached the cathode dark space the discharge would be extinguished and hence the sputtering cease (30). This is now referred to as earth shielding and is an integral part of most modern designs. It was Seitz in 1949 who started looking at the target to establish what effects the bombardment was having on the structure of the surface. He concluded that there was a surface altered layer which was caused by implantation and
disordering due to ion bombardment and heat (31). One of the earliest uses of a transparent conducting coating was for the front surface selenium photodiodes. This was first reported by Preston who produced a conducting cadmium oxide by accident (32). Cadmium was sputtered in argon but the deposit was transparent and conducting, this was thought to be due to residual gas in the vacuum system or to an air leak or to the sputtering of a surface oxide layer from the target. Preston tried to repeat the process for both zinc and tin but although both resulted in a transparent oxide neither were conducting. Veszi who also reported the success of the conducting cadmium oxide coined the title 'reactive sputtering' for the process (33). In order to produce conducting tin oxide a post annealing process was used to produce the desired non-stoichiometry (34). This technique was first used much earlier in order to oxidise chemically deposited films (35). There were other developments of interest at this time. A further report on the work carried out at the National Physical Laboratory by Preston refers to using a thin oxide layer either side of a thin transparent gold coating resulting in improved transmission over a plain gold coating. This is what is now known as anti-reflection but then was described as being analogous to the 'blooming' of camera lenses (36). The other report of interest was of the coating of plastic for decorative purposes, this consisted of a base layer of lacquer, an aluminium evaporated coating and then a further layer of lacquer (37). The use of a seeding layer was also tried as a method of improving the adhesion (38,39). Much of the deposition of conducting coatings was not well detailed, the first to systematically try different oxides with controlled amounts of oxygen was Holland and Siddall in 1953 (40). Guenttherschultze considered not only his own
work but also much that was published in the early part of the century and gave specifications of how experiments should be conducted to avoid producing dubious results. He also found that two conclusions held true even for the early experiments, these were that the sputtering rate is independent of the temperature and that the sputtering rate is dependent upon the gas species (27). During the latter half of the 1950s several workers detailed experiments that conclusively proved the 'hot spot' theory to be wrong. From the initial work by Fetz in 1942 (41) on the dependence of sputtering yield with the angle of bombardment Wehner carried out a series of experiments using single crystals and demonstrated from the pattern of the deposited film that the sputtering was related to the crystal orientation and structure (42). The major fact that disproved the 'hot spot' theory was that to produce a film that was related to the target crystal structure would be impossible for a liquid. Thus Wehner established the collision theory as the only possible mechanism for sputtering. Keywell concluded that the sputtering ratios (the number of sputtered atoms to the number of bombarding ions) are small and are of the same order of magnitude for all gas-metal combinations (43). He also concluded that electrons did not cause sputtering. Their energy is transferred to conduction and bound electrons as evidenced by the production of secondary electrons and heat. He also confirmed that there is no (or only a very slight) temperature dependence in sputtering. The energy to remove an atom is the energy required to overcome the binding energies between the atom and its neighbours. These were calculated to lie between 18-43 eV and the thresholds for sputtering were found to be between 18-180 eV which would be compatible with the calculations (44,45). The difference between the calculated
sputtering ratios and the measured sputtering ratios was taken to be due to contamination of the target surface by residuals in the vacuum system (46), this being due to the low sputtering rate and thus a high contamination rate from the residuals, this factor still applies today. The pressures at which sputtering can occur means that to deposit a high purity film a very high deposition rate is required. Work continued to decide how many collisions would occur in the target, how deep the ions would penetrate, which ions had the best sputtering yield and what species were present at each part of the process (47-50). Two good reviews made at this time were by Colligon (51) and Kay (52). The latter endeavoured to use the term 'impact evaporation' rather than sputtering, this followed a suggestion by Guentherschultz (27), but it was to no avail. The sputtering of alloys that had first been done by Crookes (5) was brought up to date by Gillam (53) who used electron diffraction to establish the depth and composition of the 'altered' layer created by the different sputtering coefficients. He proved the thickness of the layer to be a constant for a constant bombarding energy once the full depth of the layer had been established. Holland as well as working on the deposition of coatings detailed the use of the bombardment of surfaces as a cleaning technique (54). In this paper he also reported on the cracking of the hydrocarbons from the pumping system and the resulting deposition of organic films. This is now an area of considerable effort and interest as a technique for producing a new set of materials different in properties from the conventional metals or polymers. There was still occasional interest in the use of magnetic enhancement or confinement. Vineyard carried out some theoretical work on the trajectories that charged particles would describe in crossed magnetic and electric
fields (55). A much more practical application was demonstrated by Kasaev and Pashkova who used the magnetic field to constrain the motion of the cathode spot in a mercury arc lamp (56). They constrained the motion to both a circular path and a square path (with rounded corners). Had they used less current the discharge would have been abnormal rather than an arc and they would have had the first planar magnetron sputtering source, as it was that did not occur for a further fifteen years. Kay did work on using a magnet, generally a quadropole, to reduce the working pressure but this configuration produced very uneven coatings and was primarily of academic interest (57). His other magnetic configuration he referred to as an inverted magnetron and required an electromagnet to provide an axial field between the coaxial anode and cathode (58). This work led to a number of patents which were used as references in later patent applications. Knauer and Stack worked to improve the Penning discharge configuration and in doing so produced a post magnetron made up of a series of smaller magnetrons (59). This configuration with the addition of water cooling and a reciprocating motion to the magnets to improve the target utilisation is still in production today. The rate of magnetron development increased with most of the work being carried out on cylindrical configurations (57-60). The major change came in 1974 with the design patented by Chapin (61) which used a flat sheet of material as the sputtering target and was easily able to be scaled up as required (62). It is this type of magnetron design that is most used today. Penfold and Thornton continued to develop the cylindrical, inverted or post magnetron designs (63) but their use is limited to batch processes whereas the planar magnetron was developed to handle air-to-air semi-continuous production coaters improving the cost effectiveness of
the vacuum deposition process. There appears to be an increase of interest in sputtering from the 1960s onwards. Wolsky attributes some of this to the 'space programme' which required information on damage by particle bombardment in vacuum, ion propulsion sources, improved coatings for wear resistance and better electronics equipment (64). Work was still done to try to understand the fundamental mechanisms (65) and with increasing applications for sputtered coatings there was a steady flow of review papers (66-74) which highlighted gaps in the theory. There were improvements made in the deposition of coatings, reactively sputtered conducting tin oxide was produced by doping the sputtered target, this then avoided the need to post anneal the coating (75). A patent was taken out for a heat reflecting glazing system which had glass with a gold alloy coating on it, this was designed to allow transmission in the visible but reflect heat (76). This patent coupled with the earlier observation that sandwiching the metal with oxides improved the transmission (36) was taken up by Fan and a high performance filter was produced and the process later patented (77,78). With the increase in the differing applications of the coatings came the need to look at the structure and adhesion of them. Movchan and Demchishin detailed how the structure of the growing film would change with the substrate temperature (79), this was later modified by Thornton to include the deposition pressure as well (80). Weaver and later Mittal worked on the problems of adhesion of coatings (81-83) or rather the problems in testing adhesion of coatings. Closely related to the problems of adhesion is surface cleaning of substrates which also has a problem when trying to define the term clean and when trying to find a suitable technique to determine the surface contamination. Mittal recommended the use of surface analysis to study the
chemical composition of the interface (83), similarly Mattox (84) recommended various surface analytical techniques to assess the chemical composition of the surface. The study of particle bombardment and the measurement of energetic products has also led to the application of carefully controlled techniques to produce chemical information about the surface being bombarded. This opened up a new industry which was also able to study and develop the technology which has also expanded the literature and knowledge of the processes involved (74). The development of the magnetron also enabled many more products to be manufactured in plastics and then coated by vacuum deposition to look metallic. Thornton highlighted the advantages of magnetron sputtering of reduced substrated heating, excellent adhesion, good throwing power and also suitable to be scaled up to be compatible with large production vacuum coaters (85). With the greater understanding of the mechanisms of sputtering has come an increasing number of alternative methods of sputtering. In 1962 Frerichs described the use of asymmetric a.c. sputtering (86), this was the forerunner of what is now called r.f. sputtering. The sputtering process is a higher energy process than evaporation and so efforts were also made to increase the deposition energies in evaporation systems. Mattox coined the phrase 'ion plating' which used a d.c. bias for activation (87). This inevitably led people to use biasing techniques for sputtering deposition as well with combinations of both d.c. and r.f. for sputtering and biasing i.e. d.c. sputtering plus d.c. bias (88), d.c. sputtering plus r.f. bias (89) and r.f. sputtering and r.f. bias (90). Two other techniques also gained favour, the first was the ion beam deposition technique which used an ion beam to sputter the target and could also use a second ion beam to add energy to the
depositing film (91). This was a well defined process, in terms of energies, but slow by comparison to the r.f. sputtering process which in turn is at least half the rate that can be achieved by d.c. sputtering. The other technique that appeared was chemical vapour deposition (C.V.D.) which used the energy within the plasma to breakdown volatile chemicals such that part is pumped away as a gas and the rest is deposited out as the desired coating (92). This was for inorganic coatings but effectively was the same process as Holland had observed earlier (54). The introduction of chemistry into the field has probably contributed to the use of plasma etching, this is really a chemical sputtering process with the substrate forming the target, generally with a mask over the material to produce a selective etch due to different sputtering rates. The chemistry involved can be complex and the products detrimental to the vacuum system, its use is primarily in the semiconductor industry (93). The specifications of the semiconductor industry coupled with the huge amounts of money involved have further developed the deposition processes to the point of being able to deposit single crystals and with specific doping and or multilayers (94). However there are orders of magnitude difference between the cost of the process and the cost of the product when compared to the magnetron sputtering process which in general is required to be a cheap process for a large volume production because of the strict limitation on the product selling price. The recent developments on magnetron sputtering have mainly been in the area of process control and with the increased control the ability to produce coatings not previously possible. One area currently of interest is the use of the emission lines of the plasma in order to directly relate their intensity to the rate and composition of the
depositing film. This was first used by Crookes (5) but has only recently been reintroduced initially by Stirling and Westwood to look at the oxide contamination of surfaces (95) and as a form of process control by Greene (96), Schiller (97) and Enjouji (98). Schiller (99) has detailed several methods of controlling processes for the production of coatings for optical filters, decorative coatings, corrosion resistant coatings, microelectronics applications, wear resistant coatings (particularly for cutting tools) as well as still using sputtering to coat mirrors as was done by Wright in the first practical application for the process (4). Large area applications are at present primarily for coatings onto glass for solar control purposes with a smaller amount being produced for heat gain energy conservation purposes (100,101). These coating plants can handle glass of up to 3m x 4m sheets and produce of the order of 100,000 square meters per year per coating system. The coating of large area plastic is predominantly the metallising of rolls of plastic although it is without doubt only a matter of time before the control of the process allows technically more sophisticated products to be deposited onto large areas of flexible polymer sheet.
2.1 References.

1). W.R.Grove. Trans. Roy. Soc. 142 (1852) 87
   Proc. Roy. Soc. 9 (1858) 146
   Am. J. Sci. 14 (1877) 169
   London. (1839) paragraphs 1544-1560
   (reprinted New York 1965)
   Proc. Roy. Soc. A, 84 (1911) 562
   Z. Electrochem. 15 (1909) 509
13). A.W.Hull. Phys. Rev. 18 (1921) 31
14). W.Gaede. Deutches Reichspatent 286404 (1913)
16). W.Gaede. reported by O.E.Meyer.
   Verh. Deutsch. Phys. Ges. 10 (1907) 753
   Ann. Phys. 46 (1915) 357
19). T.Baum. Z. Physik. 40 (1929) 686
23). F.M.Penning. Physica (Utrecht) 3 (1936) 873
   Wetensch. 43 (1940) 41
26). E.Warburg. Weid. Ann. 31 (1877) 545
   Weid. Ann. 40 (1890) 1
27). A.Guentherschultze. Vacuum 3 (1953) 360
   U.K. Patent 632,256 (1949)
   (U.S. Patent (1942))
36). Anon. The Chemical Age Lndn. 67 (5 July 1952) 13
37). S. Bateson. Vacuum. 2 (1952) 365
       Vacuum. 3 (1952) 35
       B, 65 (1952) 649
       U.S. Patent 2,482,054 (1949)
       U.S. Patent 2,628,927 (1953)
       J. Appl. Phys. 30 (1959) 1762
       J. Appl. Phys. 31 (1960) 1583
       Fisica Corso 'Enrico Fermi' xiii
       (1960) 158
   U.S. Patent. 3,884,793 (1975)
70). C.Weissmantel. Proc. 7th Internet. Vac. Cong. & 3rd
72). W.D.Westwood. Prog. in Surf. Sci. 7 (1978) 71
73). B.N.Chapman. 'Glow discharge processes'.
74). R.Behrisch. Ed. Sputtering by particle bombardment
(Pub. Springer (Berlin, Heidelberg, N.Y.)
U.S. Patent. 3,400,006 (1968)
Fiz. Metal. Metalloved. 28 (1969) 83
81). C.Weaver. Faraday Special Discussions of the
Chemical Society 2 (1982) 18

J. Appl. Phys. 36 (1965) 237


R.C.A. Rev. 29 (1968) 566


J. Appl. Phys. 42 (1972) 2953


J.Vac.Sci. & Tech. 16 (1979) 391


2 (1983) 47


Thin Solid Films. 7 (1971) 1


3. The D.C. Glow Discharge.

If a tube containing a cathode and anode, which are positioned some distance apart, is evacuated and an inert gas is introduced it is possible, when a potential is applied to the electrodes, to stimulate the system and for it then to become luminous. This emission of light is referred to as a glow discharge or a plasma. The basic process that occurs is that when enough potential is applied electrons collide with atoms to produce ions and more electrons initially in a cascade process and subsequently in a steady state process. The electrons that are generated are called secondary electrons and may be further accelerated by the applied potential such that they too are able to undergo ionising collisions. The ions being substantially larger than the electrons are unable to be accelerated at the same rate as the electrons, this differing mobility being seen in the plasma as areas of differing light intensity.

The process of ionisation is where an electron collides with an atom (or molecule) and removes an electron from the atom. For this to occur the kinetic energy of the impinging electron must be at least equal to and probably greater than the work done to remove electrons from the outer shell of the atom against the force of attraction of the nucleus, these are known as inelastic collisions. The remaining kinetic energy is divided between the primary or incident electron and the secondary or liberated electron. Due to the large discrepancy of masses between the electron and the ion virtually no energy is transferred to the ion. If the kinetic energy of the electron is too low for an ionisation collision to occur the electron is usually reflected with negligible loss of energy, this is known as an elastic collision.
The energy transferred during a collision can be expressed as the energy transferred from the incident particle \( m_i \) to a stationary particle \( m_s \) and where \( \theta \) is the angle of collision with respect to the line between the particle centres at the point of impact. This is expressed as:

\[
\frac{E_s}{E_i} = \frac{1/2 m_s u_s^2}{1/2 m_i v_i^2} = \frac{4m_i m_s \cos^2 \theta}{(m_i + m_s)^2}
\]

This is a maximum when \( \cos^2 \theta = 1 \) and when \( m_i = m_s \)

\[
4m_i m_s
\]

The energy transfer function \( \frac{4m_i m_s}{(m_i + m_s)^2} \) when \( m_i = m_s \)

but when the masses are very different with the incident mass being small the function approximates to:

\[
\frac{4m_i}{m_s} \quad \text{for} \quad m_i \ll m_s
\]

This describes an electron striking a gas atom, little energy can be transferred. If the collision is the other way round with the heavy particle hitting the light particle then this approximates to:

\[
\frac{1/2 m_s u_s^2}{1/2 m_i v_i^2} = \frac{4m_s}{m_i} \quad \text{for} \quad m_i \gg m_s
\]

\[
u_s = 2v_i
\]

This means that a light particle would be bounced off the
heavier one and would depart with twice the impact velocity.
The form that the ionisation collision would take would be;

\[ \text{e} + \text{Ar} \rightarrow 2\text{e} + \text{Ar}^+ \]

If the generation of electrons by collisions plus the generation of electrons by ion bombardment of the cathode is greater than the loss of electrons from the plasma then the discharge will be sustained. If the losses become greater then the discharge will decay and extinguish. It is apparent from the emission that there is a high degree of activity in the discharge however even where the luminous intensity is greatest the bulk of the gas is neutral and in the electronic ground state.

If we now look at the process in more detail we can separate the discharge into different parts which are as a result of differing processes within the plasma.

Figure 3.1.1 shows a typical plasma obtained from an inert gas in a discharge tube at a pressure of 130 Pa (1 Torr) and Figs. 3.1.2 - 3.1.5 show schematics of the characteristics of the plasma. To describe the plasma in detail it is necessary to use the current versus voltage characteristic which is shown in Fig.3.2. As can be seen this can be divided into six different sections, the Townsend discharge, the breakdown region, the transition region, the normal glow discharge, the abnormal glow discharge and the arc.

3.1 The Townsend Discharge.

If a potential is applied across a discharge tube between the cathode and anode there will initially be a very small current passed of the order of 10^-16 A. This is as a result of the generation of a few ions and electrons by some external source i.e. ultra violet
A typical glow discharge in an inert gas atmosphere at a pressure of typically 135 Pa (1 Torr). Also shown are schematics of the plasma characteristics.

**Figure 3.1** A normal glow discharge in an inert gas atmosphere at a pressure of typically 135 Pa (1 Torr). Also shown are schematics of the plasma characteristics.
Figure 3.2

Voltage

Sustained by Ultra-violet light
Townsend

Self sustaining
Breakdown
Transition
N A Arc

Glow discharges
N = Normal
A = Abnormal

Current (Amps)
radiation, x-rays, cosmic rays, etc. This current will disappear if the external source of stimulation is removed. The small current will saturate, if we look at Fig.3.3.5 (which is a repeat of Fig.3.2) this is denoted by point A. The potential distribution across the discharge tube is shown by line A in Fig.3.3.2. The potential distribution stays linear in the Townsend discharge region as can be seen by lines B & C also on Fig.3.3.2. The points B & C on Fig.3.3.5 correspond to an increase in the applied potential and as we can see beyond point B the current is starting to rise evermore rapidly. Point C represents the limit of the Townsend discharge region. The applied potential is still not quite enough to accelerate a secondary electron ejected from the cathode to a fast enough velocity whereby it can undergo an ionising collision before it reaches the anode. Townsend produced an equation which enables a calculation of the current the system will pass to be made, this was of the form:

\[
I = \frac{I_0 \exp(\alpha d)}{1 - \gamma(\exp(\alpha d) - 1)}
\]

where \(I_0\) is the primary electron current generated at the cathode by the external source (e.g. U.V. radiation), \(\alpha\) is the number of ions per unit length produced by the electrons, \(d\) is the electrode spacing and \(\gamma\) is the number of secondary electrons produced per incident ion. As the applied potential is increased so \(\alpha\) and \(\gamma\) increase and \(I\) tends to infinity, thus breakdown is said to have occurred. This is seen in Fig.3.3.5 beyond point C where at a constant voltage the current rises by several orders of magnitude.
3.2 The Breakdown Region.

The voltage at which breakdown occurs is generally denoted as $V_B$. The breakdown voltage is dependent upon the mean free path of the secondary electrons and the distance between the anode and cathode. Somewhere of the order of $10^{-20}$ ions must be produced by each secondary electron for the avalanche to occur. The onset of the rise in current is that the potential reaches a voltage such that an electron can be accelerated to a high enough velocity to be able to take part in an ionising collision. Initially where the applied potential ($V_A$) equals the ionising potential ($V_I$) the current will be small. As ($V_A > V_I$) the position of the discharge tube at which an ionisation collision becomes possible moves closer to the cathode and so the distance left where an ionisation collision is possible increases. When ($V_A = 2V_I$) it is possible for the two electrons from the first collision to both be accelerated to a high enough velocity to both make ionising collisions producing then four electrons thus there is the start of a cascade process which at ($V_A \gg V_I$) breakdown occurs. As the electrons are accelerated to the anode the ions are accelerated to the cathode and when the ions bombard the surface they are able to liberate secondary electrons to add to the electrons producing ionising collisions and thus add to the fast increasing current. At the point of breakdown each electron liberated from the cathode is able to undergo enough ionising collisions such that the subsequent bombardment of the cathode will result in the generation of another electron. Inelastic collisions can also modify the electron state without necessarily ionising the atoms, the decay of the atom's excited electron back to its ground state results in the characteristic light emission. Thus at the breakdown voltage and beyond the plasma is known as a self
sustained glow discharge. The value to which the current rises at this breakdown voltage is governed by the external circuit resistance. If either the pressure or the electrode spacing is too low then the electrons cannot undergo enough ionising collisions during their flight from cathode to anode. At the other extreme if the pressure is too high and the electrode separation too large the ions produced by the collisions will undergo inelastic collisions which slow them down such that in striking the cathode they do not produce secondary electrons. These two features are the basis of Paschen's law (Fig.3.3.6) which relates the breakdown voltage to the product of pressure and electrode separation. The precise value of \( V_B \) is different for different material-gas combinations. The different materials have different secondary electron coefficients and different electron binding energies which result in the slightly different curves. The pressure times separation normally found in sputtering systems is generally to the left of the minimum and so the striking voltage is usually high. A procedure to reduce this striking voltage is to introduce a pulse of gas into the chamber, as can be seen on Fig.3.3.6 this would have the effect of pushing the required voltage towards the minimum. Once the discharge is established the pressure can be allowed to drop back to the working pressure.

A change in the potential distribution also occurs at the breakdown voltage, this is depicted in Fig.3.3.3. As our position on Fig.3.3.5 has progressed from point C to point D so has the potential distribution altered from line C to line D in Fig.3.3.3. The line D is convex upwards which is an indication that the space charge is positive. This is due to there being more ions than electrons in the cathode dark space of the discharge because of the higher mobility of the electrons than ions.
(1). As the discharge is struck there appears a light on the anode from the negative glow region of the discharge and similarly a glow on the cathode called the cathode glow formed by the neutralisation of incident ions. The region between the negative glow and the cathode is called the cathode or Crookes dark space. The precise distance of the cathode dark space is often hard to measure due to a diffuse negative glow region. In this case it can be measured from the potential distribution curve. It is the distance from the cathode to the 'knee' of the curve, an example being shown in Fig.3.3.3 for curve D with the cathode dark space equalling distance \( x_p \) (the cathode glow is very thin and can be neglected).

The cathode dark space is an important region of the plasma. The cathode dark space is inversely proportional to the gas pressure and the product of the pressure and distance is slightly less than the pressure times distance constant for the minimum breakdown voltage. Within the cathode dark space most of the ions and electrons that carry the current are produced (2), the negative glow is highly conducting and so has only a slight potential fall which indicates that the number of ions and electrons is almost equal. The area of influence of the cathode is related to the cathode dark space, the ions start to be accelerated from the edge of the negative glow and the ejected electrons are accelerated towards it resulting in a region with the most ionising collisions. Of the current reaching the cathode about 90% of it is carried by the ions and 10% is carried by the ejected electrons (3). This only true in the vicinity of the cathode because the ratio of ions to electrons changes throughout the cathode dark space. The source of ions changes from ionising collisions in the cathode dark space in a normal glow discharge to acceleration of a greater proportion of those produced in
the negative glow for an abnormal glow discharge (4).

The start of the negative glow corresponds to a point the mean distance travelled by an electron from the cathode before it makes an ionising collision. The cathode dark space is only faintly luminous and with highly luminous regions on either side any light emission tends to go unnoticed this is shown on Fig.3.1.2.

3.3 The Transition region.

After the breakdown has occurred the potential difference across the gas falls with increasing current. This region of negative resistance is known as the transition region. This region corresponds with a reduction in the cathode dark space. The potential falls to a slightly higher value than that of the ionising potential. There is little drop in potential in the negative glow and so the edge of the negative glow can be thought of as the virtual anode with the bulk of the potential drop occurring between it and the cathode. When the discharge is established the cathode dark space is virtually the whole distance between the cathode and anode. As the current is increased the potential distribution is changed with the virtual anode (negative glow) being brought closer to the cathode. The glow beyond the dark space is highly conducting due to the large number of ions and electrons and so for a small potential difference the current flow is maintained. In its extreme form this is the normal glow discharge and is shown by curve F. The space charge is effectively only operating on the first part of the curve up to the distance \( x_D \) from which point the rest of the potential distribution is small and linear. This is shown in Fig. 3.3.7 where the virtual anode is at points D, E and F corresponding to points D, E and F on Fig.3.3.5 and lines
D, E and F on the potential distribution curve Fig.3.3.3.

3.4 The Normal Glow Discharge.

In the normal glow discharge the current density is constant, the current rise is due to the area of the cathode layer increasing to a point where the whole of the cathode surface is covered. Once this point has been reached for a continued rise in current there must be an increase in the voltage. This increase in area is shown on Fig.3.3.7 at point F with a corresponding minimum value on the breakdown voltage versus pressure times electrode separation curve (Fig.3.3.6) and by line F on the potential distribution curve Fig.3.3.3.

3.5 The Abnormal Glow Discharge.

Following the normal glow discharge there is a region of increasing current for increasing voltage called the abnormal glow discharge. The reason for the increasing voltage is due to the virtual anode position becoming less than the cathode dark space. Thus for the necessary number of electrons to perform enough ionising collisions to give an increase in current the acceleration of them has to be increased by raising the potential. This is shown on Fig.3.3.4 with line F being for the normal glow discharge and lines G and H for two points in the abnormal glow region shown on Fig.3.3.5. The position of the virtual anode is shown on Fig.3.3.8. On the plot of breakdown potential versus pressure times electrode spacing (Fig.3.3.6) it can be seen that as the virtual anode is moved the breakdown potential goes through the minimum at the normal glow discharge and then rises sharply. This is why for sputtering systems which generally operate in the abnormal glow discharge region a high striking voltage is required.
3.6 The Arc.

As the current of the abnormal discharge is increased it reaches a point where it changes character, the current rapidly increases and the voltage decreases and the plasma shrinks to a spot but with very high intensity. The current density can be up to $1000 \text{ A/cm}^2$ with the current being proportional to the cross sectional area of the spot. The cathode and anode are heated up by the increased electron and ion bombardment as the current passed through the plasma is increased. This can result in the evaporation of the anode with the plasma then becoming sustained in an atmosphere of the anode material rather than the surrounding gas. Other features of the arc are the thermal ionisation which is the ionisation of molecules by other molecules, the emission of positive ions directly from the anode, high field emission due to the steep potential gradient at the cathode which is related to the very short cathode dark space which is somewhat less than 1mm and thermionic emission. Any of these factors will tend to increase the current which will intensify the problem and thus change from the abnormal glow discharge to the arc is very rapid and uncontrollable. This is shown in Fig.3.2, with the operating voltage of the arc typically only a few volts (i.e. 10 - 20 V). With magnetron sputtering sources working well within the abnormal discharge region to obtain the maximum deposition rate if there is any distortion in the electric field or the electron emission because of a surface state change then there is a strong probability of an arc occurring. This was recognised by Chapin as he developed the planar magnetron (5) and work has been done in improving the power supplies to handle such arcs (AEI Inc).
3.7 References.


3). B.N.Chapman. 'Glow discharge processes'

     12 (1940) 140


6). D.Schatz. MDX Power supply 'Application notes'.
     Fort Collins. CO 80525 U.S.A. (1985)

The cathode, often called the target, when acting as the electrode in a glow discharge is subjected to bombardment by positive ions. These ions are energetic enough to eject particles and radiation from the surface. The largest proportion of the ejected particles are neutral target atoms (1) of which those that escape from the cathode zone will be condensed out onto all surfaces in the vacuum systems. The deposition from a polycrystalline target will be in the form of a cosine distribution (2), this means that if the anode is opposite the cathode the thickest film would be deposited onto that. The bombardment of ions onto the cathode can also generate secondary electrons, positive and negative ions, desorbed gases, x-rays and photons. There will also be other effects such as reflection of incident particles, heating, chemical effects, backscattering, implantation, bulk diffusion and also some alteration of the structure of the surface (3,4). If the process is a simple d.c. sputtering one the anode and hence the growing film will be subjected to electron bombardment which will modify the nucleation and growth of the film. Thus from our simple view of the process of ejecting an atom and depositing it elsewhere the real processes involved are very complex. The processes that are possible are shown schematically in Fig.4.1.

4.1 The emission of neutral particles.

The number of atoms ejected from a target surface per incident ion is defined as the sputtering yield (5). Using the energy transfer function as used in the ionising binary collisions and adding the nuclear stopping power ($E_s$) which it has been shown is
(1). Incident ion.
(2). Chemical surface reactions.
(3). Topographical changes due to sorption.
(4). Sorption etching.
(5). Sputtered particles, neutral or charged.
(6). Reflected particles.
(7). Desorbed gas.
(8). Secondary electrons.
(9). Radiation.
(10). Implantation.
(11). Defect induced disorder.

Figure 4.1 Processes possible for a cathode subjected to ion bombardment.
proportional to the sputtering yield and incident ion energy we obtain the expression; where \( k = \text{constant} \)

\[
\frac{m_im_s}{(m_i+m_s)^2} \quad \frac{E \times k}{(m_i+m_s)^2}
\]

This can then be used to predict the sputtering yield \( (S) \) from the following expression;

\[
3 \alpha \cdot \frac{4m_im_s}{4\pi} \frac{E}{\left(\frac{m_i+m_s}{2}\right)^2} U_0
\]

This also incorporates the surface binding energy of the target material \( (U_0) \) & the factor \( (\alpha) \) which was a function of \( \frac{m_t}{m_i} \). This expression applies for bombarding energies of up to 1 KeV (6).

There is a threshold value for sputtering in the energy range 5 - 25 eV, this has been found to be approximately equal to the heat of sublimation (7). The yield is dependant upon the ion mass and the incident ion energy. There are many tables of sputtering yields for inert gases of a variety of energies (8,9,10). The deposition rate is not the same as the yield or erosion rate due to losses of material such as backscattering and condensing elsewhere other than the substrate. It is worth noting that the d.c. plasma is not monoenergetic and the target may not be pure and may thus be depositing reactively and so the deposition rate can be vastly different to the predicted yield. Ions once accelerated from the edge of the negative glow may suffer collisions and one estimate was that as much as 50% of the energy reaching the cathode was from neutrals knocked there by the incoming ions (11). It has been estimated that only
1% of sputtered particles are charged (12). There is little known about the neutral ejected particles because of the difficulty in analysing them (1) although attempts have been made (13,14) using absorption analysis. Stirling and Westwood also used emission spectroscopy and concluded that a significant proportion of the sputtered particles were molecular.

4.2 Secondary electron emission.

When generated these are accelerated away with an initial energy equal to the cathode potential. Once they arrive in the plasma they are regarded as primary electrons capable of producing ionising collisions and helping to sustain the plasma. The secondary electrons in the case of a normal glow discharge produce the largest fraction of ions by collisions in the dark space but in an abnormal glow discharge the largest proportion are produced in the negative glow region instead (15). Eventually they will arrive at the anode with the electron bombardment being the source of the anode heating (16). Ball found that 40% of the applied power density was seen as heating by secondary electrons at the anode with of the order of 1% anode heating being caused by the condensation of the sputtered neutrals. It was also found that a substantial number of electrons had not undergone ionising or thermalising collisions but arrived at the anode with their full target potential.

4.3 Positive and negative secondary ion emission.

The positive ions produced constitute less than 1% of the sputtered flux and the high negative target field does not allow them to escape but draws them back to the target surface. The negative ions produced will be accelerated away from the surface by the negative target field there is evidence that by the process of electron
stripping collisions they arrive at the anode as energetic neutrals (17). These too constitute only a small fraction of the flux compared to the flux of high energy electrons.

4.4 Reflected incident ions.

A proportion of the incident ions are neutralised at the target surface and are reflected as atoms away from the surface (18). Under low primary energies the reflection is higher than for higher primary energies (19).

4.5 The desorption of gases.

All surfaces that are in contact with the plasma will desorb gases, this is known as 'outgassing'. This takes several forms, initially the adsorbed gas layer will be sputtered away, in the case of the chamber walls the bombardment will then lead to heating followed by further desorption. The target which undergoes bombardment by ions would then sputter chemisorbed gases (20), followed by the removal of occluded gases which can be both sputtered and thermally desorbed. R.f. sputtering, which often requires the use of hot pressed or powder targets, is particularly prone to desorption of occluded gas which can cause film contamination and arcing problems (21). There are also cases where the bombardment causes a change in the chemical composition of the surface such that the volatility changes and the sputtering rate is increased over that of the pure metal, this is known as chemical sputtering.

4.6 The emission of photons and x-rays.

The sputtered particles can leave the target in an excited state, as these undergo resonance or Auger type transition photons will be emitted. These are of both
ultra violet and visible radiation. This radiation coupled with the emission of the excitation and decay of the gas atoms leads to a technique for controlling the deposition process. The intensity of the spectral emission from the target material and the gas or gases can be ratioed, if they are kept constant throughout small changes in other parameters then the thickness and stoichiometry of the depositing film can be more accurately controlled \((22,23)\). The ultra violet radiation may be a problem if sensitive substrates are used. This is also true for x-ray radiation which may be emitted at up to the bombarding ion energy \((4)\).

4.7 Surface modification.

There are various forms of surface modification all related to the bombardment of the surface. If the target is an alloy the components of the alloy will sputter at different rates. To obtain a homogenous deposited film it is necessary to sputter the target for a period of time prior to deposition to bring the target surface to equilibrium. If the sputtering power is altered the depth of the altered layer will change and a further period of stabilisation will be required \((24)\). The heating of the target may allow diffusion to take place, this would further affect alloy targets and in an extreme case the surface layer may not stabilise because the component of the alloy being sputtered fastest would be replaced by diffusion from the bulk. With good water cooling the diffusion effects are generally localised.

There will be a layer on the surface where interstitials and vacancies will be generated by the impinging ions as they undergo subsurface multiple collisions. If the bombarding ion is of sufficient energy it may then be implanted. These implanted ions will be below the layer altered by vacancies and
interstitials. Thus with further erosion of the surface the implanted gas atoms will be uncovered, also contributing to the structurally changed surface layer.

4.8 Chemical effects.

In an alloy target it is possible to heat the target sufficiently such that one of the elements will sublimate. This can occur even with water cooling, it is possible to put enough power into some targets such that the front surface of the target melts, the power limit being set by the thermal conductivity of the target material (and backing plate if used). This can be catastrophic unless the target used sputters vertically upwards and in this event one of the benefits of sputtering, which is the ability to sputter in any orientation, is lost. If the target is an alloy it is also possible to bombardment with enough energy such that the compound will dissociate. This too is common with r.f. sputtering, with the result that there is a need to add a reactive gas to regain the original stoichiometry. D.c. sputtering cannot be used to sputter non-conducting targets, which compounds often are, and so dissociation is rarely a problem.

Chemical sputtering is largely used for etching purposes but is now being seriously looked at to improve the lifetime of sputtering targets used in a reactive process. Over a period of time it is possible to build up a non-conducting compound on the surface of the target which impairs the deposition process. It is possible to add a volatile chemical that will react with the surface producing a volatile product which will be desorbed from the surface. This may take place faster than the sputtering rate of the target element (25). The ion bombardment may also help the reaction as well as providing the sputtering action. This process is used at
present in the semiconductor industry for etching samples but may find further use in the future for the target cleaning prior to a reactive sputtering process.

4.9 Reactive D.C. Sputtering.

The sputtering of an elemental or alloy target in an atmosphere containing a reactive gas produces a deposit on the substrate containing both the target material and the reactive gas as a single compound. Thus from a metal target it is possible to produce oxides, nitrides, carbides, etc. or combinations such as oxyfluorides, carbonitrides, etc. (4). The reactive gas will react with the sputtered particles primarily at surfaces. Even with the reactive gas carefully directed towards the substrate it will also reach the target and will also react with the clean surface there. This is a problem with most materials because the sputtering rate of the compound is usually lower than that of the element except in the cases where chemical sputtering takes place (10). Often the compounds on the target surface have a higher secondary electron emission than metals which is seen by a reduction in the sputtering voltage at a constant current. For there to be a higher emission it can be inferred that more of the ions are producing secondary electrons than previously and so will sputter less (4). There will also be a reduced sputtering rate due to less efficient sputtering by the reactive gas than by the noble gas (26,27) and a further reduction due to the lower yield from the compound than the elemental form of the target (28). In an extreme case such as titanium or aluminium a small increase in the partial pressure of the reactive gas can poison a significant area of the target such that due to the large difference in sputtering rate of the compound compared to that of the metal there is a
reduction in the utilisation of the reactive gas leaving a further excess of the reactive gas. Thus there is a runaway situation which can occur within a fraction of a second for a reactive target or over a period of seconds for a more stable one. With the current held constant to reverse the process and clean the target it is necessary to reduce the oxygen to well below the partial pressure which caused the catastrophic event. If this is plotted out it can be seen as a hysteresis loop of reactive gas pressure versus voltage or deposition rate. This hysteresis loop is also current, time and total pressure dependant. If the process can be controlled then d.c. reactive sputtering offers a very much higher rate of deposition than r.f. sputtering. Also there are no hot filaments as in electron beam or resistance heated evaporation to be protected from which makes the d.c. reactive sputtering a more consistant process for production.

4.10 Planar Magnetron Sputtering.

Limitations of the d.c. sputtering process are that the deposition rate is low compared to electron beam evaporation and the substrates are heated by the electron bombardment. The use of a hot filament electron emitter to produce the extra electrons to add to the plasma to balance the larger losses incurred when reducing the operating pressure, increased the deposition rate. This is called triode d.c. sputtering. Its limitation was the problems encountered when trying to produce evenly distributed plasmas when scaling up and the same problem as the electron beam evaporation in that the hot filament was very sensitive to reactive gases. The major breakthrough came in 1974 when the planar magnetron sputtering source was described (29). There had previously been work done to incorporate a magnet to
improve the low pressure characteristics of the discharge systems (30,31) but this development had the advantage that it enabled sources to be built for both small laboratory and large industrial coaters using the same design. These sources had a uniform deposition profile for all but the ends of the target with high deposition rates and much reduced substrate heating. The basis of the design was the use of the magnet to restrict the path of the electron motion and also increase the distance they travel thus increasing the chance of undergoing ionising collisions. The motion of an electron in an electric field is a straight line from the cathode to the anode. In a magnetic field it travels in a circle around the lines of magnetic force. If the two are combined with the magnetic and electric fields parallel the electrons will spiral towards the anode, thus increasing the distance they travel for the same cathode to anode spacing. This in itself would allow the discharge to run at lower pressures. For a glow discharge tube run without a magnetic field the electrons would undergo a number of collisions, if the magnetic field were added the electrons would undergo considerably more collisions due to their increased path length. Alternatively to make the same number of collisions the pressure could be reduced. Thus we have the lower pressure operation of the source but still the substrate heating. If however we have the magnetic field arranged parallel to the target surface i.e. perpendicular to the electric field, the electron will describe a cycloidal path perpendicular to both the magnetic and the electric fields. The average motion of the electrons is known as the Hall drift. The electron motion is such that they would eventually be lost from the target surface and off over the edge. To prevent the electrons being lost the magnetic field is taken into the target surface, the
electrons then become reflected because of the cathode potential which they see when approaching from a magnetic field line that at the point of re-entry is almost parallel to the electric field. The area defined by the points of entry and exit of the magnetic flux is the area of erosion of the target. Thus to gain the benefit of the magnetron there is a sacrifice to be made in the form of the reduced target utilisation. If the design is made such that the erosion area is of the form of a closed loop there will be the minimum loss of electrons. Electrons can be lost, they describe their complex spiral type motion and everytime they undergo a collision they are knocked onto a different magnetic flux line. The magnetic flux is such that it is strongest across the front of the target near to the poles and weaker as it gets farther from the target. Thus an electron near the target is unlikely to be lost but one far away from the target may be knocked far enough away from the magnets such that they are unable to retain them. Thus there will still be some electron bombardment of the substrate but this will be minimal compared to the diode sputtering source. A schematic of a planar magnetron sputtering source is shown in Fig.4.2. Sputtering is not an energy efficient process, it is estimated that between 70 - 95% of the supply power is dissipated as heat with the rest being shared between the ejected particles and emitted radiation (32,8).

The motion of the electron in a uniform electric and magnetic field is given by the expression;
Figure 4.2 A planar magnetron sputtering source.
Electron motion in an $E \times B$ field.
\[
\begin{align*}
\frac{dv}{dt} &= \frac{e}{m} (E + v \times B) \\
\end{align*}
\]

where \( v \) = velocity, \( t \) = time, \( e \) = particle charge, \( m \) = particle mass, \( E \) = electric field and \( B \) = magnetic field. The magnetic field used in the magnetron is low such that only the electrons and not the ions are affected by it. When the electric field is zero and the magnetic field is uniform then the electrons will drift along the field lines with a speed \( v_a \). This speed is unaffected by the magnetic field. The electrons will also circle the field lines such that a combination of the two produces a helical path of a radius given by (33);

\[
\frac{m_e}{v_p} = \frac{r_L}{e \cdot B}
\]

where \( v_p \) = perpendicular velocity, \( m_e \) = electron mass and \( r_L \) = the Larmor or gyro radius.

The cyclotron frequency \( \omega \) is given by;

\[
\omega = \frac{eB}{m_e}
\]

This is shown schematically in Fig. 4.3. It is not a true helical path described by the electron, the electric field across the dark space is not constant due to the variations in the space charge (34) and the magnetic flux is not truly parallel across the front surface of the target. These two factors thus lead to variations both in the gyro radius and the pitch of the helical path that the electron will follow (34).
The electrons ejected from the target surface move in a series of loops around a 'race track' or 'erosion zone'. The shape of the loops depends on the initial electron energy, if it is high the loop will be more circular (trochoidal) than if it is ejected with low energy. This is shown in Figs.4.4.1 & 4.4.2 (33,35). The electrons that perform these loops have a mean free path of \( \approx 1 \text{m} \), they arrive back at the target surface with the same energy as they departed with which is too low to cause secondary electron emission. This means that the only way the electrons can be lost is by collision or when the magnetic field is not parallel to the surface in which event they set off in their spiralling motion. The region of the target where there is a magnetic pole has a proportion of the magnetic flux that leaves perpendicular to the surface and is lost to the system. Electrons also travel along these lines and are similarly lost from the magnetron plasma. This is particularly noticeable in a circular magnetron configuration where under certain conditions it is possible to see a column of light from the emission due to the collisions of the escaping electrons. This column is separate to the confined magnetron plasma and due to its origin being from the centre of the deposition source it is directed at the substrates. Hence many designs of magnetron have an anode or earth cap to block off the fast high energy electrons and so reduce the substrate heating. It will not collect all the electrons that drift into the column but they will only be the low energy electrons lost from the fringe of the plasma and will do relatively little damage to the substrate.

The magnetic confinement can be regarded to some extent as an alternative to a localised pressure increase. The deposition rate for a constant current is pressure dependant. As the pressure increases the ion
density rises and the sputtering rate increases, this will peak beyond which point any further pressure rise produces greater than 50% of backscattered material. For a d.c. diode sputtering source this peak is approximately 13 Pa (100 mT) whereas for the d.c. planar magnetron sputtering source it is approximately 1.3 Pa (10 mT) (8,35). This lower pressure of operation is useful in that the contamination rate by residual gases is reduced. Coupled with the higher deposition rate of the magnetron source this makes the coatings made by planar magnetron sputtering very much more pure than could previously be obtained.
4.11 References.


5). G.K.Wehner & G.S.Anderson. Chapt.3.'Handbook of thin


8). L.I.Maissel. Chapt.4.'Handbook of thin film technol'

Pub. Wiley Interscience.


16). D.J. Ball. J. Appl. Phys. 43 (1972) 3047


32). R.J.Hill. 'Physical vapour deposition'
'Thin film processes' (1978)
Eds. J.L. Vossen & W. Kern.

34). P. Parker. Chapt. 15 'Electronics' (1950)
Pub. Edward Arnold.


Eds. J.L. Vossen & W. Kern (1978)
5. System considerations.

5.1 Introduction.

In embarking on the research several decisions had to be made regarding the design of the apparatus to be used. Work carried out previously highlighted some limitations of the apparatus (1). The earlier work had been carried out using an electron beam evaporation system for depositing metals onto glass and polymer substrates and a separate d.c. planar magnetron sputtering system for depositing an oxide seed layer if required.

The electron beam evaporator had several limitations proving to be suitable only for the deposition of metals, the oxygen required to regain stoichiometry from a compound material or required as a partial pressure to produce the oxide from the metal also contaminated the filament of the electron beam source. This would have required extensive modification in the form of a double pumped system to reduce the filament contamination. Other problems associated with the electron beam deposition process such as the control of rate and stoichiometry would also have had to be extensively improved. The high deposition rate led to some problems in the attempts to deposit very thin films, the movement of the shutter out of the deposition zone and back again led to graded thickness films because the deposition time was so short that the time to move the shutter was a significant proportion of the total deposition time. The other major limitation was the deleterious effect the radiant energy had on the polymer substrates. It was decided that if major alterations were necessary then improvements obtained by changing techniques should also be considered.

To improve the film uniformity, to obviate the need for a shutter and to enable some form of continuous
monitoring to be carried out it was decided that a dynamic rather than a static substrate was preferable. In the first instance this was a rotating cylinder with the substrates mounted on it. This enabled polymer, glass or single crystal substrates to be coated. Once the benefits had been confirmed a decision was made to concentrate on the deposition of coatings onto flexible polymer substrates. In order to maximise the productivity of the system a roll-to-roll coating unit was built enabling over 100m of polymer to be coated in a single evacuation.

In order to reduce the contamination of the film between successive coatings more than one deposition source was included in the system. To fit three sources and the roll-to-roll coater into the system it was necessary to use alternatives to the electron beam evaporators such that the sources could be positioned at any orientation. The alternatives were d.c. planar magnetron sputtering sources. These sources also enabled better uniformity to be obtained as well as much better control over the deposition rate.

Thus it was decided what deposition system was required and an old ex-commercial resistance heated evaporation system was taken apart and rebuilt to incorporate these features. The precise system details shall now be considered in full.

5.2 The pumping system.

Initially the pumping system consisted of a conventional oil diffusion pump and rotary pump combination. The diffusion pump was of an old design which had poor performance with a low pumping speed and which showed evidence of backstreaming of oil. To improve the pumping performance the diffusion pump was replaced by a cryopump.
Cryopumping has in recent years found favour in the vacuum industry, particularly where very clean vacuum environments are required. The cryopump utilises helium which is used as a refrigerant to produce cold surfaces in a vacuum pump where the surfaces freeze out any gases that are incident on the cold surfaces. There are two 'pumping' surfaces, one at 77K and the other at 20K, these are bolted to the two stages of the refrigerator. The panel at 20K has a coating of charcoal over its inner surface to adsorb the gases not condensed on the other two surfaces. With the gases condensing out at differing temperatures on different surfaces it is not surprising that the pumping speed is species dependent (2). With the system being sealed there is nothing to contaminate the vacuum chamber as other systems have. The only possible source of contamination has then to be from the rotary pump which should have a filter between it and the vacuum chamber. If for any reason oil does get from the rotary pump into the vacuum system then not only can the oil be cracked by the plasma but also can contaminate the cryosorption panels thus stopping the adsorption of the low molecular weight gases and restricting the pumping performance.

Initially the residual gas analyser (RGA) was connected on to the cryopump using the regeneration port. The purpose of this was to directly monitor the gases during deposition and with fast response. To a limited extent this worked. If the throughput was restricted by throttling down the system the pressure in the cryopump was in the upper part of the top range of the total pressure channel. Thus when scanning through the atomic mass numbers there was plenty of sensitivity available. Unfortunately the throttling of the system was not compatible with the way the system was required to be used. Our experience was that the throughput needed to be
as high as possible, this has also been put forward by others (3,4). When high throughputs were used the RGA could not measure total pressure only the partial pressures. This was different to the diffusion pumped unit which at higher throughputs still had a low enough pressure at the top of the pump to be able to measure the total pressure using the RGA. The reason for not leaving the RGA on top of the diffusion pump was that there was some backstreaming of oil, not during deposition but when the pump was not under heavy gas load. Also there was, under moderate gas load, some contamination by the rotary pump oil which having backstreamed to the diffusion pump it appeared to be more volatile than the diffusion pump oil and was not condensed out at the top of the diffusion pump. Thus although the RGA could be used on the cryopump in some circumstances it was decided it would be better to use it with its own pumping system.

The cryopump suffered another limitation with the high throughputs that were used during the deposition process and this was that the pumping speed did not remain constant. The cause of this was that the loading of the cold panels in the cryopump depended not only on the amount of gas requiring to be condensed but also on the temperature of that gas. Thus if the system were pumping at a maximum amount at a particular temperature then if the temperature increased then the loading would rise and the temperature of the cold finger would rise. This would not only slow down the rate of condensation but in an extreme case would cause the previously condensed gases to evaporate and thus stall the pump (this is known as dumping the pump). Recovery when dumping a pump was also very inconvenient in that it would require to be completely regenerated and then cooled down before being able to pump out the system. This procedure took of the order of 10 hours to complete making the system much less
tolerant than that of a diffusion pumped system. To try to keep within this limit the system required some throttling, even so for a 1° temperature rise within the laboratory the pumping speed would change such that there was a 0.009Pa (0.065 mT) pressure change. This is a small enough change when depositing metals but in a reactive process this can be critical, particularly as the gas flow was controlled to 0.1% and a 0.009Pa pressure change represents a 0.15% change in the total working pressure. To eliminate this the laboratory temperature would have had to have been kept constant to < 0.5° variation which was somewhat impractical. The draught or turbulence caused by a person walking past the equipment caused a greater fluctuation than that and with other water cooled vacuum systems in the laboratory plus windows and doors all creating turbulence and temperature variations it was estimated that the best that could be obtained in terms of control of the pumping speed would be 0.5%. This defeated the efforts made to control the input of gases and the supply for the magnetron source to 0.1%. Since the time of our difficulties with the pump it has been drawn to our attention that the manufacturers have modified the design of their pump to reduce some of these effects. The temperature and capacity of the top panel have been improved thus reducing the effects of the temperature changes at the same throughput however it does not eliminate the problem. Over a longer time scale there are other problems with cryopumps, the ice layer from the condensing gas can be dense at low throughputs or open structured at high throughputs. The changes in the form of the ice layer also alters the pumping performance such that the performance is dependant upon the previous history and the throughput is not constant.

These factors are detrimental to any reactive deposition process where the gas admittance is required to
be balanced carefully against the needs of the growing films. It would be possible with considerable pumping overcapacity to stabilise the process but the cost would be prohibitive and as there are alternatives unnecessary. Thus the cryopump was removed and a high performance diffusion pump and liquid nitrogen cold trap were put in its place. The liquid nitrogen trap was rarely used but it provided a permanent baffle helping to reduce any backstreaming of diffusion pump oil. Figures 5.1 and 5.2 show schematics of these two systems.

The diffusion pump has none of the limitations of the cryopump but inevitably has one or two of its own. There is the possibility of backstreaming not during deposition when the gas loading is high but during the initial pumpdown as the chamber approaches its base pressure. The time the system is left at its base pressure should be limited and a dummy gas load used if the deposition system is to be left for any length of time. This gas load is also useful in that it helps the outgassing process by its scrubbing action and leads to a cleaner system (4). In using high throughputs it is as well to be aware that there will be significant loss of oil over a period of time. On one system which has a charge of oil of 700 cc and was used for deposition on average 2-3 times per week for 8 hours per time was seen to loose 100 cc of oil per month which is approximately 1 cc per hour of use at high gas throughputs. The limitation in terms of throughput was the stall pressure of the diffusion pump, unlike the cryopump if the pump if the pump was stalled it was only a matter of a few minutes to correct the situation. In its stalled state the backstreaming was more of a problem but with the optical baffle in the form of the liquid nitrogen cold trap plus the tortuous path above the trap it was never found to be a problem. A further advantage of the
Figure 5.1

Magnetron

Roll Coater

Plate valve

RGA

Cryopump

Rotary Pump

Figure 5.2

Magnetron

Roll Coater

Plate valve

RGA

Rotary pump

Diffusion pumps

1

2
diffusion pump was that it could be operated up to its stated limit without any need to throttle it.

The base pressure that could be obtained was dependant upon the leak rate into the system and the outgasing from both the system furniture and from the roll of polyester substrate. The base pressure obtained with in excess of 100 m of polyester in the roll coater was $0.27 \times 10^{-3}$ Pa ($2 \times 10^{-6}$ T). This could all be regarded as the effective leak rate into the system and could be expressed as a mass flow of $79 \times 10^{-3}$ sccm.

5.3 The planar magnetron sources.

The constraints of the vacuum system in terms of the shape and size limited the configuration that could be used to one where the target material was mounted vertically and thus sputtered horizontally onto the roll of polyester which was transported in the vertical plane across the deposition zone. In order to be able to produce a multilayer coating more than one source was required and to give some flexibility to the system it was decided to incorporate three planar magnetron sources into the design. The resulting design is shown in the schematic diagram in figure 5.3. Commercial magnetrons were not available in a size or form such that they could be incorporated into such a system and so magnetrons of our own design and construction were used.

The design of commercial magnetrons appeared to be simple enough and the magnetic configuration of a small pot magnet and so as a simple test a small pot magnet was purchased (5). It was mounted on an electrical leadthrough in the vacuum system with a copper disc bolted to the front across the poles. It worked, the pressure of operation was high $> 10$ Pa (80 mT). With no earth shielding the plasma covered all the surfaces although the
Figure 5.3 A schematic of the planar magnetron sputtering source and vacuum leadthrough. The source is shown without the earth shields.
The diameter of the pot magnet was 38mm, the dimensions were too small to obtain an accurate magnetic flux profile. There was no water cooling which limited the input power. The operating pressure was of the order of 10 Pa. (80 mT).

Target showing an approximation of the erosion profile.

Commercial pot magnet

Figure 5.4
magnetic confinement did mean that the copper was the major material sputtered. The lack of water cooling led to the copper target glowing red hot. Figure 5.4 shows the dimensions of the magnetron and the flux strength as measured by a digital handheld flux meter (gaussemeter). The first magnetron that was built used the largest cylindrical magnet that was commercially available as a stock item. The magnet was positioned on a mild steel backing plate and with a mild steel centre pole. Across the front a copper plate was bolted, this copper plate had a water cooling tube soldered to it. The whole surface with the exception of the target surface was covered with an earth shield. Figure 5.5 shows a drawing of this design with a plot of the flux and a plot of the erosion profile taken from the copper target. This design was very instructive and led to a redesign of the magnetron. The water cooling was never sufficient using the soldered tube, it was in the wrong place, it needed to be directly behind the area of maximum erosion to have the best effect and ideally there should be the minimum amount of material between the water and the target. The water cooling pipes left the body of the magnetron in the vacuum and used a separate vacuum leadthrough to the electrical leadthrough that the magnetron was mounted on. This led to problems, the pipes were also at the magnetron potential and so needed earth shields. Unfortunately the shape was difficult and invariably there was a gap between the pipes and the earth shield that was not less than the mean free path of the electrons and arcing could and did occur. When this happened carbon from the coated polymer water cooling pipes or metal from the arc itself could cause a short to earth and the magnetron would have to be shut down leading to poor productivity due to high down time. The inner pole of mild steel was far too large making the gap between the inner and outer pole very small.
The magnetic and erosion profiles of the magnetron shown below.

Figure 5.5 A schematic of an early design of magnetron source.
The magnetic and erosion profiles of the magnetron shown below.

Figure 5.6 A schematic of the improved design of magnetron sputtering source.
Figure 5.7 A photograph of the magnetron in use showing the beam of electrons escaping from the centre.
highlighting the 'v' shaped erosion of the target. The centre pole needed only to be large enough to be able to bolt the target to and large enough such that it was not saturated by the magnetic flux from the cylindrical permanent magnet. Figure 5.6 shows the modified design with the narrower centre pole, the water cooling pipes being taken through the same electrical vacuum leadthrough that the magnetron is mounted on and the direct water cooling of the target across the whole of the erosion zone. The additional modification was that the poles were brought out to be in front of the target material with the aim of producing a broader erosion zone. Figure 6 also shows a plot of the magnetic and erosion profiles. When this magnetron was used it was noticeable that there was a column of plasma extending from the centre of the magnetron which was energetic enough to severely damage the polyester substrate. The reason for this was that in bringing the poles out in front of the target the magnetic path was so well defined between the poles that at the centre of the magnetron electrons left quite a large area as a well defined slightly divergent beam perpendicular to the surface. This is seen in the photograph in figure 7. One major factor in using the magnetron was that by confining the electrons the temperature of the deposition process could be kept low and so coatings could be put onto plastics that previously could not because of the temperature involved. Thus it was necessary to stop the beam to prevent it adversely affecting the polyester, this was done using a water cooled earth cap which was placed approximately 5 mm away from the centre pole. One other limitation of having the poles in front of the target is that it may be possible to sputter from the poles and thus contaminate the depositing film. This can be reduced by using the correct shape to the pole pieces and by the use of close earth shielding. Sputtering can be reduced from
the pole surfaces if the flux exiting the surface is arranged to be perpendicular to that surface such that the electrons are removed as rapidly as possible from the vicinity of the poles such that the ionisation and thus the bombardment is minimised in that area. This can be done on the centre boss by turning the bolt head to be hemispheric and further making sure the overhang on the outer pole is minimised such that the flux exits from the inner edge directly towards the centre pole. Other designs lead to higher erosion of the poles. Poor design may go unnoticed if the sputtering yield of the target material is much higher than that of the pole material. In this situation there can be net deposition onto the pole pieces from backscattered material. If the pole design cannot be altered or there is still an unacceptable level of contamination then a close fitting earth shield could be added to suppress the plasma and consequent bombardment from that area. It is often a trade off between possible contamination of the growing film against the amount of target utilisation.

Inevitably in modifying the magnetrons the resulting design was more complicated than envisaged. Since building the magnetrons new magnetic materials have been produced with higher performance. It was decided to assess the qualities of the new magnets to see if it would be possible to improve the performance of the magnetron and at the same time simplify the design. It was also decided to try to further improve the coating uniformity by fabricating an oblong planar magnetron source. The major loss of flux is across the air gap between the poles, this is so dominant that the other losses are insignificant. The shape of the magnetron was constrained such that the minimum air gap has to be between the poles, if it were anywhere else the magnetic circuit would be shorted. Thus the magnetron had to be deeper than the
Figure 5.8
A schematic of an oblong magnetron built using small ferrite block magnets.
pole separation. The new magnetic materials are based on the rare earth -transition metal alloys such as SmCo and NdFe (5). To make effective use of these expensive new magnetic materials the magnets used would have had to be very thin and tall. At present it is not possible to produce them in such a form, they are difficult to manufacture and in one case extremely brittle. To use them in the presently available sections the volume required would make the cost prohibitive. For one of the magnetrons built using anisotropic barium ferrite blocks costing £83.70 the equivalent SmCo blocks were priced at £1179.00 for which it was not expected to have any significantly better performance. A schematic of the design of the oblong planar magnetron sputtering source is shown in figure 5.8.

5.4 The magnetron d.c. power supplies.

The initial power supply was a commercial d.c. diode sputtering source supply designed to produce an output of 3kV 1A. This supply was poorly regulated and could not operate for sustained periods of time at full power. At anything less than full power the sputtering rate was too slow creating difficulties in winding the polyester slow enough to obtain desired thickness coatings and further difficulties in cleaning up the surface of highly reactive metal targets on which there is a tenacious low sputtering rate native oxide layer. To be able to reproducibly produce coatings of the same thickness or to produce long lengths of the same thickness coating it was already known that the supply had to be stable. The deposition rate is proportional to the current when the voltage is held constant. The sputtering yield is only slowly increased by increasing the voltage (seen as the ion energy) but the sputtering yield is directly
related to the number of particles striking the surface i.e. the current density (7). Therefore to control the deposition rate well the current had to be precisely controlled rather than as was common in commercial supplies the power kept constant. It is possible to have constant power from a supply where the voltage is swinging by hundreds of volts due to the rectification of the input. To keep the power constant there must of course be a corresponding current swing and thus a corresponding fluctuation in the deposition rate. It was decided to build a power supply to cater for our specific needs.

Figure 5.9 shows a schematic of the rectified three phase power supply that was built. The output of the supply was capable of delivering 10A at 1kV and it was subsequently run at 10A successfully. The laboratory supply was three phase with 415 volts between phases and fused at 30A. A three phase transformer was wired up in the manner of delta in and star out. The reason for this was that if the output load changed, which during arcing would occur, the load imbalance would not be passed on to the input side of the transformer, each phase of the input would maintain a balanced load throughout the load change. The variable transformer was interlocked to prevent starting the supply at anything other than 0V. This was because if the load were applied suddenly the inductance of the windings would be enough to blow the supply fuse of the phase which was on peak loading at that instant. The rectification was the weak link in the supply, it was possible for one of the diodes to have all the voltage across it in certain circumstances in which event the diode would fail. The diodes were used in a series of three in each phase and for both polarities with in total eighteen diodes being necessary. As a series of three the maximum voltage that each diode would normally be expected to have across it would be a third of the total voltage of
Figure 5.9

A schematic of the rectified three phase magnetron power supply.
1kV. In practice if an arc occurred there was a back e.m.f. from the large inductor in the circuit for which a bridging resistor is placed across the diode as protection. When the load is again reversed the speed of reaction of the diodes being finite and individually very slightly different it was possible for two of these diodes to respond quicker than the third which would then see the whole of the voltage and would fail. Reducing the resistor value would help but with the penalty of less protection from the reverse voltage. These diodes are due to be replaced by high voltage diodes (upto 30kV) as used for laser power supplies to further improve the tolerance of the supply to major arcing. The current in the circuit was limited by the inclusion of six firebars in parallel with each other and in series with the magnetron. The response of the firebars being that as they pass more current their resistance increases thus limiting the current they can pass. Similarly the voltage is dumped into the light bulbs, placed in parallel with the magnetron, rather than it damaging other more sensitive equipment such as the electronic constant current supply. The constant current supply either a 150V 7A or a 40V 10A bench supply which was placed in series with the magnetron in the earth line. The choice of voltage is dependant upon the material to be deposited, the difficult highly reactive materials require the higher voltage such that if the reactive gas supply is pulsed the voltage swing is still within the capabilities of the supply and it can thus still deliver constant current. There was an additional diode positioned on the positive terminal of this supply to further protect the supply from reverse loading. In order to monitor the current and voltage a handheld multimeter was placed in series to measure the current to another across the supply to measure the voltage. The voltage measuring multimeter on occasions
suffered problems due to the arcing which occurred fast enough to damage the multimeter in spite of it having internal protection. As an attempt to further protect the multimeter a spark gap was added across the terminals such that it would spark (breakdown) at a lower voltage than would damage the meter. To an extent this worked although each spark would slightly alter the gap and breakdown voltage and over a period of time this was enough to fail to protect the meter. Finally a 30kV attenuating probe was used which to date has worked without mishap. The final part of the circuit was the large inductor (10H) placed in series in the earth line of the supply. This large inductance was used to resist any change in the operating conditions such as the change from the abnormal glow discharge to the arc. Once an arc is struck it also resists the change back to normal running, this meant that the supply had to be shut down to extinguish any arc. The rectified supply, without the inductor or the constant current supply, had ± 5% ripple on the voltage this was reduced by the inductor and with the constant current supply in the circuit the response was such that the voltage was stable to ± 0.3% for a highly reactive target and better than ± 0.15% for less reactive targets.

An alternative supply was also built which had increased stability but this was limited in voltage such that the maximum current available for most materials was in the range 1 to 2 A. This supply used motor generators as a high stability d.c. source. Large generators were driven by three phase motors and the output was controlled by varying the voltage to the field windings. The use of the low voltage control of the field windings enabled an automatic cut out to be used to help control arcing. A 100W 1Ω wire wound resistor was positioned in series in the earth line which was wired to a relay in the field
Figure 5.10

A schematic of the magnetron power supply based on the motor generators as a stable d.c. power source.
windings supply line such that when the system arced the relay would be tripped and would cut off the voltage to the windings and thus the supply to the magnetron would be shut off. Once the arc was extinguished the relay would reset itself and the field winding supply be reinstated and thus the magnetron supply would be restored. The large inductor meant that the changes were resisted and the supply in reality decayed and when reinstated was ramped up rather than the instantaneous change that occurred at the field windings supply relay. The rest of the supply is virtually the same as the rectified three phase supply with there being the same electronic constant current supply to give additional stability, the firebars and lamps for protection and the multimeters for the necessary high accuracy monitoring of the voltage and the current. A schematic of this circuit is shown in figure 5.10. This supply had better stability than the three phase supply with 0.1% control of the current. The limitation of the supply was the voltage available from the supply, even with the higher rated electronic supply of 150V 7A the maximum voltage was limited to < 700V. This limited voltage also limited the current available, this could be improved by the inclusion of a second electronic supply adding a further 150V to the 700V already available. It is worth noting that this type of supply is very expensive and not available for commercial coating facilities, it was modified from existing equipment and as such its cost to us was nominal. New solid state power supplies are now available which would be cheaper and much more compact than the motor generator sets (8). The rectified three phase supply would be the cheapest supply but would not have the versatility of the solid state supplies or the same control and arc suppression facilities.
5.5 The gas supply and gauging system.

The vacuum systems all had a common gas source, large cylinders of gas filled to a pressure of $17.25 \times 10^6$ Pa (2,500 P.S.I.) with a regulator fitted to reduce the pressure in the gas lines to $34.5 \times 10^3$ Pa (5 P.S.I.). It had been previously noted that changes in the flow of gas made in one system would affect the other systems and also that there were gas pulses originating from the operation of the regulator. The frequency of the pulses depended upon the rate of use of the gas. The regulator valve was a flap valve which when the line pressure dropped below a minimum level would open allowing the gas at the higher pressure to enter until the line pressure reached the correct line pressure at which point the valve would shut. With the cylinder gas being at a very much higher pressure the gas enters as a pulse which is transmitted down the supply lines and, if the needle valves were open, into the vacuum systems. In an attempt to minimise this problem the vacuum systems were fitted with mass flow controllers and valves which operated fast enough to limit the effects of the pulses and also limited the interference from the other systems. In depositing metal films the problems of pulsing go unnoticed, it is only when a highly reactive target is being sputtered reactively that the process acts as an amplifier and the pulses are more of a problem. This is shown in the output from the chart recorder of the total pressure taken during the reactive deposition of tin oxide which is reproduced in figure 5.11. These pulses take the control from being at the same level as the power supplies at 0.1% to $\approx 0.3\%$. The pulses were further restricted by the inclusion of a 1.2μm pore size filter in each of the supply lines, which successfully brought the control back to 0.1%. Adding a
High Pressure

FSD = 3sccm

Low Pressure

Pulse height = 0.3sccm

Figure 5.11 A copy of the chart recorder trace showing the effect of the pressure fluctuations caused by the high pressure valve.
large volume to the lines did not remove the pulses only reduced the frequency of them.

The gauging was an integral part of the gas control system particularly for the reactive deposition process. The requirement was for the gauge to be stable, unaffected by the plasma and with an operating range covering the pressures used for sputtering. The pirani gauge covered the operating range, was stable and suitably positioned was unaffected by the plasma. To obtain the necessary resolution the output signal from the pirani gauge was amplified and offset to produce a final signal in the range 0 - 1 V which provided the control signal for the mass flow equipment. The response of the pirani gauge is species dependant but as a method of controlling the partial pressure of the reactive gas this is unimportant. The mode of operation was to use a constant flow of inert gas (argon) giving a constant pressure, the reactive gas could then be added giving a small partial pressure increase, the total pressure would then be kept constant. If the pressure changed for any reason such as the increased growth rate of the depositing compound using more of the reactive gas the total pressure would reduce causing the controller to open the valve and increase the reactive gas flow to compensate. Similarly if the target were to partially oxidise reducing the metal deposition rate and thus reducing the reactive gas requirement the pressure would then rise and the controller would shut off the valve to compensate keeping both the inert gas flow constant and the excess partial pressure of reactive gas constant.

It was possible to use any 0-1 V signal to control the reactive gas flow. This voltage could be from a residual gas analyser, an optical transmission monitor, a plasma emission spectrometer (9) or from the magnetron supply. Taking the signal from the residual gas analyser was less
Figure 5.12 A schematic of the three options that could be used as the feedback for the gas control system.
successful than taking it from the pirani gauge because of the time lag between the instant that any change was made in the vacuum system and the instant that it was detected by the residual gas analyser. This led to the gas admittance always over correcting for any change in the system which was seen as a cyclic change in the partial pressure. If this were exaggerated it could be used as a pulsing technique to produce some of the more difficult compounds. The use of pulsing has been described as a technique for producing compounds from highly reactive targets. With a partial pressure of reactive gas these targets tend to form the compound on the target surface reducing the deposition rate significantly. The reactive gas line was provided with a diverter valve between the mass flow valve and the chamber, this enabled the gas to be introduced into the system either at the substrate or into the throat of the pumping system. This kept the flow away from the target surface long enough to clear it of the compound without and detrimental effects on the growing films. This will be considered further when considering the production of oxide coatings in chapter 9. A schematic of the gas control options is shown in figure 5.12.

5.6 The roll-to-roll coater and film monitoring equipment.

A schematic of the roll-to-roll coater assembly is shown in figure 5.13. The polyester is wound from roll to roll over a large drum around which the magnetrons are placed. There are two driving motors which act against each other to provide not only the motion of the polyester film but also the tension in it. This was important, if there was insufficient tension in the polyester it would drop on the takeup roll and bind in the system which in
ROLL-TO-ROLL
COATING SYSTEM

Electrical feedthrough
Quartz crystal monitor
Optical feedthrough
Water cooled anode/earth
Reactive gas admittance
R.F. coil
Pirani gauge
Argon gas admittance
Baseplate

To pumps
Optical transmission monitor
Mass spec. pressure gauges
Roll coater drive feedthrough
Resistance monitor

Fig. 5.13
an extreme case would stop the motion completely. The tension was also important in maintaining good thermal contact between the polyester and the large drum as it passed through the deposition zone. If the plastic is not in intimate contact it is possible with a slow web speed to damage the plastic by overheating it. This shows up as wrinkled plastic in a mild form and charred plastic in a gross form. The feedback monitoring for the motor drives was also from this centre drum, the rotational speed was sampled and kept constant at any desired level. A flexible drive connected the drum via a rotating vacuum leadthrough to a mechanical positional counter which enabled any point on the polyester to be located. The roll coater assembly also acted as a base on which various other items could be secured. The coating was monitored by two types of transmission monitor and by a resistance monitor (figures 5.14 and 5.15).

The optical monitors were positioned after the deposition zone between two idler rollers such that the plastic always moved at a tangent to each roller and therefore could be maintained perpendicular to the optical path easily. Were the monitor to be between an idler roller and the windup roll the diameter would change and thus the polyester would not remain perpendicular to the monitor and there would be a further loss to the system. One of the idler rollers also acted as a support to the polyester such that two sprung probes could be pushed into contact with the coating to measure the resistance. This form of measurement damaged the film under the probes as it moved past but in spite of this the measurement was consistent and reliable. The probe separation was > 60 mm which left the most uniform area of the coating available for other tests to be carried out. The resistance was monitored in one of two ways depending upon the conductivity of the coating. If there was reasonable
Figure 5.14
Dragging probe resistance monitor.

Figure 5.15
Optical monitors.
conductivity as one could expect from the semiconducting oxides or from pure metals then a multimeter capable of measuring of up to 10 MΩ was generally adequate. Other coatings which had poor conductivity were measured by an insulation tester (megohm meter). This meter measured upto 10 GΩ. This widened the scope of using the resistance of the film as a control parameter, there being a number of films which had resistivities of > 10 MΩ but which changed resistivity by a few orders of magnitude with changing deposition conditions.

There are two optical monitors both measuring the transmittance of the coating. The first of the monitors uses four light emitting diodes and four silicon detectors giving the transmittance at four points in the visible electromagnetic spectrum (11). The relative intensities changed as the coating changed and once calibrated for a coating it could be used to maintain a constant coating performance. This monitor had not the sensitivity of the resistance monitor but did provide backup information and because particularly for multilayer coatings where the repeated damage of the coating by the resistance probes made the resistance measurements suspect. The response of the light emitting diodes had to be calibrated for each deposition because the diodes and the detectors were sensitive to the pressure in the vacuum system. The second of the transmittance monitors was based on the use of a rotating grating monochromator. A white light source was connected to a fibre optic system, the light was split with one fibre going to a monochromator & the silicon detector output was displayed on a storage oscilloscope. The second fibre was taken into the vacuum system then through a prism, the coated polyester, a second prism and then out of the vacuum system and to a second monochromator. The response from this monochromator was displayed on the second channel of the
oscilloscope. The two traces could be compared and modifications made to the deposition parameters to suitably alter the coating to produce the desired response. The limitations of this system were similar to the other system in that it was not as sensitive as the resistance monitor. The number of joints in the fibre optic that monitored the coating was high and degraded the quality of the signal available such that the information obtained was regarded as backup information but rarely was it used for controlling the process.

The rotating grating monochromator could also be used to monitor the plasma in front of the planar magnetron source, which once well characterised could form the basis of a technique for controlling the process that did not involve any contact with the coating (12). Once the characteristics had been obtained the plasma could be monitored by two large area silicon detectors. With suitable filters cutting off overlapping information these signals could be used to monitor the proportions of metal and oxygen in the plasma. This could form an extremely cheap and simple form of process control, it could not however be used for troubleshooting or characterising plasmas which the more expensive equipment could do.
5.7 References.

1). C.A. Bishop. 'Ion-assisted evaporation of thin films'
M.Phil. Thesis. Loughborough University
2). P.D. Bently Vacuum. 30 (1980) 145

3). G.A. Shirn & W.L. Patterson.


Nov (1985) 667


7). L.I. Maissel. Chapt.4 'Handbook of thin film tech.'
Eds. L.I. Maissel & Glang. (19 )

New Jersey 08003. USA. (1985) p 4


The polyester was loaded onto the roll coater outside the vacuum system on a separate rig. Of the order of 150 m was loaded at any one time, about half this would be coated and removed and then the second half would be coated. The roll coater was then fitted into the vacuum system and the optical and resistance monitors added followed by checks made on the optical and electrical continuity. The position counter could then be set and the system pumped down. For a new load of polyester it required approximately 15 hours to reach a pressure of better than $6.7 \times 10^{-4}$ Pa ($5 \times 10^{-6}$ Torr). If it was the second half of the loaded polyester being pumped down for the second time then the pumpdown time could be considerably reduced. This was because the polyester would lose some of the included water and also some of the gas trapped when the polyester was wound onto the roll before the first pumpdown and would not have had sufficient time between pumpdowns to regain the moisture or gas. Also there would only be half the material which would also reduce the potential problem. Thus it was possible to pump the system down to a similar pressure in approximately 5 hours. When the system had reached the desired base pressure a small amount of gas was bled into the system to give a gas load such that the chances of backstreaming was reduced. (This assumes the system to be pumped by a diffusion pump and for there to be no liquid nitrogen cold trap in operation.)

In order to start the deposition process it was first necessary to choose the required magnetron and route the gases to that position, the reactive gases being directed onto the substrate opposite the magnetron in use. The water then needed to be turned on followed by the electrical power. The inert gas could then be fed
into the chamber and the pressure allowed to stabilise. The inert gas used was argon, 70 sccm was fed giving a pressure of approximately 2.5 mT which was enough to be able to start up the magnetron. The power to the magnetron could then be turned up to the desired level (generally 1 A). The roll coater was then set in motion, by not moving the polyester straight away a thick deposit was formed on the polyester substrate which if for any reason registration of the polyester was lost could be used to reset the position counter. The transmission monitors would also decay to zero with the deposition of thick metal on the polyester and the light emitting diode transmission monitors could then be set to zero. It is worth noting that both the light emitting diodes and the detectors were in the vacuum system and either one or both were sensitive to changes in pressure and so it was necessary to set both the 100% and the 0% readings whilst at the working pressure immediately prior to deposition. The resistance readings were also altered with the onset of deposition from reading infinite resistance on the uncoated polyester to a few ohms on the metal coated polyester. If an oxide or nitride coating was required once the target conditions had stabilised the reactive gas could be added, this had to be done slowly otherwise the target surface would become contaminated and this would radically alter the deposition rate. This could be used to give information on the required deposition parameters. As the reactive gas was fed in the transmission and resistance monitors were observed, the gas could be fed in until a predetermined value was reached or as a series of steps over a region of interest to try to bracket the optimum conditions. At each step the values of flow, current, voltage, pressure, transmittance, resistance, speed, tension, position, etc. were recorded.
If a single layer coating was all that was required then the shut down procedure was the reverse of the start up procedure. If a second coating was required then the magnetron was shut down, the polyester rewound, the gases rerouted and the deposition process repeated with the new source material. This could be repeated for any number of coatings but for a limited number of different materials based on the limitation of only having three magnetrons and hence only three target materials available in the chamber at any one time. The resistance monitor consisted of two dragging probes and as such it could damage the coating, this limited the use of the resistance monitor to the first layer although it would read some value of resistance for subsequent layers this value was also a measure of how much damage it had done to the previous coating as well as the resistance of the second coating. Most of the coatings that were of interest to us were transparent which meant that the optical monitors were operating only in the region 80% - 100% transmittance and so were of limited use because of their restricted sensitivity. The resistivity was much more sensitive to changes often changing by several orders of magnitude which meant that although the value of resistance could be wrong it was still possible to use it as a control parameter.

After shutting down the power supplies, isolating the pumps, turning off the water and bringing the system up to air the polyester could then be indexed. This was done by winding to each point of interest as established by the recorded measurements and the position counter and then marking the polyester. Once this had been done the roll coater was then detached from the vacuum system and taken back to the rig where the polyester was loaded and the coated polyester was removed by winding onto cardboard cores. The roll coater was then available to
be returned to the vacuum system and the coated polyester could be examined at leisure. By winding the polyester from one core to another measurements could either be made in situ or samples could be cut out and measurements made elsewhere, i.e. on the electron microscope, scanning spectrophotometer, etc.
7. Techniques for thin film examination.

7.1 Optical microscopy.

The coatings were deposited onto either glass slides or optical quality polyester (I.C.I. 'O' grade Melinex). These substrates both exhibit few surface defects and as thin coatings closely match the surface and the crystal size is too small to resolve there was nothing to see by optical microscopy. Where the optical microscope is useful was in looking at gross defects such as pinholes, cracking of the films due to stress, corrosion or degradation and abrasion scratches. It was possible to look for pinholes in opaque metal films using transmission microscopy. This was less successful with the transparent oxide coatings and the surfaces had to be examined by electron microscopy. The cracking of the transparent oxides was visible but with difficulty due to the lack of contrast. The corrosion or degradation and the scratching of the films were generally visible to the naked eye and were easily examined under the optical microscope.

To improve the contrast in the more difficult cases the Nomarski technique was used (1). Figure 7.1 shows the components of a Nomarski interference microscope. The key part of the design was the use of the modified Wollaston prism which produced a pair of oppositely sheared wavefronts. Each ray passes through the prism twice, once on the way to the specimen and once back from the specimen passing through the prism an equal distance away from the axis as the incident ray but on the opposite side. The summation of the two optical paths was equal to passing through the centre twice. The effect of the prism was to produce a lateral displacement of half a wavelength and the contrast was produced by the resultant coherent interference. When a white light
Figure 7.1
The Nomarski interference microscope.
source was used there was a series of interference colour bands visible. These could be adjusted to cover the whole of the field of view with one with any height deviations appearing as a different colour to the background. Any slope to the specimen appeared as a brightness change with the appearance becoming brighter or darker depending upon the sign of the slope. The contrast could be optimised by moving a lever which gave a lateral movement to the prism allowing the whole range of Newton colours to be utilised. This technique was referred to as differential interference contrast microscopy. The minimum displacement that could be measured was of the order of half the wavelength divided by the fraction of the band separation that could be resolved. This could be a height of 15 nm for a step on a flat specimen with a monochromatic light source but would be nearer 25 nm for a white light source and worse still if the specimen were poor.

7.2 Scanning electron microscopy (SEM).

The principle of electron microscopy are well documented (2), essentially it is a technique where by using shorter wavelength radiation a resolving power considerably better than that of optical microscopy is obtained. The resolving power of any microscope is limited by the wavelength being used to form the image. The optical microscope used visible radiation with a wavelength of approximately 400 nm and has a resolution limit of 200nm. The electron microscope used electrons for illumination, these are of much shorter wavelengths of the order of 0.006nm and gave correspondingly better resolution of 3nm. The limit to the resolution then being limited to the lens aberrations not solely wavelength. This higher resolution enabled us to look at the surfaces of both the glass or the polyester. As they
were essentially featureless the electron microscope was best used in conjunction with other techniques.

7.3 Transmission electron microscopy (TEM).

This technique was closely allied to scanning electron microscopy using much of the same hardware. The difference is that in the scanning electron microscope the information was obtained by reflection and in transmission electron microscopy as its name suggests the samples are examined in the transmission mode. To be able to pass the electron beam through the sample the thickness of the sample had to be < 100nm thick. If the sample were thicker than that little information could be obtained due to lack of contrast because few electrons were able to penetrate the specimen. At the other extreme very thin specimens allowed too many electrons through and saturated the phosphor screen also giving little information. With very thin films it was also common for the electrons to physically damage the thin films which were unable to dissipate the heat created by the electron bombardment. The films deposited onto glass substrates could not be examined because they could not be separated from the glass and release agents were not used for fear of altering the nucleation process. A technique for removing the polyester and leaving the coating resting on a microscope grid had been previously developed (3) and permitted these coatings to be examined. In looking directly at the coated films it was possible to obtain structural information. In some cases a measurement of crystal size could be measured directly from the micrographs. If this was not immediately available then the electron diffraction patterns could be obtained to give a crystal size and look for any preferential growth in specific crystal orientations by comparison to standard measurements (4).
Where the deposited coatings were too thick to be examined even after the polyester had been etched away it was possible to use a microtome (5) to produce thin sections. The method of operation of the microtome is shown in figures 7.2 and 7.3. This technique was primarily used on the electrochromic coatings where there were successive thick coatings deposited and as shown in figure 7.3 a section of the whole structure could be obtained. As can be imagined to produce a section of few different materials each with their own vastly mechanical properties was fraught with difficulties and the technique was used sparingly.

7.4 Energy dispersive analysis of x-rays (EDAX).

Figure 7.4 shows schematically some of the possible events that can occur when a specimen is subjected to irradiation by an electron beam. In reflection the important events are the backscattering of primary and secondary electrons, the emission of Auger electrons, x-rays and light and in transmission there are unscattered, elastically and inelastically scattered electrons. Most of these are sources of information, the electrons could be utilised in electron microscopy as could the light emission by cathodoluminescence also the Auger electrons could provide chemical information. Similarly the x-rays could be used to obtain some chemical information.

The x-ray energy spectrum that is produced during the electron bombardment is characteristic of the elements so bombarded. There were a considerable number of correction factors which were programmed into the microprocessor controlled equipment, these were designed for bulk samples. The samples that were examined were thin generally <100nm thick. These samples gave rise to a number of problems in the analysis of the films. The corrections for the microscope were constant from when
Figure 7.2 The operation of the microtome.

Figure 7.3 The production of samples.

Coated polyester thin slice in the cast matrix.
The various types of radiation produced by an incident electron beam.
the instrument was calibrated and could be regarded as having a small error associated with them. The variable error derived from the three corrections that were made for the atomic number (Z), the absorption (A) and the fluorescence (F) effects. These were fixed in the computer and assumed the films were thick enough for the absorption and fluorescence to have had a significant effect. When used in the transmission mode it was different in that because the films had to be thin to be able to transmit the electron beam then there was little effect from the fluorescence or absorption. Thus for the same coating there were different percentage compositions given dependant upon which of the two electron microscopes the analysis had been carried out on. This problem was highlighted when the coatings were being examined for contamination by iron and aluminium during the design and modification of the magnetron sources. The visual appearance of the sources was that the iron poles and the earth shields had been coated with sputtered target material, the analysis of the films indicated ≈5% aluminium and ≈3% iron content when the analysis was carried out on the scanning electron microscope. Until that time no contamination had been apparent. Modifying the sheilding had little effect and following analysis of the same sample by energy dispersive analysis on the transmission electron microscope and then by Auger electron spectroscopy the same sample was shown to have < 1% aluminium content. It was from this sample that the limitation of the commercial software used was discovered. This is not to say that the technique was unusable for thin film coatings but just that there needed to be care in the interpretation of the results.
7.5 SEM plus EDAX.

If the films were scratched prior to being loaded into the electron microscope then initially the edges of the scratches could be examined (6-8). If the edges were clean and sharp then the film could be regarded as well adhered to the substrate. If the scratch had however a ragged edge then the failure was taken to be between the film and the substrate with the ragged edge being produced by the film deforming in front of the abrading particle when lifted off the substrate and then the tearing occurring as the particle continued to push forward. If the coating was a multilayer coating then it was possible to use the analysis technique to confirm the interpretation from the scanning electron microscope alone. An analysis of the bottom of the scratches and of the ragged edges could often confirm where the failure took place using the residual material to aid the interpretation.

7.6 TEM plus EDAX.

This could only be used on the single layer coatings and was used to check on the impurity levels of coatings after any structural alterations to the planar magnetron sputtering sources. The limitation of the technique was the need to remove the substrate from the coating. It was usually more cost effective to examine the samples by Auger electron spectroscopy and the results were more consistant particularly if the problems of the energy dispersive correction programme were taken into consideration.
7.7 **Electron microprobe microanalysis.**

This technique could be used to give the chemical composition of a coating. A fine beam of electrons was accelerated to 15kV and a focused spot of electrons was scanned across the surface. This bombardment produced three main sources of information, backscattered electrons, the absorbed electrons and the characteristic x-rays. These could be produced as a picture of the surface in the same way as was done on the scanning electron microscope. If the spot of electrons were stopped on a selected area then either a wavelength or an energy dispersive spectrometer could be used to detect the characteristic x-rays which were metered by a proportional counter. This was compared to a standard specimen analysed under the same conditions and after suitable corrections were carried out, similar to those used in the scanning electron microscope with the energy dispersive technique, quantitative information could be extracted (9). The limitation of this technique when analysing thin films was that the x-ray information was generated from a 'bubble' 1,000 nm diameter starting approximately 100 nm below the surface. The films examined were < 100 nm thick and so the bulk of the signal was generated from the substrate not the film (10). This could be regarded as a poor signal to noise ratio and thus the errors in the percentage composition were increased. In trying to obtain confidence in other techniques (particularly the energy dispersive analysis carried out on the electron microscopes) this technique was of use but in general other techniques revealed more information and more easily.

7.8 **X-ray diffraction.**

This technique was used to obtain not only the crystal size and orientation but also by the displacement
and broadening of the diffraction peaks it was possible to obtain a measurement for the stress in the film (11-15). This technique was also limited by the lack of volume of film under analysis. The film signal intensity was low and with the substrate signal approximately five orders of magnitude larger there are significant errors introduced in identifying the minor peaks. Some of these minor peaks were situated on the slope of the substrate signal further introducing errors. The thin films on the polyester substrates were under stress following the deposition of the coating, this was apparent from the curvature of the sample. To enable the sample to be mounted in the machine it had to be held flat, this process added to the stress the film was already under. The peak shifts and the peak broadening looked for to give accurate results were further complicated by the flexible nature of the substrate. Due to the precise nature of the information sought and the low intensity of the signal available the procedure for examination of the films was considerably extended, as such the technique could not be regarded as a quality control one. The warm up and stabilisation time of the x-ray machine was of the order of two days and the sampling time was approximately 30 hours to obtain the minimum of six peaks required for the analysis of each sample. The minimum film thicknesses required were also thicker than were generally deposited and specially grown samples 100 to 300 nm thick had to be produced to be able to use this technique. The use of the x-ray diffractometer was limited to samples which could not be suitably prepared for examination on the transmission electron microscope where electron diffraction patterns could be produced much more readily although without the possibility of obtaining any information about the amount of stress in the film.
7.9 Auger electron spectroscopy (AES).

A 10 mm by 10 mm sample of coated polyester was placed into a high vacuum system and by bombarding the sample with low energy electrons (1 - 10 keV) Auger electrons were ejected, collected and analysed (16). The impact of the electrons caused some of the atoms within the sample surface to be ionised followed by an electron re-arrangement within the atoms. A typical process could be as follows; if an electron from the K energy level were ejected an electron from the L2 level could fill the vacant site releasing an amount of energy $E_k - E_{L2}$. This energy could then be transferred to another electron possibly in the L3 level, which could then be ejected from the solid. This latter electron would be the characteristic Auger electron and its energy $E$ would be given by;

$$E = E_k - E_{L2} - E_{L3}$$

As $E_k$, $E_{L2}$ and $E_{L3}$ are all characteristic of the particular element concerned it was possible by measuring the energies and the number of Auger electrons to determine the chemical composition of the sample surface. The energies of the Auger electrons were low (20 to 1000 eV) and as such even if they were generated from far within the sample it would only be the ones from the first two or three atomic layers ($\approx 2$ nm) that could escape with their energy intact. The detection limit was $\approx 0.1\%$. Hydrogen and helium could not be detected because they do not have enough energy levels for the Auger transitions to occur. Even so this technique was more attractive in its performance than the energy dispersive alternative technique where the detection limit started at atomic number 10 or thereabouts. Also as the technique deliberately examines only the surface there was no problem caused by the thickness of the coatings as there was in other techniques (17). The
information was presented as a plot of the secondary electron distribution \([N(E)]\) versus the energy \([E]\) from zero up to the primary beam energy. The Auger electron contribution to the secondary electron distribution was very small and the background signal had a small gradient associated with it and so the spectra were differentiated electronically and the spectra plotted as \(\frac{dN(E)}{dE}\) versus \(E\). This suppressed the background and enhanced the peaks due to the auger electron contribution making the interpretation considerably easier. With the high surface sensitivity of this technique any surface contamination would form a significant proportion of the information presented. It was possible to still obtain information about the coating by ratioing the relative peaks or an alternative was to remove the surface in the same high vacuum system producing a fresh clean area suitable for examination. Ion bombardment was used to remove the surface, unfortunately this could with some compound materials lead to some dissociation of the compound and allowances had to be made in those cases (18). As a technique Auger electron spectroscopy could be used to give a composition profile through an interface by alternately plotting spectra and ion etching the surface. This was useful in looking for contamination between deposited layers.

7.10 X-ray photoelectron spectroscopy (XPS).

This technique was also known as electron spectroscopy for chemical analysis (19). The surface was irradiated by low energy x-rays, photoionisation of the sample surface occurred and the resultant ejected photoelectrons were collected and analysed. The ejected photoelectrons had a kinetic energy distribution which was determined by both the x-ray energy and the binding energy of core electrons. The information was presented
as a plot of kinetic energy (eV) versus the counts per second. This plot gave the composition of the surface. Other more detailed plots of the binding energy (eV) versus the counts per second gave the chemical information, e.g. stoichiometry. The atom, by reason of ejecting a core level electron was in an excited state and would decay to the ground state. This was done either by releasing energy as a photon (x-ray emission) during the filling of the core hole by another electron or by transferring the energy to a third electron which was released into the vacuum as an Auger electron. These Auger electrons also appeared in the plot of kinetic energy and could be used to provide additional information. The information obtained in the plots was only from the surface atoms (~ 2 nm) with a detection limit of 0.5 At.%. All surfaces become contaminated even under vacuum and the surface contamination was seen all too well, usually as carbon and to a lesser extent oxygen. This contamination could distort or hide the information that was sought and so the surface few layers were usually removed by ion beam etching within the ultra high vacuum system in which the analysis was carried out. As in the Auger electron spectroscopic system this could create problems of dissociation of the compounds under the bombardment of the ion beam. One of the techniques advantages is that the area of the sample irradiated and thus the source of the information is large (> 1 mm) and this produced a more typical bulk chemical composition than could be obtained from a more localised information source. The other extra information that it can produce over the other techniques described so far is that of the stoichiometry of compounds. However as an analysis tool it was still best when used in conjunction with other techniques rather than as the sole technique.
7.11 Laser ionisation for mass analysis (LIMA).

The laser ionisation for mass analysis instrument (20) was also capable of providing not only the elemental composition but also some information regarding their chemical form (21). A schematic of the equipment is shown in figure 7.5, it comprises a 'Q' switched Nd:YAG laser quadrupled in frequency which was fired at the sample. The laser pulse volatilised and ionised a cylinder of material, these ions were then mass analysed by a time of flight mass spectrometer. The information was sampled by a transient recorder and could be plotted as the mass to charge ratio ($M/Z$) versus the relative intensity ($I$). The laser produced a 4 ns pulse of light that had a power density variable from 0 to $10^{11}$ Wcm$^{-2}$ at 265 nm and could volatilise a cylinder of material of 2μm to a depth of up to 0.5μm. The information produced was qualitative not quantitative however with a detection limit of 10 ppm and when used in conjunction with the x-ray photoelectron spectrometer and comparing to known standards semi-quantitative information could be obtained (22). Both positive and negative ions were produced and either could be selected for analysis. The mass sensitivity was such that the isotopes of the elements could be resolved and thus overlapping peaks produced by the binding energy spectra plot from the x-ray photoelectron spectrometer could be better analysed. This technique could be used on the multilayer coatings and by suitably adjusting the conditions the top layer could be analysed separate from the second layer.
Schematic diagram of LIMA 2A

Fig 7.5
7.12 Ellipsometry.

It has long been established that thin films could be detected on surfaces by the change in ellipticity of a beam of light following its reflection from the surface (23). Since this early observation the technique has been developed and has become a standard technique for thin film analysis with equipment and computer software being readily available all based on the same basic theory (24). The information required to obtain the film characteristics of thickness and refractive index are the values of $\Delta$ and $\Psi$ (25-27). These are the relative phase difference for the two polarisations and the ratio of the reflected amplitudes for light polarised in the plane of the incidence to that polarised normal to the plane of incidence ($\tan \Psi$). These cannot be measured directly and have to be derived from other measurements. Figure 7.6 shows a schematic of the equipment used. The light source was a helium:neon laser which initially passed through a depolariser which converted the beam to one of circularly polarised light. The beam then passed through a polarising drum containing a prism which converted the beam into linearly polarised light. The beam then passed through a compensator containing a $\lambda/4$ plate which produced the elliptically polarised beam that was required. With the surface in one particular orientation this elliptically polarised beam when reflected would once again linearly polarised such that in passing through an analysing prism the beam could be extinguished. The final beam intensity was measured by a photo detector which enabled the conditions to be set that produced the minimum response by the detector. To reduce the systematic and random errors two pairs of readings from the polariser and analyser were taken, $P_1$ and $A_1$ to give the first minimum and then at a
Lasernepolariser or Extinction meter

Analyser Angle of Compensator

Elliptically polarised light

Linearly polarised light

Laser Depolariser Polariser

Attenuator

Compensator Elliptically polarised light

Angle of incidence

Film

specimen

Figure 7.6 A schematic of the ellipsometer.
complementary position $P_2$ and $A_2$ where $P_2 \approx P_1 + 90^\circ$ and $A_2 \approx 180^\circ - A_1$ (28). The technique for finding the initial values $P_1$ and $A_1$ is to set the polariser and analyser drum to $45^\circ$ and to rotate the polariser positively and negatively from this starting point to see in which direction the photo detector minimum lay. Once the signal had been minimised the same procedure was carried out using the analyser drum. The two drums were then moved alternately and the signal would progress towards a minimum. Once obtained and the values noted the same procedure was repeated to obtain the complementary values but starting near to the expected values to reduce the time taken to reach the minimum. The four values obtained were then inserted into the computer along with the angle of incidence and the programme would then calculate either the real and imaginary parts of the refractive index or the refractive index and thickness. To find the refractive index and thickness for the transparent films it was first necessary to determine the refractive index of the substrate. The glass slides were thin and needed to be mounted on a thicker block of glass with a matching fluid to eliminate the reflection from the back surface of the slide from interfering with the front surface reflection. The polyester substrates were more of a problem, the polyester by its method of production is biaxially oriented and this produces variations in the refractive index depending upon orientation. If the above procedure was adopted and samples were measured along the same film orientation a reliable substrate refractive index of $n = 1.74$ and $k = -0.01$ (or less) could be obtained. If the value of $k$ was much higher than $-0.01$ then it was an indication that there was still some contribution from the back reflection reaching the photo detector or that the film was absorbing at that
wavelength. Further to this the back surface reflection also needed to be eliminated. This was done by abrading the back surface with wire wool to roughen the surface and then coating the area with dense black ink. This effectively removed the back surface reflection from contributing to the signal reaching the photo detector. There is a lower limit to the thickness below which dielectric films on polyester substrates produce unreliable results, this appears to be $\approx 25$nm and around this value there may be an additional error due to operator competence. The errors in this technique have been estimated to be $< 1\%$ for films with a high reflection at 632.8nm and which are free from absorption (29). The technique has its primary application in the semi-conductor industry where using well defined substrates, i.e. single crystal silicon slices, the native oxide of $\approx 2$nm could accurately be measured. The solution to the equations could be plotted graphically as a series of ellipses where the tip of each ellipse would virtually pass through the same point. Each of these ellipses represented the change in the film thickness at a single refractive index, the different ellipses are represented the differing refractive indices. Thus if the thickness of the film were enough at a particular refractive index it would be possible to produce a false answer. This is because the answers for $\Delta$ versus $\psi$ were a periodic function of the thickness and thus an alternative method of finding the thickness was required to be used to confirm which of the thickness solutions was correct. From figure 7.7 it can be seen that there was also an area centred around the point $180^\circ, 120^\circ$ where the discrimination was poor. To avoid that area it was necessary to either alter the angle of incidence or by defining the refractive index by a different technique it was possible to use a modified
The response from the ellipsometer to changes of refractive index and thickness. Each complete loop represents a thickness that would give a change in $\delta$ of $180^\circ$. (ref 25).
programme where using a fixed refractive index a single solution of thickness could be obtained. Ellipsometry was used extensively on the deposited films but not without these problems occurring. To be able to avoid the anomalies the refractive index of films were on occasions also obtained from the interference pattern in the plotted response of the scanning spectrophotometer (30) or by the Brewster angle extinction technique as developed by Abeles (31) and by obtaining the thickness of the films from measurements using the Talystep or once again from the spectrophotometer.

7.13 The scanning spectrophotometer.

A scanning spectrophotometer which covered the wavelength range 170 - 2600nm was used. The wavelengths of interest were primarily in the visible and near infra-red (340 - 2600nm) where the light source used was a tungsten lamp and the detectors used were a photomultiplier tube for the visible and lead sulphide (PbS) cell for the near infra-red. Figure 7.8 shows the optical layout of the equipment. The layout is for examining samples in transmission, to examine the films by reflection an attachment shown by the broken lines in figure 7.8 is fitted. This attachment simply reflects the beams up to a reference mirror and the sample and then redirects the beam back to the original optical path. The reference mirror for the reflection measurements was a freshly deposited aluminium front surface mirror. This equipment was used on virtually all the coatings to provide details of the changing performance of the reflectance, transmittance and absorptance throughout the visible and near infra-red spectrum to compare with changing deposition parameters.
Figure 7.8
The optical layout of the scanning spectrophotometer.
The far infra-red performance was determined using a different instrument of similar design which covered the spectral range 2000 -15000 nm. This was seldom used, for a working measure of the far infra-red performance an emissometer was used instead.

7.14 The emissometer.

The emissometer comprised a head with a heater and detector mounted in it and a heatsink block. The samples to be tested were placed on the heatsink block with water used to obtain a good thermal contact. The head was first placed over a low and then a high emissivity standard and the instrument calibrated. The emissometer head could then be placed over the sample to be tested and a direct value of the emissivity would be displayed. The radiation detector (a differential thermopile) responds only to radiation heat transfer and was designed to have an output linear with emissivity (± 0.01 units) and has a near constant response to thermal wavelengths (3 - 30 μm). In the far infra-red it was assumed that there was 0% transmittance (the spectrophotometer confirmed this) thus the reflectance plus absorptance were equal to 1. When in a steady state all the radiation absorbed would also be emitted thus by measuring the emissivity a good measure of the reflectance was obtained (32).

\[ A + R + T = 1 \] for \( T \approx 0 \) in the infra-red

\[ A = 1 - R \quad \& \quad E = A = 1 - R \] in steady state
7.15 The thickness measurement by the stylus technique.

This measurement is also referred to as the Talystep measurement after the manufacturer's name of the machine. This technique used a fine pointed stylus (≈ 50μm radius) which was moved in an arc over the surface. The vertical movements were measured by an electrical displacement type pick-up where the transducer was a variable air gap differential inductance. The stylus force was variable and was dependant upon the scale of magnification used. The output was presented as a graph and a calibration block was used to check the performance. This instrument could measure films on glass slides of 5nm ± 1nm reliability (33) providing there was a suitably sharp step to traverse. Measuring the coatings on the polyester substrates was more difficult because of the flexible nature of the substrate. This was overcome by sticking the polyester onto a glass slide and using a very light load, even so the thinnest films measured were 10nm with an error of ± 2nm.

7.16 The Hall effect measurements.

When a metal or semiconducting coating is placed in a magnetic field and a current passed along the sample perpendicular to the magnetic field there is an electric field generated perpendicular to both the current flow and the magnetic field. This is known as the Hall effect (34). For a film of length (l), width (b) and thickness (d) the resistance is given by the equation;

\[ R = \frac{p l}{d b} \]

where (p) is the resistivity.
if \( l = b \) then \( R = \frac{V}{ld} = R_s \)

where \( R_s \) is known as the 'sheet resistivity' and is expressed in 'ohms per square'.

The method for measuring the sheet resistivity \( (R_s) \) directly is to measure the resistance of a rectangular film and divide by the number of squares of the film material that lie between the end contacts. The samples cut in our system had a ratio of \( l:b \) of 4.5:1 and thus \( R_s \) becomes:

\[
R_s = \frac{V}{ld} = \frac{V}{l \times 4.5}
\]

The shape of the cut sample also included the contact points to carry out the Hall measurements. The various other parameters are related as follows:

\[
\frac{1}{R_h} = \frac{\mu}{N e B I} \sigma p
\]

where \( B \) = magnetic field \( I \) = current
\( V_h \) = Hall voltage \( \mu \) = mobility
\( e \) = charge on an electron
\( N \) = carrier concentration

The current through the electromagnet gave a maximum field of 0.65 Tesla.

This technique was used to compare the electrical properties to the optical properties for the very thin metals and to determine the performance of the transparent conducting oxides. There was a requirement
for the films to have a sheet resistivity of \( \approx 200 \ \Omega/\square \) or less to make full use of the technique. Above this level the heating of the film could cause errors in the results limiting the reproducibility and accuracy of the technique. Thus preventing measurement of the mobility and carrier concentration.
7.17 References.


4). A.S.T.M. Index


8 The optical performance of dielectrics and thin metals.

The differing electronic structure of materials enables materials to be categorised as insulators, metals or semi-conductors. If the energy bands are filled and there is an energy band gap as part of the structure then when an external electric field is applied to the material it will not conduct and is then classed as an insulator. If the energy bands are incompletely filled then the electrons are free to move and the material will conduct and the material is classed as a metal. Between the two is the class of semi-conductors where either the energy band is virtually full or has only just started being filled and some conduction can take place, this is also temperature dependant and the semi-conductors become insulators at absolute zero.

The electrons in the insulators are referred to as being bound whilst in the metals they are referred to as free electrons. The bound electrons can be thought of as being small masses bound to an infinite mass (nucleus) by a spring which is analogous to a restoring force or damping factor. This means that under excitation the electron can vibrate but with the phase slightly altered. The free electrons are undamped and under excitation will oscillate but the sense of oscillation will be opposite to the driving force. This results in the reradiation of interfering waves which rapidly decay any refracted wave. This is shown schematically in Figure 8.1 which is a plot of the dielectric function $\varepsilon(\omega)$ versus the frequency ($\omega$). To the left of the dotted line where $\varepsilon$ is negative the light will be attenuated, this is seen as reflection as would be expected for metals in the visible region of the electromagnetic spectrum. To the right of the dotted line where $\varepsilon$ is positive the light is transmitted as would be expected for dielectrics.
Figure 8.1 The dielectric function of a free electron gas versus frequency in units of the plasma frequency. Electromagnetic waves only propagate when the dielectric function is +ve.

Figure 8.2 A schematic band diagram for a metal.
The absorption of energy by materials is by a quantum-mechanical process whereby a photon loses its energy to an electron which makes a transition to a higher level. There are two main types of transitions available, the intraband and interband transitions, figure 8.2 illustrates these.

The intraband transition is the optical excitation of an electron from below the Fermi energy to another state above the Fermi energy but within the same band. This is shown by line A in figure 8.2. Transitions are described mathematically by the free electron model as first put forward by Drude (1). As the transitions take place within the same band there is no threshold energy required for the transition to occur. It does not require the material to have only partially filled energy bands such that the electrons can make the transitions within the same band, this limits the material to being a Metal.

The interband transition is the optical excitation of an electron into another band as illustrated by line B on figure 8.2. This transition is known as a direct or vertical transition and requires a threshold energy in order to be able to occur. This threshold is analogous to the excitation of an electron across the band gap in an insulator. These transitions are described by the Lorentz model which characterises the contribution to the absorption by the bound electrons.

There is also a third contribution to the absorption which is often assumed to be unity, this is the lattice or high frequency dielectric constant.

8.1 The Lorentz model.

The electrons will vibrate at a natural or resonance frequency without any driving force, this resonance frequency is represented by \( \omega_0 \). In fact there is more
than one resonance frequency as shown in figure 8.3 there are three. The one in the visible / ultraviolet region of the electromagnetic spectrum is the one we shall assign \( \omega_0 \) to, the others are far enough away from the frequency range of interest that they can be thought of as a constant and included in the theory as part of \( \varepsilon_L \) the lattice dielectric constant.

If the material is thought of as having the small mass (electron) bound to the infinite mass (nucleus) by a spring then the effect of an electromagnetic wave on the electron oscillator can be explained by the Lorentz oscillator model.

The motion of the electron is described by:

\[
\frac{d^2r}{dt^2} + \frac{dr}{dt} + m\omega_0^2 r = -eE_{loc}
\]

where \( m \) = the electronic mass, \( e \) = the magnitude of the electronic charge, \( E_{loc} \) = the applied electric field acting as the driving force. The term \( m\Gamma (dr/dt) \) represents the damping force or energy loss due to radiation damping. The term \( m\omega_0^2 r \) is a Hooke's law restoring force. This model uses two assumptions, one is that the nucleus mass is infinite and the second is that it is assumed that there is no interaction between the electron and the magnetic component of the light wave. If the local field is taken to vary as \( e^{-i\omega t} \), then equation 1 can be solved (2) to give:

\[
\alpha(\omega) = \frac{e^2}{m (\omega_0^2 - \omega^2 - i\Gamma\omega)}
\]
Figure 8.3  Frequency dependence of contributions to the polarisability.

Figure 8.4  Dispersion curves.
Full line = the effect of damping neglected
Dotted line = damping included.
where \( \alpha(\omega) \) = the frequency dependant atomic polarizability which is complex due to the inclusion of the damping term which means that at any frequency the polarization differs in phase to the local field.

From Maxwell's equations (3) various parameters can be defined:

\[
\begin{align*}
    c & \quad c \\
    n = - & \quad v = --- \\
    v & \quad n = \sqrt{\varepsilon \mu}
\end{align*}
\]

where \( n = \) the refractive index, \( c = \) the velocity of light in vacuum, \( v = \) velocity, \( \mu = \) the magnetic permeability and \( \varepsilon = \) the dielectric constant. For non-magnetic materials \( \mu = 1 \). The materials that are of concern for optical filters are non-magnetic and the above equations can be simplified thus:

\[ n = \sqrt{\varepsilon} \quad \text{and} \quad \varepsilon = n^2 \]

The refractive index of a material that has some absorption is not quite so simple and the parameters \( n, \varepsilon \) and \( v \) are complex and of the form:

\[
\begin{align*}
    c & \quad c \\
    n = n - ik = - & \quad = \sqrt{\varepsilon} \\
    v & \quad \sqrt{\varepsilon}
\end{align*}
\]

where \( n = \) the real and \( k = \) the imaginary part of the refractive index with \( k \) being known as the extinction coefficient. Further developing the Lorentz model it can be shown that;
\[ N\alpha \]
\[ \varepsilon = \varepsilon_L + \frac{1}{\varepsilon_0} \]

where \( N \) = the number of atoms per unit volume. Using equation 2, equation 3 can be rewritten as:

\[ \frac{\varepsilon^2}{\varepsilon_L} \frac{1}{\varepsilon_0} \quad \varepsilon = \frac{\varepsilon}{\varepsilon_0} \quad \frac{1}{\varepsilon_0} \quad m\varepsilon_0 \quad \left( \omega_0^2 - \omega^2 \right) - i\gamma \]

Further use of Maxwell's equations can produce an expression for the interaction of light with a medium which defines the complex dielectric function as being related to the complex refractive index in the following way:

\[ \varepsilon = \varepsilon_1 + i\varepsilon_2 = n^2 \]

and \[ \varepsilon_1 = n^2 - k^2 \]

\[ \varepsilon_2 = 2nk \]

\( \varepsilon_1 \) and \( \varepsilon_2 \) are not independant quantities. They are also related to \( n \) and \( k \) as defined by the Kramers-Kronig dispersion equations (2). Using the separate dielectric functions and equation 4 we can write:

\[ \frac{\varepsilon_1}{\varepsilon_L} \frac{1}{\varepsilon_0} \quad m\varepsilon_0 \quad \left( \omega_0^2 - \omega^2 \right) \quad \left( \omega_0^2 - \omega^2 \right)^2 + r^2\omega^2 \]
As has been mentioned there is a third term that should be included in the complex dielectric constant and that is \( \varepsilon_1 \), the lattice constant. This is approximated in the above equations as unity but when separating out the component from measurements made on real filters it can be seen to be other than unity (4). The term \( \Gamma \omega \) refers to the region of anomalous dispersion, if we look at figure 8.3 it can be seen that there are three regions of dispersion. These arise from the orientation of a) the permanent dipoles (microwaves), b) the ionic lattice vibrations (infra-red) and c) the displacement of electrons (visible / ultraviolet). The region of interest is region (c), this is also shown in figure 8.4. If there is no damping of the system considered there will be a true singularity at \( \omega_0 \), the resonance frequency. As there is damping there will be a maximum and a minimum to the curve joined by a line with negative slope. This region with the negative slope is referred to as the region of anomalous dispersion and is characterised by rays of shorter wavelength being refracted less than those of longer wavelength giving a reversal of the conventional sequence of prismatic colours. Re-arranging equations 5a and 5b \( n \) and \( k \) can be expressed as follows:

\[
\varepsilon_2 = \frac{Ne^2}{2nk} \frac{\Gamma \omega}{m \varepsilon_0} \frac{(\omega_0^2 - \omega^2)^2 + r^2 \omega^2}{(\omega_0^2 - \omega^2)^2 + r^2 \omega^2}
\]

As mentioned, there is a third term that should be included in the complex dielectric constant, and this is \( \varepsilon_1 \), the lattice constant. This is approximated in the above equations as unity but when separating out the component from measurements made on real filters, it can be seen to be other than unity (4). The term \( \Gamma \omega \) refers to the region of anomalous dispersion. If we look at figure 8.3, it can be seen that there are three regions of dispersion. These arise from the orientation of a) permanent dipoles (microwaves), b) ionic lattice vibrations (infra-red), and c) the displacement of electrons (visible/ultraviolet). The region of interest is region (c), which is also shown in figure 8.4. If there is no damping of the system considered, there will be a true singularity at \( \omega_0 \), the resonance frequency. As there is damping, there will be a maximum and a minimum to the curve joined by a line with a negative slope. This region with the negative slope is referred to as the region of anomalous dispersion and is characterised by rays of shorter wavelength being refracted less than those of longer wavelength, giving a reversal of the conventional sequence of prismatic colours. Re-arranging equations 5a and 5b, \( n \) and \( k \) can be expressed as follows:
If a light is incident upon a surface at an angle θ the reflection and transmission of the light can be expressed using the Fresnel equations:

\[
\begin{align*}
\text{t}/ &= \frac{2n_1 \cos \theta_i}{n_2 \cos \theta_i + n_1 \cos \theta_t} \quad 8a \\
\text{t}_\perp &= \frac{2n_1 \cos \theta_i}{n_1 \cos \theta_i + n_2 \cos \theta_t} \quad 8b \\
\text{r}/ &= \frac{n_2 \cos \theta_i - n_1 \cos \theta_t}{n_2 \cos \theta_i + n_1 \cos \theta_t} \quad 8c \\
\text{r}_\perp &= \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t} \quad 8d
\end{align*}
\]

where \( n_1 \) = the refractive index of medium 1 on the incident side of the interface, \( n_2 \) = the refractive index of medium 2 on the transmission side of the interface, \( \theta_i \) = the angle of incidence of the light, \( \theta_t \) = the angle of refraction of the light in medium 2.
Snell's law can be written as:

\[ n_1 \sin \theta_i = n_2 \sin \theta_t \]  

Using equation 9 we can rewrite the Fresnel equations 8a-d and if the angle of incidence of the light is taken to be normal to the surface the equations can be further simplified to give;

\[
\frac{4n_1n_2}{(n_1 + n_2)^2} \quad T_{//} = T_\perp
\]

\[
\left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \quad R_{//} = R_\perp
\]

Where the films are absorbing and the refractive index has both a real and an imaginary component to it the reflection can be written in general terms in the following way;

\[
\frac{(n - 1)^2 + k^2}{(n + 1)^2 + k^2} \quad R_{//} = R_\perp
\]

Using these equations it is possible to see the contribution that \( \varepsilon_1, \varepsilon_2, n & k \) make to the resultant reflection (R) versus frequency (\( \omega \)) plot as shown schematically in figure 8.5.
The plot can be divided into four regions:

i) \( \omega < \omega_0 \) Transmitting

ii) \( \omega \approx \omega_0 \) Absorbing

iii) \( \omega > \omega_0 \) & \( \omega < \omega_p \) Reflecting

iv) \( \omega \gg \omega_0 \) & \( \omega > \omega_p \) Transmitting

In the regions (i) and (iv) \( \varepsilon_1 = n^2 - k^2 > 1 \) and \( \varepsilon_2 = 2nk \approx 0 \) and thus \( k \approx 0 \), \( n > 1 \) and \( \varepsilon_1 \approx n^2 \). This characterises the region as being transparent. Region (ii) is highly absorbing and may also have high reflectivity.

Region (iii) is characteristic of the response from free electrons and there is a metallic reflectance. For dielectrics this is well into the ultraviolet range such that region (i) is in the visible. The onset of region (iv) is defined as \( \omega > \omega_p \) where \( \omega_p \) is at \( \varepsilon_1 = 0 \) and is hence defined as:

\[
\omega_p^2 = \frac{N e^2}{\varepsilon_0 m \varepsilon_L}
\]

Semi-conductors have a band gap, once an electron gains enough energy to cross the gap it then has all the different levels in the next band to move in and so it then behaves like a free electron and can move throughout the crystal, its optical performance will then be similar to a metal.

In using the semi-conductor films as optical filters it can be seen that the electrical conductance becomes an important factor. The density of free carriers determines the transition point as given by the plasma frequency and the mobility as defined by the free carrier
relaxation time determines the sharpness of the transition.

Work done on conducting oxide films by Morris et al.(5) has shown the limit to the mobility to be dependent upon phonon or ionised impurity scattering. A model suggested by Anderson from work done on silicon (6) was adapted by Morris et al.. This is shown schematically in figure 8.6. The free electron density within the grain boundary (n2) is taken to be less than within the grain (n1) with the difference in concentration leading to a potential barrier of (q0) separated by the grain dimensions (d). In practice the barrier height was found to be small (< 1mV) and the variations in Hall mobility were fitted to a grain boundary barrier model (7). Indicating that due to the oxygen diffusing in, along and away from the grain boundary the grain boundary scattering was small. The scattering due to phonon and grain boundary scattering being small the impurity scattering was predominant. This implied that reducing the impurity scattering ought to be a priority rather than increasing the grain size as had initially been thought.

8.2 The Drude model.

For the metals the Lorentz equations can be simplified, this is because there is no restoring force term required. This gives two new equations for ε₁ and ε₂;
Figure 8.6
\[
\varepsilon_1 = \varepsilon_L \left( \frac{\omega_p^2 \tau^2}{1 - \frac{1}{1 + \omega^2 \tau^2}} \right)
\]

\[
\varepsilon_2 = \varepsilon_L \frac{\omega_p^2 \tau^2}{1 - \omega^2 \tau^2}
\]

\[\text{Equations 12a & 12b are the Drude equations and can be derived from the Lorentz equations by setting } \omega_0 = 0. \text{ The properties of the free electron metal are similar to those of an insulator at frequencies } (\omega) > \omega_0. \text{ The equivalent plots of } \varepsilon_1, \varepsilon_2, n & k \text{ and the resultant plot of the reflection } (R) \text{ versus the frequency } (\omega) \text{ are shown schematically in figure 8.7.}\]

\[\Gamma \text{ has been replaced by } \tau. \text{ } \Gamma \text{ can be shown to be equal to the relaxation time } (\tau^{-1}) \text{ which is the mean free time between collisions (when the collision length is } \approx \text{ excursion length). Equations 12a & 12b are the Drude equations and can be derived from the Lorentz equations by setting } \omega_0 = 0. \text{ The properties of the free electron metal are similar to those of an insulator at frequencies } (\omega) > \omega_0. \text{ The equivalent plots of } \varepsilon_1, \varepsilon_2, n & k \text{ and the resultant plot of the reflection } (R) \text{ versus the frequency } (\omega) \text{ are shown schematically in figure 8.7.}\]

The figures 8.5 and 8.7 represent the idealised behaviours of insulator and metals respectively.

If we look at the solar spectrum as depicted in figure 8.8 it can be seen that there could be a useful solar gain made in houses if the window could be made to be transmitting in the visible part of the electromagnetic spectrum and reflecting at the infra-red wavelengths. From the Lorentz and the Drude models it is apparent that thin metals of Gold (Au), Copper (Cu) or Silver (Ag) and some semi-conductors could have their plasma edge in an area that would approximate to this type of performance (8), data has been tabulated that confirms this (9). Early coatings used gold (10,11), the transparent conducting metal oxides came later and were initially developed for other purposes (12) although note had been made that they could be used as an alternative to the Gold films (13). The earliest reference to using
Fig. 8.7

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ N} \\
\text{T} & \text{ R}
\end{align*} \]

\[ \begin{align*}
\epsilon_1, \epsilon_2 \\
\epsilon_1, \epsilon_2
\end{align*} \]

\[ \begin{align*}
\text{R} & \text{ R} \\
\text{T} & \text{ N}
\end{align*} \]
Figure 8.8 A schematic of the ideal transmittance and reflectance performance.
the gold films also noted that there was an improvement in the transmittance of the coating by the addition of transparent oxide layers on either side of the metal layer (11), this was described as being analogous to the coatings that were used for the 'blooming' of photographic lenses. Slightly later the theory of induced transmittance was put forward explaining why the transmittance of absorbing films with high \( k/n \) ratios could be greatly enhanced (14). Of the available metals they specified Silver films as being particularly responsive. Thus from these papers all from the 1950's the basics of the 'heat mirror' type filters was already in existence and for very specialised uses already in production. These coatings became of greater interest in the 1970's with the possible applications of them for energy conservation applications (15,16). The applications of the filters are in domestic windows (17), in solar collectors and as an energy saving mechanism in tungsten filament light bulbs (18). The requirements of these different applications led to the development of different filters with different optimisations to match the requirements but all based upon the above optical principle and thus using the thin metals or wide band gap semi-conducting metal oxides.

8.3 Reflectance and transmittance of thin films.

If we imagine a parallel beam of light of unit amplitude and of wavelength \( \lambda \) which falls on a plane parallel sided, homogeneous, isotropic film of thickness \( d \) and refractive index \( n \). If the film is supported on a substrate of refractive index \( n_2 \) and the refractive index of the first medium is taken to be \( n_0 \) and at an angle of incidence \( \theta_0 \) the amplitudes of the
reflected and transmitted beam can be written in terms of the Fresnel equations as shown in equations 8a-d.
The amplitude of reflectance can be written as:

\[
\frac{\sqrt{R_1 + R_2 e^{-2i\phi}}}{1 + \sqrt{R_1 R_2 e^{-2i\phi}}}
\]

where \( R_1 = |r_1|^2 \) & \( R_2 = |r_2|^2 \)

\[
r = \frac{n_0 - n_1}{n_0 + n_1} \quad R = \left( \frac{n_0 - n_1}{n_0 + n_1} \right)^2
\]

\[
t = \frac{2n_0}{n_0 + n_1} \quad T = \left( \frac{2n_1 n_0}{n_0 + n_1} \right)^2
\]

and when \( n_1 \) is complex \( n_1 = n_1' - ik_1 \)

\[
r_1 = |r_1| e^{i\lambda} \quad r_2 = |r_2| e^{iB}
\]

\(|r|\) then becomes:

\[
|r| = \left( \frac{(n - n_1)^2 + k^2}{(n_0 + n_1)^2 + k^2} \right)^{\frac{1}{2}}
\]
and \( A \), the phase change on reflection becomes:

\[
2n_0k \\
A = \tan^{-1} \frac{2n_0k}{(n_1 - n_0)^2 - k_0^2}
\]

and similarly for \( B \) with \( n_1 \) and \( n_2 \).

When a film is absorbing then there is a change in amplitude and phase and it can be shown that:

\[
e^{-2i\delta} = e^{-4\pi k d} e^{-i4\pi n_1 d} = e^{\theta e^{i\gamma}}
\]

where

\[
\delta = i\beta + \gamma \quad \text{and} \quad \delta = \frac{2\pi nd}{\lambda}
\]

The equation 13 for \( r \) thus has to be modified if there is an absorbing layer included in the filter such that it includes the phase change at the interface as well as through the medium. To produce an equation for \( R \) we multiply by the complex conjugate and the final equation becomes:

\[
R = \frac{R_1 + 2\sqrt{R_1 R_2} \cos(C) e^{-\beta} + R_2 e^{-2\beta}}{1 + 2\sqrt{R_1 R_2} \cos(D) e^{-\beta} + R_1 R_2 e^{-2\beta}} \quad 14
\]

where \( C = B - A - \gamma \quad \text{and} \quad D = B + A - \gamma \)

In the particular case of a very thin metal coating, covered with a dielectric which was required to be antireflecting, the thickness of the dielectric was calculated as follows:

The system was taken to be; a beam travelling in air of refractive index \( n_0 = 1 \) was incident upon a dielectric film of refractive index \( n_1 = 2.5 \) & \( k_1 = 0 \).
which antireflects a silver film of refractive index \( n_2 = 0 \) & \( k_2 = 3.5 \).

For the reflectance to be a minimum when there is no absorption in the dielectric coating \( \cos(C) \) in equation 14 had to equal to \(-1\).

\[
C = B - A - \gamma = \frac{2n_0 k_1}{A = \tan^{-1} \left( \frac{(n_1 - n_0)^2 - k_1^2}{2n_0 k_1} \right)} = 0
\]

\[
A = \frac{2n_0 k_1}{(n_1 - n_0)^2 - k_1^2} = 0 \Rightarrow \pi
\]

Thus;

\[
C = B - \pi - \gamma = \cos^{-1}(t) = \pi
\]

\[
B - \gamma = 0 \Rightarrow B = \gamma
\]

\[
B = \tan^{-1} \left( \frac{2n_1 k_2}{(n_2 - n_1)^2 - k_2^2} \right) = \tan^{-1} \left( \frac{14}{6.25} \right) = 65^\circ
\]

As \( n_1 \) approaches the value of \( k_2 \) so \( B \) the phase change on reflection approaches 90 Degrees.

Thus the coating thickness has to also approximately equal 90 degrees to produce a total phase change of 180 degrees.

And

Thus \( \gamma = 90^\circ \) --- for light passing through the \( \lambda \) dielectric and back.
So for a dielectric coating of a refractive index \( n = 2.0 \) the thickness equals;

\[
d = \frac{\lambda}{\pi}
\]

Which is different to the usual antireflection coatings of thickness \( \lambda_\alpha \) (19).

8.4 The loss in performance due to scattering.

Deposited films are not perfect in structure, they are under stress and hence have a high density of dislocations, they are not completely dense and have grain boundary disorder and generally contain inclusions (often gaseous). This is highlighted in that the conductivity varies with many parameters including the coating thickness (20 - 24). It is thus hardly surprising that the coatings deposited fall short of the theoretical performance.

It is possible to derive a relationship between the mean polarisability \( \langle \alpha \rangle \), the dielectric constant \( \varepsilon \) and the number of molecules per unit volume (25). In its simplest form this has been attributed to Clausius-Mossotti (26-27).

\[
\frac{N\alpha}{3\varepsilon_0} = \frac{\varepsilon - 1}{\varepsilon + 2} = \frac{n^2 - 1}{n^2 + 2}
\]

If \( \varepsilon \) is taken to be the dielectric constant of the composite the expression can be expanded to take in the separate contributions of each of the components of the composite. The expression then becomes;
\[
\frac{\varepsilon - 1}{\varepsilon + 2} \frac{\varepsilon_a - 1}{\varepsilon_a + 2} \frac{\varepsilon_b - 1}{\varepsilon_b + 2} = f_a + f_b \quad 16
\]

This expression is the Lorentz–Lorenz effective medium expression (28,29). This expression makes the assumption that there is a vacuum as the host medium where \( \varepsilon = 1 \), this is not a practical assumption and the expression is better written in the form;

\[
\frac{\varepsilon - \varepsilon_h}{\varepsilon + 2\varepsilon_h} \frac{\varepsilon_a - \varepsilon_h}{\varepsilon_a + 2\varepsilon_h} \frac{\varepsilon_b - \varepsilon_h}{\varepsilon_b + 2\varepsilon_h} = f_a + f_b \quad 17
\]

This expression can be further modified, if there is a mixture of two materials of which it is not certain which is the host material and thus (\( f_a \)) and (\( f_b \)) are comparable it is possible to make \( \varepsilon = \varepsilon_h \) in which case equation 17 would become;

\[
\frac{\varepsilon_a - \varepsilon}{\varepsilon_a + 2\varepsilon} \frac{\varepsilon_b - \varepsilon}{\varepsilon_b + 2\varepsilon} = f_a + f_b \quad 18
\]

This expression is known as the Bruggeman equation or the effective medium approximation (30).

In the case of the thin film coatings that were deposited it seemed more appropriate to think of the films as made up of well packed spheres with a small amount of porosity. This eliminates the Bruggeman expression which takes the components to be comparable. If we take the Lorentz–Lorenz expression (equation 16) and take component (b) to be the dilute phase then we can make \( \varepsilon_h = \varepsilon_a \) which would result in a new expression;
Similarly if component (a) were the dilute phase then 
\( \varepsilon_a = \varepsilon_b \) could be inserted and a complimentary 
expression obtained;

\[
\frac{\varepsilon - \varepsilon_a}{\varepsilon + 2\varepsilon_a} = \frac{\varepsilon - \varepsilon_b}{\varepsilon + 2\varepsilon_b} \quad 19a
\]

\[
\frac{\varepsilon_b - 2\varepsilon_a}{\varepsilon + 2\varepsilon_a} = \frac{\varepsilon_b - 2\varepsilon_b}{\varepsilon + 2\varepsilon_b} \quad 19b
\]

These expressions (19a & 19b) make up the Maxwell Garnett effective medium equations (31). This expression 
was used to look at the modified film optical performance due to the film density being less than unity. A comparison was also made to the change in optical performance that would be associated with any change in the free electron density (32). This will be considered further in Chapter 9.
8.5 References.

      Ann.d.Phys. 3 (1900) 369

2). F.Wooten. 'Optical properties of solids' (1972)


7). L.A.Kazmerski 'Properties of polycrystalline and amorphous thin films and devices'


11). Anon. The Chemical Age (London) 67 (1952) 13


13). L.Holland & G.Siddall Vacuum 3 (1953) 375


18). J.Brett et al. I.E.S. Trans. (July 1980) 197


20). M.Faraday. Phil.Trans. 147 (1857) 145


22). C.E.Ells & G.D.Scott J.Appl.Phys. 23(1) (1952) 31


26). R. Clausius.  Mech. Warmetheorie. 2 (Braunschweig  
2nd Edn.) (1879) 62


Phil. Trans. R. Soc. Lond. 203 (1904) 385
Phil. Trans. R. Soc. Lond. Ser. A. 205 (1906) 237

13 (1986) 175
9. The deposition of coatings, results and discussion.

This can be broken down into various areas, the single layer transparent coatings, the thin metal coatings and the oxide-metal-oxide sandwich coatings and finally the electrochromic coatings.

The first category, the single layer transparent coatings, includes both conducting and non-conducting coatings. These coatings are invariably sputtered from an elemental target in a reactive atmosphere of argon and oxygen with the result that oxide coatings are produced. The oxide coatings that are conducting are useable by themselves for applications such as, electromagnetic shielding, heating panels and as heat reflecting filters (1-3). The conducting oxides as well as the non-conducting oxides can be used as seed layer coatings, protection coatings or as anti-reflecting coatings for other filters (4-6) or can be used in various combinations to produce other effects such as the electrochromic devices, the base layers of liquid crystal display panels or special filter stacks (7-9).

As such to be able to control the deposition of the oxides was every bit as important as the deposition of the thin metals but was often considerably more difficult (10).

9.1 Single layer transparent coatings.

Within the range of elements it is possible to use to reactively deposit oxide coatings it is inevitable that some will be easier than others. The techniques used to deposit the coatings were different for the most reactive targets compared to the more inert targets and as such will be considered separately. The transparent
conducting oxide coatings, as the most useful of the coatings, is worth considering first. This group of coatings generally have low melting point targets as the material source and are not classed as highly reactive.

9.2 The transparent conducting oxide coatings.

To date it has been found that the coating with the highest conductivity of all the semi-conducting oxide coatings is that of indium oxide (11). Any alloying of the oxide reduces the conductivity however it also increases the tolerances of the deposition parameters such that it is more easily made using equipment with less strict control of the critical parameters.

The equipment dedicated to the production of the indium oxide had a tolerance of $\approx 0.05\%$ control on both the reactive gas admission and the constant current supplied to the planar magnetron sputtering source. The essential features of the process were the control of the current and the gas supply and the monitoring of the coated polyester soon after the deposition zone such that the partial pressure of the reactive gas could be optimised to give the best conductivity to the deposited coating.

The reasons for the partial pressure of the reactive gas being so critical are firstly to produce the correct stoichiometry in the deposited film and secondly to minimise the contamination of the target surface such that the deposition rate was maximised. The process started with the sputtering of pure metal and the deposition of pure metal. The condensing metal is energetic and reactive, also the target is clean and highly reactive. As the reactive gas is introduced it will react with both the condensing metal and the metal
target surface. If the oxygen flow is increased so the oxidation of the surfaces increases to a point where the production of pure metal is less than the supply of oxygen. When this point is reached the pressure in the system will rise corresponding to the oxygen excess. This can be a problem, the sputtering rate of oxides is lower than that of metals (12). Once the oxygen partially oxidises the target there would be a proportion of metal sputtered and a proportion of oxide sputtered, the result was a reduction in the oxygen requirement. As the flow had been increased there was an increase in the partial pressure i.e. excess oxygen which would further oxidise the target and further reduce the requirement and thus a runaway process was set in motion. In the case of a lower reactivity target this could occur in anything from a few seconds to many tens of seconds. The argon was fed into the system as a fixed flow and the total pressure was monitored, if a small excess flow of oxygen was induced there was an increase in the partial pressure and hence an increase in the total pressure. If this was taken to be the 'mean' signal it could be fed into the flow controller electronics as a feedback signal and any deviation from the mean value would then produce a response in the oxygen flow. Hence if the total pressure increased due to the partial oxidation of the target there would be a reduction in the oxygen mass flow to compensate thus limiting the runaway effect. Similarly if the target oxidation were to be reduced then the oxygen utilisation would increase and the pressure drop which would then measured would prompt the response of the oxygen flow being increased. There were controls on the electronics of the mass flow controllers to prevent oscillation of the response. The response of the system was governed by the sensitivity of the pressure gauge, its position in the chamber and the
distance that the mass flow meters and valves were away from the gas inlet and the reactive zone within the chamber. All these factors gave a time constant to the system, if the time constant was smaller than the speed in which the runaway effect could occur then the reactive process could be controlled effectively. If the time constant were longer than the speed of the runaway effect then sooner or later the target would be fully oxidised and control of the process would be lost. The conducting oxides were slower to react than the time constant of the system and the process could be controlled, this was not so with some of the highly reactive non-conducting oxides as will be seen later. To improve the sensitivity the pressure gauge signal was amplified and offset to give the pressure control of \( \approx 0.05\% \). The other half of the control requirement was the magnetron supply, this was required to have the same degree of control as the gas flow. If we look at the voltage versus the current characteristic as shown schematically in figure 9.1 (13) it can be seen that for much of the curve there was a nearly linear relation between the voltage and current. For a small change in the voltage there was a large current change. It was decided that to have the best control of the supply the current should be kept constant and power supplies were built to have this facility. There was also a linear relationship between the current and the deposition rate (14) which was useful in predicting the coating thickness.

To further discuss the reactive deposition of coatings it is worth using a figure published by Schiller et al. (15) and reproduced here in figure 9.2. The graph plots various parameters against the partial pressure of oxygen with the current held constant. It can be seen that for an increase in oxygen partial pressure there was an increase in the magnetron
Figure 9.1

The voltage / current characteristics shown for three metals. The system was operated at a pressure of 0.55 Pa which corresponds to an argon flow of 28 sccm.
Figure 9.2
voltage, an increase in the refractive index and a decrease in the absorption index and the condensation rate. The stoichiometry that was desired was at the peak in the refractive index and minimum in the absorption index. This does not give the maximum deposition rate i.e. that of the metal, but it was significantly higher than that from a wholly oxidised target. The plot is typical of most reactive deposition processes, although the one depicted is particularly severe. In terms of the oxygen partial pressure it can be seen how critical it was to have full control of it, the slightest deviation from the mean and the rate changed dramatically as did the coating properties. An additional problem was that if the deviation from the mean was large due to an excess partial pressure of oxygen then the voltage would pass over a peak and then drop reversing the characteristic of the voltage slope. If the voltage or total power was used for the control parameter (15) the reversal of the voltage characteristic would be disastrous with the response preventing recovery back to the correct operating conditions. In terms of control of the process the conductivity of the coating was far more sensitive than any of the other properties of the coatings. The conductivity could change by a few orders of magnitude for very small changes in the oxygen flow. Thus for full control to deposit the conducting oxides it is necessary to use the conductivity as a feedback to establish the mean partial pressure setting and then the total pressure to control the oxygen flow about this mean with the current fixed as a constant.

With a target current of 1A and a voltage of 600V with a gas flow of 70 sccm of argon and 18 sccm of oxygen a conducting transparent indium oxide coating of refractive index of 1.95 and thickness of 14.7 nm was
produced with a resistivity of $4.7 \times 10^{-6}$ $\Omega$m as a matter of routine.

If the current was fixed and the oxygen flow increased a graph of the voltage versus the oxygen flow could be plotted. This is shown in figure 9.2 where it can be seen that with increasing oxygen flow there was a voltage peak followed by a voltage drop. If the oxygen flow were then to be reduced in a similar series of steps it can be seen that the voltage does not follow the same characteristic but produces a hysteresis loop. A typical one is shown in figure 9.3 which is for a chromium target. This further highlights the need to prevent the target from being oxidized over, recovery of the target may lead to a different deposition rate for the same current and oxygen partial pressure. Further to this there was a time dependance also related to the target reactivity, figure 9.4 shows two hysteresis loops for the same target, with the same current and same range of oxygen flows the difference being in the change in time between alterations in the oxygen flow. As could be predicted from the difference in sputtering rates between the metal and the oxide the target is fast to oxidize but much slower to sputter the oxide clear and reach an equilibrium state. Successive quick increases of oxygen oxidised the target at a flow much lower than was obtained by a much slower increase in the oxygen flow. This was a further reason for not using the voltage as a source of control. It was particularly noticeable when depositing tin oxide in that for the second attempt to produce a conducting tin oxide the oxygen flow was quite quickly brought up to just below the pressure where the conducting oxide had been found during the first experiment. As the oxygen was then slowly increased through the expected critical zone (or 'window') there was no evidence of any conducting oxide being produced.
Figure 9.3

A hysteresis loop caused by the different sputtering rates of the metal and the oxide.
Figure 9.4 Two hysteresis loops showing the rate of change dependance.
With hindsight it was apparent I had already passed the 'window' of conditions simply by introducing the oxygen at too fast a rate. In fact there were a number of lessons to be learned from the deposition of the conducting tin oxide coatings.

The production of transparent conducting oxides has been well established (16-18), a number of the references referred to transparent conducting tin oxides (19-21). There had been a number of attempts within our research group to reproduce these results but without having to resort to either using hot substrates or to having to carry out an annealing process after deposition. These initial attempts failed. It was easy to produce highly transparent tin oxide but invariably it was non-conducting. The optical performance was acceptable for use as part of a thin metal heat reflecting type filter but there was concern regarding the reproducibility of the coating especially as there was not a sharply defined parameter to control the process by. One solution used commercially has been to sputter the tin in a 100% oxygen atmosphere, this was not thought suitable because the deposition rate would be slower than could be achieved using a mixed atmosphere and sputtering from a partially metallic target. Also from a safety aspect using a high proportion of oxygen is not advised due to its possible reactivity with the pumping oil, an inert gas dilutant would be required in the backing line and this was rejected as an option. The experience of depositing indium oxide led to the belief that if we could produce a conducting tin oxide we would be able to use the conductivity for the very fine control of the deposition process. The technique for establishing the deposition parameters for a 'new' oxide was to start by sputtering the metal and slowly in small steps increase the oxygen flow to a point where there stopped being any
change in the observed coating. In the case of the indium oxide the resistivity fell to a minimum and then rose again and a level of oxygen flow was reached where the resistance was beyond the standard measuring equipment (>20MΩ on a digital voltmeter (DVM)) and as the area of interest had been passed the experiment was stopped. This had been done with the tin oxide to little or no effect, there had been no drop in the resistivity and the optical measurement of transmittance had remained constant over a large range of oxygen flows. It transpires that there were several reasons why I had failed to find the conducting oxide. One was the expectation that the required oxygen flow would be similar to that required to produce indium oxide. Another was that the target history was important and finally the conditions where the oxide was to be found was very much more sharply defined than the indium oxide. The size of steps of the increase of oxygen flow coupled with the hysteresis meant that the optimum conditions were passed in one step and the resistivity was too high on either side to see any variation using the standard DVM. The width of this 'window' of oxygen flow was approximately 2 sccm in a total flow of 100 sccm. Also there was a chance that in recording other parameters the short time that the resistivity was within range and changing rapidly was missed. Having collated the previous information on tin oxide and gained further experience in depositing other oxides the second attempt was approached with a much more thorough method. The general approach was similar, the oxygen flow was increased in a series of steps as before albeit smaller ones, this time however no upper limit was set on the oxygen to argon ratio and less attention was paid to the measurement of transmittance, most of the attention being concentrated upon the resistivity. This approach was
more successful and a slightly conducting transparent coating was produced. Repeating the coating was almost as difficult as finding it in the first place, it was found that the target 'conditioning' or 'history' was important and the target ages affecting the position of the 'window' of required conditions. This aging is as a result of a change in the proportions of oxide and metal on the surface leading to a difference in the proportion of oxygen required and a slow change in the thickness of the surface oxide layer. This was seen as as the older the target was the higher the oxygen flow that was required. This was also true during the deposition that as the time of the deposition increased so to did the requirement for oxygen. This could not be explained away as purely changes in the surface area of the target or changing proportions of oxide to metal on the surface. This change in oxygen flow was as much as from $<20\%$ up to $>50\%$ during one deposition. This change in the required oxygen flow was matched by a change in the resistivity it was possible to achieve. The resistivity was minimised at the start of the deposition procedure and as the target aged and required more oxygen it had to be minimised again. This second minimum was lower than the first and the third was lower than the second and this would continue throughout the deposition process which would take place over a period of several hours. One change that would be consistent over a long period of time was the rising temperature of the roll coater during the deposition process. The roll coater had no water cooling and so over a period of time the whole assembly heated up, this could be to a temperature of $>500\, ^\circ\text{C}$ over a deposition time of $>8$ hours. This is a temperature rise of $>30\, ^\circ\text{C}$ which is not an insignificant amount. This opens up a field of further work that is required which is to optimise the conductivity of conducting oxide
films at different carefully controlled substrate temperatures. In retrospect there was another indication that there was a significant change in the deposition process and that was from the chart recording of the total pressure. The total pressure was monitored by a pirani gauge, the output of which was amplified and offset to give a 0 - 1V signal that had a resolution of better than 0.1% (to match the power supply performance). This can be better seen in figure 9.5, the oxygen flow has been taken up to 45 sccm and the 'window' of conditions where the film was conducting had been missed and the chart shows the total pressure output as the oxygen flow is reduced. In each case as the flow is reduced, after the initial pressure drop, the pressure rises. This occurs until we reach a flow of 17.5 sccm where after the initial pressure drop there is a further slower fall in the pressure, this is also the case down at 14.9 sccm and as the flow was again increased at 16.0 and 17.0 sccm. When the flow was further increased to 18.0 sccm the slope once again reverses and after the initial pressure rise there was once again a slow pressure rise. When collating the information after the experiment it became apparent that the region where the film showed signs of conductivity was where this turning point in the total pressure occurred which was related to the way the oxygen affects the target. Once some experience of producing the conducting form of the oxide was obtained methods of speeding up the optimisation of conditions was sought. The most productive modification was instead of measuring the resistance of the film using the hand held multimeter was to connect the probes to an industrial insulation testing meter (known as a megaohm meter). The change of resistance was over several orders of magnitude for only a 1 or 2 sccm change in oxygen flow, much of it was
Figure 9.5
A graph of the output from the Pirani gauge of the system total pressure. The region of interest is where the pressure is constant following a change in the oxygen input. This is near the change 16-17 sccm.
off scale when using the DVM and by replacing it with the megaohm meter effectively widened the 'window' of conditions over which the deposition could be controlled. Had this meter been used initially the deposition conditions where the coating was conducting would probably not have been missed. Once an approximate minimum had been obtained using the megaohm meter the DVM with its better resolution could be reinstalled and fine tuning of the coating was then possible. The next step was to reproduce the conducting oxide in other vacuum systems, this was done both within our own laboratory but also was successful when following publication (22) a commercial concern was able to, for the first time, successfully produce transparent conducting tin oxide. The initial films produced had a sheet resistivity of 2,800 Ω/□, when the process was transferred within our laboratory films of 400 Ω/□ were produced and then the company obtained films of 200 Ω/□. This is not as low as obtained by using hot substrates or an annealing process but it is getting closer all the time, the use of a heated drum such that the polyester could be held at a slightly elevated temperature may well be enough to match the best films produced on glass (23). Other observations on the deposition of the tin oxide is that on the early films there was a trend that with increasing conductivity there was increasing stress and that the colour of the films went from a deep brown to a very vivid yellow to a pale brown to a fully transparent film with increasing oxygen flow from the oxygen deficient side. The refractive index is approximately 2.10. Analysis of the conducting films was done by the Tin Research Institute using Mossbauer spectroscopy and it was reported that the films were of SnO2 with no other oxides present. The difference in the oxygen content between different coloured samples was too small to be
measured by this technique. This was not a problem the difference in resistivity was enough to control the production of different coloured films. The requirement for the oxide-metal-oxide coatings only required the oxide to have a high refractive index and to be transparent throughout the visible with minimum absorption. Further analysis using the electron microprobe microanalysis machine did show an impurity content, this was \(<1\%\) of iron. This was not unlikely, the pole pieces of the magnetron are mild steel and this could easily be a source of contamination.

To control the thickness of the coatings it is common to find a quartz crystal monitor placed in the deposition zone, it is particularly useful in the case of evaporation where the volume of material in the source changes with time and for constant power the evaporation rate also changes. Magnetron sputtering is different, the sputtering rate being a linear function of the current supplied to the magnetron makes the condensation rate highly predictable. However in the reactive mode where the proportion of metal to oxide on the target surface can change it can be useful to have an alternative technique to determine the film growth rate. Figure 9.6 shows the hysteresis loop for the tin oxide and above it is the response from the quartz crystal monitor. The monitor works by oscillating a crystal and monitoring the change in frequency as the coating changes the crystal mass, this is compared to a reference crystal. The system had to be calibrated which in itself presented a problem, as its proposed use was to monitor the change from metal to oxide deposition, the masses of the metal and oxide were different and as such the calibration was different. As it was impossible to accurately calibrate the frequency to deposited thickness it resulted in simply plotting the rate of frequency
The hysteresis loop for both voltage and condensation rate from the rate monitor.

Figure 9.6

The hysteresis loop for both voltage and from the rate monitor.
change versus the oxygen flow. Over the critical range the rate of frequency change mirrors the hysteresis loop but at the low powers used the resolution of the rate of frequency change was worse than could be obtained by monitoring the voltage. Also the sampling time gave a lag to the response which further prevented the response being used to control the process.

The transmission monitors were also of limited use for the deposition of this type of coating. Figure 9.7 shows the visible transmittance of three films with resistivities measured during deposition of 30kΩ, 300kΩ and 10MΩ. The variation in resistivity all occurs with the oxygen to argon ratio being held constant highlighting the need for some form of feedback control to maintain the optimum conditions. In terms of the transmittance the light emitting diodes (LEDs) only showed a 3% change between the highest and lowest resistivity coatings. The wavelength of each of the LEDs was, Green = 565nm, Yellow = 583nm, Red = 635nm and the Infra-red = 940nm. All of these wavelengths were too far away from the large change which occurred in the blue region of the visible spectrum. The white light source with the monochromator used as a transmission monitor was little better, the white light had little output in the blue and the silicon detector response fell off rapidly in the blue such that the information sought was always in the least sensitive region. Since this work a blue LED has been obtained and it may possibly improve the sensitivity, although it is nearer in wavelength to the region of interest the light output of the LED is much less than the others and so may prove to be no more sensitive.

With the shortcomings of the other monitoring techniques the one left with the greatest sensitivity was still the resistance monitor. This monitor which when
Figure 9.7 The graph shows the rapid change of resistance of the oxide coatings for a relative small change in the optical properties with the gas flow held constant.
used in conjunction with the megaohm meter as well as the DVM gave over ten orders of magnitude range in which to look for changes of resistance during the deposition process.

Indium oxide was easier than the tin oxide having a wider 'window' of deposition conditions that could be used to produce a conducting coating and less dependance on the target 'history' or system temperature. Once obtained the conditions were reproducible in other systems. In being reproducible such that they could be used as dielectric layers for the oxide-metal-oxide heat reflecting type filters or in the case of the highly conducting indium oxide as the conducting layer for the electrochromic devices.

9.3 The transparent non-conducting oxide coatings.

Non-conducting oxide coatings can present further problems in that monitoring the resistivity is no help in producing optimised coatings. It was necessary for this class of coatings to be optimised from information obtained after the polyester had been removed from the vacuum system. This was inevitably less satisfactory than being able to optimise the coatings during the production process.

There was some interest in the deposition of chromium oxide onto polyester and this was attempted. Initially the voltage versus oxygen flow hysteresis loop was produced as shown in figure 9.3, from this a range of oxygen flows were chosen likely to produce films of interest. The area of interest in other films was over the range of flows that produced a voltage that approached the peak voltage from the 'metal' side, from figure 9.3 this would be over the range 7 - 9 sccm of
oxygen. The coatings having been produced were then examined and any variation in properties was compared to the changes in the deposition conditions in an attempt to isolate a controlling parameter. In this case there was nothing that could be pinpointed as critical however it became apparent that there was a variation in the optical appearance of the films relating to the winding speed of the polyester. This was not as pronounced as the oxygen deficient tin oxide that was a very vivid yellow but it was easily noticeable. The films varied from a dark brown to a pale metallic brown, with it possible to produce the pale metallic brown films even after the target had been oxidised and where there was a considerable excess of oxygen present. Bearing in mind the apparent connection with the rolling speed it was thought that possibly an impurity from the polyester was creating the problem. The most likely impurity being one of those trapped there when the film was wound. As the winding was done in air the most likely impurities were nitrogen or water. From other deposition experiments which used nitrogen as the reactive gas it was thought unlikely that the nitrogen would react preferentially to the oxygen which left the water as the most likely impurity. Hence it was thought possible that there was the possibility of some hydroxide being formed in the growing films. The amount formed would largely depend upon the the rate of arrival of the water relative to the rate of arrival of the depositing chromium. Thus it was thought possible that the content would vary up to a limiting percentage where no matter how much more water was present no additional hydroxide would be formed in preference to the chromium oxide. To look for this type of information samples were taken and examined using the x-ray photoelectron spectroscopy surface analytical machine (XPS).
The XPS analysis was carried out using a V.G. Escalabs spectrometer. The experimental conditions used were; an x-ray source of aluminium Kα radiation at 1486.6eV, with an anode voltage of 15kV and an anode current of 20mA over an analysis area of 0.4 cm² and the emission angle normal to a surface giving a response of up to 10⁵ counts per second. The ion beam parameters for etching the surface were, argon ions of 5keV and an ion current of 20μA at a gas pressure of 10⁻³ Pa.

The coating being non-conducting and the polyester also being non-conducting the first problem encountered was an odd scatter to the relative intensities of the elements in the broad XPS scan. This was found to be due to the charging up of the sample surface distorting the measured kinetic energy. Figure 9.8 shows a plot of this, with the instrument set to measure the response of one of the peaks the graph shows the change in response with time. It can be seen that it takes of the order of 10 minutes to produce a stable response. In order to obtain accurate information from the scans taken it was necessary to allow the equilibrium to be reached before the measurements were taken. A broad scan XPS spectrum was taken of each of the samples and the relative peak heights were measured and compared to a Johnson Matthey Specpure standard sample of Cr₂O₃. Table 9.1 shows the atomic percentages of carbon, chromium and oxygen obtained from a series of six coatings deposited onto polyester at different rolling speeds. The position of the chromium 2p³/₂ and oxygen 1s peaks were measured relative to the carbon 1s peak at 284.6eV and these positions were also compared to those found in the standard sample. The oxygen/chromium ratio was calculated from the atomic percentage composition and it can be seen that there was an excess oxygen, the standard sample gave a ratio of 1.4(9) whereas all the others
### Table 9.1: The percentage composition of the Cr$_2$O$_3$ and O/Cr ratio.

<table>
<thead>
<tr>
<th>Sample No.</th>
<th>C</th>
<th>O</th>
<th>Cr</th>
<th>O/Cr</th>
<th>Speed (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>51116</td>
<td>38.6</td>
<td>39.4</td>
<td>22.0</td>
<td>1.79</td>
<td>0.025</td>
</tr>
<tr>
<td>51099</td>
<td>33.7</td>
<td>41.8</td>
<td>24.5</td>
<td>1.71</td>
<td>0.050</td>
</tr>
<tr>
<td>50940</td>
<td>26.8</td>
<td>44.4</td>
<td>28.8</td>
<td>1.54</td>
<td>0.225</td>
</tr>
<tr>
<td>50418</td>
<td>28.2</td>
<td>45.0</td>
<td>26.8</td>
<td>1.68</td>
<td>0.126</td>
</tr>
<tr>
<td>50940</td>
<td>27.1</td>
<td>45.2</td>
<td>27.7</td>
<td>1.63</td>
<td>0.225</td>
</tr>
<tr>
<td>50550</td>
<td>28.2</td>
<td>44.8</td>
<td>27.0</td>
<td>1.67</td>
<td>0.127</td>
</tr>
<tr>
<td>50852</td>
<td>25.9</td>
<td>45.6</td>
<td>28.6</td>
<td>1.59</td>
<td>0.117</td>
</tr>
<tr>
<td>50400</td>
<td>28.8</td>
<td>45.0</td>
<td>26.2</td>
<td>1.72</td>
<td>0.075</td>
</tr>
</tbody>
</table>
Figure 9.8 Time (minutes)

Evidence of the sample charging during the XPS analysis.
were substantially higher. One possibility was that the oxygen was bonded to the carbon on the surface of the sample. Figure 9.9 shows a plot of the carbon atomic percentage versus the oxygen/chromium ratio. At 0% carbon if the coatings were stoichiometric the oxygen / chromium ratio would be 1.5. As the graph is nowhere near passing through that point it was assumed that the chromium oxide was not stoichiometric and that the oxygen had to be bonded in another form as well and not just to the surface carbon. It was known that the oxygen could be found bound as the hydroxide. To find any evidence of it in such a form would confirm the hypothesis that some had been formed in the deposition process. High resolution scans were taken of the samples to look at the symmetry and the position of the peaks. As can be seen in Table 9.1 the positions of the peaks were slightly shifted and the peaks also appeared skewed on the high resolution scans, the reason for the asymmetry could have been two overlapping peaks. The position of each peak was taken relative to the carbon 1s peak. The oxygen peak was plotted and the position where the hydroxide would occur was added, the peak was digitised and a curve stripping routine used to separate out the peaks. If there was a contribution from a second peak then subtracting the second peak would return the main peak to full symmetry. This was done and the removal of the second symmetric peak did indeed produce a correctly positioned and symmetric peak thus indicating the oxygen was present in the form of the hydroxide. This technique was not really suitable for producing highly accurate values for the percentage contribution of each of the components however it was estimated that there could have been as much as 30% hydroxide present in the worst case. It was not possible to use the curve stripping to correlate the proportion of hydroxide to the
rolling speed of the polyester during the deposition process. To further try to establish the hydroxide content a standard sample of chromium hydroxide was sought, unfortunately it appeared to be unobtainable and so one was made. Some chromium was dissolved into solution and then precipitated back out and then thoroughly washed. The sample produced an oxygen/chromium ratio of 2.8(8) which was encouraging, the sample was not completely pure, there was some residual salts left from the Soxlet extractor water supply which had been used to wash the chromium hydroxide. Apart from confirming the hydroxide peak position and proving that the chromium hydroxide could be produced this exercise proved to be of little help.

Other coatings sensitive to water in the wound polyester have since been produced, so far no solution has been found to control the incorporation of water in the films other than to wind the polyester through once whilst under vacuum prior to the deposition process to remove as much water as possible. This was fine as a laboratory process but as a commercial process would be expensive. Instead of a separate process it would be possible to have additional pumping and baffling such that the water removal and the deposition process could occur during the same cycle but it too would add to the cost because of the additional pumping. However if these coatings are required to obtain the best coatings possible this type of preconditioning would have to be included.

Of the non-conducting oxides one of the most important was titanium dioxide. This was because of its high refractive index (=2.4 at 633nm onto polyester and =2.5 at 633nm onto glass) which when used in conjunction with a low refractive index material was the basis of the manufacture of many types of multilayer filter (24,25).
The heat mirror coatings described by Fan et al. (26) utilised its high refractive index as an optical match for the thin silver films to optimise the transmittance in the visible. Unfortunately the titanium target was one of the most reactive ones, which was not surprising as titanium is used as a getter material in some types of vacuum pump (27). This high reactivity was a severe limitation in the production of a reproducible high refractive index material in long lengths onto the polyester substrates. The graph shown in figure 9.2 shows the desired region of operation of the deposition process, what cannot be shown is the difficulty in maintaining the stability of the process in that region. When operating at the correct conditions, which again could only be set with reference to previous depositions and the resultant coatings, any random change in the operating pressure would change the proportion of oxidation of the target. This was exactly the same process as in the other materials except in the case of these highly reactive targets the change occurred very very rapidly (<<1s). The target would runaway and 'lock-on' to either its metallic or its oxide state, neither of which was of any use. As these processes could occur faster than any of the process parameters could be manipulated a controlled disruption was introduced. The two 'guaranteed' set of conditions were the metallic target or the oxide target. If the conditions were varied such that the target always returned to one of these states some degree of control could be obtained. The fully oxidised target was not suitable because of the much slower sputtering rate of the oxide compared to that of the metal thus the target was always brought to the metallic state. The method of doing this was to pulse the reactive gas input. There were several ways that this was done, one simple way was
allow the controls on the mass flow controllers to be adjusted such that they did not damp the response of the reactive gas to the pressure change. By doing this any change of the reactive gas flow would overshoot the desired point and have to be corrected by making a further adjustment in the opposite sense which again would overshoot and thus an oscillation could be set up.

This was not really controlled in either magnitude or frequency, it only worked because of being brought back to some base point i.e. the metallic target state. A more precise version of this was to derive the control signal not from the total pressure but from the oxygen partial pressure as could be obtained from a residual gas analyser (RGA). The time lag between the response of the RGA and the change in oxygen flow gave the frequency and the controls on the mass flow controller could limit the magnitude. This was an improvement but to maintain control it also required to have the target brought back to the metallic state. The most successful technique used was to fix the flow of both the argon and the oxygen with the oxygen input being controlled through a diverter valve. A signal from a square wave signal generator operated the diverter valve such that the oxygen was fed either onto the substrate by the conventional gas ring or into the throat of the pump. The use of the diverter valve was to keep the pumping speed constant by keeping the gas load constant but selectively controlling the activity of the oxygen. The diverter valve was operated at a frequency \( \approx 1 \) Hz. This technique allowed an oscillation about any part of the voltage versus oxygen flow curve to be followed and if the system had been calibrated, with the voltage of the target in the metal and oxide state being known, the mean position of the oscillations would accurately give the properties that could be expected from the deposited coating. This
A technique was used to some effect by others in the deposition of nitride coatings (28). It was the best of the techniques available, the others required complex pumping or baffling which limited the deposition rate or required to be sputtered from an oxide target such as by radio frequency sputtering. These other techniques were either slow or could not produce the highest refractive index coatings (29). Work done by Suzuki (30) suggested that working from a titanium target in a reactive deposition mode was preferable to working from an oxidised target. Figure 9.10 is a schematic of the voltage variation versus the oxygen flow and the variation of the voltage versus time as the gas was pulsed. The voltage variation versus time is shown for three different regions of the voltage versus partial pressure curve. On the first curve the target was too metallic, the 'hiccup' on the curve is where the gas flow switching signal has been superimposed on the oscilloscope trace. The second curve was where the conditions were just on the oxide side of the optimum, as can be seen the curve has changed from concave to convex, at the optimum point the curve was virtually of the form of a sine wave. The third curve was when the target was well oxidised, the voltage change was reduced and the response of the voltage was slower than in the other curves.

Although this technique worked and the high refractive index titanium dioxide (rutile) coatings could be produced the difficulties of the process prevented it from being widely used in the production of heat reflecting filters. Even with the less reactive targets the conditions generally chosen make the process less of a reactive one and more a matter of sputtering an oxidised target. The reactive gas being present to correct the stoichiometry of the condensing coating
Figure 9.10 A schematic of the gas pulsing technique
should the sputtering process cause any decomposition of the oxide sputtered from the oxidised surface.
9.4 The thin metal and oxide-metal-oxide filters.

Ever since early work done by Faraday (31) there has been interest in the properties and performance of metals deposited as thin layers such that they are transparent. There have been four materials regularly cited as being of particular interest (32-36), these are gold, silver, copper and aluminium. The combination of high visible transmittance and high infra-red reflectance can be found in the Drude type materials where there is a suitably positioned reflectance edge. Filters with this type of performance are known as heat mirror type filters. The earliest ones were generally made using thin gold coatings (37) the gold being inert but soft. Improvements were made by predepositing an oxide layer and also postdepositing a top oxide layer (38). The purpose of sandwiching the thin gold with an oxide layer (usually bismuth oxide) was for improved adhesion, to protect the soft gold and to enhance the transmittance. It had been observed that the resistivity of gold films of the same thickness were different and dependant upon the substrate. If an oxide coating was deposited onto the substrate prior to depositing the gold the conductivity of the gold film was improved (39) or the minimum thickness of a conducting film was lowered. This was also observed to be true for sputtered films when compared to evaporated films (40,41). This improvement being attributed to the change in the relative surface energies in the first case and the increase in the deposition energy in the latter case. The gold films were soft and the harder top oxide protected the films from damage. It was observed that the top oxide coating also improved the transmittance of the film. An early reference to a reflectance reducing effect was by Lord Rayleigh (42) and not long afterwards further comment was
made on the phenomenon by Taylor (43) who also developed a technique to artificially tarnish his lenses to produce the effect. It was however much later that the true nature of the effect was discovered (44,45). One report on the work of Preston (46) referred to the effect of the top oxide coating on the top of the gold as being analogous to the 'blooming of photographic lenses'. The top oxide coating was named after its function, an antireflecting coating.

The use of the heat reflecting filters was initially in small areas for use in windows for furnaces however gold films or substantially gold coatings were later developed for architectural glazing units (47). These units were meant to be made with high transmittance, heat reflecting coatings but there had been a step backwards in that the gold or gold alloy coating directly onto the glass and the top coating was only used for protection. The alloying of the gold was to promote improved adhesion. In the 1970's came a burst of activity in research into ways of saving energy. There were two areas of interest, the main one was for the coating of glass for domestic glazing (48,49) and the second one was the coating of lamps to reduce the energy requirement without loosing any light output (50,51).

The possibility of saving some of the energy put into illumination had been recognised by Hoffman (52) as early as 1922 and as only $\approx 10\%$ of the energy input was seen as visible light output it was surprising that little appears to have been done about it for a further fifty years. Gold and copper were used as a heat reflecting coating for sodium lamps with a zinc sulphide antireflecting coating (53) in 1965 but it was not until work on utilising solar energy promoted the use of silver thin films sandwiched between high refractive index oxides that the tungsten filament lamps could be improved
The requirements of the coating for this application are much more stringent than for the window applications and are worth further consideration. Figures 9.11 and 9.12 refer to the theoretical transmittance of the four metals gold, copper, aluminium and silver in the thin film form (55,56). The films are calculated to have a common sheet resistivity of $1\Omega/\square$ and to be deposited onto polyester sheet 50$\mu$m thick with a refractive index of 1.7. The use of a common sheet resistance is a convenient way of giving the films an equivalent performance in the infra-red where the resistivity is related to the infra-red reflectivity and hence the emissivity. It can be seen in figure 9.11 that the peak transmittance for the copper and gold lie well into the visible region towards the red whilst the silver peaks at the blue end and into the ultra-violet. The figure highlights the poor performance of the aluminium, the reason for this was the relatively high value of the real part of the refractive index ($n$) giving a high adsorption. In a paper on induced transmission Berning and Turner established that the performance of the thin film was related to the $k/n$ ratio and that for the highest transmission the $k/n$ ratio had to be as high as possible (57). As a very approximate guide the $k/n$ ratios of the four metals of interest are silver $\approx 60$, gold and copper $\approx 30$ and aluminium $< 10$. This factor was further developed by Macleod (58) for multilayer filters. Figure 9.12 is for the same films as in figure 9.11 but with the thin metal films sandwiched between two dielectric layers of refractive index 2.0 and of 20$\mu$m thickness. These two figures show the highest transmittance can be obtained from the silver films. As the light output from the tungsten filament lamps was considered as a governing factor it meant that only the silver films were good enough to be used. Berning in his
Figure 9.11

The visible transmittance of thin metal films of thicknesses chosen to give a sheet resistance of 1Ω on a polyester substrate 50μm thick with a refractive index of 1.7.
Figure 9.12
As in figure 9.11 except that the metal film is now computed as sandwiched between two dielectric layers of refractive index 2.0 and thickness 20nm.
study of design of architectural coatings also demonstrated other features of filter design. These were that so long as the dielectric film refractive index was greater than 2.0 the peak transmittance was relatively insensitive to the refractive index. He also demonstrated that the transmittance bandwidth was dependant upon not only the silver thickness but also the refractive index. What was aimed for as a filter was a thin continuous silver layer between a dielectric coating of a refractive index as high as possible. The ideal dielectric refractive index \( n \) being approximately equal to the value of the imaginary part of refractive index of the silver \( k \) \((51)\). The value \( k \) for silver was \( \approx 3.6 \) and the highest value of \( n \) for a dielectric that was robust and transparent in the visible was that of titanium dioxide where \( n \approx 2.5 \) \((48)\).

The deposition of coatings for architectural purposes does not require the maximum visible transmittance, indeed due to marketing policy would seem to preclude it. Each different climate in the world requires a slightly different window performance, in hot countries it is permissible to reject much more of the light than in a cold climate where every bit of energy requires to be gained. The performance of the titanium dioxide - silver - titanium dioxide can be as high as 89% transmittance at 0.5\( \mu m \) and 98% reflectance at 5.0\( \mu m \) \((59)\). For the energy saving windows the lowest performance that was acceptable has been reported as 83-85% transmittance in the visible and 93% reflectance in the infra-red \((60)\). In order to produce coatings for energy conservation the process has to be a high rate one and to keep the unit cost down the coatings have to be available in large areas. In terms of true energy saving, expressed as a time to recoup the cost of fitting the windows, double glazing would not be expected to repay the cost in under
fifty years (61). An alternative could be the retrofitting of coated polyester, here the major cost would be for the frame, the cost of the coated polyester would be low and the payback time has been estimated at only one winter (62). The deposition process onto large areas is not without problems. The highly reactive nature of the elemental metals that could be used to produce the highest refractive index oxides eliminates them from being considered for inclusion in existing large area deposition systems. Even with the lower refractive index oxides the oxide coatings are produced by a process that is classed as reactive sputtering but which in reality is the sputtering of the oxide in an atmosphere which is substantially oxygen. The large excess of oxygen fully oxidises the target lowering the sputtering rate but making the process more easily controlled. A further consequence of the large excess of oxygen was the damage it caused to the thin metal (usually silver) layer. Oxygen promotes agglomeration of the thin metal (63,64), this reduces the performance of the filter quite dramatically. Nucleation ought really to be considered alongside agglomeration, the number of nucleation sites affects the minimum thickness at which continuity could be obtained and was governed by the energy of the depositing species (63). The agglomeration was also energy related and was affected by the thickness of the deposited film and structure (64). Figure 9.13 is a plot of the changing film characteristics with increasing thickness. The transmittance curve decreases to an initial minimum, the film at the minimum had an island structure and there was no electrical conductivity, the minimum being due to a maximum in the scattering caused by the individual silver islands in an air matrix. Any further increase in the amount of condensed silver, seen either as further nuclei or as an
Thin silver film on PET

--Transmission

Mobility, $\mu$, cm$^2$ v$^{-1}$ s$^{-1}$

Mobility $\rightarrow$

Sheet resistance

Film thickness increasing

Fig. 9.13

Input power to silver target, in Watts
increase in the area of each island resulted in a reduction in the scattering and a rise in the transmittance up to a new maximum. Between the minimum and the maximum was the percolation threshold (65), the point where the film was just conducting. The coating aimed for was at the maximum point or just beyond in terms of thickness. Less than the maximum and the film would contain voids which would act as agglomeration initiators. The decrease in transmittance beyond the maximum was purely due to the film thickness not scattering.

Figure 9.14 is a micrograph of a three layer coating where agglomeration has taken place. A flap of the top oxide has been conveniently lifted up clearly exposing the agglomerated silver film, the agglomeration is visible under the rest of the top oxide coating too but is less clear. By careful positioning of the area to be sampled it was possible to analyse the different areas to confirm that the flap lifted up was just the oxide and that the agglomeration in the hole was purely the silver with the bottom oxide being still intact. As oxygen acts as an initiator of agglomeration the deterioration of the silver could occur as early as during the deposition of the final oxide coating where the excess oxygen in the presence of the plasma would be highly reactive and damaging. The effects could be readily seen, a damaged film would appear to be 'blotchy' with a different coloured appearance perceived where the agglomeration had occurred. In the industrial glass coating process a protection layer had to be deposited between the silver and the top oxide. This sacrificial layer was generally of a highly reactive metal such as titanium, aluminium, tantalum etc. and was deposited as a very thin layer of \( \approx 0.5 - 3\text{nm} \) (66). The purpose of the sacrificial layer was for it to be either sputtered or oxidised in preference to the silver during the time between the
Figure 9.14 An electron micrograph of an oxide-metal-oxide filter which has deteriorated. The picture size is 17 x 20 \( \mu m \).
substrate being first passed into the final deposition zone and a monolayer of the top oxide being deposited which in itself would provide the necessary protection. The coating had to be thin enough that any residual material left was fully transparent and thus non-absorbing such that it did not affect the filter performance. The precise control of the oxygen partial pressure in the roll coating system enabled the three layer filters to be deposited without the aid of any sacrificial intermediate layer. This was attributed to the limitation on the excess oxygen available during the reactive deposition process and to the high rate of deposition of metal from the planar magnetron sputtering source. As the oxidation of the metal is unlikely to occur in flight and from the deposition rate the metal has been sputtered it is likely that the oxidation of the metal takes place on the substrate surface also locally depleting the surface of free reactive oxygen.

The thickness of the dielectric coatings to act as an antireflecting coating for the thin metals was not calculated by the usual $\frac{\pi}{4}$ theory (67). Fan gives an alternative calculation for finding the optimum thickness depending upon the desired performance (58). The reason for the difference in the thickness to the standard antireflection coatings was that the metal had a refractive index (n) of almost zero whilst there was a substantial extinction coefficient (k), this made the phase change on reflection closer to 90° rather than the 0° or 180° as was more usually found in multilayer filter design. The thin metal coatings therefore require an antireflection coating thickness of $\approx \lambda/8$ of a high refractive index material (68).

Figure 9.15 shows a series of curves for copper based thin film heat reflecting filters, the emissivity giving the performance of the filter in the infra-red.
Figure 9.15 R & T curves for three samples of Cr$_2$O$_3$/Cu/Cr$_2$O$_3$ coated onto polyester for which the emissivities are: A; E=0.22  B; E=0.33  C; E=0.60
Figure 9.16 R & T curves for three samples of Cr$_2$O$_3$/Ag/Cr$_2$O$_3$ coated onto polyester for which the emissivities are: A; $E=0.16$; B; $E=0.29$; C; $E=0.41$. 
Figure 9.17 R & T curves of SnO₂/Ag/SnO₂ deposited onto polyester. The tin oxide has a refractive index of 1.92 @ 633nm. The emissivities were A; E=0.19  B; E=0.08  C; E=0.05
The transmittance can be seen to peak towards the red in the visible region of the electromagnetic spectrum. The solar performance of these filters would be limited by the absorption of the films towards the blue of the visible spectrum and the low transmittance and low infra-red reflectance. The change in performance with the thin metal thickness changes as has been reported elsewhere (69,70) with the thicker the film the closer the reflectance edge starts to rise in the visible. Figure 9.16 complements figure 9.15, it is the same outer dielectric coating but with silver as the thin metal film. These films are considerably better with the peaks being closer to the peak eye sensitivity at 550nm and very much broader. The infra-red reflectivity is also considerably improved for comparable visible performance. There is still some absorption towards the blue end of the visible spectrum this is a result of the absorption in the chromium oxide which had a hydroxide content. Figure 9.17 shows an improvement in the transmittance at the blue end of the visible spectrum which is due to the change to using tin oxide as the dielectric coating. The silver thickness has been further increased here and the reflection edge has been brought well into the visible.

The requirement of a high visible transmittance and a high infra-red reflectivity can be used to present the performance of the filters produced. Figures 9.18, 9.19 & 9.20 are plots of the performance of the thin metal layer filters with changing thickness presented as a plot of the principal visible transmittance at 550nm (which is approximately the peak eye sensitivity) (71) compared to the reflectance at 12µm, appropriate for the heat emissivity measurement. Figures 9.18, 9.19 & 9.20 show the results for gold, copper and silver films on their own and sandwiched between a tin oxide dielectric.
The dotted line (---) indicates the theoretical performance for a single layer of gold. The triangles (Δ) indicate the corresponding experimental results.

The full line (—) indicates the theoretical performance for a gold film with an optically matched dielectric film of tin oxide on either side. The squares ( ) indicate the corresponding experimental results.

TRANSMITTANCE (550nm) vs REFLECTANCE (12000nm)

Figure 9.18  AU
TRANSMITTANCE (550nm) vs REFLECTANCE (12000nm)

Figure 9.19 Cu

Line notation as per fig.9.18
Figure 9.20 Ag
Line notation as per fig.9.18
actual performance is compared to the theoretical performance for both conditions. It can be seen from the figures that the films did not achieve the value of transmittance that was predicted, this was particularly so for the gold and copper. Figure 9.21 shows the prediction using the Maxwell-Garnett theory (72,73) with the packing density varying from 1.00 (ideal) down to 0.955. Figure 9.22 shows the prediction for the silver film performance when there was a reduction in the free carrier concentration (74). This was calculated by first separating out the contributions to the performance of the dielectric constants (75). The lattice contribution could then be kept constant and the free carrier concentration indexed giving the change in the optical performance. Comparing the two figures 9.21 and 9.22 both could account for the changes in the performance however it only required a small change in the packing density for the Maxwell-Garnett plot to show large changes in performance whereas it required a large change in the free electron concentration to give a comparable change. The changes in properties seen in films made under varying conditions in our laboratory, appropriate to the nucleation phenomena previously observed, suggests that the Maxwell-Garnett prediction to be the more likely explanation (76,77,34,68). Further support would come from the fact that all films will contain some impurities whether they be argon gas inclusions or solid impurities. These can be beyond the resolution limit of the analysis techniques used but nevertheless a source of scattering in the thin films (78-80). Figure 9.23 shows the spectral characteristics of a silver film, which has a performance closest to that predicted, sandwiched between a matching tin oxide dielectric onto polyester sheet, at room temperature and without the need for any sacrificial layer to be included.
The effects of reductions in the packing density in the silver film as predicted for \( \text{SnO}_2/\text{Ag}/\text{SnO}_2 \) systems. The density change is from 1.0 to 0.955.

Figure 9.21 Maxwell-Garnett plot for Ag
The effect of reduced carrier concentration in the silver on the performance of a SnO$_2$/Ag/SnO$_2$ filters.
Figure 9.23  The spectral characteristics of the 'best' filter.

(-) Theoretical prediction.  (---) actual performance.
9.5 The electrochromic coatings and devices.

An electrochromic device is one in which there is a change in the transmittance of the device when a potential is applied across the electrodes of the device. The process is reversible by reversing the polarity of the applied potential. The action of reducing the transmittance is known as colouration and the action of increasing the transmittance is known as bleaching. Figure 9.24 is a schematic of an electrochromic device and figure 9.25 is a schematic of the changing transmittance with applied current and voltage versus time.

The interest in this type of device has grown recently with the improvement of deposition techniques. There is now the possibility of depositing such devices onto glass in large areas which could compete in both the displays market and as switchable windows for both energy saving and aesthetic applications. The electrochromic devices in general display the following attributes, they have excellent contrast and a wide viewing angle, they only require a low operating voltage and they have an open circuit memory (81). The memory property is important in that once switched and coloured the potential need not be continued as with liquid crystal display devices or organic electrochromic devices (82). Electrochromic devices based on a reflective surface have already been manufactured in very small areas (83,84). One limitation to increasing the size of the devices is the switching current, for one of the products the peak current requirement is 4mA/cm² which would correspond to a current requirement of 40A for a 1m² device. To reduce this limitation the conductivity of the transparant conducting layer has to be improved such that the current requirement can be minimised and also such that with a large area device there is a minimised
1. Substrate.
2. Transparent conducting electrode.
3. Electrochromic layer.
4. Fast ion conductor or electrolyte.
5. Transparent counter electrode.
6. Optional top cover.

FIGURE 9.24 Composition of an electrochromic device.
Figure 9.25  The characteristics of an electrochromic device.
voltage drop between the edge and the centre of the device. Any voltage drop would result in less of a
colouration and a slower reaction when the device was
switched. This is at present a limitation for devices
deposited onto glass and it is more so for devices when
deposited onto polymers where elevated substrate
temperatures have not been possible during the deposition
of the transparent conducting oxide layer due to the
thermal sensitivity of the polyester. New polymers are
being investigated at present where a modest substrate
temperature can be tolerated (up to 200°C) which is
expected to greatly improve the conductivity of the
transparent oxides to more nearly match the coatings that
are currently available on glass.

The electrochromic layer was one of the transition
metal oxides, of these oxides the one that has generated
the most interest was that of tungsten trioxide (85-90).
This group of materials form the basis of what are known
as 'tungsten bronzes'. This name originated from the
colour of some reduced acid sodium tungstate (91) which
took on a metallic lustre. The term now applies to a
ternary metal oxide of the form $A_xB_yO_z$ where $B$ is the
transition metal, $A$ is the dopant and $x$ is in the range
$0 < x < 1$ (92). The tungsten bronze $A_xWO_3$ has a
perovskite structure as shown in figure 9.26. The unit
cell has a tungsten atom at the centre of a cube
surrounded by six oxygen atoms at each of the face
centres, this produces eight 'interstitial' sites at the
cube corners which can be occupied by the dopant atoms.

There are a number of transition structures available
depending upon the number of sites occupied (93). The
dopants for the tungsten trioxide are usually sodium or
lithium. Other group Ia dopants such as potassium,
rubidium and caesium can be used but as the size
increases the mobility in the lattice is reduced and they
are not preferred. The other dopant that can be used is
A plane of $\text{WO}_3$ octahedra from the crystal lattice.

Figure 9.26

The perovskite structure of a doped tungsten oxide.
hydrogen. This is almost in a separate category in that the dopant level attainable is lower at \( 0 < x < 0.3 \) (84,94). Doping tungsten trioxide by sodium or lithium during deposition was not practicable, a sputtering target was not available suitably doped and the geometry of the chamber did not lend itself to using evaporation in combination with the sputtering of the tungsten. Thus it was decided that to dope the tungsten trioxide hydrogen would be used, the hydrogen being introduced using a third mass flow controller to give precise control of the doping levels. The electrochromic effect in tungsten trioxide was first observed by Deb (95) although it was already known that different stoichiometries or doping levels caused colour changes (96,97). Faughnan et al. later suggested that the colouration process of the tungsten trioxide was due to the modification of the oxide into the bronze producing colour centres associated with the electrons localised on the tungsten ions (98). The injected electron is trapped by a \( \text{W}^{6+} \) atom forming a \( \text{W}^{5+} \) atom with the charge being kept neutral by a positive ion resting in one of the interstitial sites. The optical absorption is due to a resonance transfer of the electron amongst the equivalent tungsten atoms i.e.

\[
\text{W}^{5+}(A) + \text{W}^{6+}(B) \xrightarrow{hv} \text{W}^{6+}(A) + \text{W}^{5+}(B)
\]

The operation of the device occurs thus; a potential is applied across the device between the outer electrodes, there is electron and ion injection into the electrochromic layer, the electrons are associated with the tungsten and the ions are injected to maintain the charge neutrality. In our case the electrons are supplied from the transparent conducting oxide and the ions are of hydrogen and are donated from the
electrolyte. The operation speed is dependant upon the ion injection rate (99). Increasing the driving force does not necessarily increase the injection rate, indeed it can hamper the process by gas generation at the electrolyte - electrochromic layer interface. The rate of injection can be modified by the structure of the electrochromic layer in that the more porous the layer the faster the operation. There is at present some conflicting evidence as to what is required of the structure of the tungsten trioxide, it has been reported that only amorphous films will work (87,88,90,100,101) however other authors have produced crystalline films which also work (102-104). It is more likely that the crystal size is small especially considering the techniques used to deposit them such as flash evaporation and planar magnetron sputtering and that there are differing ideas on what is crystalline and what is not. Further it is more likely to be the stoichiometry that is critical not its crystallinity (105).

The roll coating apparatus with the precise control of the deposition parameters was suited to being able to step through a range of stoichiometries of the electrochromic layer. This was done, the argon flow was held constant and the oxygen flow increased in a series of steps. The deposition conditions were as follows; 1A and 640V supply to the magnetron, an argon flow of 70 sccm and an oxygen flow of from 20 to 40 sccm giving working pressures of 0.44 Pa to 0.56 Pa. Figure 9.27 shows the transmittance of three films through the visible spectrum. The visual appearance is marked on each of the traces. The changing flow of oxygen inevitably produced a variable sputtering rate depending upon the oxidation state of the target, this was monitored by use of the ellipsometer at 633nm. These results are shown in figure 9.28, where it can be
Figure 9.27 Tungsten trioxide deposited using three different oxygen mass flows produced different spectral characteristics.
Figure 9.28 A plot of the refractive index (@ 633nm) and thickness versus oxygen mass flow.
seen that there is a dip in the refractive index and a peak in the thickness within the region where the samples appeared blue. This colour was encouraging in that the reported colour of tungsten trioxide was blue (94), however the refractive index was some way from the reported index of 2.5 for single crystal tungsten trioxide (95). At the peak in the thickness the refractive index did however came close to the value of 2.2 as reported by Deb (95). Although the colouration of the tungsten trioxide was associated with an increase in the conductivity no evidence was seen of this on the digital volt meter connected to the dragging probes in the vacuum system. Analysis of the film was then done, initially by x-ray photoelectron spectroscopy (XPS).

The XPS analysis was carried out using a V.G. ESCALABS spectrometer. The experimental conditions used were; an x-ray source of aluminium Kα radiation at 1486.6 eV, with an anode voltage of 15kV and an anode current of 20mA over an analysis area of 0.4cm² and the emission angle normal to the surface giving a response of up to 10⁵ counts per second. The ion beam parameters for etching the surface were, argon ions of 5keV and an ion current of 20μA at a gas pressure of 10⁻³ Pa.

Figure 9.29 shows the broad scan XPS spectrum of one of the tungsten oxide films. The surface has a carbon peak that disappeared upon etching as can be seen from the other two scans of the surface after 3 and 6 minutes etching. The tungsten peak at 35.7 eV (binding energy) on the XPS broad scan was looked at using a high resolution scan. This is shown in figure 9.30. As can be seen after the 3 minutes etching the doublet has disappeared and the peak has been moved towards a lower oxide state indicating the film has undergone dissociation under the ion beam bombardment, indicating that to obtain accurate chemical information from the XPS...
A broad XPS scan of the surface of the tungsten trioxide film and the surface after an ion etch of three and six minutes.
Figure 9.30

A high resolution XPS scan of the Tungsten Trioxide showing the effects of the ion etch.
analysis care has to be exercised in the interpretation of the graphs obtained. Tungsten peaks also appear at higher binding energies on the XPS broad scan (see figure 9.30), the 4p doublet does not overlap as the 4f doublet does but is separated by more than 50 eV. The tungsten 4p3/2 peak is the more intense of the pair and is used to ratio to the oxygen peak. The tungsten peak is also much closer to the oxygen peak than is the 4f doublet and so the photoelectron escape depth is more comparable to that from the oxygen. The oxygen/tungsten ratio for the films characterised in figure 9.27 at 0, 3 and 6 minutes etching are presented in table 9.2. Compositions were obtained using unsputtered Johnson Matthey Specpure Tungsten Trioxide as a standard. As can be seen at 0 minutes etch (i.e. at the surface) the oxygen peak is too high. This can be accounted for by the carbon contamination on the surface which will also hold a proportion of oxygen. As the film is etched the ratio drops rapidly but consistently with increasing dissociation. The films produced at higher oxygen partial pressures also show a higher ratio of oxygen to tungsten. This too is consistent after the 3 and 6 minutes of etching.

The scan of binding energy of the tungsten peak shown in figure 9.30 is for the blue film made with a flow of 30 sccm of oxygen. Although we are not able to etch the film we can obtain information by comparing samples. The dotted line on the left hand shoulder of the tungsten trioxide doublet was representative of the modification of the doublet for samples with a lower oxygen flow during deposition.

The XPS had not really produced enough information to optimise the deposition of the tungsten trioxide and so further analysis using laser induced mass analysis (LIMA) was carried out. The LIMA is a surface
Table 9.2

<table>
<thead>
<tr>
<th>O₂ Flow (sccm)</th>
<th>Etch time (mins)</th>
<th>0</th>
<th>3</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td></td>
<td>3.27</td>
<td>1.42</td>
<td>1.19</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>3.65</td>
<td>1.86</td>
<td>1.48</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>3.75</td>
<td>2.12</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Ratio of $O/W$ by XPS analysis.
analytical technique capable of providing both the constituents of coatings and also some chemical information. The LIMA comprises a 'Q' switched Nd:YAG laser quadrupled in frequency which is fired as a pulse at the sample. This laser pulse volatilises and ionises a cylinder of material. The ions are then mass analysed by a time of flight mass spectrometer. The information is sampled by a transient recorder and can be plotted as the mass to charge ratio versus the relative intensity. The laser produces a pulse of light with a power density variable from \(0\) to \(10^{11}\) W cm\(^{-2}\) at 265 nm and can volatilise a cylinder of material of 2 \(\mu\)m to a depth of 0.5 \(\mu\)m. Both positive and negative ions are produced and either may be selected for analysis (106).

The same samples as examined on the XPS instrument were analysed using the LIMA machine, this included the J.M.Specpure standard sample. Figure 9.31 shows two scans of the tungsten oxide deposited. As can be seen on the lower of the two traces the peaks are in two groups of three, the first of group corresponds to \(\text{WO}_2\), \(\text{WO}_3\) and \(\text{WO}_4\). The individual spikes within each of the peaks are from the different isotopes. The areas under the peaks were recorded and the changes in proportions of each peak was plotted as a function of the oxygen flow during deposition. This is shown in figure 9.32. The position of the \(\text{WO}_3\) peak observed using the LIMA corresponds well to the minimum in the refractive index, and the change in relative proportions of \(\text{WO}_2\) and \(\text{WO}_4\) correspond well to the XPS results in that the higher the flow of oxygen during the deposition the higher the oxygen content within the film as analysed.

The second (upper) scan on figure 9.31 is from a film that was deposited with a partial pressure of hydrogen as well as the argon and oxygen with the intention that the hydrogen would be included in the film as a dopant. The scan shows an additional peak at the
Figure 9.31 Two spectra from the LIMA of tungsten trioxide, the top trace being of the hydrogen doped sample.
Figure 9.32 The relative proportions of $WO_2$, $WO_3$ & $WO_4$ as measured using LIMA.
mass to charge ratio of 362-366 due to a complex ion of tungsten. Checks were made on the calibration of the instrument, the peak was close to where the W₂ peak would have been found and had the calibration not been correct the peak could have been misinterpreted. The calibration was correct and the peak was not from any free tungsten but had to be from a multiply charged tungsten oxide complex ion. The other check that was made on the peak was to check on the isotope 'fingerprint'. All elements have characteristic isotope ratios and the peak contained too many isotopes to consist only of tungsten. Further interpretation of the peak was hindered by the lack of resolution of the time of flight mass spectrometer. This peak was only seen in the traces of films that had been deposited with a partial pressure of hydrogen. The peaks of WO₂ and WO₃ were reduced by 1.5% and 3.5% respectively and the WO₄ was enhanced by 5% when compared to the equivalent films deposited without any hydrogen flow.

To endeavour to find out precisely what effect the hydrogen had upon the coating the samples with the hydrogen doping were also returned for analysis by the XPS instrument. There was one established form of tungsten oxide which incorporated hydrogen which was sought, this was H₂WO₄ (107), the peak positions on the high resolution binding energy scan proved that this particular form was not present. There was however evidence that the hydrogen was present as a hydroxide. Samples were analysed with differing flows of hydrogen (0 to 20 sccm) and constant flows of oxygen and argon (28 and 70 sccm respectively), however it appeared that the hydrogen had an immediate effect on the films and further increases in the flow of hydrogen had no further effect. The formation of the hydroxide fits well with the report made by Faughnan (98) on the fact that the
hydrogen preferred to form an OH bond and suggested that the formula $H_xWO_{3-x}$ should more correctly be written as $(OH)_xWO_{3-x}$.  

These experiments demonstrated the possibility of depositing films of controlled stoichiometry from substoichiometric metallic films all the way through to superstoichiometric oxygen excess films. However there was no feedback control on the system that was sensitive enough to monitor the films as they were being deposited. To prevent the deposition becoming a statistical exercise in producing good films a feedback control was required. The fact that changes in the conductivity had frequently been reported gave the clue, the range of the digital voltmeter only went up to 20MR. In one reference (96) the resistivity was reported to have changed by over six orders of magnitude up to 2000MR. An insulation tester was connected in place of the digital voltmeter and sure enough there was over six orders of magnitude change in the film resistivity. This is shown in the figure 9.33, with the minimum resistivity being well to the substoichiometric side of the flow that produced the tungsten trioxide peak as determined by the LIMA analysis. With this feedback control any stoichiometry of tungsten trioxide film could be deposited onto the polyester reproducibly and in long lengths.

Samples of tungsten trioxide were coated onto polyester that had previously (but in the same pumpdown) been coated with conducting indium oxide. These samples were tested for any electrochromic effect. The stoichiometric oxide samples showed no effect and the undoped oxides also showed no observable effect. The hydrogen doped substoichiometric tungsten trioxide showed some evidence of electrochromism with a 30% change in transmittance. This small compared to the published
Figure 9.33 The resistivity of the tungsten oxide film as measured in the vacuum system soon after deposition and plotted against oxygen flow.
performance obtained by others when coating onto glass substrates (108). The electrolyte may have been one source of the poor performance. The simplest and most convenient electrolyte available was lithium perchlorate in solution in propylene carbonate (109). This had the lithium providing the active ions and so in the hydrogen doped film injection of the lithium was attempted which may account for the poor response. The subject of electrolytes was not pursued. There appear to be two groups of electrolyte, the solid state electrolytes and the wet electrolytes. Of these two groups there was a further sub division, those that require water and those that do not. As such there are many permutations that may be tried to try to optimise the performance of the electrochromic device. The deposition system proved capable of providing the control necessary to enable any stoichiometry of tungsten oxide to be produced in long sample lengths.
9.6 References.


    41(7/8) (1983/84) 225


    21(2) (1980) 101

8). I.A.Shanks Electronic Eng. (Aud 1974) 30


24). P.H.Berning in 'Physics of thin films' 1 (1963) 69


31). M.Faraday Phil.Trans.R.Soc.Lond. 147 (1857) 145


   Sept 13 (1949)
   U.S. Patent No. 2,628,927
   Feb. 17 (1953)


40). C.E. Ells & G.D. Scott  J. Appl. Phys. 23(1) (1952) 31


43). H.D. Taylor  "The adjustment & testing of telescope


46). Anon  The Chemical Age (Lond) 67 (Jly 1952) 13

   (Sept. 3 1968)

49). S. Selkovitz report LBL 7829 (March 1978) Lawrence Berkeley Laboratory Univ. of Calif. U.S.A.


   CEP. Consultants Ltd. U.K.

   G. Hass & R.E. Thun 2 (1964) 239


71). S. Stenstrom 'Optics and the eye' Pub. Butterworths (1964)

   203 (1904) 385
   205 (1906) 237
76). C.A. Bishop  'Ion assisted evaporation of thin films'  
M.Phil. Thesis. (1980) 
Pub. CEP. Consultants. U.K.
78). C.W. Pitt  Proc. IPAT (1977) 149  
7(3) (1970) 453
84). F.G.K. Baucke & J.A. Duffy  Chemistry in Britain  
21(7) (1985) 643
85). W.C. Dautremont-Smith Displays (Jan 1982) 3
     Displays (Apl 1982) 67


     Lawrence Berkeley Laboratory
     Univ. of Calif. U.S.A.

95). S. K. Deb Phil. Mag. 27 (1973) 801


97). C. J. Smithells 'Trietise on tungsten' 3rd Edn (1952)

98). B. W. Faughnan et al. RCA. Rev 36 (1975) 177

100). B. J. Molnar et al., J. Vac. Sci. & Tech. 15(2) (1978) 261


10. Conclusion.

A vacuum system was built to enable rolls of polymer sheet to be coated with a variety of films. The system included home-built planar magnetron sputtering sources which provided a highly reproducible source of elemental material which was deposited onto the polyester sheet which was transported in-front of the magnetron sources on a roll-to-roll system. By suitably balancing the mass throughput of inert and reactive gas versus the sputtering rate reactive deposition was possible and oxides were produced. The deposited films were monitored during the deposition process soon after the deposition zone. This on-line monitoring enabled coatings to be optimised and produced with highly controlled stoichiometries and in long lengths.

This highly sophisticated machine enabled coatings normally associated with high temperature processes to be deposited onto heat sensitive substrates, such as polyester, at ostensibly room temperature. The coatings produced demonstrate this capability.

Conducting oxide coatings were produced and using the deposition parameters coatings were later produced on other deposition systems. This included an industrial roll coater which had previously only been able to produce non-conducting transparent tin oxide.

The 'heat mirror' type coatings were deposited with enough control of the oxygen mass flow that an intermediate protection layer normally associated with the production of these filters could be dispensed with.

Other oxides which had less a less stable source material were produced by using the additional facility of pulsing the reactive gas supply. This was particularly useful for the deposition of titanium
dioxide. This technique whilst not maintaining the precise balance of sputtering rate to reactive gas admittance which could be achieved with other less reactive targets nevertheless did enable one of the most difficult oxides to be deposited.

Finally, tungsten trioxide was deposited with a wide range of controlled stoichiometries onto both polyester and polyester previously coated with conducting indium oxide. Whilst these coatings have not yet been optimised they do show the potential of this equipment for the precise control of the stoichiometry coatings that can be achieved by reactive planar magnetron sputtering onto polymer substrates.
11. Appendix.

Publications.


5). 'Factors influencing the nucleation of silver on plastic'. C.A.Bishop, R.P.Howson & M.I.Ridge. Inst.of Physics Conf.Ser. No.54


8). 'Ion plating onto plastics'.
J.N. Avaritsiotis, R.P. Howson, M.I. Ridge & C.A. Bishop
Trans. Inst. Metal Finishing. 57 (1979) 65

9). 'Solar heat mirrors produced by ion plating'.
R.P. Howson, J.N. Avaritsiotis, M.I. Ridge & C.A. Bishop


21). 'Transparent and heat reflecting coatings'.
   R.P. Howson, M.I. Ridge & C.A. Bishop.

22). 'Design and performance of novel planar magnetron
   sputtering sources for ion plating'.
   A. Bishop, M.I. Ridge & R.P. Howson.

23). 'Reactive planar magnetron sputtering of oxides'.
   R.P. Howson, M.I. Ridge & C.A. Bishop.

24). 'The control of high rate reactive sputtering

25). 'Transparent conducting oxides of metals and alloys
   made by reactive magnetron sputtering from
   elemental targets'. R.W. Lewin, R.P. Howson,
   C.A. Bishop & M.I. Ridge. Proc. IPAT (1985) 284
   Vacuum 36 (1-3) (1986) 95

26). 'Large area optical filters using sputtered metal
   films'. C.A. Bishop & R.P. Howson.

27). 'The performance of large area optical filters using
   d.c. magnetron sputtered metal thin films in oxide-
   metal-oxide sandwiches'. C.A. Bishop & R.P. Howson.

28). 'The deposition of large area coatings for energy
   efficient systems'. C.A. Bishop, R.P. Howson &
   R.W. Lewin. Accepted for Pubn. SPIE (1986) Vol 652-05


31). 'Optical properties of tungsten oxide films as a function of their stoichiometry as determined by LIMA and XPS'. C.A.Bishop, G.Edge, I.Sutherland & R.P.Howson. Presented at Vacuum 86 & to be published in Vacuum.
The End.