Imaging Photon Detectors and their use with Single and Multiple Fabry-Perot Etalon Systems for Atmospheric Wind Measurements

by

Ian McWhirter

Department of Physics and Astronomy
University College London

Thesis presented for the degree of Doctor of Philosophy
of the University of London

July 1993

This thesis

is dedicated to the memory of

Hilda McWhirter

ABSTRACT

Imaging Photon Detectors are extremely sensitive imaging devices capable of detecting single photons of light. The Fabry-Perot etalon, a multiple beam interference device, is capable of detecting very small changes in the wavelength of light and is thus well suited to the determination of Doppler shifts.

The combination of these two devices has enabled the realisation of a series of operational interferometers for the measurement of wind velocities in the upper atmosphere. This is achieved by measuring Doppler shifts in optical phenomena occurring at high altitudes. The instruments have been successfully deployed in Northern Scandinavia, North America and other locations world-wide.

A full description of the component parts of the IPD and its associated electronics is presented. The theory and practical limitations of the device are discussed, together with a critical performance analysis of the complete imaging system. In addition to the IPDs for the Fabry-Perot interferometers, special types have been built for high time-resolution atmospheric lidar and a rocket-borne auroral imager.

The Fabry-Perot etalon is described and the practical aspects of incorporating it into an interferometer are considered. The instruments are required to run unattended for extended periods, so particular care has been paid to long term aspects of stability, reliability and safe operation.

Etalons can be tuned using piezo-electric transducers to vary the cavity length, in conjunction with capacitance sensors which determine the precise amount of movement. Such devices are termed capacitance stabilised etalons. These etalons can be combined in multiple etalon systems which provide greatly improved optical filtering. This allows measurements to be made against the higher background illumination encountered during daylight hours. Triple etalon interferometers have been built which have been flown on balloons in Texas and operated from the ground in Northern Sweden.

CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

CHAPTER 8

CHAPTER 9

CHAPTER 10

CHAPTER 11

CHAPTER 12

CHAPTER 13

LIST OF ACRONYMS AND ABBREVIATIONS

ADC	Analogue to Digital Converter
ALOMAR	Arctic Lidar Observatory for Middle Atmosphere Research
APL	The Atmospheric Physics Laboratory, University College London
CAT	Circular Arc Terminated anode
CCD	Charge Coupled Device
CSE	Capacitance Stabilised Etalon
DAC	Digital to Analogue Converter
DDDI	Double Differentiation Double Integration pulse shaping
DIS	Doppler Imaging System
DPM	Dual Port Memory
DQE	Detected Quantum Efficiency - of a photon counting system
EHT	Extra High Tension (high voltage)
eV	Electron volts - a unit of energy
FIFO	First In First Out memory
FPI	Fabry-Perot Interferometer
FSR	Free Spectral Range
FWHM	Full Width at Half Maximum - a measurement of resolution
GaAs	Gallium arsenide
GenII	Generation II image intensifier, with S20/25 photocathode
GenIII	Generation III image intensifier, with GaAs photocathode
HAK	Helium-Argon-Krypton calibration lamp
HOT	High Output Technology microchannel plates
HRDI	High Resolution Doppler Imager, on UARS.
HRE	High Resolution Etalon
ICSU	International Council of Scientific Unions
ILC	International Light Corporation Inc., U.S.A.
IPCS	Image Photon Counting System
IPD	Imaging Photon Detector
IR	Infra-red
IRF	Institutet för Rymdfysik - Swedish Institute for Space Physics
ITL	Instrument Technology Ltd., U.K.
ITT	International Telephone and Telegraph Company Inc., U.S.A.
KGI	Kiruna Geophysical Observatory, Sweden, now known as IRF
LC	Inductor-capacitor (time constant)
LRE	Low Resolution Etalon
MCP	Microchannel Plate
MDAC	Multiplying Digital to Analogue Converter

MRE	Medium Resolution Etalon
NEA	Negative Electron Affinity
OPD	Optical Path Difference
PHR	Pulse Height Resolution
PID	Proportional-Integral-Derivative temperature controller
PSD	Phase Sensitive Detector
PSF	Point Spread Function
PZT	Piezo-electric Transducer
QE	Quantum Efficiency
RC	Resistor-Capacitor (time constant)
RMS	Root Mean Square - the square root of the mean of the squares
RQE	Responsive Quantum Efficiency - applies to photocathodes
S20	A multi-alkali photocathode
S25	A multi-alkali photocathode with extended red response
SCOSTEP	Scientific Committee on Solar-Terrestrial Physics (ICSU)
SPU	Signal Processing Unit
TIDI	TIMED Doppler Interferometer
TIMED	Thermospheric Ionospheric Mesospheric Energetics and Dynamics
TTL	Transistor-Transistor Logic
UARS	Upper Atmosphere Research Satellite
UCL	University College London
UV	Ultra-violet
WITS	World Ionospheric/Thermospheric Study (Published by SCOSTEP)

CHAPTER 1 - INTRODUCTION

1.1 Overview

This thesis charts the development of two devices: the Imaging Photon Detector and the Fabry-Perot etalon. Their realisation and subsequent incorporation into a variety of instruments, followed up by ensuring their optimum operational performance in the field has constituted the major part of the author's work at the Atmospheric Physics Laboratory, University College London.

Imaging Photon Detectors (IPDs) are extremely sensitive imaging devices capable of detecting single photons of light.

The Fabry-Perot etalon is a multiple beam interference device which can detect very small changes in the wavelength of light, and is thus well suited to the determination of Doppler shifts.

The marriage of the IPD and the Fabry-Perot etalon has produced a series of operational interferometers for atmospheric wind velocity measurement by remote sensing of Doppler shifts in optical phenomena occurring at high altitudes. These instruments have been successfully deployed in Northern Scandinavia, North America and various other locations world-wide. Two triple etalon interferometers have been flown on balloons from Palestine, Texas, and a design study has been undertaken to incorporate a similar instrument in a polar-orbiting satellite. These Fabry-Perot interferometers are described in detail, together with their control systems and data handling. The instruments run unattended for extended periods, so particular care has been paid to long term aspects of stability, reliability and safe operation.

1.2 The aim of this thesis

The instrumentation described in this thesis has been used by many researchers and it has constituted the enabling technology for the work on which numerous research papers and several Ph.D. theses have been based. These are listed at the end of this volume. In the course of many discussions with those who have used the instruments it became clear that they frequently adopted a "black box" approach and were unacquainted with the physics of the tools that were being used to

gather their data. It also emerged that there was no single source to which researchers could be directed in order to clarify their queries. The information that they required was often unpublished, or thinly dispersed throughout a multitude of papers and references.

This thesis, therefore, has been developed not only as a record of my own design, development and use of the instruments, but also as an account of the diverse technologies involved. It is presented in a form which, it is hoped, will be of help and interest to those using the instrumentation, so that a better understanding may enable them to further their scientific goals.

1.3 The primary scientific objectives

Winds in the upper atmosphere must be measured by remote sensing. The region is too high for direct balloon measurement and too low for a stable satellite orbit. Early forms of remote sensing used chemical trails which were released from rockets and photographed from the ground. Some information was also obtained by studying the decay of satellite orbits, and by observing meteor trails.

The prime objective of the instrumentation developed at the UCL Atmospheric Physics Laboratory has been the measurement of upper atmosphere winds by determining the Doppler shift of airglow and auroral emission lines. These lines are extremely faint. Even a strong emission may have an intensity of only a few tens of kilorayleighs (*Aruliah 1992*). The Rayleigh is a unit used to quantify the intensity of radiating layers in the atmosphere and is defined as an apparent emission rate of 10^6 photons/second/cm^2 column.

The principal lines of interest are the 630nm and 557.7nm OI emissions. The 630nm line is used for measurements of the upper and middle thermosphere (150 to 250 km). The 557.7nm line will allow measurements of the lower thermosphere (100 km) to be made when conditions are favourable. However, any auroral activity results in 557.7nm emissions over a very wide height range, which limits the usefulness of the observations. The development of detectors with gallium arsenide photocathodes, which are sensitive to the near infra-

red, has enabled the observation of the OH emission at 843nm to be measured, thus providing data for the upper mesosphere (around 85 km). In addition, balloon-borne instruments built at UCL have successfully observed Doppler shifts in absorption lines (*Rees et al, 1982a*). These are too broad to be measured when observed at ground level, but can be measured from satellites (*Hays et al, 1992 & 1993*).

1.4 The requirements for the instrumentation

The light gathering power of an optical instrument is defined by a quantity termed the etendue, U. This is the product of two parameters, the light collecting area of the instrument, A, and the solid angle of the field of view, Ω. For any instrument the product A Ω is constant throughout. It is inherent in the design of the instrument and cannot be increased by any change in the subsequent optical configuration. For example, an increase in image magnification is met by a corresponding decrease in the angle of the light cone at that point.

The Fabry-Perot etalon has a high etendue for its size and cost. In practice, A is defined by the diameter of the etalon and Ω by the required resolution of the instrument. The interference fringes created by the etalon consist of concentric circles of increasing angular diameter. The angular diameter of the fringe pattern varies inversely with the resolution of the etalon. Also, the more fringes observed, the greater the value of Ω, but the less the resolution with which they will be measured.

If I_R is the intensity in Rayleighs, the observed intensity is:

$$I = \frac{I_R . 10^6}{4\pi} \text{ photons}^{-1}.\text{cm}^{-2}.\text{s}^{-1}.\text{sterad}^{-1} \qquad [1.1]$$

and the number of photons per second, n, incident on the detector of an instrument can be quantified as:

$$n = T.I.A.\Omega \quad \text{photons}.\text{s}^{-1} \qquad [1.2]$$

Ω is equal to $2\pi(1-\cos\alpha)$, where α is the half angle of the field of view of the instrument. T is the transmission of the optical system. This includes losses at the lens and window surfaces and also the

attenuation, typically 50%, through the band-pass filter used to exclude continuum which would otherwise mask the signal. For an imaging Fabry-Perot interferometer it also includes the transmission profile of the etalon in image space. The resulting signal from a typical emission line may only amount to a few thousand photons per second.

1.4.1 Detector development

Until recent years it has been difficult to take advantage of the full light-gathering potential of the Fabry-Perot etalon. The extremely low light levels involved require the use of a detector which is sensitive to individual photons.

Measurement of atmospheric Doppler shifts was originally accomplished some years ago (*e.g. Armstrong, 1956*), but these measurements have been restricted by the use of a conventional photomultiplier tube for the detector. Such devices are not capable of imaging, and can only measure the intensity of one part of the fringe pattern at a time. They are usually placed on the optical axis of the instrument and a sequence of measurements is made whilst the fringe pattern is scanned through the collecting area. Because such detectors have no spatial resolution, any increase in the size of the collecting area results in a blurring of the measured profile, and the aperture is, therefore, commonly restricted to a pinhole.

In the 1970s the Atmospheric Physics Laboratory at UCL embarked upon a programme of development of a new type of detector, known as the Imaging Photon Detector, or IPD. This device is similar in concept to a photomultiplier tube in that it is sensitive to individual photons of light. The incident photons generate individual electrons within the device by photo-emission. These photoelectrons are then used to initiate a cascade of many electrons, which is sufficiently large to be measured electronically. The IPD uses a special type of electron multiplier called a microchannel plate. This enables each electron cascade to be restrained laterally, thereby preserving the positional information of the incident photon. The resulting cloud of electrons is collected by an anode, which is an electrode held at a more positive potential. The anode can be designed to be position

sensitive, that is to say, the electronic signals measured at the edges of the anode can be used to determine the exact position of the incident charge cloud. One way of achieving this is to make the anode from a sheet of electrically resistive material.

Research on this class of detector has been proceeding at a small number of other institutions, principally at the Space Sciences Laboratory at the University of California, Berkeley (*Lampton and Paresce, 1974*) and a degree of collaboration was achieved during the early stages of the work. However, the specific requirements of the UCL instruments necessitated that the development be pursued on largely independent lines. The systems needed to be small, rugged, portable and capable of operating outside a laboratory environment. Also, financial constraints meant that in house development of the electronic systems, tailored to the precise requirements of the instruments, would be far more cost effective for a programme which, ultimately, would result in a proliferation of instrumentation world-wide. A collaborative development programme was instigated with Instrument Technology Ltd., St. Leonards-on-Sea, Sussex, a company specialising in vacuum technology

1.4.2 Etalon development

In order to realise the full potential of the Imaging Photon Detectors a programme of development was initiated to produce high quality large aperture etalons. The challenges involved in making high quality etalons of up to 150mm diameter are considerable, and the work of a small South London company, I.C. Optical Systems, in manufacturing these devices has been instrumental to the success of the UCL interferometers. Major improvements in the production of dielectric coatings have also contributed greatly to the performance of the etalons.

The involvement of UCL with these etalons began in the 1970s with the development of stable and rugged etalons for rocket flights and for the interferometer flown on the Dynamics Explorer satellite (*Rees et al, 1982c; Killeen et al, 1982*).

The requirement for multi-etalon systems required a mechanism for accurate tuning of the etalon cavities. This led to the development of the capacitance stabilised etalon. A similar developmental programme was undertaken by Queensgate Instruments Ltd. (*Hicks et al, 1974 & 1976*) but, for the reasons outlined in the previous paragraph, an in house programme tailored to the precise needs of the UCL instruments was deemed preferable to the purchase of comparatively expensive commercial systems.

1.5 The subjects discussed in this thesis

This thesis presents the underlying technology of the detectors and etalon systems used at UCL and chronicles the author's design and development of the instrumentation. It also provides a critical assessment of the instrument performance, without which the integrity of the scientific data cannot be assured.

A list of the UCL instruments currently operating world-wide is given in chapter ten, which describes the UCL interferometers. A detailed description of each station may be found in *Rees, McWhirter, Aruliah and Batten (1992)* which is the UCL entry in the *SCOSTEP World Ionosphere / Thermosphere Study (WITS)* handbook.

A comprehensive list of publications resulting from the scientific data obtained from the instruments in which the author was involved is to be found at the end of this thesis.

1.5.1 The Imaging Photon Detector

A description of the component parts of the IPD and its associated electronic signal processing is presented. Chapters on the microchannel plate electron multiplier, photocathode technology and the position sensitive anode are followed by an account of the design and development of the complete detector and the evolution of the Signal Processing Unit (SPU) and its associated electronics. The SPU is the electronic system which analyses the positional information contained in the low level signals produced by the detector and which, photon by photon, assembles a two dimensional computerised image.

The theory and practical limitations of the imaging system are discussed, together with a critical performance analysis.

There follows a discussion of special types of IPD which have been built using anodes with discrete elements. These can accommodate higher illumination levels than the resistive anode. and some are also capable of very high time resolution. Such detectors have been used in conjunction with powerful pulsed lasers for the evaluation of techniques for atmospheric lidar, and are currently being refined for use in an Arctic lidar facility: an Arctic Lidar Observatory for Middle Atmosphere Research (ALOMAR).

1.5.2 The Fabry-Perot Interferometer

The theory of the Fabry-Perot etalon is presented, together with the important equations which describe its operation, with special reference to its use in the UCL interferometers. This is followed by a discourse on the design, construction and operation of the UCL instruments.

1.5.3 Capacitance Stabilised Etalons

A chapter on capacitance stabilised etalons describes the development of the sensitive electronics built to control the cavity length and parallelism of a Fabry-Perot etalon to an accuracy and stability of better then one thousandth of a wavelength of light. This technology is an essential part of the realisation of multiple etalon interferometers.

1.5.4 The Triple Etalon Interferometer

This may be seen as the ultimate Fabry-Perot interferometer. The improved optical filtering obtained by three etalons in series allows measurements to be made against the higher background illumination encountered during daylight hours. This section describes the design and construction of two balloon-borne instruments for observation of absorption spectra and the evolution of a ground based interferometer for daytime measurements of emission lines. The extremely demanding performance criteria are analysed.

1.5.5 Future developments

The thesis concludes with an account of future work. This covers the continual refinement and expansion of the ground-based network of observing stations, the construction of instruments for the ALOMAR atmospheric lidar project, and involvement in the TIDI (TIMED Doppler Interferometer) to be flown on the NASA TIMED spacecraft.

1.6 Division of responsibilities

The author's principle area of responsibility has been the design and development of the electronics systems described in this thesis, including the testing and bringing to operational status of all the IPD and capacitance stabilised etalon systems. The author has also been involved in a consultative role in many aspects of the mechanical engineering, optical design and instrument software. He has taken a major part in the instrument assembly and optical and electronic alignment, both in the laboratory and in the field, performing on-site modifications to hardware and software as required.

The electronic units described were all designed by the author, with the exception of the dual port memory, which was designed by Gavintech Ltd, who also assisted with electronic assembly. Paul Hammond joined the group in 1985 and took over the design of the printed circuit boards and much of the electronic assembly. Keith Smith and Jim Percival were responsible for the mechanical design and construction of the instruments. Kevin Page contributed to the manufacture and assembly of the more recent interferometers. Software development has passed though a number of hands. Peter Rounce designed and programmed the on-board computer for the balloon instruments and Dan Wade wrote and continually maintains the multi-function Halo programme for the PC computer controlled instruments. David Rees took responsibility for the overall concept of the instruments and the optical design, and also obtained the funding to make it all possible.

CHAPTER 2 - THE MICROCHANNEL PLATE

2.1 Introduction

The microchannel plate, or MCP, is the heart of the Imaging Photon Detector. It is the device which provides the amplification necessary to enable the positions of individual photons of light to be determined. It consists of a parallel array of electron multipliers in the form of a disc. A particle such as an electron or high energy (ultra-violet) photon incident on the input face gives rise to a pulse of many electrons from the output face at the same lateral position. The device can, therefore, be used for two dimensional image intensification. If the incident electrons are generated by a photocathode, it can be used to image low energy photons in the visible and near infra-red region. The electron pulse can be fed directly to a phosphor screen to give a visible image, or to a position sensitive anode from which the co-ordinates of each pulse can be determined. Microchannel plates are particularly well suited to Imaging Photon Detectors. They exhibit low noise and have low power requirements for operation. They are small in size, light in weight, and relatively immune to magnetic fields.

This chapter gives a brief description of their fabrication and the theory of their operation. A more detailed account can be found in the literature, for example *Wiza (1979)* and *Hamamatsu (1982)*. Different methods of combining microchannel plates for increased gain are discussed and the practical limitations of MCPs are considered.

There are two major problem areas when using MCPs. Firstly, ion feedback (*Adams and Manley, 1966*), which results in instability and can cause photocathode degeneration, and secondly, channel recovery time (*Eberhardt, 1981*), which limits photon count rates for point sources on the image.

2.2 Theory of operation

The MCP is a development of the channel electron multiplier, or channeltron, which can be thought of as a continuous dynode electron multiplier. Both devices are made from a special formulation of glass which is heavily lead-doped. Such glass exhibits resistive and good secondary emissive characteristics. Whereas the channeltron consists of a single tube of this special glass, the microchannel plate is an array of millions of capillaries each with an internal diameter in the region of 10 to 20 microns and each forming an independent channel electron multiplier. Figure 2.1 shows a diagrammatic representation of an MCP. An MCP can also be seen in figure 4.3, which is a photograph of some component parts of an IPD.

The array typically takes the form of a disc, 18 to 40 mm in diameter and 0.5 to 1.0 mm thick. A thin film of metal, usually nichrome or iconel, is deposited on each face of the disc in the area between the open ends of the channels. When a voltage is applied between the opposite faces of the MCP an electric field is created along each channel and a small current, the strip current, flows through the plates. An electron entering the negative end of a channel and hitting the wall will produce secondary electrons which are accelerated by the electric field. They in turn produce further secondary electrons. The process is repeated many times along the length of the channel until finally a cloud of electrons emerges at the opposite end. This is illustrated in figure 2.2. The gain of each channel is dependent on the ratio of the length to diameter, α, which is typically 40.

2.3 Fabrication

MCPs are made by fibre drawing techniques. Single fibres are made which have a central core of glass which is soluble in a chemical etchant, surrounded by a cladding of lead-doped glass. The fibres are packed into a hexagonal array, drawn to the required diameter and sliced. The core is then dissolved away and the plates are baked in a hydrogen furnace which reduces the lead oxide on the channel walls to lead.

The array is usually sliced at an angle - the bias angle. This ensures that primary electrons do not pass straight through without hitting the walls, and in a multi-stage device the plates can be arranged with the bias angles opposed to discourage ion feedback. This is known as the "chevron" or "V" configuration for two plates and the "Z" configuration for three plates, as shown in figure 2.3. The bias angle is typically between five and fifteen degrees.

An obvious problem is that the walls of the channels have a finite thickness and therefore electrons may be collected by the ends of the channel walls instead of entering the channels. The ratio of the area of open channel to the total cross-sectional area of the plate is called the open area ratio, A_o. This is often increased by funnelling the input to each channel, so that A_o can be effectively as large as 85%.

2.4 Gain and plate configuration

The gain of a microchannel plate is defined in terms of the ratio of emergent electrons to incident electrons. It will be seen that the gain of a single plate is limited, although higher gains may be obtained by combining plates into a stack.

2.4.1 Single plates

The secondary emission coefficient, δ, is defined as the average number of electrons emitted by an electron incident on the channel wall. Assuming that an electron entering the channel produces δ secondary electrons, then the second stage of multiplication produces δ^2, and the third δ^3, up to δ^n when the end of the channel is reached. Each channel can thus be thought of as a conventional n-dynode electron multiplier (*Eberhardt, 1979*), with n having a typical value of 20. This is a useful model, although it should be emphasised that there are many possible electron trajectories through the channel, and this is only an average value.

As the applied voltage increases so does δ because the collision energy is increased. However, the number of collisions, n, will decrease because of the higher electron velocity, which gives a greater distance between collisions. This leads to a flattening of the gain/voltage curve. A typical curve of gain versus applied voltage is shown in figure 2.4.

The pulse height distribution from a single MCP takes the form of a negative exponential. The upper limit of gain is determined by the onset of ion feedback. When the gain is raised beyond 10^4, ions are formed by electron collisions with residual gas molecules. The ions drift back to the channel input and can produce additional secondary electrons, causing instability through regenerative feedback. Additionally, in devices which use a photocathode to provide the electrons to the MCP, ions can accelerate back to the photocathode and quickly degrade it. The use of an ion barrier to prevent this is discussed later in this chapter.

2.4.2 Two and three plate assemblies

Higher gains may be achieved by mounting plates together in a stack. With the bias angles oriented in the chevron or Z configurations it is difficult for ions to travel backwards past the plate junctions. Gains of up to 10^7 can then be realised. The increase in gain is encouraged by the fact that the channels will not be perfectly aligned with each other, and the charge will, therefore, be spread over a number of channels. Statistically, the most likely mode is for a channel from the first plate to share its electrons amongst three channels of a second plate, which in turn will feed seven channels of a third plate. This is illustrated in figure 2.5.

As the gain of a channel plate stack is increased it is found that the pulse height distribution changes from a negative exponential to a quasi-Gaussian shape. This is the result of space charge saturation near the channel output (*Bryant & Johnstone, 1965; Schmidt and Hendee, 1966*). The high space charge density reduces the kinetic energy of the

electrons until the secondary emission coefficient, δ, is reduced to unity. The secondary electrons emitted from the walls are retarded by the lateral space charge field, and instead of traversing the channel are returned to the same side of the wall, without, in some cases, having gained enough energy to produce further electrons. This occurs when the electron energy falls below the region of 50 eV. This is illustrated in figure 2.6.

For pulse counting applications, such as the IPD, it is essential to run the channel plate stack in the saturated mode. The processing electronics can then be optimised so that a very high proportion of the charge pulses are of the correct amplitude for the position computing circuit. The negative exponential mode has a large proportion of pulses at very low amplitude which are indistinguishable from amplifier noise in the processing electronics. Also, as the channel plate gain, unlike the subsequent electronic amplification, is essentially a noise free process, it is most desirable to obtain the maximum gain possible from the plates.

Three plates in the Z configuration give the maximum gain realisable from a straight stack of plates, although four plates have routinely been used to improve reliability. The channel resistance has a negative temperature coefficient and any channel breakdown tends to be regenerative and catastrophic. A four plate stack gives a gain similar to the three plate stack, but with lower individual plate voltages. However, the pulse height distribution is somewhat broader which is undesirable for the processing electronics.

The quasi-Gaussian pulse height distribution is conveniently quantified by measuring the pulse height resolution, PHR. This is given by the width of the distribution curve at half of the maximum, normalised to the pulse height value at the maximum of the curve. It is referred to as the full width at half maximum, or FWHM value. Values of 100% to 150% are typical for three and four plate devices. An improved configuration which uses five microchannel plates can achieve figures of better than 70%. Figure 2.7 shows the unsaturated

negative exponential distribution. Figure 2.8 is a good example of a saturated quasi-Gaussian distribution, measured with an ITT five plate detector which shows a FWHM of 65%.

2.4.3 Five plate assemblies

It will be seen later that the position of the charge cloud is determined by a method which computes the centroid of its footprint on the anode of the detector. One advantage of this charge centroiding approach is that it is quite acceptable for the output pulse to be spread over a large number of channels, as long as there are sufficient electrons for the centroid to be determined accurately. A technique used for increasing the gain of the assembly is to spread the charge over more channels. This is achieved using a stack of five microchannel plates. Two plates in a chevron configuration supply the initial gain. The charge pulse then crosses a gap of about 1 mm, which enables it to spread out before being further amplified by a three plate Z stack. The gap is given a reverse bias of between 50 and 200 volts. Overall gains in excess of 10^8 can be realised.

The device needs careful tuning to optimise gain and pulse height distribution, but affords a considerable improvement in resolution because the signal-to-noise ratio of the anode signal is improved in direct proportion to the total MCP gain. This technique was first described by *Firmani et al (1982)*. Such devices, manufactured by ITT Electro Optical division, have been used by the author. The configuration is illustrated in figure 2.9 and the operating conditions are described fully in chapter six.

2.4.4 Curved microchannel plates

An alternative approach to realising a higher gain is to use a curved microchannel plate (*Timothy, 1981*). The distance an ion can travel before striking the wall of the channel is substantially reduced. The ion momentum is thereby kept to a low value and the probability of ion induced secondary electrons is minimised. The pulse height resolution is around 40% - much narrower than the 70% to 120% typical of straight

channel assemblies (*Laprade & Cortez, 1983*). This is because the charge is confined to a single channel. The gain of a single curved channel plate can be as high as 10^5. This is still considerably lower than that obtainable from a Z plate assembly. Curved plates are relatively expensive and difficult to manufacture, and have not been utilised as extensively as conventional straight plates.

2.5 The ion barrier

Although the above methods of minimising ion feedback permit stable operation of the microchannel plate stack, there will still be some ions emerging from the front plate which will cause the gradual destruction of the photocathode. A technique for minimising this is to put a thin film of aluminium oxide across the face of the first plate. This provides an effective ion barrier whilst allowing the passage of electrons. However, the film, which is 5 to 7 nm thick, is a poor transmitter of electrons unless a sufficiently high potential is applied between the photocathode and the microchannel plate assembly. Typically a potential of 400 to 700 volts is required for efficient transmission. The transmission decreases rapidly below 100 volts.

The values often quoted for transparency are those given by *Csorba (1979)*. However, these figures give an extremely pessimistic view of the situation. This is because the higher accelerating voltage also increases the secondary emission coefficient. The curve given by *Csorba* is derived by measuring the output current of the MCP, which in the linear mode used for image intensifiers is proportional to the secondary emission yield of the incident photoelectron. In the case of the pulse counting mode used for IPDs the yield is not important as long as each photoelectron generates a cascade. The curve will, therefore, level off at the point where the majority of photoelectrons generate at least one secondary electron.

This point was made in *Rees, McWhirter et al (1981a)* where we stated that any improvement beyond 400 volts accelerating potential is masked by the increased secondary emission coefficient. *Clampin et al (1988)*

have taken this to imply that we claim that there is no improvement beyond 400 volts, which is not the case. It is just not quantifiable with the particular arrangement used by *Csorba*. The figures given by *Read et al (1986)* are more in accord with the findings in *Rees, McWhirter et al (1981a)* and give a transmission value of 50% at 400 volts, levelling off to 60% at 600 volts, compared with 70% efficiency for an unfilmed plate. The lack of further improvement in transmission beyond potentials of 500 to 700 volts has been routinely observed when commissioning IPD tubes at UCL. Some variability has been noted between detectors. This is due to differences in the film thickness, and probably also to variations in the secondary emission coefficient at the mouth of the channels.

The presence of the film affects the overall quantum efficiency, not simply because there will always be some attenuation through the film, but also because secondary electrons liberated from the area between the channels will tend to be absorbed by the film rather than finding their way down the channels. It has been established (*Laprade et al, 1990b*) that the efficiency of an unfilmed plate will be up to 20% greater than that of a filmed plate. This finding is born out in *Read's* comparison.

There is some controversy as to whether the ion barrier is actually required for tri-alkali S25 type photocathodes. Some experimenters maintain that it is not necessary if sufficient care is given to electron scrubbing of the plates to remove residual gas (*Clampin et al, 1988*). Others conclude that the absence of a film results in progressive photocathode deterioration (*Durand et al, 1987*). Photocathode lifetime is, in any case, highly indeterminate, so it is hard to ascertain whether or not the demise of a specific device is directly attributable to the absence of an ion barrier.

At UCL an important requirement has always been to achieve the maximum possible lifetime for every device operating in the field and it is clear that the best chance of achieving this is by using the additional protection afforded by filmed plates. They have, therefore,

always been specified for the UCL detectors. Filmed plates are mandatory for gallium arsenide photocathodes as they are very sensitive to contamination, requiring a vacuum three orders of magnitude lower than the S25 devices.

2.6 Microchannel plate lifetime

Catastrophic voltage breakdown or photocathode degeneration (the former often causing the latter) is the usual failure mode for Imaging Photon Detectors. However, it is important to consider the useful life of the microchannel plates in the absence of the above failures.

The lifetime of an MCP is determined by the deterioration of the secondary emission coefficient of the channel walls due to the electron scrubbing, especially in the high gain region. Tests have shown that the gain degrades to 50% of its original value after an accumulated charge output of about 3 x 10^{-2} coulombs/cm^2. (*Hamamatsu, 1982*). In the author's experience, low channel gain has never been the cause of a device reaching the end of its useful life, although a small increase in voltage from time to time is helpful in keeping the gain up to the optimum value.

2.7 Magnetic field immunity

The immunity to magnetic fields is extremely good, two to three orders of magnitude better than conventional photomultiplier tubes (*Hamamatsu, 1982; Lo et al, 1977*). This is because of the small dimensions and correspondingly high electric field strengths which give short electron transit times between electrodes. In the limit, the radius of the cycloid motion of the incident electron becomes comparable with the inter-electrode gaps and the gain of the device drops dramatically. *Hamamatsu* and *Lo et al* have demonstrated that this happens at field strengths in the region of several hundred Gauss for devices of this type. However, for imaging detectors, as opposed to non-imaging photomultipliers, image distortion will occur at rather lower levels of magnetic field. For a single stack MCP device there are two inter-electrode gaps to consider. The first is the

photocathode to MCP gap, typically 0.5mm, which has a very high electric field of 1000 to 1500 v/mm. The second is the MCP to anode gap, typically 1 mm, which has an electric field of 100 v/mm. If extra immunity is required, this field can normally be increased without affecting the operation of the device. However, it will be seen in chapter four that a certain class of position sensitive anodes, known as Wedge and Strip, needs careful adjustment of the field. This is because the spreading out of the charge cloud is critical for correct operation. A larger MCP to anode gap (typically 3 to 5 mm) is also required, so magnetic immunity is somewhat less for devices incorporating these anodes and it may be necessary to provide a mu-metal magnetic screen around the device.

2.8 Dead time

When a pulse of electrons leaves a channel, an equal positive charge is created at the channel output. This suppresses the amplification of any following electron pulses until it is neutralised by the strip current of the IPD. The charge to be replaced comes mainly from the last dynode equivalent of the channel, or approximately 1/20 from the end. It is replenished by the strip current which charges an effective capacitance at this position. The storage capacitance of a single channel, C_c, has been estimated for single plates using a high current flooded input test pulse method (*Eberhardt, 1981*), and has been found to be 2.5×10^{-16} farads/channel.

This figure may also be estimated by considering the microchannel plate as a parallel plate capacitor of capacitance C_p, where

$$C_p = A(\epsilon_r(1-Y) + Y).\epsilon_0/l \qquad [2.1]$$

A is the MCP area, Y is the open area ratio $\simeq 0.65$, and l is the plate thickness = 0.5 mm, ϵ_r is the dielectric constant of the glass, which is 8.3 for Corning 8161, and $\epsilon_0 = 8.854 \times 10^{-12}$ $F.m^{-2}$.

This yields a value of 6.3 pF/cm^2, or 12.2 x 10^{-18} F/channel, given a channel density of 5.1x10^5 channels/cm^2 (for 12.5μ holes on 15μ centres).

Assuming most of the charge comes from the last dynode, the equivalent storage capacitance is given by n.C_p, where n is the number of notional dynodes, taken as 20. C_c is thus again found to be in the region of 2.5 x 10^{-16} F/channel. However, in the single photoelectron mode of operation as opposed to the flooded input mode, the axial capacitance, C_c, is augmented by a lateral capacitance between the group of active channels and the adjacent quiescent ones. It has been estimated (*Eberhardt, 1981*) that C_c is thus augmented by a factor of up to 50.

The equivalent circuit for the channel is shown in figure 2.10.
The recharge resistance is given by:

$$R_{rec} = \frac{R_c}{n+1} \qquad [2.2]$$

where R_c is the channel resistance, and n is, again, the number of notional dynodes.

Given a typical strip resistance of 8x10^8 ohms/cm^2,

$$R_c \simeq 4 \times 10^{14} \text{ ohms}$$

and so $$R_{rec} \simeq 2 \times 10^{13} \text{ ohms.}$$

The RC time constant of the recharge circuit, which is the recovery time to 63%, is, therefore:

$$50 \times R_{rec.} C_c = 50 \times 2\text{x}10^{13} \times 2.5\text{x}10^{-16} \text{ seconds}$$

$$= 250 \text{ milliseconds.}$$

This implies a surprisingly low maximum rate in the region of 4 counts/second per channel group, which is a serious limitation of the IPD when used for imaging applications involving bright point sources. Values of the measured maximum count rate per pixel are given in a

later chapter where it will be shown that the above estimate is not unrealistic for a worst-case situation. The calculations are, of course, based on a much simplified model, but they are nevertheless very useful for assessing microchannel plate performance to within an order of magnitude.

In general, the effect of channel recovery time on pulse counting systems is the lowering of the modal pulse height so that anode pulses from bright areas fall below the lower acceptance threshold of the electronics. This can manifest itself in a "doughnut" effect, where bright point sources are depicted as rings with dark centres. To counteract this, the dynamic range of the processing electronics should be as wide as possible and the lower electronic threshold set as low as possible, consistent with reliable position computing and noise levels.

It is obviously advantageous to use channel plates which have a low channel resistance when maximum count rates or channel gains are required because channel recovery time will be reduced in direct proportion to the increased strip current. However, the low resistance can result in excessive thermal dissipation (*Soul, 1971*). Special plates have now been developed which overcome this difficulty. They are known as High Output Technology (HOT) plates (*Laprade, 1989*).

2.9 Dark count

It is possible for electron cascades in the microchannel plates to be generated by events which are not originated by the input signal. These are usually indistinguishable from signal induced events except that they are present when the device is not illuminated. Such events are known as "dark counts" or collectively as "dark current". In the case of a detector with a photocathode, the majority of dark counts are usually caused by thermionic emission from the photocathode material. This effect will be discussed in chapter three. However, there are other mechanisms, associated with the microchannel plates, which can give rise to dark counts (*Laprade and Wheeler, 1990a*).

The most common sources of MCP generated dark counts are field emission and breakdown. These are due to contamination or defects in the MCP assembly. It is, therefore, important to exercise extreme care in the manufacturing process in order to minimise these problems. If the source of the events is a defect in the middle or towards the end of the MCP stack, the resulting output charge pulses will be smaller as the full gain of the stack will not be obtained. It may, therefore, be possible to exclude some of these events by careful adjustment of the threshold controls in the processing electronics. These controls allow only a certain range of pulse heights into the processor and can be set so that smaller than average pulses are rejected. However, the overall detection efficiency will be reduced as some signal induced events will inevitably be excluded. Another possible source of dark counts is thermionic emission from the channel surfaces. This is extremely low because of the high work function of the lead glass.

If the above causes can be eliminated, a residual dark current of around one count/cm^2.s remains. This is primarily caused by beta particles emitted by radioisotopes contained in the microchannel plate glass. This is negligible when compared with the thermionic emission from most visible light photocathodes, so it is not a major concern for the UCL projects. However, it is quite significant for other applications, such as X-ray astronomy, where very low signal intensities are being observed. Development work is currently underway to reduce this residual current by an order of magnitude (*Feller et al, 1989*).

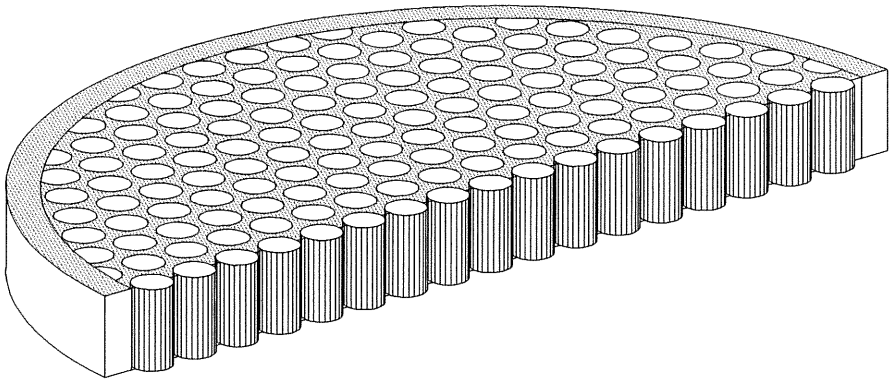


Figure 2.1

A section through a microchannel plate
(channel diameters much enlarged for clarity)

Figure 2.2

Electron trajectories

in a single channel

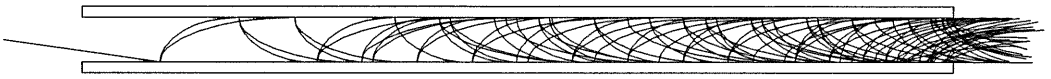


Figure 2.3

Microchannel plate configurations

Chevron or V

Z

Figure 2.4

Gain characteristics of a microchannel plate

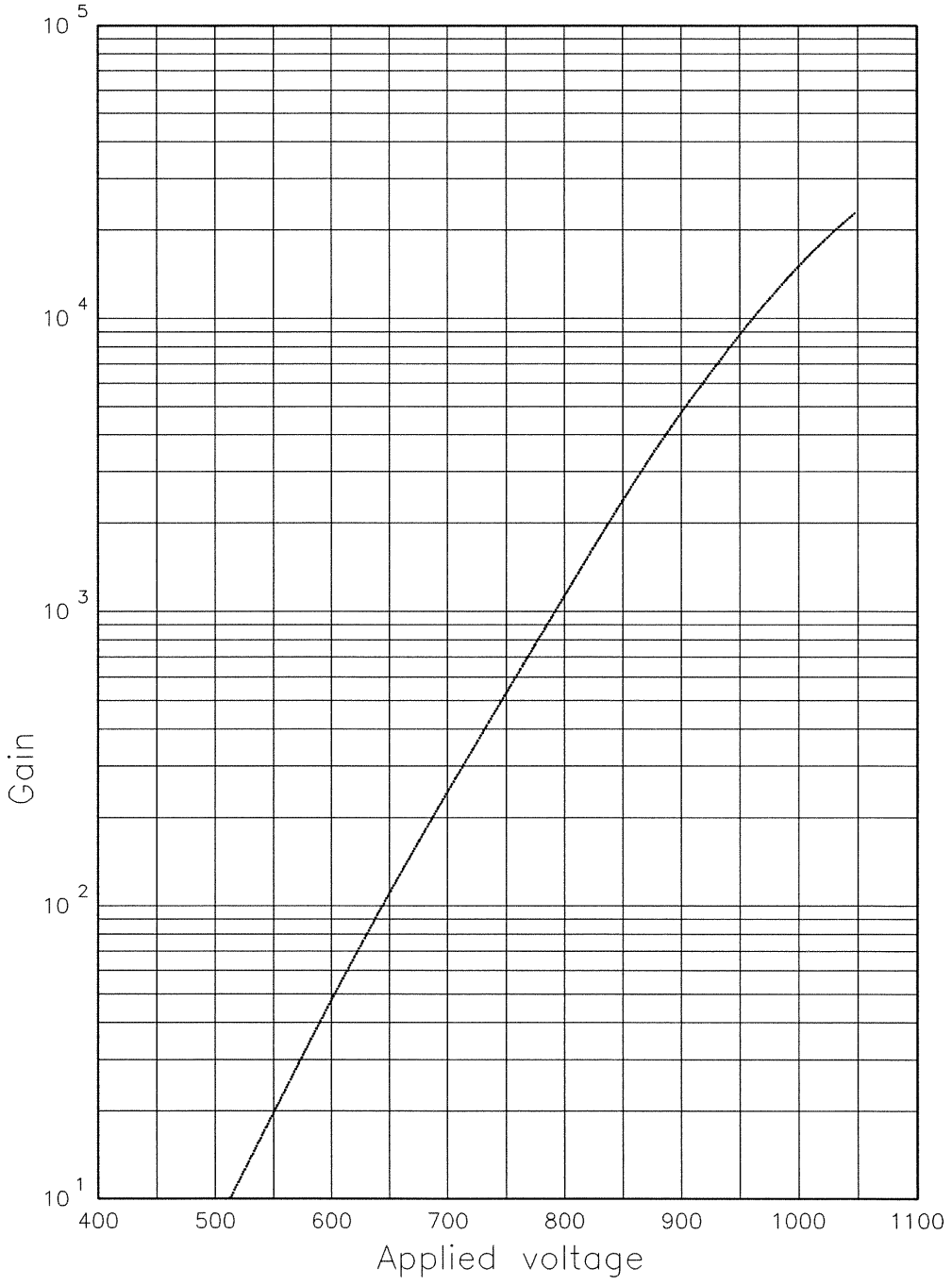


Figure 2.5

Charge sharing between channels of successive plates

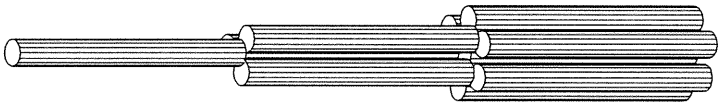
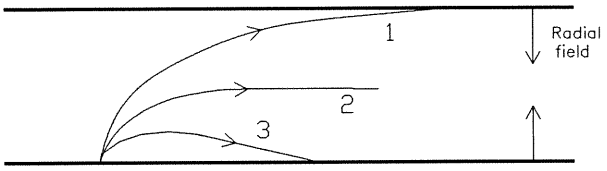


Figure 2.6

Gain saturation in a channel due to space charge



1. Electron able to cross channel and produce secondaries
2. Electron no longer able to cross channel
3. Electron returned to same side of channel with insufficient energy to produce secondaries

From Bryant and Johnstone

Figure 2.7

An unsaturated negative-exponential pulse height distribution

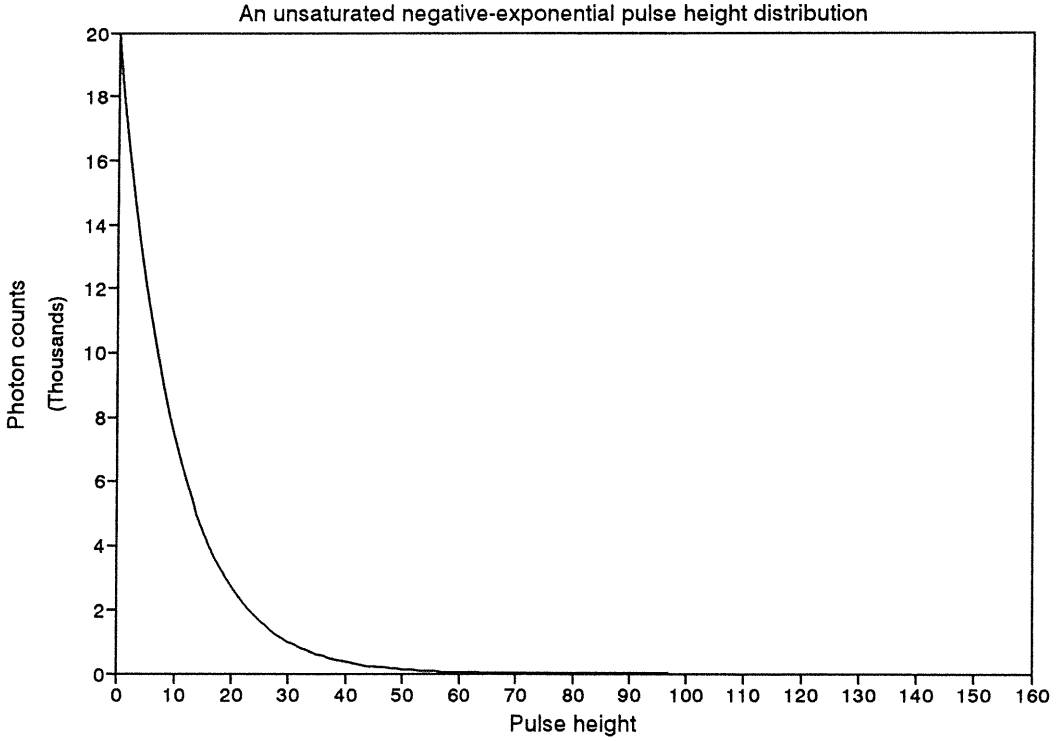


Figure 2.8

A saturated quasi-Gaussian pulse height distribution

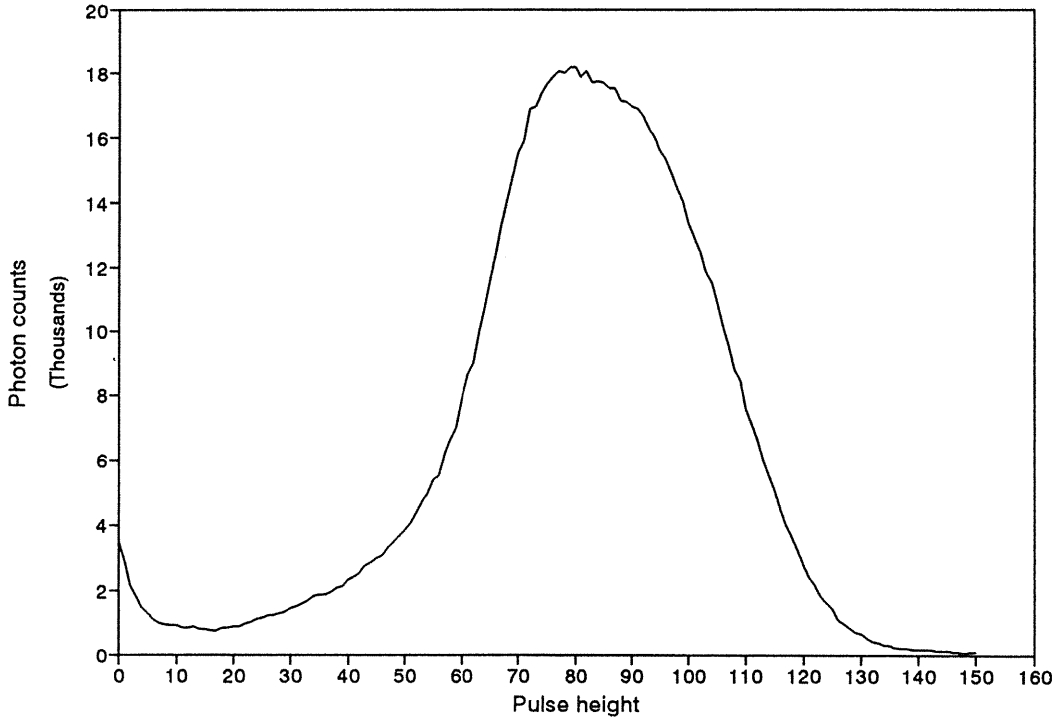


Figure 2.9

Five microchannel plate double stack configuration

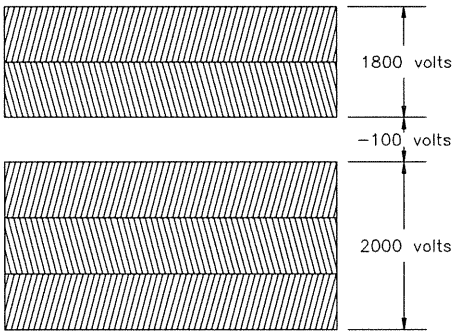
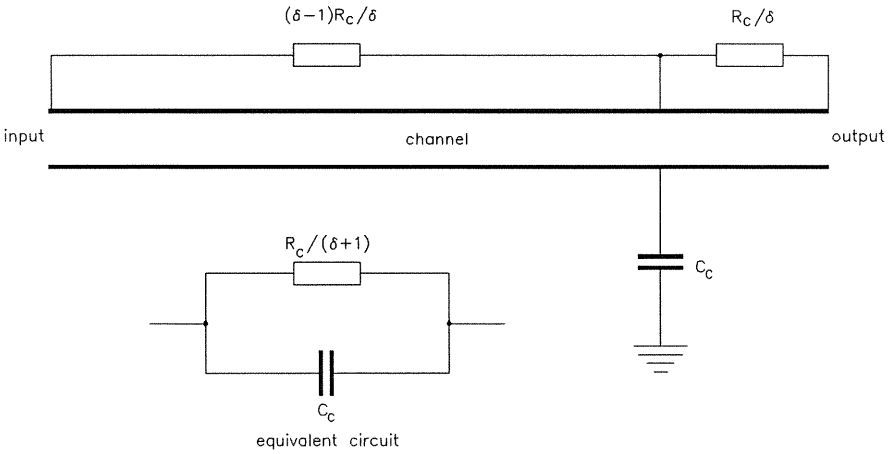


Figure 2.10

Channel recovery equivalent circuit



CHAPTER 3 - THE PHOTOCATHODE

3.1 Introduction

High energy radiation, such as ultra-violet photons and x-rays, can directly initiate a cascade of electrons in a microchannel plate. However, when MCPs are used for imaging radiation in the visible and infra-red regions, it is necessary to use a photocathode to convert the incident photons into electrons. This consists of a thin, carefully controlled layer of photoemissive material under vacuum from which electrons are emitted in response to photon stimulation.

Photoemission was first observed by Hertz in 1887, but it was not until 1905 that the work of Einstein led to an understanding of the process when he applied the quantum theory of Planck to the phenomenon. This theory relates the energy of a photon of light to its frequency. A photon arrives at the photocathode surface with an energy $h\nu$, where ν is the frequency of the light source and h is Planck's constant. This energy may be transferred to a free electron within the photocathode material. If sufficient energy is imparted to the electron it may then be able to escape from the photocathode to the vacuum.

Photocathodes may be classified as opaque or semi-transparent. In the opaque photocathode, the light is incident on a thick layer of photoemissive material and the electrons are emitted from the same side as that struck by the radiant energy. In the second type, the photoemissive material is deposited on a transparent substrate so that the electrons generated in the photocathode material are emitted from the side of the photocathode opposite the incident radiation.

A full account of the phenomenon of photoemission may be found in the literature (e.g. *Dance, 1969*). In this chapter, different types of

modern photocathode are described and several critical aspects of photocathode performance are discussed. These include:

1. Quantum efficiency, with its dependence on both the structure of the photocathode and the frequency of the incident light.
2. Thermionic emission, which imposes a lower limit on the useful sensitivity of the device.
3. Field emission, which can cause instability and "hot spots" on the image.

The chapter concludes with a discussion of the image degradation caused by the lateral dispersion of the emitted electrons as they cross the gap to the input face of the microchannel plate stack.

3.2 Photoemission in metals

According to quantum theory only one electron can occupy a particular quantum state, and in a single atom these states are separated into distinct shells. However, in metals and crystal lattices the states are modified by interaction with neighbouring atoms, resulting in continuous bands of permitted energy levels. The Fermi-Dirac energy distribution function gives the probability of occupation for a quantum state of energy E, and is expressed by:

$$p(E) = \frac{1}{1 + \exp[(E-E_f)/kT]} \qquad [3.1]$$

The quantity E_f is defined as the Fermi level, and when $E = E_F$ the probability of occupancy is one half. The expression is seen to be temperature dependent. The Fermi level is the maximum permitted energy level at absolute zero, but at higher temperatures thermal energy permits a few electrons to have slightly higher energies. The Fermi function is plotted for different temperatures in figure 3.1.

The energy diagram for the surface of a metal is shown in figure 3.2. The work function, Φ, represents the energy which must be given to an

electron at the top of the distribution in order to overcome the potential barrier at the metal-vacuum interface.

To summarise: in order to cause an electron to be emitted from the photocathode surface, the energy of the incident photon must be sufficient to raise the energy level of the electron to the Fermi level and overcome the work function of the photocathode material. Any excess energy is retained by the electron as kinetic energy.

3.2.1 Long wavelength response limit

The kinetic energy of an emitted electron is thus given by the equation:

$$E = \frac{mv^2}{2} = (E_i + h\nu) - (E_f + \Phi) \qquad [3.2]$$

where E_i is the initial energy of the electron in the photocathode. From equation [3.2] it can be seen that an electron at the Fermi level will just be able to escape when the energy of an incident photon, $h\nu$, is equal to Φ. There exists, therefore, a threshold photon energy below which electrons cannot be liberated. In practice, as the metal will not be at absolute zero, there will be some electrons above the Fermi level and the cut-off will not be abrupt. There is a corresponding threshold wavelength which dictates the long wavelength limit to the response of photocathodes, given by:

$$\lambda_c = \frac{hc}{\Phi e} \qquad [3.3]$$

(The numerical value of the electronic charge, e, is required to convert the value of Φ, expressed in electron volts, into joules)

This wavelength dependence is critical when selecting photocathodes for specific applications. For example, photons of the shortest visible wavelength, 380nm, will liberate electrons from materials with a work function of less than 3.2 eV, and an appropriate photoemitter

is relatively easy to realise. Photons in the near infra-red are only capable of imparting about 1.5eV of energy, so this region requires a more specialised photocathode.

Typical metals have work functions of 4 to 5 eV, and so are only of use as ultra-violet photoemitters. The alkali metals, however, have particularly low work functions as the single electron in the outermost orbital is relatively easily stripped from the atom. Their work functions range from 1.8eV for caesium to 2.4eV for sodium. They can, therefore, be used for visible light photocathodes. The response of caesium extends furthest into the red and it is commonly used to extend the red response of photocathodes.

3.2.2 Quantum efficiency

Only a fraction of electrons incident on a photocathode will produce electrons. The ratio of emitted electrons to incident photons is termed the Quantum Efficiency, and is usually expressed as a percentage. A number of factors conspire to decrease this efficiency. Many incident photons may be lost by reflection at the surface of the material. Additionally, only electrons liberated close to the surface stand a reasonable chance of escaping before their energy is lost to the surrounding material. This is due to collisions, either with other electrons (electron scattering) or with the lattice (phonon scattering). This constraint is quantified by a parameter known as the Escape Depth, which is the maximum distance beneath the surface from which an electron can escape.

For this reason the thickness of semi-transparent photocathodes is critical. If the film is too thick, much of the incident radiation is absorbed at a distance from the vacuum interface greater than the escape depth, whereas if the film is too thin, the incident radiation will be lost by transmission.

3.2.3 Thermionic emission

The lowering of the work function of a photocathode material has an unfortunate side effect. Electrons are also liberated by virtue of their thermal energy, gained from the bulk of the material. A low work function makes it easier for these thermal electrons to escape. This effect is known as thermionic emission and electrons so generated are indistinguishable from photoelectrons. Photocathodes optimised for the best response at the red end of the spectrum have the highest thermionic emission. This can be as high as several thousand electrons per square centimetre of photocathode at room temperature. It is important to bear this in mind when selecting a photocathode for a particular application. If a device with good red sensitivity is used for the detection of blue light it will have a higher thermionic emission than is strictly necessary.

It is possible to reduce thermionic emission by cooling. A number of IPDs have been fitted into a purpose-made Peltier cooled housing, which will be described in a later chapter. Cooling the photocathode to -20°C can reduce thermionic emission dramatically, to a few tens of electrons per square centimetre.

3.3 Semiconductor photocathodes

Photoemission technology progressed rapidly following the invention of the silver-oxygen-caesium photocathode in 1929 (*Kroller*), and a number of effective photoemissive compounds were subsequently discovered. It was, however, an empirical science, and it was not until the more recent advances in solid state physics that the nature and behaviour of these compounds became understood. They are, in fact, semiconductors.

The difference between metals and semiconductors is described in terms of the band theory of solids, which relates to electrical conduction and photoemission. According to quantum theory, electrons in a lattice are contained within an energy band which is sub-divided into an inner "valency" band and an outer "conduction" band. Electrons in the

valency band cannot move from the atom since the band is fully populated and there are few vacant states in neighbouring atoms for the electrons to move to. If, however, the electrons can transfer to the higher energy conduction band they are free to move and can contribute to electrical conduction. In a metal the valency and conduction bands actually overlap, as shown in figure 3.2, so the electron mobility is high and the material is a good conductor. In other materials, however, there is a forbidden gap between the two bands. Electrons cannot become mobile until they receive enough energy to cross the energy gap into the conduction band. If the gap is greater than about 2eV then the material is an insulator. If less, then the material has special properties and is known as a semiconductor. The energy band model for a semiconductor is shown in figure 3.3. At absolute zero the valency band is full and the conduction band empty. At higher temperatures there are as many electrons thermally excited into the conduction band as there are empty states in the valency band, so the Fermi level is half way between the two. There are some effects which cause E_f to vary slightly, but for most purposes it can be taken as being at the middle of the gap for an intrinsic (un-doped) semiconductor. The work function is now the sum of two components: the band gap, E_g, and the distance from the bottom of the conduction band to the vacuum level, known as the electron affinity, E_a.

Semiconductors are much more efficient photoemitters than metals for three reasons:

1. They are less reflective, so fewer incident photons are lost.
2. The escape depth is greater. Energy loss by electron scattering is low as there are relatively few free electrons. Phonon scattering is the predominant loss mechanism.
3. Materials can be synthesised with lower work functions than metals, typically 0.8 to 2.2eV.

Photocathodes are now manufactured to agreed commercial formulae, one designation being the JEDEC S-series. The silver-oxide-caesium photocathode, the first important formulation to be used commercially,

is designated S1, and another popular type, Cs_3Sb is known as S11. One of the most useful photocathodes for the scientific programme at UCL has been the multi-alkali Na_2KSb:Cs (*McCarroll et al, 1970*). The designation :Cs indicates a trace quantity of the element. This formula is designated S20 or S25. It is used in Generation II image intensifiers and may be referred to as a Gen. II photocathode. The S25 has an extended red response, at the expense of reduced blue response and higher thermionic emission. This type of photocathode is more difficult to prepare and less reproducible than the S11 type because the ratios of the alkali metals must be very carefully controlled. In practice this means that the photosensitivity is monitored during production. Typical photocathode responses are shown in figure 3.4.

3.4 Fabrication of multi-alkali photocathodes

These photocathodes are usually formed by evaporation. A layer of antimony, about 5nm thick, is deposited onto the substrate. The remaining elements are then introduced in vapour form as the device is heated in order to allow the elements to react. The alkali metals are added very carefully until the required responsivity is achieved. The final spectral response is highly dependent on the precise conditions under which the photocathode is manufactured.

3.5 Negative electron affinity photocathodes

Dramatic improvements have been made in the efficiency of photoemitters by the development of techniques which modify the energy band structure at the surface of the material. A condition is created known as "negative electron affinity" (NEA) (*Rouget & Baud 1979*). This describes a state where the vacuum level is lower than the bottom of the conduction band. NEA photocathodes generally use gallium arsenide as the photoemitter. They are used in Generation III image intensifiers and are, therefore, often called Gen. III photocathodes. The energy levels at the surface are modified by the addition of a molecular layer of caesium and oxygen. The nature of the Cs-O layer has not been precisely determined and several possibilities are discussed by *Rouget and Baud*.

The effect of NEA is to increase the escape depth by as much as 100 times. An excited electron drops to the bottom of the conduction band in a very short time, about 10^{-12} seconds. However, it can stay there for much longer, about 10^{-10} seconds, before dropping into the valency band. During this time the electron will still be able to escape to the lower vacuum level. The escape depth is equal to the minority carrier diffusion length, in the region of 3 μm (*Csorba, 1979*). The escape depth for electrons in the valency band is dictated by the hot carrier diffusion length, typically a few tens of nanometres.

3.6 Fabrication of negative electron affinity photocathodes

The technique used for manufacturing NEA photocathodes is quite different from the evaporation method used for multi-alkali photocathodes. The photocathode is formed on a crystal of gallium arsenide which is made by epitaxial growth techniques. The active GaAs layer is approximately 1μm thick. A suitable substrate is required with similar lattice parameters and good optical transparency. This may be gallium aluminium arsenide or gallium phosphide. The substrate is then bonded onto the glass window of the device. The substrate transparency drops off rapidly in the green and blue regions of the visible spectrum and limits the short wavelength sensitivity of the gallium arsenide IPD. The assembly is subjected to heat cleaning in high vacuum prior to exposure to caesium and oxygen. The manufacture is complicated by the need to maintain a vacuum of 10^{-10} Torr. This is necessary in order to create and maintain the very thin (0.1 nm) caesium-oxygen layer, which amounts to only 4 to 5 molecules. A detailed account of the fabrication of NEA photocathodes is given in *Csorba (1985)*.

3.7 Measurement of photocathode sensitivity

The sensitivity, or radiant responsivity, of a photocathode is conveniently expressed in terms of milliamperes of photoelectron current per watt of incident radiation at a specific wavelength. This is measured by illuminating the full area of the photocathode with a collimated beam of light derived from an incandescent source. The

light is first passed through a narrow-band filter of about 10 nm bandwidth. The photocathode is held at a potential of around fifty volts negative with respect to the microchannel plates, the electrodes of which are connected together for the purposes of this measurement. A sensitive current meter placed in series with the microchannel plate connection gives the photoelectron current. In the APL test facility, the intensity of the light is measured with a sensitive radiometer, the International Light Corporation IL1700, which is calibrated for the appropriate wavelength in accordance with the manufacturer's data. A set of filters covers the visible and near infra-red regions of the spectrum at 50 nm intervals. This enables a complete response curve to be obtained. A curve is usually supplied for each IPD by its manufacturers who perform the measurement in a similar manner.

The quantum efficiency of the photocathode is obtained by calculating the ratio of incident photons per second, n_p, to liberated photoelectrons per second, n_e.

The incident power, P, at a given wavelength, λ, is given by:

$$P = n_p.h\nu = \frac{n_p.hc}{\lambda} \qquad [3.4]$$

The photocurrent, I, is given by:

$$I = n_e.e \qquad [3.5]$$

where e is the electronic charge.

Combining [3.4] and [3.5] gives:

$$\frac{n_e}{n_p} = \frac{I}{P} \cdot \frac{hc}{e\lambda} \qquad [3.6]$$

Assigning values to the constants gives the convenient expression:

$$QE\ \% = \frac{124.S}{\lambda} \qquad [3.7]$$

where λ is in nanometres and S is the photocathode sensitivity in milliamps per watt.

The efficiency thus measured is sometimes referred to as the RQE, or the Responsive Quantum Efficiency. This is to distinguish it from the DQE, or Detected Quantum Efficiency, which is the ratio of incident photons to photon events detected by a complete system, and includes losses incurred in the microchannel plates and processing electronics.

Sensitivity is also sometimes expressed in the photometric unit of microamps per lumen. This is the luminous responsivity, the lumen being the unit of luminous flux. Photometry has its origins in visible light measurement, and photometric measurements are weighted in accordance with the response of the human eye, for which a standard response has been agreed in terms of lumens per watt for each wavelength. This means that a single figure can be given which has a meaningful value with respect to the perceived responsivity of the human eye. However, most applications of the IPD are not concerned with the human eye response, and the form of photocathode sensitivity curves is not only quite unlike that of the eye, but differs greatly from one type of photocathode to another. The photometric sensitivity is, therefore, only of use for general comparisons between photocathodes of similar type and is no substitute for a full radiant response curve. It is particularly important to examine the full response curve when selecting a photocathode for a specific application.

3.8 The proximity focussed lens

It is appropriate at this point, having considered the nature of the photocathode and discussed the microchannel plates in the previous chapter, to consider the interface between the two, which is known as the proximity focussed lens. This is a curious designation, as it is not really a lens at all and there is no attempt to focus in the true sense. It is simply a gap across which lateral movement of the photoelectrons is minimised. This is done by making the gap as small

as possible and applying the highest possible accelerating voltage, thereby maintaining the integrity of the image. This gap is a major cause of image degradation in the IPD system, so it is important to quantify it.

3.8.1 The Gaussian distribution

Most forms of image degradation caused by random fluctuations follow a Gaussian distribution. This describes the probability distribution of any quantity which arises out of the sum of the effects of a large number of separate contributions. It is given by:

$$F(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \cdot \exp(-x^2/2\sigma^2) \qquad [3.8]$$

where

$$\sigma^2 = \overline{x}^2 \qquad [3.9]$$

The quantity x is the deviation of an event from the mean value, and σ is the root mean square value, or RMS value of the deviation.

Although the RMS deviation is the quantity usually derived from theoretical calculations, practical measurements are more easily obtained in terms of Full Width at Half Maximum, or FWHM, the term introduced in paragraph 2.4.2. It is, however, a simple matter to convert from one to the other.

The value at the centre of the distribution, defined as $x = 0$ is given by:

$$F(0) = \frac{1}{\sigma\sqrt{(2\pi)}} \qquad [3.10]$$

For the half-maximum points of the distribution:

$$F(x) = 0.5 \cdot F(0) \qquad [3.11]$$

thus, at the half maximum points:

$$\exp(-x^2/2\sigma^2) = 0.5 \qquad [3.12]$$

$$\frac{x^2}{\sigma^2} = \ln 0.5 \qquad [3.13]$$

so the half width, x, is given by:

$$x = 1.1774\ \sigma \qquad [3.14]$$

Therefore:

Full Width at Half Maximum = 2.3548 x RMS deviation [3.15]

3.8.2 Radial energy

The image degradation due to lateral motion of the photoelectrons as they cross the photocathode to MCP gap is a function of the average radial velocity of the electrons leaving the photocathode. This, in turn, depends on the radial component of the kinetic energy of the emitted electrons. The distribution of energies has been investigated by *Eberhardt (1977)*. It is highly dependent both on the wavelength of the incident light and the type of photocathode. *Eberhardt* has determined that the average radial energy ranges from 50meV for red light to 300meV for blue light.

3.8.3 Electron ballistics

The image degradation can be assessed by addressing the electron ballistics, as described by *Eberhardt*. If two plane surfaces are separated by a gap L with a potential V between them, an electron leaving one surface is subjected to an acceleration of eV/mL where e and m are the electronic charge and mass respectively.

The equation of motion is thus:

$$L = v_{oz}t + \frac{1}{2} \cdot \frac{eV}{mL} \cdot t^2 \qquad [3.16]$$

where v_{oz} is the initial velocity of the electron along the axis of the detector.

The initial energy of the photoelectron, of the order of one electron volt, is very small compared with the energy of several hundred electron volts supplied by the accelerating potential. It is, therefore, reasonable to ignore v_{oz}, and the equation rearranges to:

$$t = L(2m/eV)^{1/2} \qquad [3.17]$$

The radial energy in terms of electron volts is given by:

$$V_r = mv_r^2/2e \qquad [3.18]$$

where v_r is the radial component of the velocity.

Combining [3.17] and [3.18] gives an expression for the radial displacement r of the photoelectron:

$$r = v_r t = 2L(V_r/V)^{1/2} \qquad [3.19]$$

The probability $p_r(r)dr$ of finding an electron with a radial displacement between r and r + dr is thus related to the probability of finding an electron with a radial energy between V_r and $V_r + d(V_r)$:

$$p_r(r) = p(V_r) \cdot \frac{dV_r}{dr}$$

$$= p(V_r) \cdot \frac{rV}{2L^2} \qquad [3.20]$$

The normalised current density distribution j(0)/j(0) for a point source of illumination is found by dividing $p_r(r)dr$ by the annular area, $2\pi r.dr$.

Thus:

$$\frac{j(r)}{j(0)} = \frac{p(V_r)V}{4\pi L^2} \qquad [3.21]$$

The normalising constant j(0) is dependent on the experimental configuration.

The point spread function (PSF) is given directly by the current density distribution, which is seen in [3.21] to be a linear function of p(Vr). Equation [3.19] can, therefore, be used to substitute r for V_r in the radial energy distribution to give the point spread function.

Taking the radial energy distribution to be Maxwellian, as demonstrated by *Eberhardt*, the probability distribution of radial energies is given by:

$$p(V_r) = p(0).\ \exp(-V_r/\bar{V}_r) \qquad [3.22]$$

where $\bar{V}_r$ is the average radial energy and p(0) is the probability of zero radial energy.
Rearranging [3.19] gives:

$$V_r = \frac{r^2V}{4L^2}$$

and thus:

$$p(V_r) = p(0)\ .\ \exp(-r^2V/4L^2\bar{V}_r) \qquad [3.23]$$

The normalised point spread function is thus given by:

$$PSF(r) = \exp(-r^2V/4L^2\bar{V}_r) = \exp\ (-r^2/2\sigma^2) \qquad [3.24]$$

This has the form of a Gaussian distribution and σ is, therefore, the root mean square deviation, given by:

$$\sigma = L(2\bar{V}_r/V)^{1/2} \qquad [3.25]$$

or in terms of full width half maximum:

$$\text{Point spread function FWHM} = 3.33L(\bar{V}_r/V)^{1/2} \qquad [3.26]$$

The RMS deviation is thus proportional to the gap and inversely proportional to the square root of the applied voltage. This is illustrated graphically in figure 3.5, which shows the relationship between resolution and applied voltage for different gaps, and figure 3.6, which depicts the resolution as a function of gap. The radial energy has been taken as 200meV, in the middle of the range determined by *Eberhardt*.

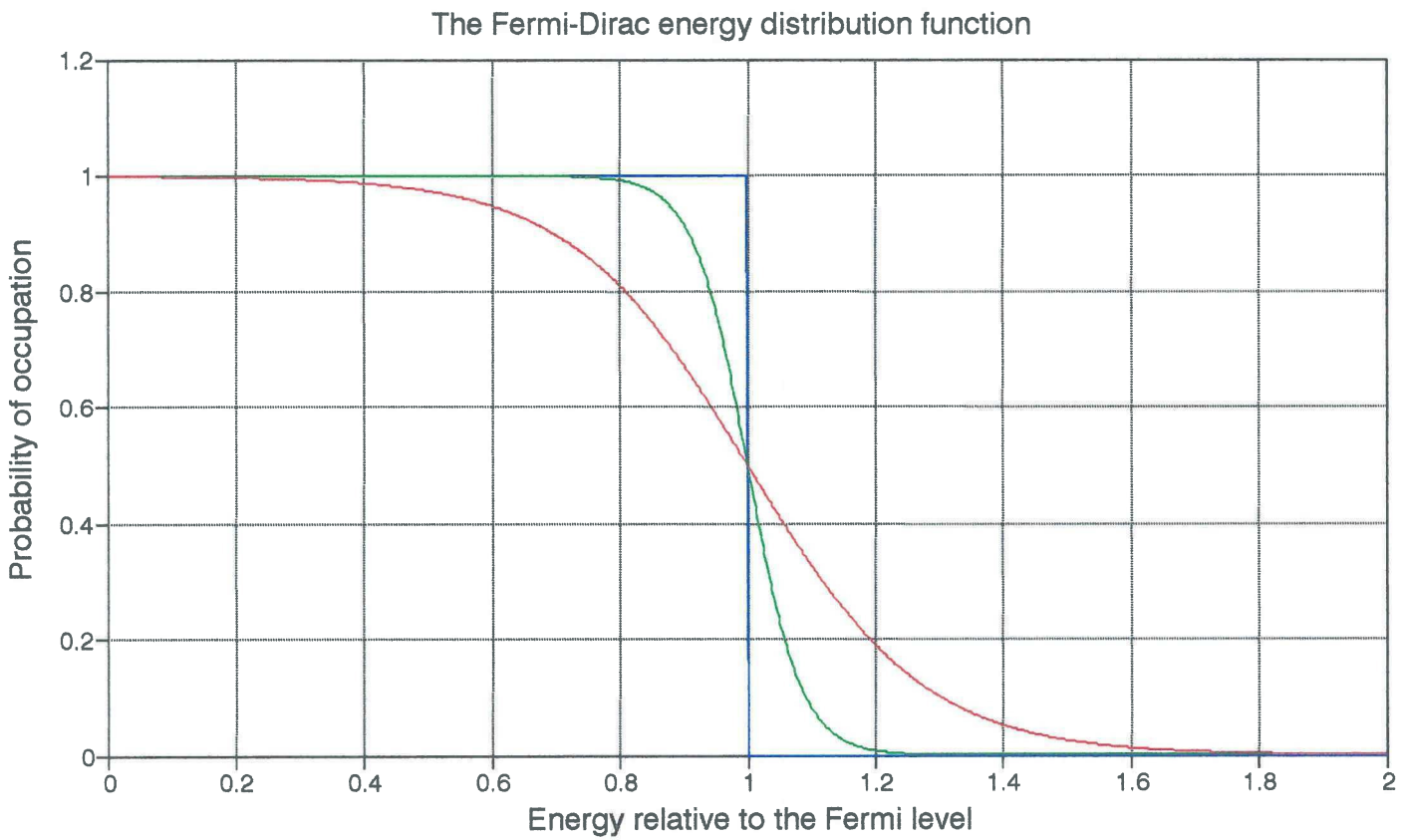
3.8.4 Optimisation of the gap size and potential

Equation [3.25] may be re-stated replacing V with Le, where e is the electric field strength. This gives:

$$\sigma = (2L\bar{V}_r/e)^{1/2} \qquad [3.27]$$

from which it is apparent that a smaller gap, L, will allow a smaller field strength, e, for the same resolution. If the required resolution is obtained by reducing the gap rather than by increasing the applied voltage, the field strength can be minimised. This is desirable because the photocathode may not be perfectly smooth, due to manufacturing defects. The electric field will be much stronger around any sharp points, resulting in field emission, This can cause "hot spots" on the image. However, when an ion barrier is used on the front of the channel plate stack, significant transmission losses can be incurred if the overall voltage is too low, as explained in the previous chapter. A potential in the region of 700 volts is needed to ensure adequate penetration of the barrier. This will result in a very high field strength if the gap is too small. A gap of 500 microns is normally a satisfactory compromise.

Figure 3.1


The Fermi-Dirac energy distribution function
Probability of occupation
1.2
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Energy relative to the Fermi level
0°K
300°K
1000°K

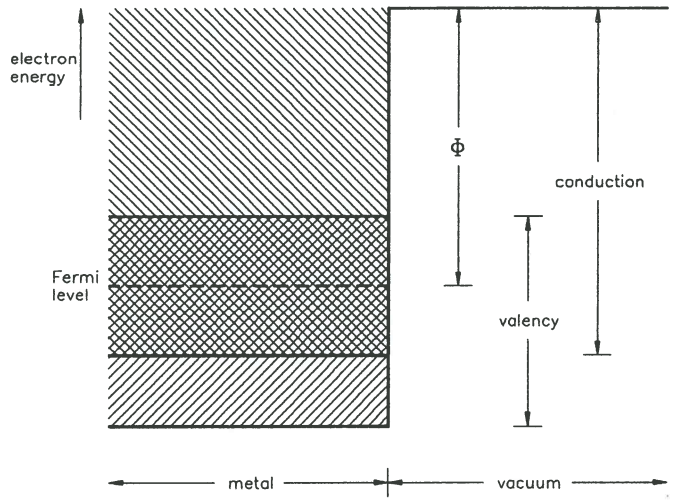


Figure 3.2
Energy levels in metals

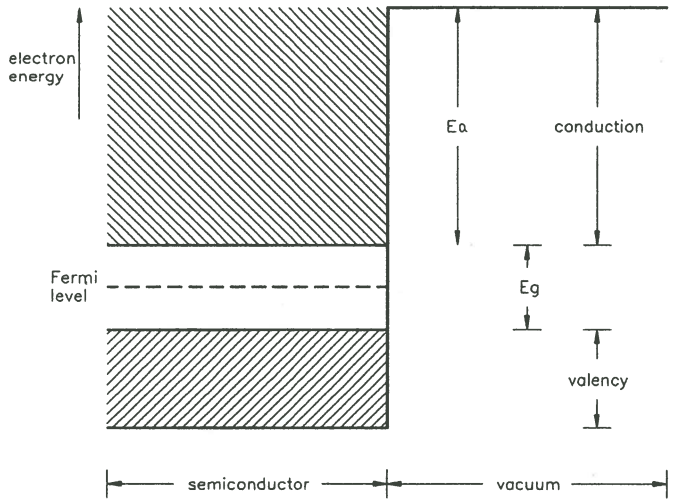


Figure 3.3
Energy levels in semiconductors

Figure 3.4
Photocathode sensitivities

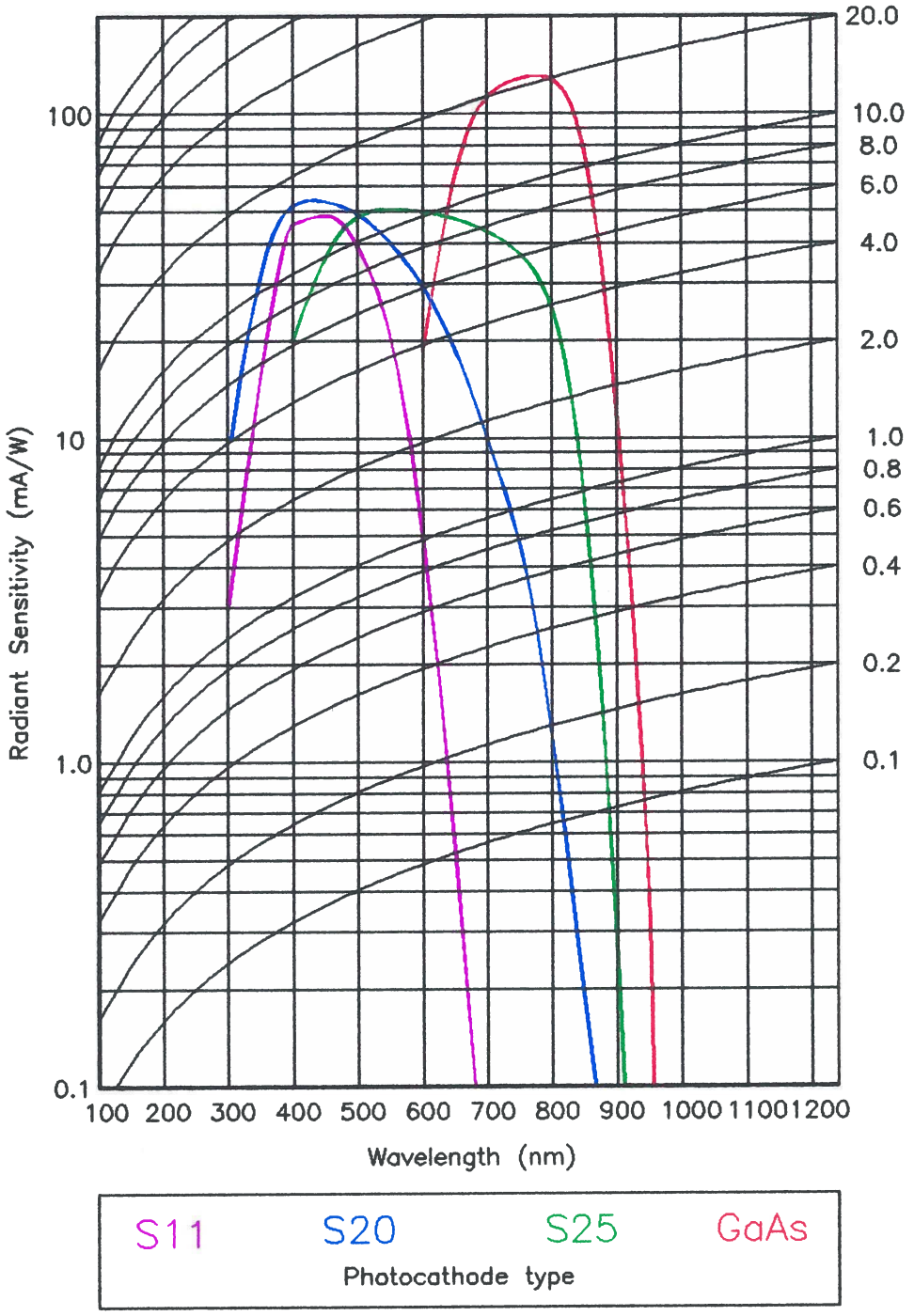


Figure 3.5

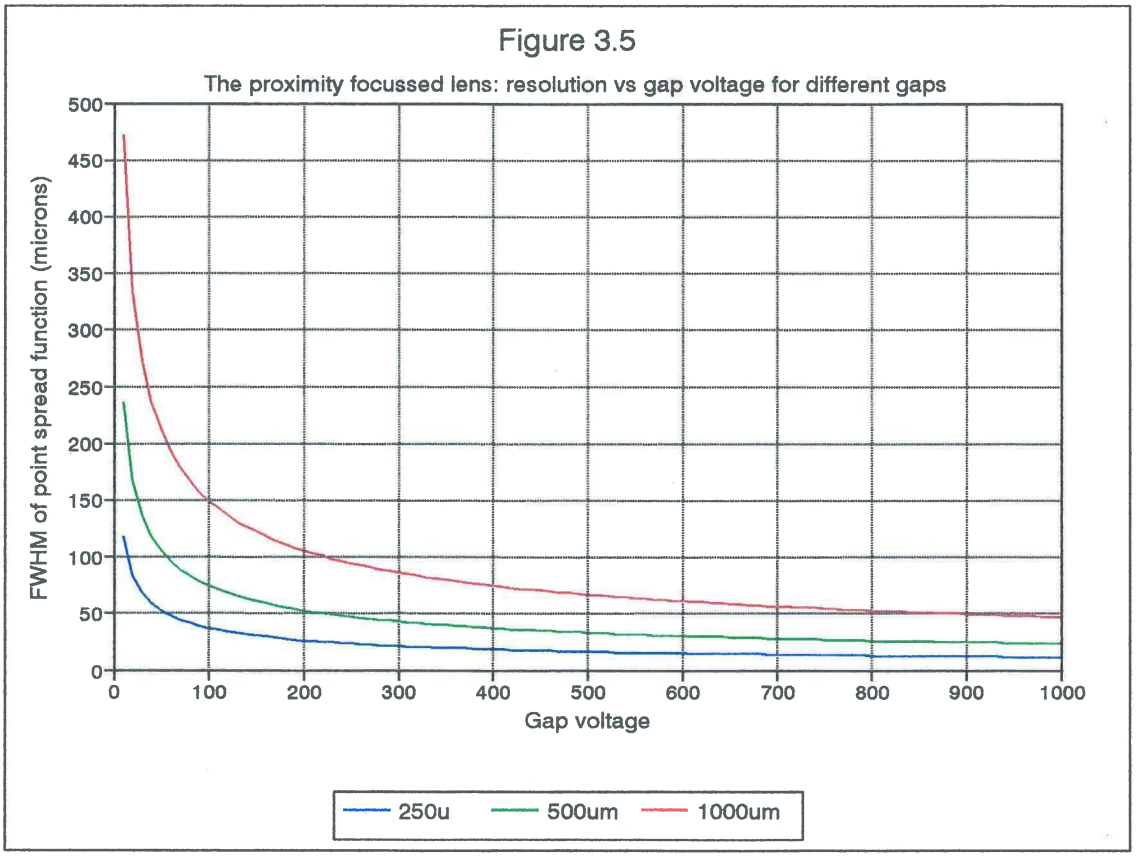
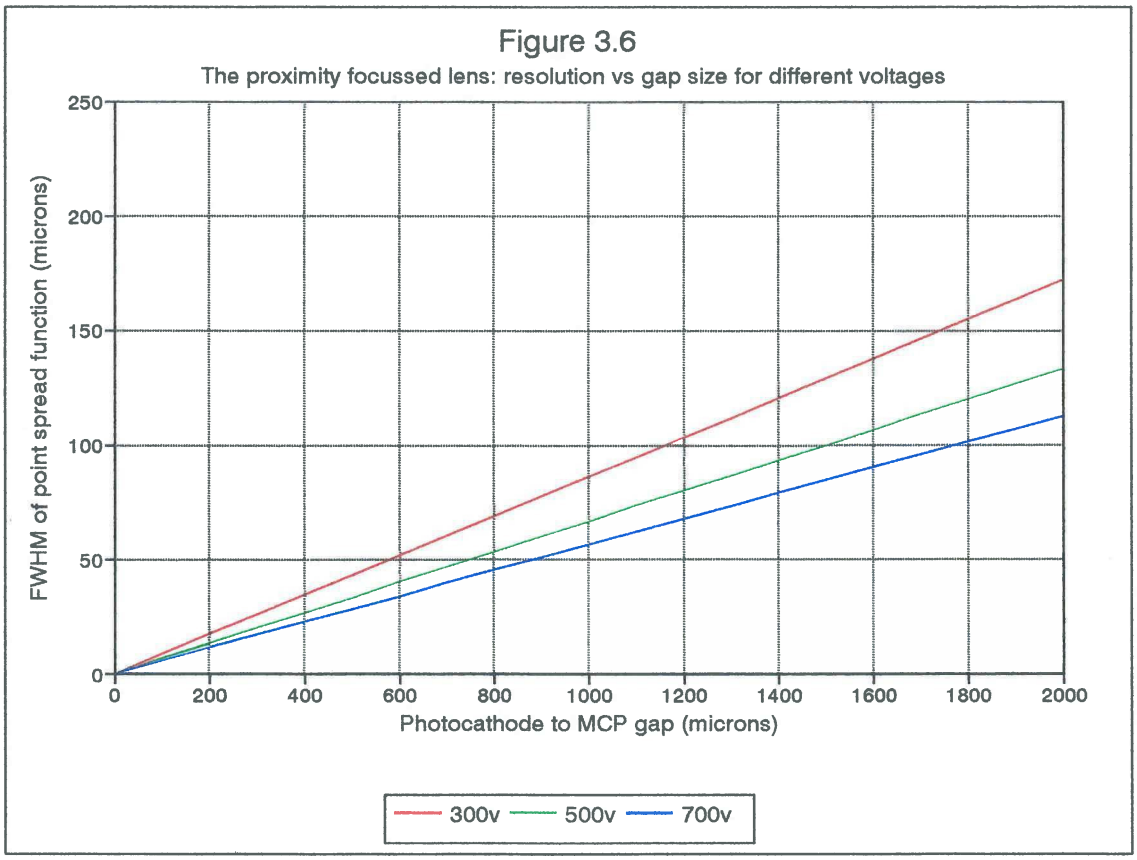


Figure 3.6



CHAPTER 4 - POSITION SENSING ANODES

4.1 Introduction

In a microchannel plate device the output charge pulse from the MCP stack may be directed at a phosphor screen to produce a visible image. Alternatively, it may be collected by an electrode held at a more positive potential, termed an anode. Ordinary photomultipliers use a metallic anode which simply records the presence of a pulse. For positional determination something more complex is required. This chapter investigates the theory of resistive anodes as position sensing devices, and introduces the concepts involved in the position decoding electronics. Image degradation due to the electronic noise inherent in the anode is discussed. An alternative position sensing anode, the Wedge and Strip anode, is compared with the resistive anode.

4.2 Resistive line theory

A variety of resistive elements have been used for position determination, the simplest being the one dimensional resistive line. There are essentially two approaches for obtaining positional information from resistive elements: rise-time measurement and charge division.

In order to derive the positional information, the signals from the anode must be amplified as they are much too small to be used directly in position computing electronics. The required characteristics of the amplifiers are unusual in that they do not need to reproduce faithfully the shape of the input signal, in fact, it is undesirable that they do so. The output signal needs simply to retain all its positional information whilst noise and pulse overlap must be minimised.

The resistive line may be considered as a transmission line with associated resistance, capacitance and inductance. A charge pulse

incident on the anode will give rise to a time-dependent voltage, described by the telegrapher's equation (*Radeka and Rehak, 1979*):

$$\frac{\partial^2 u}{\partial x^2} = LC \frac{\partial^2 u}{\partial t^2} + RC \frac{\partial u}{\partial t} \qquad [4.1]$$

where u = voltage

x = position coordinate along line

t = time

R,L,C = resistance, capacitance and inductance per unit length.

The term in LC represents electromagnetic wave propagation and the term in RC represents diffusion. The former may be ignored if the characteristic impedance of the line is significantly smaller than the ohmic resistance of the total line, i.e. $2\pi \sqrt{(L/C)} << R\ell$ (where ℓ is the line length).

4.3 The rise-time method

If a charge lands on the anode at time t = 0 then it is clear that, as time increases, charge will spread along the line, eventually resulting in the same potential at all points if the line is unterminated. If the line is terminated at each end by high impedance amplifiers, the shape of the voltage pulse measured by each amplifier will depend on the point of charge injection. The rise-time of the pulse may be taken as a measure of distance along the line. However, this method has been found to be less satisfactory than the charge ratio method, for reasons explained in the next chapter.

4.4 The charge amplitude ratio method

If the line is terminated at each end by a virtual ground connection, such as a charge amplifier, the point of injection may be determined by the ratios of charges flowing out of each end of the line.

The charge spreading is described by the diffusion equation [4.1], neglecting inductance:

$$\frac{\partial^2 u}{\partial x^2} = RC \frac{\partial u}{\partial t} \qquad [4.2]$$

This is similar in form to heat flow equations and is solved by the methods given in the text book by *Carslaw & Jaeger*.

The instantaneous current, i, is related to the voltage gradient by:

$$i = \frac{1}{R} \cdot \frac{du}{dx} \qquad [4.3]$$

and integrating with respect to time to give the charge flowing into the terminating amplifiers realises the expression:

$$\frac{Q(t,x)}{Q_{in}} = 1 - \frac{x}{\ell} - \sum_{n=1}^{\infty} \frac{2}{n\pi} \cdot \sin(n\pi x/\ell) \cdot \exp(-n^2\pi^2 t/RC\ell^2) \qquad [4.4]$$

where Q_{in} is the charge incident on the anode (*Radeka 1974*).

The bulk of the expression is seen to consist of a time-decaying summation. The remaining terms show that when the sum becomes negligible the fraction of accumulated charge is a linear function of the position of the incident charge. The time constant for the decay is $RC\ell^2/\pi^2$ and in the d.c. limit the resistive line is an ideal linear position sensor (*Kalbitzer and Melzer, 1967*).

4.5 The two dimensional resistive anode

The resistive sheet anode is an extension of the one dimensional resistive line device. The diffusion is characterised by the "heat-flow" equation in two dimensions (*Bronwell, 1953*):

$$\nabla^2 u = RC.\frac{du}{dt} \qquad [4.5]$$

However, the simple boundary conditions applied to a resistive line are no longer applicable as the sheet is terminated at four discrete points on its perimeter.

Fraser and Mathieson (1981a,b & c) have investigated the distortions arising from a variety of terminations. They concluded that all simple two dimensional linear resistive anodes are subject to distortion, which may be minimised by careful selection of the anode and signal processing parameters. A more complex resistive anode, described in the next paragraph, is comparatively distortion free.

Rise-time measurement has been successfully employed as a method of position decoding for two dimensional anodes (e.g. *Owen and Awcock, 1968; Stümpel et al, 1973*). This method was originally considered to offer significant advantages in terms of speed and simplicity, but advances in analogue electronics have enabled the fast computations required for the charge ratio method to be more easily realised. *Lampton and Paresce (1974)* have presented a comprehensive summary of the different approaches to position decoding. More recently, high speed digital processing has been used to perform charge ratio division (*Clampin et al, 1988*).

In the charge ratio method, the position is computed using the following algorithms for signals A,B,C & D derived from charge amplifiers equally spaced around the anode perimeter.

Unipolar:

$$X = \frac{2(A+B)}{A+B+C+D} \qquad Y = \frac{2(B+C)}{A+B+C+D} \qquad [4.6]$$

Bipolar:

$$X = \frac{(A+B) - (C+D)}{A+B+C+D} \qquad Y = \frac{(B+C) - (A+D)}{A+B+C+D} \qquad [4.7]$$

The bipolar algorithm is more complex to realise but, as will be seen later, offers certain advantages over the unipolar approach.

4.6 The distortionless resistive anode

A resistive anode has been devised which does meet the boundary conditions required for a linear response (*Lampton and Carlson, 1979*). A uniform potential gradient in an infinite resistive sheet of resistance per square $R_{\square}$ remains undisturbed by the cutting of a circular hole of radius a, provided that the hole is bounded by a linear resistance $R\ell$ per unit length given by:

$$R\ell = R_{\square}/a \qquad [4.8]$$

Gear (1969) pointed out that the theorem applies to any distribution of holes. This includes an arrangement of four intersecting holes surrounding a central area. The area then drops out to give a finite sheet bounded by four concave arcs. This still has the electrical characteristics of the original infinite sheet and can be used as a linear position sensing anode. It has been termed the Circular Arc Terminated or CAT anode, or simply the Gear anode. This is shown in figure 4.1, and can also be seen in figure 4.3, which is a photograph of some component parts of the IPD.

The following reasoning may be used to confirm that such a device is truly a linear position sensor. If a uniform potential gradient is set up between opposite pairs of vertices of the anode, a high impedance point probe will register a signal proportional to distance across the anode. It follows from the reciprocity theorem that if a signal is injected at the probe, then the current flowing to ground from any pair of vertices will have the same position-dependent magnitude. The same argument applies to pulses provided that the magnitude of the whole charge integrated over time is measured, taking into consideration the anode diffusion time constant. This is unaffected by the border resistors and is given by $R_{\square}C_D/\pi^2$, where C_D is the total anode capacitance (*Fraser & Mathieson, 1981d*).

4.7 Pulse amplification

As previously mentioned, the amplifiers used for the charge pulse need only to convey the positional information contained in the anode signals. The requirement for the charge division method is termed a ballistic signal (*Baldinger & Franzen, 1956*), that is, one with an amplitude which is proportional to the total charge input to the amplifier.

The rise-time position computing electronics require a high input impedance voltage amplifier. The first generation charge ratio electronics built at UCL used a current sensitive pre-amplifier, the output of which was a voltage representing the time-varying current from the anode terminals. This was used with a gated integration stage in the computing electronics to measure the total charge for each pulse.

It was soon realised that the gated integrator approach had a number of drawbacks. In particular it was highly susceptible to d.c. offset errors and the integration period for each pulse was amplitude dependent. A different method was, therefore, adopted. The current amplifier was exchanged for a charge amplifier followed by a two stage pulse-shaping filter.

The charge amplifier provides a voltage representing the time and amplitude profile of the charge collected at the anode terminal. The time profile of this pulse is dependent on the anode RC time constant. The pulse-shaping circuitry is a type of band-pass filter with matched "differentiation" and "integration" sections. The design follows an approach used by *Alexander (1971)* and is termed a "double differentiation double integration" or DDDI amplifier. The amplitude of the output pulse is proportional to the magnitude of the input charge, but the pulse length is constant, determined by the shaping time constants. The pulse is bipolar with equal energy about zero and thus can be a.c. coupled with consequent immunity to d.c. offset errors.

For optimum operation it is important that the time constants of the amplifiers are matched to the RC time constant of the anode. A large anode resistance gives slowly rising pulses. If the part of the amplifier output signal stimulated by the early part of the charge pulse starts to decay before the charge pulse has ended, the overall amplitude of the pulse will be reduced. Under such conditions the output will no longer accurately represent the charge input. This is known as "ballistic deficit", and is defined by *Baldinger & Franzen (1956)* as "the difference between the amplitude of the output voltage pulse resulting from a current pulse of finite duration and that resulting from an infinitely short pulse carrying the same total charge."

It is essential for positional linearity and optimum signal-to-noise ratio that the entire charge be faithfully measured from each anode output. It is, therefore, most important to minimise ballistic deficit. However, if the anode resistance is reduced in order to achieve this, the noise due to the anode resistance is proportionately increased, so it is counter-productive to reduce the resistance more than is necessary to avoid significant ballistic deficit. It is generally accepted (*Lampton & Paresce, 1974*) that the best compromise is to make the pulse shaping time constants equal to the anode diffusion time constant ($R_{\square}C_D$).

A further effect must be taken into consideration when designing the time constants - the effect of pulse pile-up. The amplifier output pulse takes approximately twelve shaping periods to decay to 1%. If a second charge pulse occurs before this time it will effectively ride on the tail of the preceding pulse and the positional information will be corrupted. In the event of two simultaneous pulses, the processing electronics will interpret this as one event spatially midway between the two actual events. Clearly, for a given photon rate, shorter shaping time constants will result in less pile-up. It is possible to design pulse pile-up rejection circuits which will prevent the position computing electronics from accepting pulses which arrive

before the base line has stabilised after the previous pulse. This technique is discussed more fully in chapter five, together with a detailed description of the amplifiers.

4.8 Noise considerations

The resistive anode is, by its very nature as a resistance, a source of noise, and this gives rise to a positional uncertainty. The Johnson current noise is given by:

$$\overline{i_n}^2 = 4ktBG \qquad [4.9]$$

where B is the bandwidth of the system in Hz and G the conductance of the source, in this case, the anode. However, this is a steady state current and it must be looked at in terms of its contribution to the positional information of the shaping amplifier output pulse.

The conductance of the anode is taken as $1/R_2$ where R_2 is the so-called "two terminal resistance". This is the resistance between opposite pairs of anode contacts which are shorted together. This value is appropriate because in the positional algorithms [4.6] and [4.7] adjacent terminals are electronically combined, and so R_2 is the effective noise generator for each axis.

The denominator (A+B+C+D) in equations [4.6] and [4.7] is not affected by noise. The noise components of opposite pairs, for example (A+B) and (C+D) cancel out as they are anti-correlated, being opposite ends of the same resistance.

The mean square charge noise is given by:

$$\overline{q_n}^2 = \overline{i_n}^2 \tau^2 \gamma^2 \qquad [4.10]$$

where $\tau = RC$, the amplifier time constant, and γ is a dimensionless

noise form factor, which has been shown by *Fraser and Mathieson (1981a)* to be 2.39 for DDDI pulse shaping with equal RC time constants.

Taking the bandwidth as $1/2\pi RC$ the expression becomes:

$$\overline{q_n}^2 = \frac{2kt}{\pi R_2} \tau \gamma^2 \qquad [4.11]$$

$$= 3.6kt\tau/R_2$$

4.9 Noise performance of the distortionless anode

The border resistors of the CAT anode make a substantial contribution to the anode noise. The two-terminal resistance, R_2, is shunted by the two border resistors on either side. The value of each resistor is approximately $R_{\square}d/a$ so the noise is increased by a factor of (1+2a/d). This means that there is a trade-off when choosing the optimum value of a. A small a/d ratio severely limits the useful imaging area of the anode and a large ratio incurs a greater noise penalty. A value of two is used in practice, so the noise is five times greater than that of an unbordered sheet with the same resistivity.

The quantitative effects of the anode resistance on resolution will be discussed in chapter seven, which deals with the overall imaging performance of the IPD.

4.10 The Wedge and Strip anode

Anodes of this type use a completely different technique in order to determine position by charge division. The charge cloud from the microchannel plates is spread over a pattern of interleaving electrodes. The pattern is such that the position of the charge centroid can be computed from the ratios of the pulses measured on each electrode.

The Wedge and Strip is the most well known configuration of this class of anode, but other geometries are possible. The anode consists of three electrodes termed wedge, strip and Z. The first consists of a row of identical triangular wedges aligned with the Y-axis of the anode. A row of strips is interleaved between the wedges. Unlike the wedges, the strips decrease in width across the X-axis of the anode. The Z electrode simply fills in the space between the wedges and strips.

Prototype detectors have been made in collaboration with Instrument Technology Ltd. The theory of the anode, originally designed by H.O. Anger, is described by *Martin et al (1981)*. It is an ingenious device, and is illustrated in figure 4.2.

The positional algorithms are given by:

$$X = \frac{2S}{W+S+Z} \qquad [4.12]$$

$$Y = \frac{2W}{W+S+Z}$$

These are realised in exactly the same way as the resistive anode algorithms, and the same Signal Processing Unit is used, with slight modifications to realise the new algorithm.

The advantages of the Wedge and Strip anode are generally quoted as:

1. It lacks the large inherent current noise of the resistive anode.
2. It can easily be made in large sizes.
3. It is inherently linear.

It is clear from figure 4.2 that to ensure accurate charge division the electron cloud from the MCP stack must be spread out over a sufficiently wide area so that it covers at least one full cycle of the pattern. In order to do this an electron drift region of about 15mm is provided between the MCP stack and the anode. The potential

gradient across this gap must be uniform. This is achieved by providing a series of conducting rings around the perimeter of the gap, held at intermediate potentials. Although the anode is claimed to be distortion free, problems in the electron drift area can introduce distortions, as demonstrated by *Siegmund et al (1983)*.

The dominant sources of noise are amplifier voltage noise and partition noise. The amplifier noise is dependent on input capacitance, which is much greater than with the resistive anode. This is largely caused by inter-electrode capacitance and can be very high as the Z electrode weaves its way through the wedges and strips. The length of the Z electrode means that it will also have a finite resistance and therefore an additional noise contribution.

The amplifier noise is given by:

$$Q^2 = \left[\frac{3kt}{g} + 4KTR_Z \right] \left[C_D + C_{FET} + C_S \right] \qquad [4.13]$$

where g is the transconductance of the input FET of the amplifier. C_D and C_{FET} are the capacitances due to the anode and the FET. C_S is the stray input capacitance and R_Z is the resistance of the Z electrode.

It was seen in paragraph 4.8 that image perturbation due to the noise generated by a resistive anode is independent of position, and so the resolution is the same over the full image area. This is because the dominant noise components are anti-correlated in the denominator of the algorithm, and therefore cancel out. However, this is not the case for the Wedge and Strip anode and it has been shown (*Martin et al, 1981*) that at the edge of the image the perturbation due to noise is twice that in the centre.

The partition noise is due to the statistical distribution of the

electrons between the electrodes, and is given by *Allington-Smith and Schwartz (1984)* as:

$$R_p = 2.36L \left[\frac{\sqrt{f(1-f)}}{(f_{max} - f_{min})\sqrt{Q}} \right] \qquad [4.14]$$

where R_p is the FWHM resolution component due to the partition noise and L is the length of the sides of the anode. Q is the total charge per event and f is the fraction of charge collected by the electrode, which lies between the limits f_{min} and f_{max}.

The devices made for UCL were not particularly successful and the imaging performance was very poor. This was attributed to manufacturing problems in the anode. It is clear from the literature that excellent performance can be obtained once the necessary expertise has been accrued. However, a demountable vacuum system is required for such experimentation and the lack of success with the two sealed devices made by ITL meant that further investigation of this technique was liable to be prohibitively expensive. These devices were, therefore, forsaken in favour of the resistive anode.

Resistive anodes have been harshly dismissed by some in favour of Wedge and Strip anodes. (*Allington-Smith and Schwartz, 1984; Schwartz and Lapington, 1985*). However, their comparisons of the imaging performance were made with the outmoded rise-time method of resistive anode decoding, which has several limitations. In particular, the repeated references to the temperature dependence of the resistive anode simply do not apply when the charge ratio method is employed. Also, the resistive anode charge noise is much less significant when used with state of the art high-gain microchannel plate stacks.

Resistive anodes have been found to be both reliable and repeatable, and for many practical purposes just as effective as Wedge and Strip anodes in terms of image quality, speed and linearity.

Figure 4.1

The Circular Arc Terminated Anode.

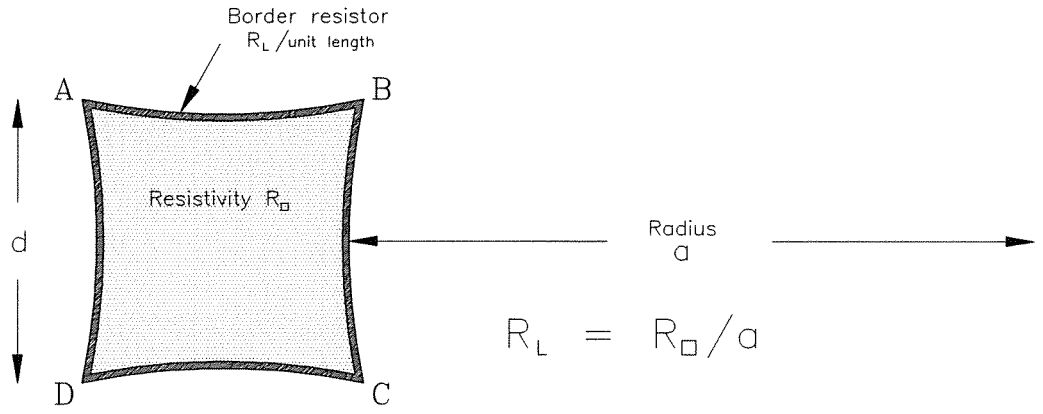
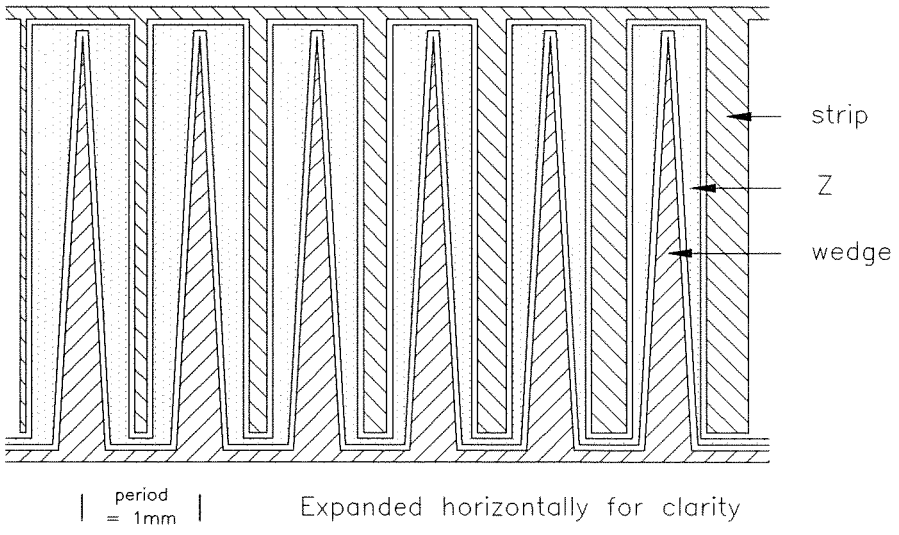


Figure 4.2

The Wedge and Strip Anode



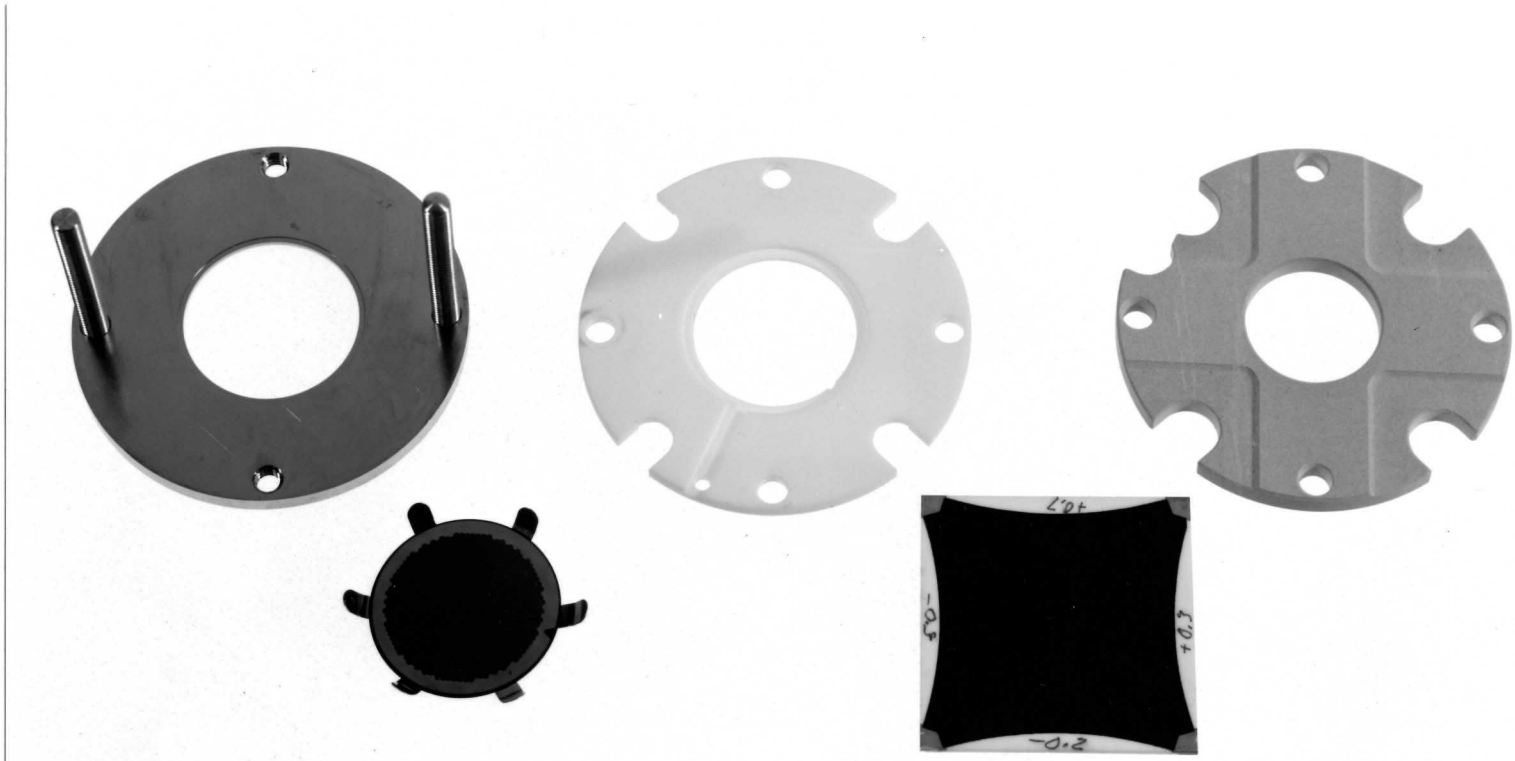


Figure 4.3 Component parts of the ITL IPD showing the CAT anode and MCP stack.

CHAPTER 5 - SIGNAL PROCESSING ELECTRONICS FOR THE RESISTIVE ANODE

5.1 Methods of signal processing

The charge pulse incident on the resistive anode may be processed in a number of ways *(Lampton & Paresce, 1974)*. As discussed in chapter four, these methods fall into two groups: the pulse rise-time method and the charge amplitude ratio method. Initially, a rise-time system was constructed, but it was later abandoned for the charge ratio method. The signal processing systems which were designed and built are detailed below. The amplifiers which convert the anode charge pulses into signals suitable for the processor are also discussed as they are an integral part of the processing system.

5.2 Development of the rise-time method

The first electronic system built was based on rise-time detection. Four pulses were obtained from voltage amplifiers connected to each corner of the anode. The rise-time of the output pulse from each amplifier was dependent on the equivalent RC transmission line between the incident charge pulse and the anode contact. The system is illustrated diagrammatically in figure 5.1. The X and Y co-ordinates are each determined by measuring the signals from two diametrically opposed amplifiers. A clock is started by the pulse from X_1 and a fixed delay is initiated by pulse X_2 after which the clock is stopped. The delay is chosen to be just greater than the maximum difference in rise-time between the two pulses, thus ensuring that the clock registers in the case when X_2 precedes X_1 by the maximum amount. This method was not very satisfactory for a number of reasons. The difference in rise-time was very small, of the order of tens of microseconds. The resolution is proportional to the clock speed, so a fast clock with accurate switching was essential. The requirement for high photon rates necessitated a short anode time-constant and compounded the need for a high clock speed. The prototype, which was built in modular form, was very prone to false triggering. Inspection of the waveforms indicated that the positional information, barely

discernible in the rise-time, was clearly visible in the pulse height variation, so it was decided that a pulse ratio method would be a more satisfactory approach.

5.3 Development of the charge amplitude ratio method

Four generations of ratio computing electronics were designed and built. They all essentially performed the same operations, which were:

1. Derive four signals proportional to the charge pulses.
2. Add, subtract and divide them to realise the positional algorithms.
3. Convert the co-ordinates to digital format, generate an output data strobe and reset the input circuits ready for the next pulse.

The positional algorithms are recalled from chapter four as:

Unipolar:
$$X = \frac{2(A+B)}{A+B+C+D} \qquad Y = \frac{2(B+C)}{A+B+C+D} \qquad [4.6]$$

Bipolar:
$$X = \frac{(A+B) - (C+D)}{A+B+C+D} \qquad Y = \frac{(B+C) - (A+D)}{A+B+C+D} \qquad [4.7]$$

An important parameter is the system dead time. This is the time during which the system is busy processing an event and is, therefore, unable to accept new events. The reciprocal of this figure gives the maximum number of photons that can be processed per second. The implications of dead time and of the choice of algorithm will be discussed later in this chapter. The following paragraphs detail the evolution of the electronics.

5.3.1 First generation charge ratio electronics

The first charge ratio electronics used current amplifiers to convert the current pulse from the anode contacts to a voltage pulse. Gated integrating circuits were then enabled for the duration of the pulse, resulting in a voltage level proportional to the magnitude of the

total charge from each anode contact. Analogue dividers were used to realise the unipolar algorithm. The circuit is shown in figure 5.2. A major problem with this method was that the pulses were unipolar and thus very sensitive to baseline drifts in the high gain amplifiers. A diode clamp stage was added prior to the integrators, which improved the performance considerably. Summing circuits provided the inputs to the analogue dividers, which were implemented with analogue multipliers operated in feedback mode. After allowing time for the multipliers to settle, an analogue to digital conversion provided eight bit co-ordinates for each axis. The "end of convert" signal then reset the integrators ready for the next pulse and generated an "output data" strobe for the computer interface. The analogue co-ordinate signals were also made available to feed to an x-y oscilloscope. This gave a real time image, with the photon events visible as a succession of dots on the screen. The full image could be seen, due to the persistence of the phosphor.

5.3.2 Second generation charge ratio electronics

It became apparent that the gated integrators were a significant source of error, as explained in paragraph 4.7. They were dispensed with and the current amplifiers replaced with charge-sensitive amplifiers, followed by pulse shaping circuits. The output signal from such an arrangement is a bipolar pulse, the height of which is proportional to the magnitude of the charge input. It is, therefore, only necessary to measure the pulse height rather than integrate the signal over the whole pulse. Furthermore, the bipolar form of the pulse means that a.c. coupling can be used in the input stages of the processing electronics, thereby eliminating d.c. offset problems in this area.

The amplifier design is similar to that described by *Alexander (1971)*. A charge amplifier, like a current amplifier, has a virtual ground at the input and hence a very low input impedance. However, the feedback loop consists of a resistance and capacitance in parallel. The resistance is very large so the step response is dominated by the capacitor. A voltage step is produced at the output which is

proportional to the instantaneous charge input. The step slowly decays due to the RC time-constant of the network.

The resulting signal has the appearance of an irregular sawtooth wave with the height of the rising edge proportional to the charge. The rise-time is dependent on the anode time-constant and amplifier speed. The base line is highly signal dependent and further processing in the form of pulse shaping amplifiers is required. The design chosen was the "double differentiator double integrator" type, so called because it has two stages, each one configured as a differentiator and an integrator combined. Intuitively it has the appearance of a band-pass filter, but the operation of the circuit is best appreciated in the time domain.

The response has been shown by *Alexander* to be a bipolar pulse given by:

$$\frac{e_0}{Q_{in}} = \frac{R_3R_5}{C_1R_2R_4} \cdot \frac{(3\tau-t)t^2e^{-t/\tau}}{6\tau^3} \quad \text{volts/coulomb} \qquad [5.1]$$

Differentiating and solving for the maxima gives:

$$t_{max} = (3 + \sqrt{3})\tau$$

and thus the height of the positive maximum is:

$$\frac{e_0(t_{+\,max})}{Q_{in}} = 0.127 \cdot \frac{R_3R_5}{C_1R_2R_4} \qquad [5.2]$$

Further inspection of equation [5.1] shows a zero crossing at $t = 3\tau$ and a return to -1% of the peak value at 12.0τ.

The resulting bipolar pulses can be a.c. coupled thus eliminating any d.c. offset errors. The pulses are of equal length regardless of pulse height. A computer simulation of the pulse is shown in figure 5.3 and an oscilloscope trace of a typical output signal in figure 5.4.

The signal processor was completely redesigned. A peak detector initiated the timing signal to fast sample-hold circuits which stored

the peak value of each input pulse. Summing circuits then provided the inputs to the analogue dividers. These were improved in that true log/antilog dividers replaced the multipliers used in the first generation. The system dead time was 40 μs, comprising 15 μs for settling of the analogue circuitry and 25μs for the analogue to digital converters (ADCs). The circuit diagrams of the charge/shaping amplifier and second generation processor are shown in figures 2 and 3 of *McWhirter et al (1982)*.

5.3.3 Third generation charge ratio electronics

The system dead time of 40μs implied a maximum count rate of 25,000 photons per second, considerably less than the potential of the detector itself. The next generation, therefore, was designed to reduce the dead time to the minimum that the state of the art would allow. The ADCs accounted for more than half the dead time, with the divider circuit taking most of the remainder. It was decided to adopt a new approach to the ratio determination and to combine this with faster ADCs. The ADCs used were the type known as "successive approximation". They work by effecting an electronic scale pan. A ten bit register is fed to a digital to analogue converter (DAC), and the output of this is compared with the input signal to the device. The most significant bit of the register is set high, with the remaining bits low. If the scale pan tilts in favour of the input signal the bit is retained, otherwise it is reset to low. The process is repeated with the next most significant bit, and so on down the line until all ten bits have been tested. The register then contains the digital equivalent of the input signal. In order to define the scale of the output, i.e. the level of analogue input required for a digital output of 1023 (all bits set high), a normalising reference voltage is used in the internal DAC and the output is of the form:

$$\frac{\text{Digital output}}{1023} = \frac{\text{input voltage}}{\text{reference voltage}} \qquad [5.3]$$

The device is clearly acting as a divider. In order to use it effectively for the IPD signals, the following criteria must be met:

1. The ADC must have provision for an external reference voltage.
2. It must operate correctly over a range of reference voltages.
3. The reference circuitry must accept a dynamic signal and be fast settling.

The Analog Devices AD759 operated with a conversion time of 3μs was found to meet these criteria. The sample-hold and summing amplifiers were improved to give a settling time of 3μs, and 1μs was allowed for re-setting, giving a dead time of 7μs, a considerable improvement. This method meant that the analogue x-y co-ordinate signals were no longer produced, so a pair of DACs was used to generate them from the digital output. This had the advantage that it provided confirmation that the analogue to digital stage was functioning correctly.

5.3.4 Fourth generation charge ratio electronics

The improvements made in the third generation meant that a dominant mechanism of degradation was that of pulse pile-up in the charge/shaping amplifiers. As explained in paragraph 5.3.2, the bipolar pulses are all the same length and decay to within 1% after twelve shaping time-constants. The time-constant used so far had been 0.5 μs. This was chosen because the best noise performance is obtained by matching the shaping time-constant to the anode time-constant (*Lampton & Paresce, 1974*). The amplifier settling time was, therefore, 6μs. This meant that at the higher count rates now available there was a significant probability that pulses would overlap. Although the processor will not accept pulses during the dead time, the amplifiers are still fully operational. A pulse accepted for processing after the dead time period has expired may be "riding on the tail" of another pulse which had occurred during the dead time. The effects of this may be minimised using pulse pile-up rejection, a technique discussed later, although this does affect the overall detected count rate. It was therefore decided to build a "high speed" version of the amplifier and reduce the shaping time-constant to 0.1 μs. The existing amplifier

was not suitable for this and a recently introduced high speed operational amplifier was used for the pulse shaping stages.

The previous amplifier would drive only short lengths of unterminated screened cable. This restriction was quite acceptable for the balloon-borne instruments and had the advantage of minimising power dissipation. However, an increasing number of ground-based applications required the detector to be operated at distances of up to ten metres from the processor and computer, and it had been necessary to build an external driver. A 50 ohm driver was, therefore, incorporated into the output stage of the new amplifier.

The existing peak position detector in the main processor was not fast enough for the high speed pulses, and the acquisition time of the sample-hold circuits was too slow and temperature dependent to hold the peak value accurately in the reduced time available. A substantial re-design was effected in these areas, using a newly available high speed analogue switch in the sample-hold section.

Figure 5.5 shows the circuit diagram for the high speed charge/shaping amplifier and figure 5.6 shows a photograph of the circuit board.

The fourth generation Signal Processing Unit is shown in block diagram form in figure 5.7 and the full circuit diagram is in figure 5.8. An internal view of the unit is shown in figure 5.9 and the unit is illustrated together with the IPD in figure 5.10.

5.4 Operation of the Signal Processing Unit

The operation of the fourth generation signal processor is described with reference to figure 5.8. The four pulses from the charge and shaping amplifiers are summed by IC1 and also fed individually to ICs 2-5. ICs 6-11 form a sample-hold circuit and ICs 12-18 form the

necessary inversion and addition to realise the following signals:

$$\text{X numerator} = (A+B)-(C+D)$$

$$\text{Y numerator} = (B+C)-(A+D)$$

$$\text{denominator} = A+B+C+D$$

The analogue to digital converters IC20 and 21 perform the division of equation [5.3]. By making V_{ref} the denominator, $V_{in(X)}$ the X numerator and $V_{in(Y)}$ the Y numerator, the positional algorithms of equation [4.7] are realised.

The summed pulse from IC1 is input to the lower threshold comparator, the peak detector and the upper threshold comparator. It is also buffered by IC19 to provide the "pulse" output on the front panel.

If an event is detected by the lower threshold comparator IC29, the peak detector comparator IC31 is enabled. The differentiator IC32 provides the input to IC31 which then produces a positive-going edge which coincides with the peak of the charge pulse. This triggers the "peak" monostable IC34a, unless inhibited by the upper threshold comparator IC30. The "peak" monostable then triggers the "hold" monostable IC33b and the "encode" monostable IC34b. The "encode" pulse prepares the ADCs IC20 and IC21 for conversion and provides a 3μs wait for the summing circuits to settle. After this pulse the analogue to digital conversion begins. When both ADCs have finished converting, the combined "end of convert" signals trigger the "output data strobe" monostable IC35b which then provides the data strobe unless the "inhibit" line is asserted by the user interface, in which case the event is lost. The "end of convert" signal also triggers the "next event" monostable IC35a which clears the "hold" monostable ready for the next photon pulse. The digital outputs from the ADCs are latched into the output buffers by the "next event" monostable. The digital co-ordinates are then input to the digital to analogue converters IC26 and IC28 to provide the analogue outputs.

The counters, rate 1 and rate 2, are fed from the lower threshold comparator and the output data monostable respectively. They each

consist of a monostable IC39 a & b, driving an analogue switch, IC40 a & b, which switches between 0 and 15 volts. The resulting pulse train is filtered and amplified by IC38 and IC41. IC37 is the EHT trip comparator which drives the latching relay RLA1. This relay switches the EHT low voltage supply on and off. It is also operated by the biased toggle switch on the front panel, and by the "auto" computer control from the "Tecmar" parallel interface.

5.5 Consideration of the positional algorithms

The first two generations of electronics used the unipolar algorithm [4.6]. However, the ADC used in the following generations has the capability for bipolar operation (i.e. a negative numerator could be used) and it was decided to take advantage of this and use the bipolar algorithm [4.7].

The main benefit of this approach is that the centre of the image (digital co-ordinates 511,511) corresponds to a numerator input of zero volts. This means that the image size can be adjusted with a single gain control on the denominator channel, which acts as a "zoom". The photocathode image can easily be expanded to fill the digital field. This is desirable for the general design philosophy of maximum flexibility in the field, with any detector being compatible with any processor. In general, this single adjustment to optimise the digitised image size to the active image size is all that is required to match an IPD to a Signal Processing Unit. The unipolar algorithm requires positional offset adjustments in both axes whenever the image size is changed. If the image is zoomed further, the central portion of the photocathode can be expanded to give improved digitisation over a reduced area.

It has been suggested (*Clampin et al, 1988*) that the unipolar algorithm gives superior results because the bipolar system introduces two additional sources of uncorrelated noise. However, the anode noise, which is the dominant noise component, is anti-correlated between opposite sides of the anode. This means that the denominator noise is

doubled. The signal is also doubled, so provided that the amplifier noise is small by comparison the signal-to-noise ratio remains the same.

5.6 Dead time and its effect on photon counting efficiency

Electronic photon counting systems are not usually able to take account of every photon event that is detected by the photon detector itself. As the process is inherently a sequential one, any stage in the processing is likely to have an associated dead time following a detected event during which no further photons can be processed. As seen in the preceding paragraphs, the IPD system is no exception, and it is important to quantify the effect of dead time on counting efficiency.

In this respect, photon counting systems fall into two categories: paralysable and non-paralysable. In a non-paralysable system the photons are processed as fast as the dead time will allow, and the maximum output rate is asymptotic to the reciprocal of the dead time. In a paralysable system, however, the dead time can be extended indefinitely by subsequent photons arriving before the end of the current dead time. In this case the output rate rises to a maximum and then actually decreases with any further increase of input rate. Different stages of a system may fall into different categories. These effects are quantified in the following paragraphs and then discussed with reference to the IPD system, the response of which is shown in figures 5.11 and 5.12.

5.6.1 Efficiency of a non-paralysable system

If a system has input and output event rates of R_i and R_o events per second for the input and output respectively, and a dead time of t_d, the system will be "dead" for a period $R_o.t_d$ each second. The total

time each second for which the system can accept photons is, therefore, $(1 - R_o t_d)$ seconds, that is to say:

$$R_o = R_i E_n = R_i(1 - R_o t_d)$$

where E_n is the non-paralysable counting efficiency.

This rearranges to:

$$R_o = \frac{R_i}{(1 + R_i t_d)} \qquad [5.4]$$

5.6.2 Efficiency of a paralysable system

An output from a paralysable system occurs if a photon is detected after a time greater than t_d following the previous pulse. The probability of this happening is equal to the probability of zero photon events occurring in a time t_d.

Given a series of random events occurring at an average rate R, the probability, P, that n events will occur in a time t is given by the Poisson distribution:

$$P(n,t) = \frac{R.\exp(-Rt)}{n!} \qquad [5.5]$$

The output rate R_o for an input rate R_i is thus given by:

$$R_o = R_i.E_p = R_i.\exp(-R_i t_d) \qquad [5.6]$$

5.6.3 Systems combining paralysable and non-paralysable stages

It is not unusual in photon counting systems to encounter a paralysable stage followed by a non-paralysable stage, for example when a pulse pile-up rejector is followed by an event processor, such as the IPD processor. If the non-paralysable dead time, t_n is smaller than the paralysable time t_p then the paralysable equation [5.6] applies. If, however, t_n is greater than t_p then the two efficiencies

have to be combined. The counting efficiency E_p (from equation [5.6]) is multiplied by the probability of zero counts during the time for which the non-paralysable time exceeds the paralysable. This is calculated by substituting the interval $t_n - t_p$ for t_d in equation [5.4].

This gives:

$$R_o = R_i . \frac{\exp(-R_i t_p)}{1 + R_i(t_n - t_p)} \qquad [5.7]$$

5.7 Count rate limitations of the IPD system

The predominant areas of count rate limitation for the IPD system are:

1. Processing electronics dead time.
2. Moderate pulse pile-up, causing image degradation.
3. Severe pulse pile-up, causing amalgamation of several pulses.
4. Microchannel plate local dead time.
5. Speed limitations in processor to computer transfer.

5.7.1 Processing electronics dead time

The IPD Signal Processing Unit is presently a non-paralysable system. This means that input signals occurring during the dead time will not extend the dead time. The system simply ignores them and takes the first event that occurs after the dead time has expired.

The output count rate is, therefore, asymptotic to the reciprocal of the dead time as the input rate increases. This is shown in figure 5.11. The first and second generation SPUs had a dead time of 40μs whereas in the subsequent versions this was reduced to 7μs. In the following paragraphs the input rate, R_i, will be defined as the number of photon pulses per second which are accepted as valid by the upper and lower threshold circuits of the Signal Processing Unit.

5.7.2 Moderate pulse pile-up

Moderate pulse pile-up does not affect the counting efficiency above that imposed by the processor dead time. It does, however, affect the quality of the data, as pulses analysed by the processor may be corrupted by the tail ends of pulses occurring during the dead time. This can be alleviated by incorporating a pulse pile-up rejector into the processor, making it into a paralysable system.

The pulse pile-up rejector prevents the processing of pulses which occur within a certain time after a preceding pulse. Once the processor has accepted an event for processing, any further events occurring within one settling period (6μs) of the end of the dead time are made to extend the dead time by a further settling period. This is repeated until a period of 6μs has elapsed with no event occurring. At this point it is "safe" for the processor to accept another event. With this system, the output count rate rises almost as fast as in the non-paralysable case up to the point where $R_i \times t_d = 1$. After this the output rate falls rapidly as the input rate increases, as shown in Figure 5.12. This approach is extremely effective at excluding corrupted events and gives a marked improvement in contrast at higher count rates, at the expense of a reduced DQE for those higher rates.

5.7.3 Severe pulse pile-up

When the pulse pile-up is so severe that a significant proportion of pulses are superimposed, severe degradation of the image occurs. The processing electronics will compute a single position in between the superimposed events, with a resulting in-filling effect between bright areas of the image. As explained in the previous paragraph, a pulse pile-up rejector will help considerably, but DQE will be affected. The upper threshold discriminator will serve to reject some of these multiple events if their combined amplitude exceeds this level. However, such high photon count rates are generally undesirable and should be avoided for the sake of the photocathode.

5.7.4 MCP local dead time

This has been discussed in detail in paragraph 2.8. In general, the main effect is a non-linearity in the intensity response. Photon events in very bright areas are not detected if the drop in the modal gain of the MCP stack means that the resulting pulses fall below the lower threshold of the processing electronics. This effect is obviously image dependent and is worse when imaging point sources. These appear blurred and diffused because the outer edges of the source are enhanced with respect to the centre.

5.7.5 Speed limitations in data transfer

The photon co-ordinates for each event are read by a parallel port and the image is built up in an array in computer memory. Each x and y co-ordinate is allocated a corresponding memory location which is incremented by one each time a photon is recorded in that position. The data rate is, therefore, limited by the rate at which the host computer can read the co-ordinates through the port and increment the image array.

This is a non-paralysable process and may be viewed as an extension of the signal processor dead time. A handshaking system exists whereby the computer outputs an inhibit signal in response to the processor's output data strobe. This prevents further strobes from being generated until the computer is ready, although the processor may still process photons during this time. Photons are, therefore, rejected if the processor dead time, t_{spu}, expires before the computer is ready to accept another event. This means that if the computer acquisition time is less than the processor dead time then the total system dead time is unaffected. If greater, the system dead time will be extended by an amount $t_c - t_{spu}$ where t_c is the computer processing time. The effective dead time, therefore, becomes t_{spu} or t_c, whichever is the greater.

For $t_d > t_{spu}$ the data acceptance rate is reduced by a factor of:

$$\frac{R_{in}}{1+R_{in}t_c} \cdot \frac{1+R_{in}t_{spu}}{R_{in}} = \frac{1 + R_{in}t_{spu}}{1 + R_{in}t_c} \qquad [5.8]$$

In practice, an average value is used for t_c, as the computer does not always take the same amount of time to process the data. Experience has shown that a 12MHz 80286 machine is sufficiently fast for the DQE to be unaffected by the computer, that is to say, t_c is less than 7 μs.

The data transfer efficiency at high rates can be improved considerably using a First In First Out register, or FIFO. Its name describes it exactly. It is a stack register, with an input port and an output port, which typically will hold up to 64 words of data. The data clocked in at one end can be clocked out at the other end in the same order. At first appearances this would seem to offer no advantage whatsoever. However, the FIFO stack acts as a smoothing device, similar to the way in which a water tank supplies a constant flow of water regardless of any fluctuations in the input flow to the tank.

The asynchronous processor events accumulate in the FIFO and are read out by the computer as fast as it is able. Events that occur during the computer dead time are not thrown away unless the FIFO is full. Instead, they are read during the periods that the computer would otherwise have been waiting for data. The acquisition rate is equal to the SPU output rate until the limiting rate of $1/t_c$ is reached.

5.8 The dual port memory

An alternative way of speeding up the image acquisition is to use a "dual port memory" approach. A dedicated block of memory sufficient to hold the image is plugged into the computer back plane. The address bus of this memory is accessed directly by the signal processor and a "read-increment-write" cycle is implemented in hardware. This is a very fast process, limited mainly by the speed of the memory used. Transfer rates of several hundred thousand events per second are readily obtainable, faster than the current signal processor can work.

After a suitable integration time, determined by the computer, the address bus is switched, or "ported" to the computer address bus, thus enabling the computer to access the image. It is, therefore, known as a "dual port memory" (DPM). The process is controlled by a parallel port incorporated into the memory board. The image data does not pass through this port, it simply appears in the computer memory as soon as the address lines are switched over.

This approach would appear to be ideal and such a board has been manufactured for the IPD. One advantage in particular is that two detector systems can be run from one computer, one using the parallel port for image transfer and one using the dual port memory. However, the DPM is a specially produced board and rather expensive. For very little more money one can now buy an additional PC computer.

The DPM is not without its problems. Firstly, it occupies two full PC backplane slots, and as the standard parallel port board otherwise required is often needed for control of other processes the computer becomes rather congested. The board could no doubt be reduced to one card using state of the art memory, but problems remain which are inherent to the IBM-PC architecture. The early PCs could address one megabyte of memory as the 8086 and 8088 processors have 20 address lines. However, MS-DOS, the operating system, is only designed to access up to 640 kilobytes. The area between 640k and one megabyte is known as "upper memory". Although not directly accessible by DOS it can be used for certain additional devices, including the dual port memory. However, an increasing number of machines use this for other applications, such as the additional memory used for the now standard VGA graphics, and conflicts often occur.

The 80286 processor used for the PC-AT compatibles has 24 address lines. This opens up a further area of memory, between one megabyte and 16 megabytes, known as "extended memory". A newly designed DPM could be located in this area, provided that it does not conflict with extended memory installed in the computer. However, as machines become

faster and operating systems more flexible, the advantages of a purpose built board begin to diminish.

5.9 Dynamic range and pulse height distribution

The signal processing electronics will only operate satisfactorily over a limited dynamic range, largely dictated at the upper end by signal clipping, and at the lower end by signal-to-noise ratio and limitations of the ADCs. The upper and lower threshold limits define a window of allowable pulse heights. The range of pulse heights coming from the microchannel plates, shown in figure 2.8, is comparable with the processor's dynamic range, so careful optimisation of signal levels is essential to ensure that the proportion of pulses accepted for processing is maximised. However, if the lower threshold is set too low the resolution will be degraded because the low amplitude pulses have an inferior signal-to-noise ratio. Also, the lowering of the modal gain of the microchannel plates at high photon rates requires the lower threshold to be set as low as possible in order to give good efficiency for bright areas of the image. There is, therefore, a trade-off between DQE and resolution and the chosen compromise will depend on the particular application.

5.9.1 Implications of the quasi-Gaussian distribution

The pulse height distribution may be approximated to a Gaussian distribution for the purposes of assessing the dynamic range requirements. The distribution is conveniently quantified by measuring the pulse height resolution, PHR. This is given by the width of the distribution curve at half of the maximum, normalised to the pulse height value at the maximum of the curve, and is referred to as the full width at half maximum, or FWHM value. As seen in chapter two, values of 70% to 150% are usual.

In practice, the processing electronics can have a dynamic range of 10:1, i.e. input signal levels down to 10% of the maximum allowable level will give valid results.

For a pulse height distribution centred between the 10% and 100% levels, the maximum of the curve comes at 55%, assuming a symmetrical distribution. Equating this point to the mean value of a Gaussian probability distribution, the deviation at 10% and 100% is +/- 45%, which, normalised to the mean, is +/- 0.82.

The probability of the input signal exceeding a specific value of x may be derived by integrating equation [3.8] between x and infinity. As the curve is taken to be approximately symmetrical the probability of the signal being less than -x will be the same.

The required counting efficiency can thus be expressed in terms of a required FWHM.

Designating the 100% level x_{max}, and the 10% level x_{min}:

$$P(x > x_{max}) = P(x < x_{min}) = 0.1 \quad \text{for } \sigma = 1.28$$

and

$$P(x > x_{max}) = P(x < x_{min}) = 0.01 \quad \text{for } \sigma = 2.33.$$

In other words, for 80% of the events to be captured these points correspond to a deviation of 1.28σ, equivalent to a FWHM of 150%, whilst the 2.33σ capture width necessary to acquire 98% of all events represents a FWHM of 84%.

It must be remembered that the pulse height distribution is only approximately Gaussian in shape and is usually non-symmetrical. In particular, there is always a residual negative exponential component which is ignored for this approximation. However, the above analysis is useful for determining the approximate dynamic range requirements of the processing electronics.

5.9.2 Gain matching

In order to ensure that pulse height distribution is matched to the processor it is necessary to adjust the combined gain of the MCP stack and shaping amplifiers to ensure that the peak height is at the optimum level. In practice, the easiest way to achieve this is by fitting a fixed gain charge/shaping amplifier and adjusting the MCP

voltage whilst observing the pulse height distribution on a pulse height analyser. As the MCP voltage is increased, the initial negative exponential distribution is replaced by a well formed quasi-Gaussian curve, which maintains its shape until a certain limit is reached. Above this gain, instability in the MCP operation results in a broadening of the curve, giving a high proportion of large amplitude events. The range over which a suitably well formed distribution is maintained is typically 5:1.

For a particular MCP configuration it is normally satisfactory to choose between a standard amplifier gain for the majority of devices, and a high gain version to suit detectors which exhibit instability and need to be run with a lower MCP gain.

The gains used for the two classes of detector are:

	Amplifier gain	Nominal MCP gain
Single stack detector:		
Standard:	7.0×10^{12} v/c	3.6×10^{6}
High gain:	1.5×10^{13} v/c	1.7×10^{6}
Double stack detector:		
Standard:	8.4×10^{11} v/c	3.0×10^{7}
High gain:	1.3×10^{12} v/c	2.0×10^{7}

In order to measure the gain with a pulse height analyser, the four anode signals have to be added together to give the total charge incident on the anode. The anode outputs cannot be used individually as they also contain positional information. This addition is performed in the Signal Processing Unit and the resultant pulse appears on the "pulse" output on the front panel. Because the individual signal levels are quite high, it is necessary to incorporate a gain reduction factor of 0.4 into the addition channel in order to provide a distribution of pulse heights appropriate to the input range of the pulse height analyser. A pulse peak output of two volts corresponds to the optimum signal level for the processing electronics.

5.10 Amplifier design considerations

The charge and shaping amplifiers are built on a square circuit board directly behind the detector. The board contains four identical amplifiers, each with three stages. The charge amplifier is a standard cascode stage, of a design which is well proven in the field of nuclear physics. The subsequent two stages are built around operational amplifiers to produce the required shaping response. The Mark I design used discrete components to realise the operational amplifiers, but later designs were able to take advantage of high speed monolithic integrated circuits. The output signal from each amplifier is in the region of one volt, so the large signal voltage gain is of prime importance when selecting a suitable device. The Harris Electronics HA2525 finally proved to be a satisfactory choice. As previously mentioned, a 50 ohm driver was incorporated after the Mark I design. The Mark II design used a discrete emitter follower which was incorporated into the feedback loop. The latest (Mark III) design uses an integrated circuit unity gain buffer.

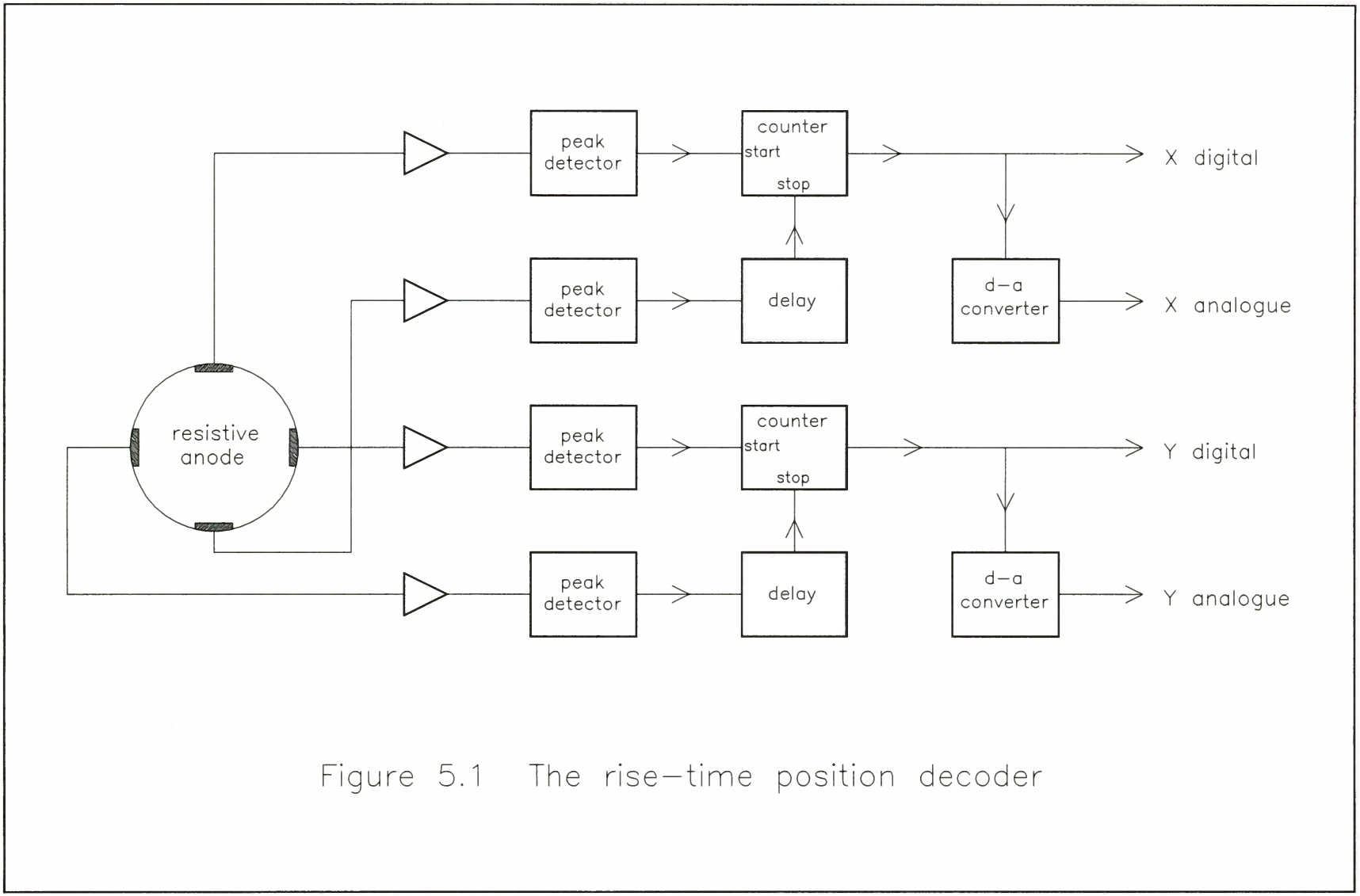


Figure 5.1 The rise–time position decoder

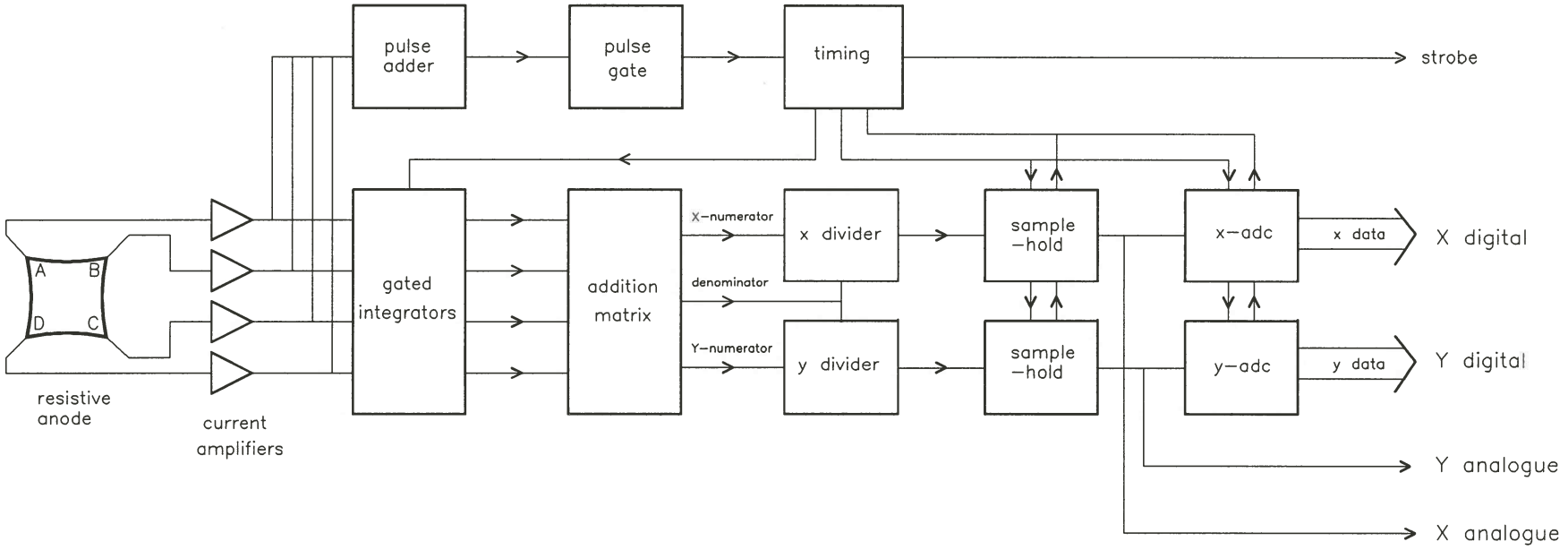
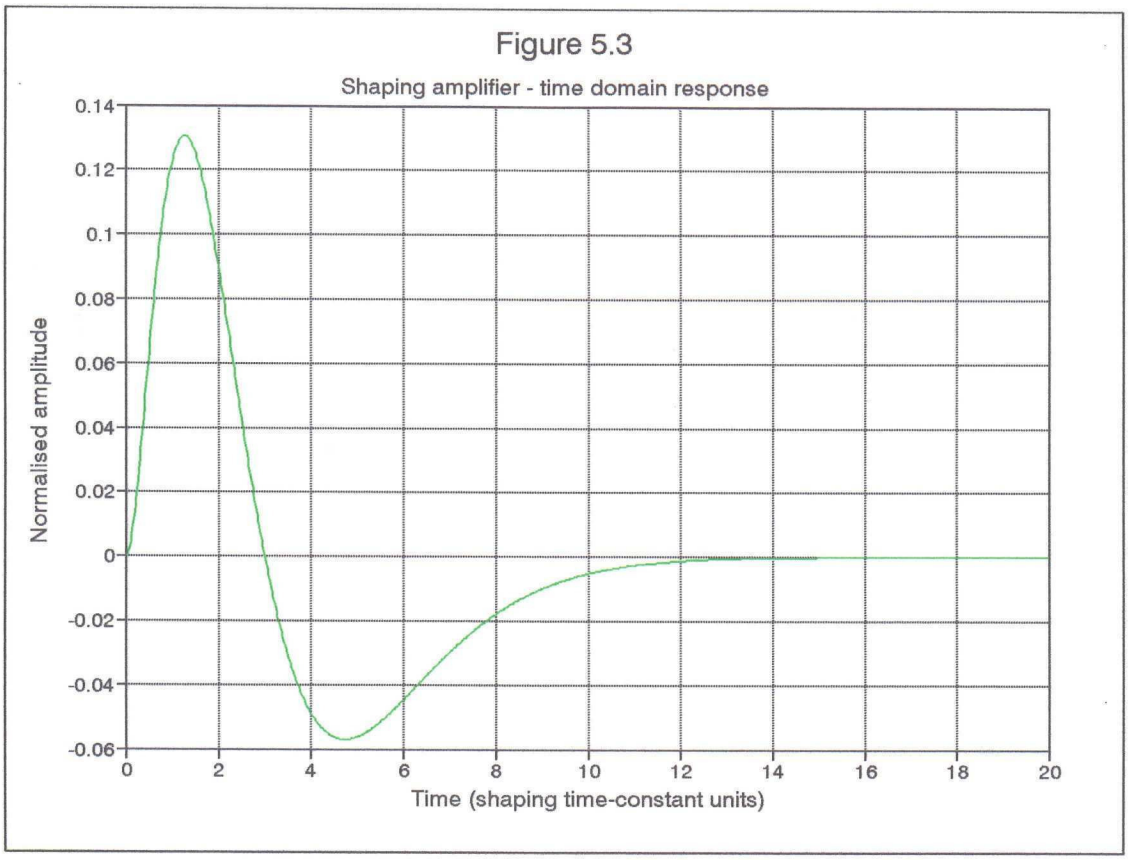


Figure 5.2 The first generation charge ratio Signal Processing Unit



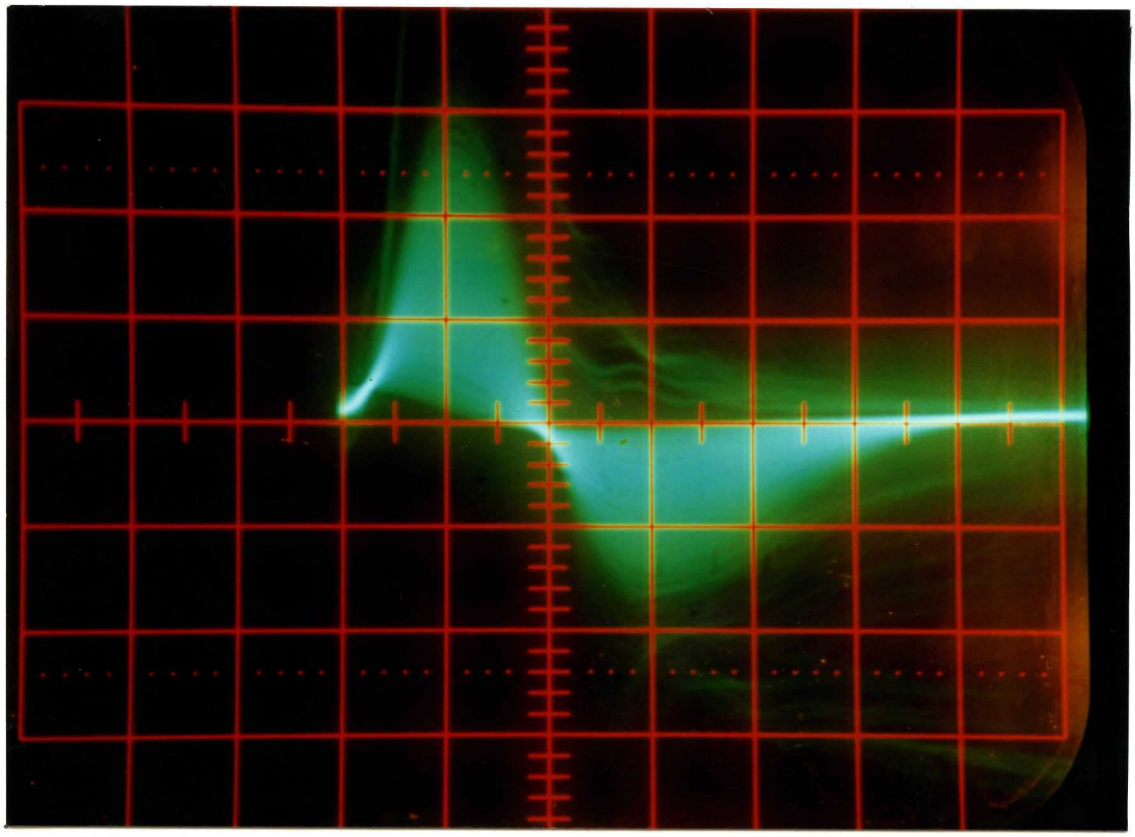


Figure 5.4 The amplifier output pulse

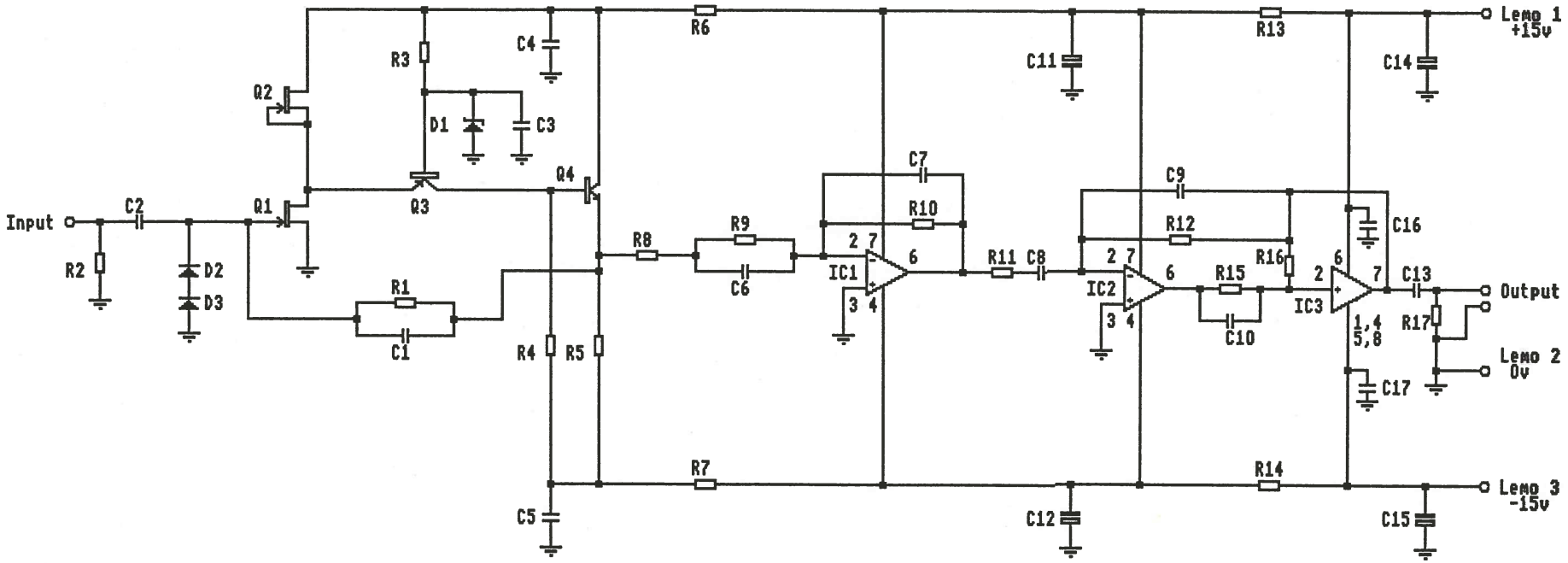


FIGURE 5.5

THE CHARGE / PULSE-SHAPING AMPLIFIER
FOR THE RESISTIVE ANODE IPD

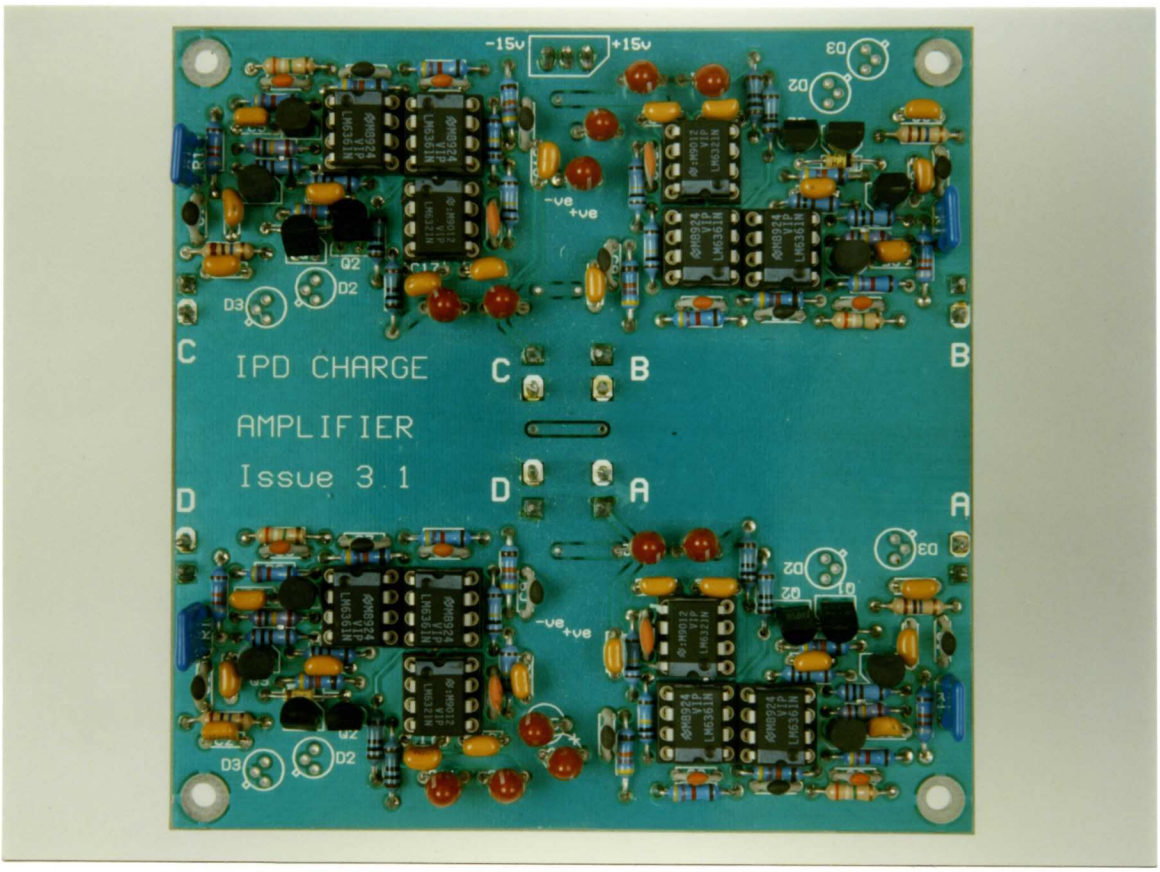


Figure 5.6

The charge and pulse-shaping amplifier

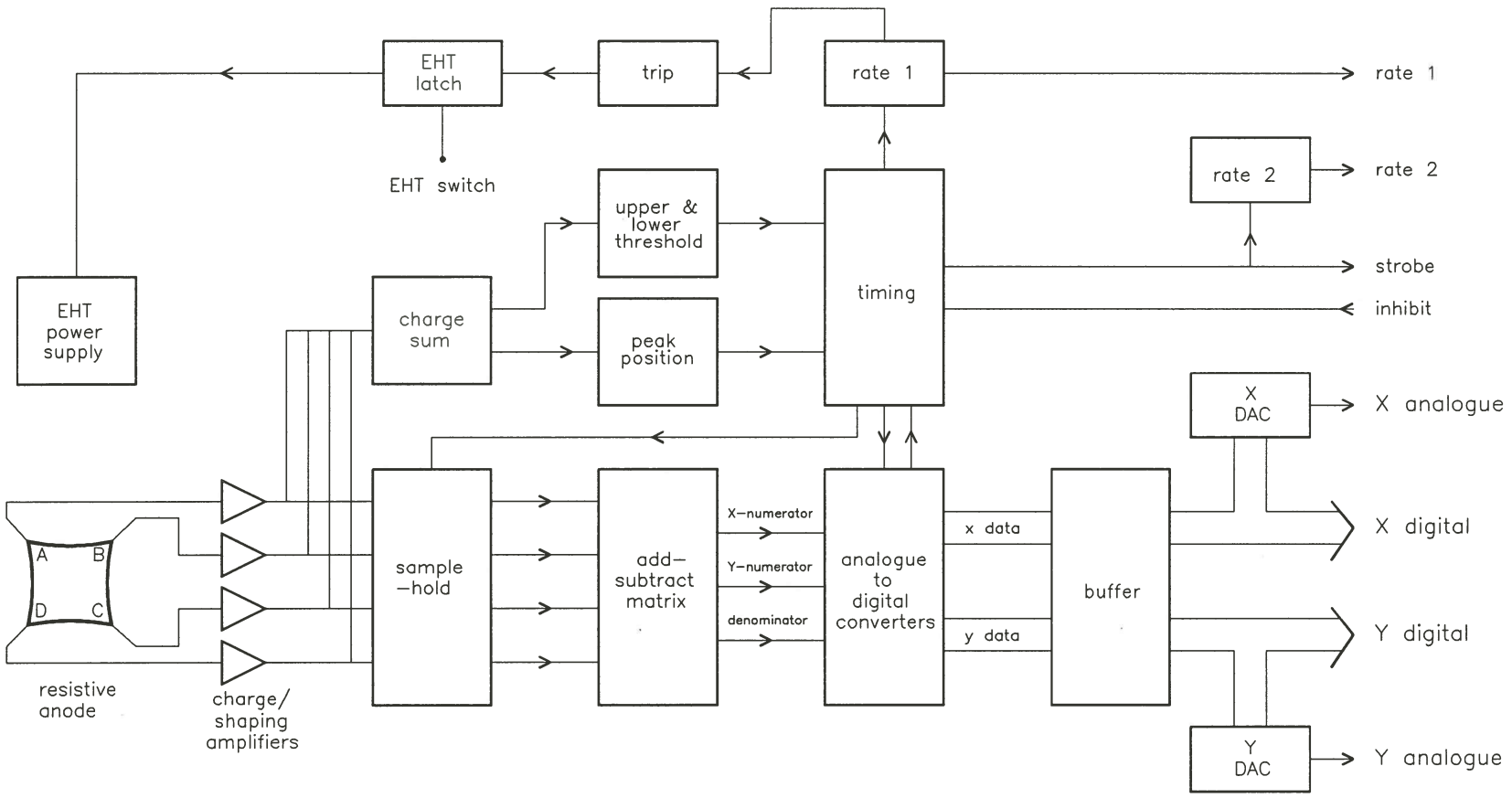


Figure 5.7 The fourth generation Signal Processing Unit

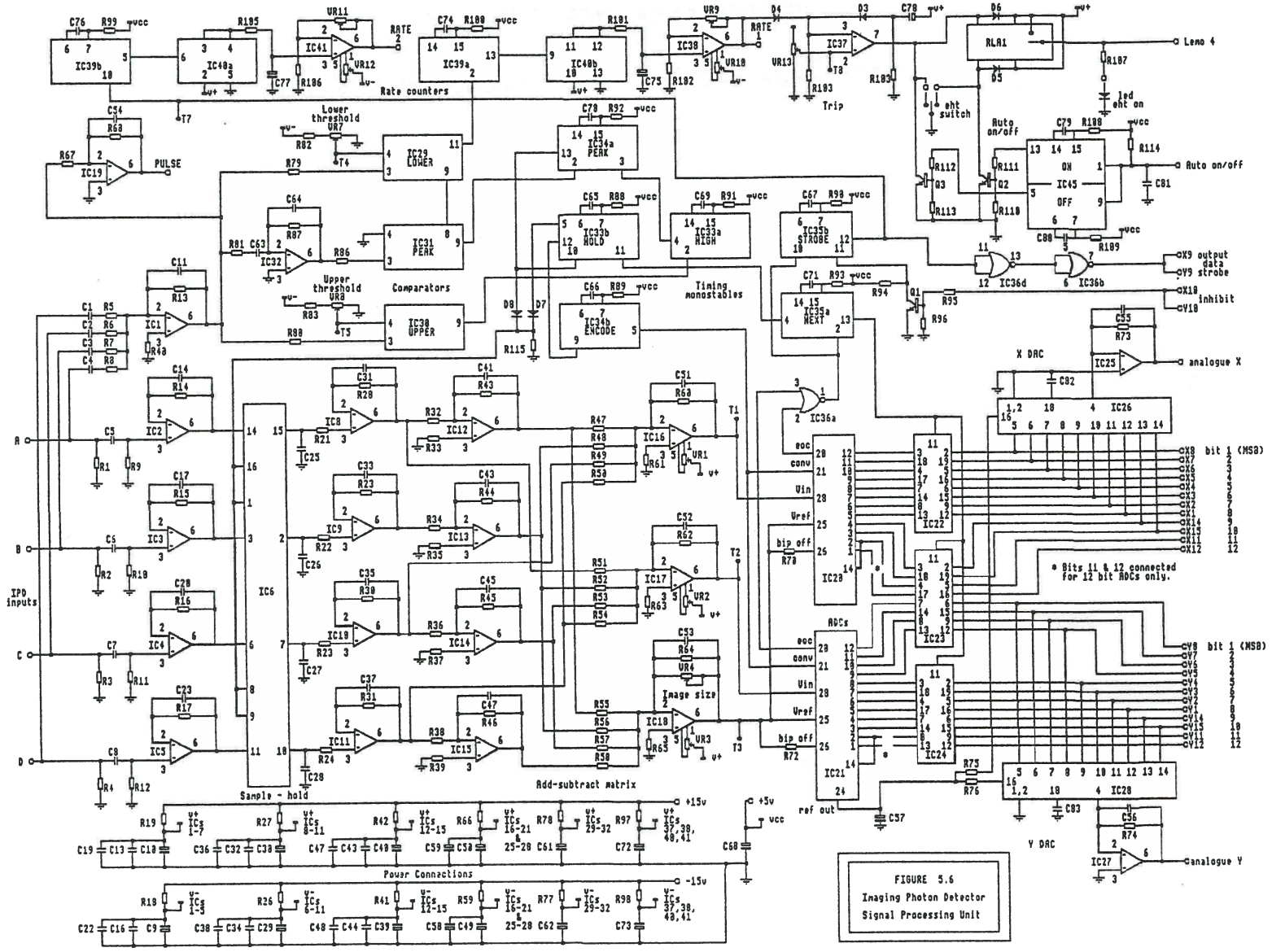


FIGURE 5.6
Imaging Photon Detector
Signal Processing Unit

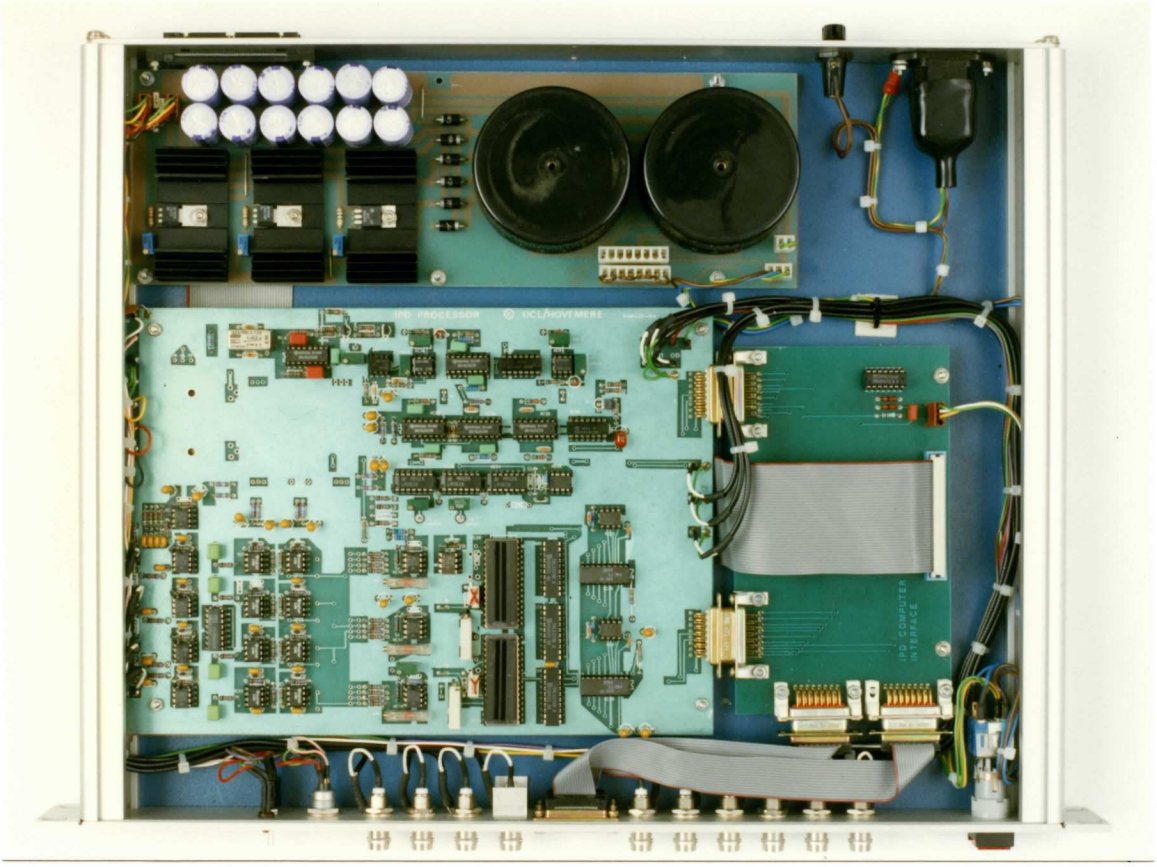


Figure 5.9 The fourth generation Signal Processing Unit

Figure 5.10 The SPU together with the IPD

Figure 5.11

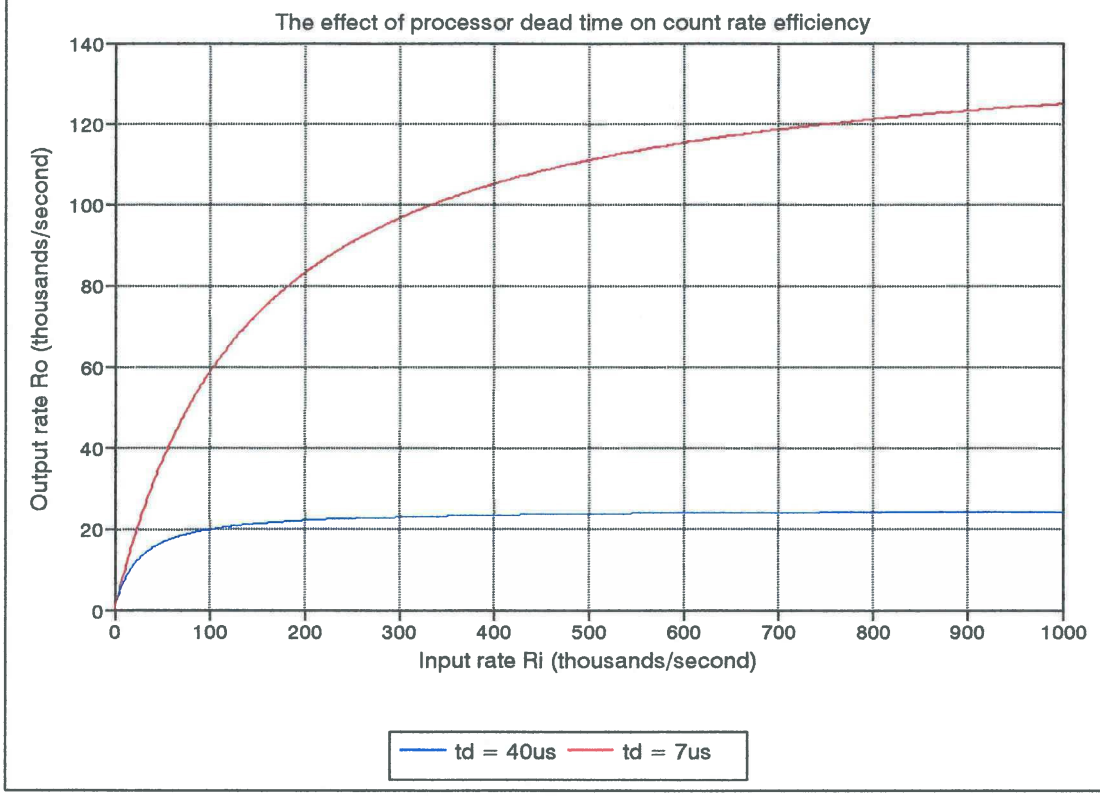
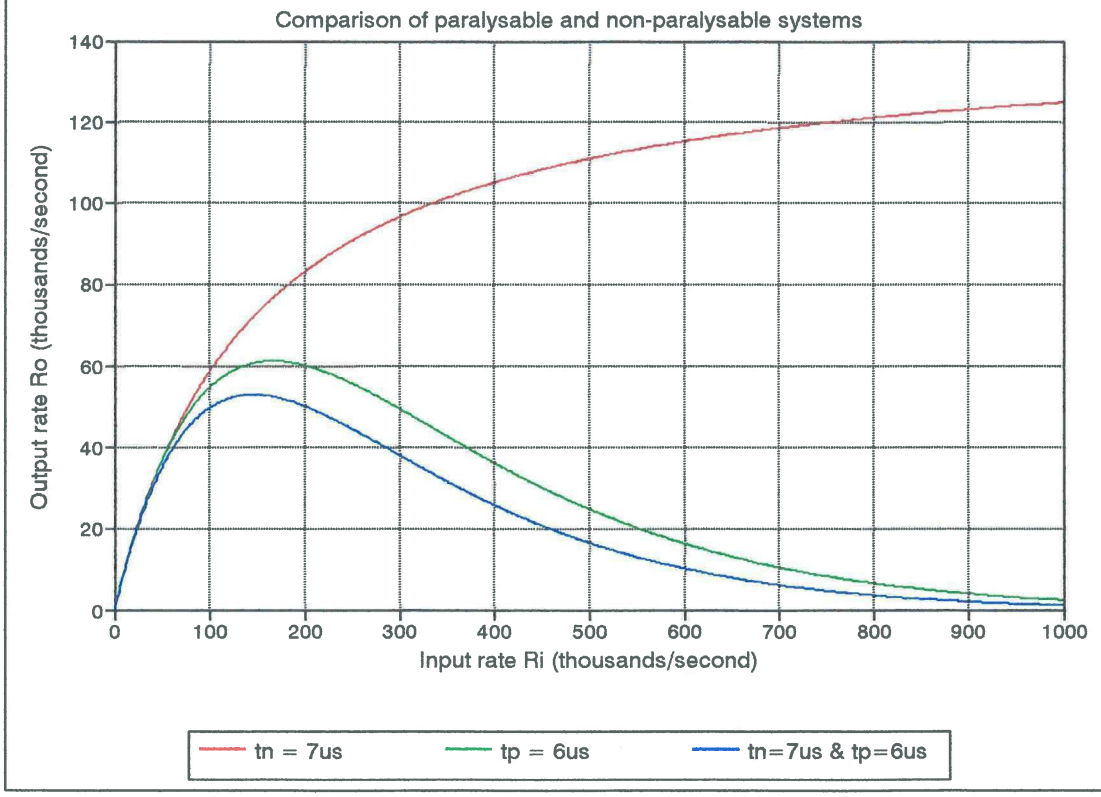


Figure 5.12



CHAPTER 6 - THE PRACTICAL REALISATION OF THE IMAGING PHOTON DETECTOR

6.1 Introduction

The preceding chapters have described in detail the component parts of the Imaging Photon Detector and its signal processing electronics. This chapter discusses the fabrication of the complete device and the practical aspects of its operation. Topics covered include the determination of suitable electrode voltages, adjustment of charge amplifier gain and incorporation into a suitable housing.

6.2 The manufacture of the detector

Although the IPD is not a standard production device, its manufacture is now a routine matter for an experienced and well equipped vacuum technologist. However, at the start of the author's involvement with these devices the tube was in its infancy. Originally, the UCL IPD tube was developed as a collaborative project between University College London and a Sussex company, Instrument Technology Ltd.

6.2.1 The prototype device

The initial problem to be overcome in the tube's manufacture was the requirement to fabricate the photocathode on a substrate away from the microchannel plate stack in order to avoid cross contamination, and then to unite the two assemblies under vacuum. This was achieved in the prototype tube by using a glass vacuum envelope large enough to contain the microchannel plate assembly and the photocathode substrate side by side. The photocathode was manufactured inside the envelope and then moved across on rails until registered over the microchannel plate assembly. This was a very clumsy arrangement and a more elegant solution was sought.

6.2.2 Devices with opaque photocathodes

Trials were carried out with tubes which used an opaque photocathode. This was deposited directly onto the face of the first microchannel

plate, on the surface in between the channel openings. The tube is described and illustrated in *Rees, McWhirter et al (1980)*, where it is explained that this approach offered a number of potential advantages. These may be briefly summarised as:

1. The construction of the device was greatly simplified.
2. The sources for the photocathode were contained within the device. It was, therefore, possible in principle to re-process the photocathode without dismantling the tube (*UCL Dynamics Explorer memo 3, 1979*).
3. The geometry meant that the photocathode was not susceptible to degradation from ion bombardment from the microchannel plates.

However, a number of disadvantages were found which meant that only limited success could be achieved with this approach. These were:

1. The open area of the channel plates was not photosensitive. As a result the effective photocathode area was only 40% of the total area of the MCP face.
2. In order to minimise contamination it was necessary to curtail the photocathode process, thereby reducing the potential sensitivity.
3. The electron optics were poor. For successful imaging, it is important that each photoelectron should enter a channel adjacent to its point of emission, but the absence of a strong accelerating potential meant that this was by no means certain, especially in the centre of the device. Furthermore, it became apparent that many of the photoelectrons were either not reaching the channels at all (being re-absorbed into the surface) or that they were entering the channels with insufficient kinetic energy to initiate a cascade. Although moderate photocathode sensitivities in the region of 80 μA/lumen could be obtained, the overall efficiency of the device was consistently very low.

6.2.3 Devices with transparent photocathodes

Two important developments meant that devices with transparent photocathodes became a possibility. Firstly, the "filmed" microchannel

plates described in paragraph 2.5 became available. These have an ion barrier on the input face. The ion barrier consists of a thin layer of aluminium oxide, typically 5nm thick. This layer is transparent to electrons of sufficient energy but provides an effective barrier to prevent ions generated by the microchannel plate collisions from reaching the photocathode.

The second development is the "transfer facility" (*Dolizy and Legoux, 1969*). This consists of an evacuated chamber in which the IPD is manufactured. The tube, made from stainless steel, is built in two halves. One half contains the microchannel plate assembly, resistive anode and lead-through connections for the electrode voltages. The other half contains the photocathode substrate, which also serves as the window for the device. The window is about 5mm thick and is usually made of glass. Other materials, such as quartz and magnesium fluoride, may be used where it is required to extend transparency into the near ultra-violet region. Alternatively, a fibre-optic bundle may be substituted for the window should the optical configuration of the system require this.

Using such an arrangement, the photocathode can be made and tested quite separately from the rest of the device. Once the photocathode performance has been assured, the two halves of the detector are brought together inside the transfer mechanism. Mating flanges equipped with a v-groove ensure accurate registration of the two halves as the detector is pressure-welded. The two halves are pressed together with great force and a layer of indium contained in the v-groove fuses to form a vacuum-tight seal. A photograph of the tube is shown in figure 6.1 and the internal configuration is illustrated in figure 6.3. This method of construction was adopted as the standard for over 50 tubes manufactured by Instrument Technology Ltd., which were evaluated and commissioned by the author at University College London.

6.2.4 The ITT double stack devices

The double stack (five MCP) detectors were made by ITT Electro-Optics at Fort Wayne, U.S.A. These were consistently of a higher standard than the ITL tubes. They also had the advantage that they did not have a metal casing at photocathode potential. Instead, the assembly was encapsulated in silicone rubber and could be mounted directly into a housing at ground potential. The main reason for switching to ITT was that they possessed the facility for manufacturing devices with gallium arsenide (GaAs) photocathodes (*Clampin and Paresce, 1989*). The enhanced infra-red response of this type of device was required primarily for the observation of hydroxyl emissions at 843nm. Six GaAs devices were purchased. Unfortunately, ITT decided to close down their manufacturing facility for these tubes, and this coincided with failures in four of them. One device failed with dramatically visible internal voltage breakdown and one with total loss of photocathode sensitivity. Another was destroyed by the failure of an external component (a bias resistor). The fourth tube developed an intermittent partial short circuit across the V MCP stack, although it is still working with reduced sensitivity. The remaining two tubes continue to work well. Two tubes with S25 photocathodes were also purchased, and one with an S11 response. An ITT tube is shown in figure 6.2 and a diagram of its internal configuration in figure 6.4. The ITT devices all had extremely good photocathode responsivity, and as a result, high thermionic emission. Consequently, they needed to be installed in the cooled housings described later in this chapter.

6.3 The commissioning process for the tubes

The commissioning process involves integration of the tube into a suitable housing with charge/shaping amplifiers and high voltage power supply, determination of the optimum electrode potentials, and adjustment of the amplifier gains if necessary.

For each tube, operating voltages are chosen which give the best compromise of resolution, pulse height distribution, counting

efficiency, low field emission and minimisation of other spurious effects, such as noise from channel plate defects.

The testing carried out at Instrument Technology Ltd. was minimal, so it was necessary to set up comprehensive test facilities at UCL for measuring parameters such as resolution, distortion and pulse height distribution. The careful evaluation of each device by the author has been a critical final stage in the quality control, and it was sometimes necessary to reject devices on the basis of their performance at UCL. Considerable experience was accrued in running the tubes and optimising their performance.

The tubes obtained from ITT Electro Optics were subjected to much more comprehensive testing by the manufacturer and it has not been necessary to reject any tubes on the basis of their initial performance.

6.4 Operating potentials for the IPD electrodes

The tubes are operated with the anode near ground potential and the photocathode at between three and five thousand volts negative, depending on the number of MCPs. When discussing inter-electrode potentials, those which accelerate electrons towards the anode are deemed to be positive and those which retard the electrons are considered negative.

6.4.1 Voltages for the single stack device

The values of the resistor chain are selected to give the required electrode voltages. A "standard" set of resistor values has evolved which is appropriate for the majority of tubes, but some tubes require variations on this. The fine tuning of the microchannel plate voltage V_{mcp} is achieved by adjusting the overall voltage once the correct resistor ratios have been determined. This makes only a small and relatively unimportant difference to the other electrode voltages. The other potentials required are: V_{pk}, between the photocathode and the first microchannel plate, and V_{an}, between the last microchannel plate

and the anode. The voltage requirements for operation of the tube are outlined below.

V_{pk}: It is recalled from paragraph 2.5 that this should be as high as possible to ensure maximum penetration of the ion barrier and minimum lateral motion of the photoelectron. However, as discussed in paragraph 3.8.4, field emission spots can become a problem and it is sometimes necessary to reduce the voltage in order to minimise this effect. Too high a voltage can also cause internal breakdown inside the tube, resulting in damage to the photocathode and microchannel plates. Internal breakdown has been a persistent problem with ITL tubes and V_{pk} is, therefore, limited to 500 volts for these devices.

V_{mcp}: This is chosen to give a good Gaussian pulse height distribution at the optimum amplitude for the processing electronics. The channel plates should be run at the highest possible gain consistent with stable operation and a well formed pulse height distribution. As this gain is similar within each class of detector it is possible to use a standard charge/shaping amplifier and to adjust the channel plate voltage to give the correct range of pulse heights as measured on a pulse height analyser. If a pulse height analyser is not available, for example in the case of repairs in the field, it is possible to set the voltage approximately by illuminating the detector and adjusting the voltage for maximum count rate from the processing electronics. Sometimes the amplifier gain has to be modified to suit individual tubes. This can happen when a tube exhibits instability at a lower than normal gain, or when the tube has to be run at a lower gain in order to minimise a hot spot or other blemish on the image caused by a microchannel plate defect.

The MCP voltage can be applied across the entire stack so that it is shared equally between all the plates, provided that they all have a similar resistance. However, the first channel plate in the stack is filmed, as explained in paragraph 6.2.3. Because these plates are manufactured separately it is not always economic to select plates which have good resistive matching to the remainder of the stack. When

matching is not achieved, an intermediate connection is fitted to the junction of the first MCP and the remainder of the stack. In this configuration the filmed plate is normally driven with one third of the total stack voltage, but this can sometimes be varied to obtain an improved pulse height distribution.

V_{an}: This is not a critical parameter as long as the voltage is not so low that the charge cloud spreads over the edge of the anode. A potential of 100 volts is satisfactory.

6.4.2 Voltages for the double stack device

These devices require two separate channel plate potentials, V_{vmcp} and V_{zmcp} for the V and Z MCP stacks. The potential of the gap between the two stacks is V_{gap}. These voltages need much more careful fine tuning than with the single stack detectors. The optimum operating voltages are assessed by ITT prior to shipment and are detailed in the data sheet supplied with each tube.

V_{pk}: As with the single stack device, this should be as high as possible. The ITT devices are particularly tolerant of high V_{pk} potentials and 700 volts is the usual value used for these tubes. Although higher potentials could probably be used this is the maximum value permitted under the conditions of the tube warranty.

V_{vmcp}: This should be as large as possible consistent with safe operation of the device, and is usually about 1800 volts.

V_{gap}: There are two modes of operation for the gap, which initially appear to be contradictory. The first is to operate with a positive potential to accelerate the electron cloud so that it covers a relatively small area on the face of the Z-stack. The second is to apply a negative potential which repels the lower energy electrons, allowing only the higher energy ones to reach the Z-stack. The rejected electrons will be those produced towards the end of the Vmcp channels which have low energy. Those emitted further back along the

V-stack channels are able to gain enough energy to overcome the gap potential before they leave the channel. They are, therefore, highly collimated and have very little lateral spread. The negative mode does, obviously, introduce a significant gain reduction, typically a factor of ten, so the Z-stack has to be run at a higher gain to compensate.

Floryan and Johnson (1989) have demonstrated that the best combination of resolution, pulse height distribution and counting efficiency is obtained with a negative V_{gap}, usually between -50 and -200 volts.

V_{an}: Again, this is not critical and is normally in the region of 100 volts.

6.5 The high voltage power supply

This unit is required to run from the 15 volt line from the IPD processor. The specific requirements for the supply are as follows:

1. It should be adjustable, and stable to within 0.1%.
2. It must be "slow-start", that is, on switch-on it should slowly rise to its set value over a period of about one second with no overshoot. This is to avoid stressing the microchannel plates and other internal components.
3. It should be low noise. Noise generated by the switch-mode power supply and other sources should be filtered to millivolt level.

The current requirement is modest. The strip current is in the order of 5μA and the resistor chain has a current demand of approximately 100μA. Two versions of the power supply are required, one for the single stack (three or four MCP) detector from ITL and one for the five MCP double stack device from ITT.

6.5.1 The Brandenburg power supply for single stack IPDs

This power supply uses a Brandenburg high voltage module type 482BN. The module has a very small physical size and is ideally suited to the

format chosen for the detector housing. As the module is not stabilised, it is necessary to add voltage feedback and slow start circuitry. The resistor chain voltage is held constant by measuring the voltage across a resistor in the ground end of the resistor divider chain. The resistor has a value of about 0.1% of the total, so the voltage across it is approximately 1 volt for each kilovolt measured at the top (photocathode) end of the chain. This arrangement means that the current through the resistors is held constant and consequently the voltage across any one resistor remains the same when one of the others is changed. This is very convenient when it is necessary to modify one of the inter-electrode potentials whilst leaving the others unaltered. The circuit diagram of the supply is given in figure 6.5. The Brandenburg module has proved to be extremely reliable, even though operated at the limits of its specification.

6.5.2 The Waytronics power supply for double stack IPDs

The five MCP arrangement requires an overall operating potential of up to 5kV, which is higher than the Brandenburg module is able to supply. This posed a problem as the detector housing would have needed to be considerably enlarged to accommodate the majority of power supplies on the market. Eventually, Waytronics Ltd. suggested that their small 2kV power supply could be safely modified to give an output of 5kV. The circuit was rather more complex than that of the Brandenburg unit. As it was stabilised and adjustable there was no need for additional circuitry. A special batch of boards was made which were the same physical size as the Brandenburg boards and included provision for the resistor network. Although broadly satisfactory, it was necessary to modify the circuit at UCL as the start up was not slow enough and tended to overshoot, which was quite unacceptable. Additional screening was also necessary to reduce the noise. Because the voltage feedback was already incorporated in the circuit it could not be derived from the resistor chain, so changes in any one resistor value affected all the electrode potentials. However, this was not a serious disadvantage. The configuration of the supply is shown in figure 6.6.

6.6 The gain of the charge and shaping amplifiers

There is normally sufficient latitude in the available adjustment of the microchannel plate gain to allow a standard amplifier to be used for each class of IPD. However, the amplifier gain sometimes needs to be increased to compensate for poor MCP performance and it was explained in paragraph 5.9.2 that a set of four standard amplifiers has been designed, comprising a standard and high gain unit for both single stack and double stack devices.

6.7 The detector housing

The housing for the IPD needed to fulfil several requirements:

1. The detector had to be mounted close to the charge/shaping amplifiers, which were to be incorporated into the housing.
2. The ITL detector can consisted of a metal body at a potential of approximately 3500 volts. This had to be insulated from the housing and the charge amplifiers, which were operated at ground potential.
3. It was essential that the housing should provide a complete screen around the detector to shield the detector, charge amplifiers and the anode contacts which operate at very low signal levels.
4. The high voltage supply had to be incorporated in the housing to avoid long high voltage cables leading to the housing. It was important to ensure that noise from this switching power supply was completely excluded from the anode circuitry.
5. The housing was designed to be as compact as possible consistent with ease of dismantling and servicing. It had to fit within the confines of the standard optical bench which was chosen for the balloon instrument and subsequently adopted for the majority of the UCL interferometers.

The housing is circular in design, with a cylindrical cover which slides off to expose the detector and circuit boards, as shown in figure 6.7. All parts are made of black anodised aluminium except for the plastic holder into which the detector is clamped. The high

voltage board is located in a screened compartment at the rear and the charge amplifiers are mounted directly behind the detector.

The detector housing is also shown in figure 5.10 from which the compact nature of the system can be appreciated. This shows the detector together with its Signal Processing Unit. Apart from the data acquisition computer, this constitutes a complete imaging system.

6.8 Cooling the IPD

It was explained in paragraph 3.2.3 that the process of extending the red response of the photocathode involves lowering the work function of the photocathode material. Consequently the thermionic emission is greatly increased. Detectors with S20 photocathodes can have a very low thermionic emission at room temperature, in the region of tens of counts per second. These can be used successfully without cooling. Extended red photocathodes, such as S25 and gallium arsenide, typically exhibit thermionic emissions as high as several thousand counts per second. Such devices are only of practical use if they are cooled to around -20°C. A cooled housing has been under continual development at UCL.

6.8.1 Methods of cooling

There are two approaches to cooling a detector. The first is to use a flow of refrigerated coolant and the second is to use a heat pump to remove the heat directly.

The first approach has the advantage of simplicity as far as the detector housing is concerned, but the main disadvantage is that a supply of liquid at the required temperature, typically -20°C, has to be maintained. This involves using an expensive circulating cooler and it is necessary to pay particular attention to the problem of condensation on the liquid supply pipes.

The second method offers more flexibility at the expense of a more complex housing. Peltier effect heat pumps are small semiconductor

modules measuring typically 30mm square by 3mm thick. The Peltier effect is a property of semiconductor materials. The current carriers in a semiconductor also transport thermal energy. The majority carriers are either electrons or holes, depending on whether the semiconductor is p-type or n-type. The Peltier module consists of two ceramic plates joined together by many semiconductor spacers which are electrically connected in series so that a current may be passed through them. The semiconductor elements are alternately p-type and n-type so that whereas the direction of the current zigzags between the plates, the heat always flows in the same physical direction. The flow of current creates a temperature differential of tens of degrees Celsius between the hot side and the cold side. If the hot side is then cooled, to ambient or below, the temperature differential will be maintained. Heat may be removed from the hot side using forced convection with a fan or by a flow of liquid coolant.

6.8.2 The UCL fan-cooled detector housing

The first type of cooled housing to be built at UCL was fan-cooled. The detector was encased in an aluminium shroud from which four Peltier modules transferred heat to heat sinks, which formed the sides of the rectangular housing. Because the shroud was at the same potential as the detector can, over 3000 volts, a layer of thermally conducting electrical insulation had to be inserted between the shroud and the Peltier modules. A fan mounted at the rear of the housing drew air past the heat sinks. The space between the detector shroud and the sides of the housing was filled with an expanding foam filler, created by mixing two liquids together. A glass window in the front of the housing enabled light to reach the photocathode. A ring of glassfibre tubing filled with silica gel dessicant was placed in the gap between the detector window and the housing window. This ensured that the air in this cavity was completely dry, thus avoiding the formation of ice on the detector window.

The fan-cooled housing suffered from serious problems. The cooling efficiency was poor as, even with the most powerful fan available, the

heat sinks remained at several degrees above ambient. The dependence of the detector temperature on the ambient temperature was also unacceptable. On a hot day, with the room temperature approaching 30°C, the detector would be cooled to only just below 0°C. This problem was aggravated by the conflict between the requirement for a free flow of air and the need to exclude light from the optical system. Unless it was possible to arrange for the detector housing to protrude from the light-tight covers of the instrument, the air temperature around the detector was inevitably above ambient. Another problem was condensation. The front face of the housing and the glass window were colder than the surrounding air as a result of conduction to the internal shroud. The continual flow of moisture-bearing air meant that condensation would form on the face of the device. In spite of the efforts to seal all the mechanical interfaces with O-rings and silicone rubber, water would sometimes seep into the housing and cause voltage breakdown between the shroud and the rest of the housing.

6.8.3 The UCL liquid-backed Peltier cooled housing

Although the fan-cooled housing had some limited success, it was far from satisfactory for the reasons stated above. It was also very difficult to assemble, the heat sinks requiring careful clamping into position prior to filling with foam. A completely new housing was designed which dispensed with the fan and instead used a flow of liquid through copper blocks clamped to each Peltier module. This approach gave a dramatic improvement in cooling efficiency. The liquid can be mains water, which is usually cooler than the air temperature and subject to less fluctuations. Alternatively, a flow of water/anti-freeze mix can be supplied by a circulating cooler which will, if required, cool the liquid to below zero, enabling detector temperatures as low as -30°C to be obtained. The inconvenience of providing a flow of coolant was far outweighed by the increase in cooling performance. The copper blocks were connected together with copper pipe, with two pipes protruding from the rear of the device for connection to the liquid supply. Considerable improvements were made to the overall design. Specially shaped pieces of machinable

insulating foam were used where possible to provide the bulk of the insulation, leaving only a relatively small cavity to be filled with the expanding foam. Particular care was paid to maximising the amount of insulation between the internal shroud and the front plate. The Peltier modules were located inside the four corner posts of the device and the copper blocks clamped onto them by grub screws in the corner posts, ensuring a controlled, even pressure. A more reliable electrical insulating material was found for the Peltier-to-shroud interface. The one used previously was a silicone rubber impregnated woven material which was prone to breaking down, although it was specified for operation at the applied voltage.

It might appear that the use of a circulating cooler offers little advantage over the simple approach of cooling with a flow of liquid coolant alone. However, the usual coolant temperature of between 5°C and 10°C is far more manageable and suitable for long term unattended operation than a coolant temperature of around -30°C, which would be required for cooling with liquid alone. A photograph of the cooled housing is shown in figure 6.8.

6.8.4 Flow requirements for the Peltier cooler

It is important to ensure that the flow of liquid for the cooler is sufficient to remove the heat generated. The heat consists of two components: the heat removed from the detector, which equals the heat flowing in through the insulation of the unit, and the ohmic heating of the Peltier modules. The first quantity is assessed from the manufacturer's data on the performance of the Peltier modules and the latter is simply the product of the applied current and voltage.

The units typically require a supply of eight volts at eight amps, which equals 64 watts of ohmic heating. According to the manufacturer's data, this will remove heat from the cold junction at a rate of approximately 40 watts. This gives a total of 104 watts, which is the required cooling capacity of a circulating cooler, should one be used. This is equivalent to approximately 25 calories per second. A

rise in coolant temperature of two degrees as it passes through the detector can be considered acceptable. It then follows that a flow of 750 millilitres per minute is required, assuming that the specific heat of the coolant is approximately that of water.

The temperature of the cooled detector will rise in step with the average coolant temperature, which is determined by taking the average of the input and output temperatures. It is, therefore, possible to use a slower rate of flow if some impairment in cooling performance is acceptable.

6.8.5 The IPD thermometer and safety cut-out

It became clear that the coolant flow was a critical single point failure. Because of its mechanical nature it is one of the most unreliable components in the system. Failure of liquid flow results in the Peltier modules becoming hotter and hotter due to the ohmic heating, so that even the cold side of the junction becomes hot enough to damage the detector. Unfortunately, this happened on one occasion, resulting in the demise of the detector. In order to monitor the cooling performance and prevent this from re-occurring, an electronic thermometer has been constructed. This reads the temperature from a precision thermistor mounted on the detector shroud. The temperature read-out is on a vertical bar of light-emitting diodes, measuring in five degree intervals from -35°C to +30°C. If the temperature reaches +30°C this is interpreted as a fault condition. An alarm circuit sounds a buzzer and operates a solid state relay to interrupt the power to the Peltier modules. The power then remains off until the problem is diagnosed and the unit is re-set.

Figure 6.1 The ITL single stack IPD

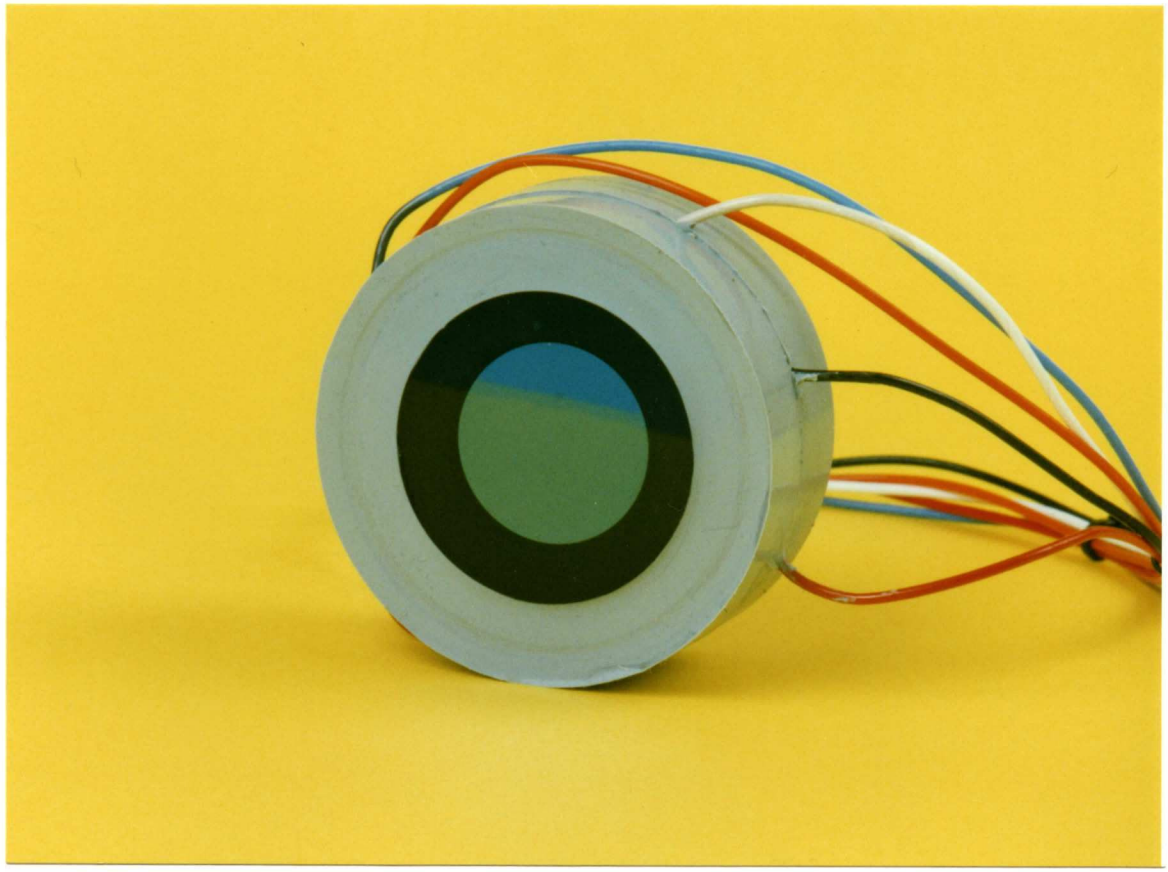


Figure 6.2 The ITT double stack IPD

Figure 6.3
The ITL single stack IPD

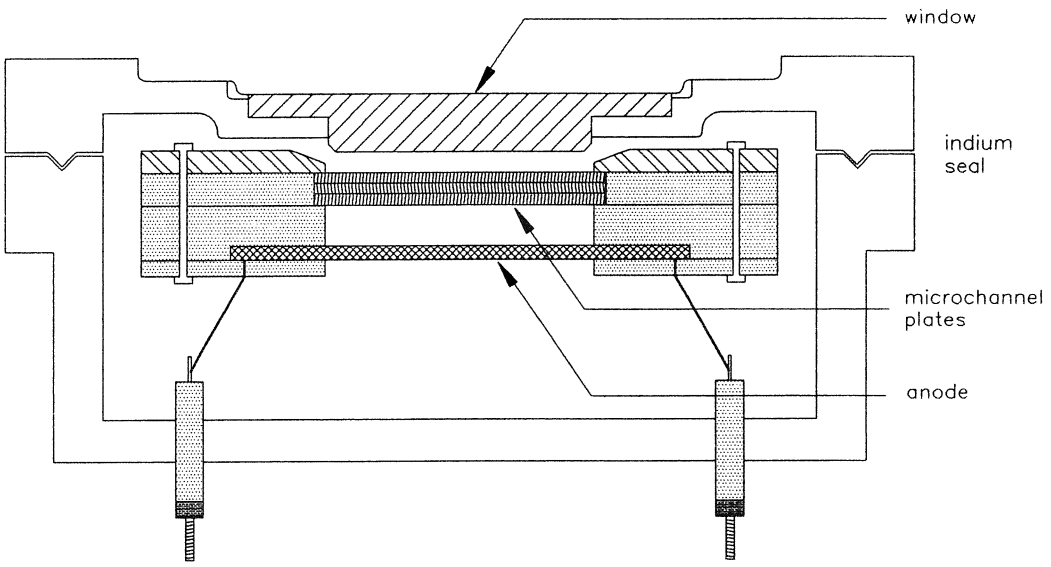
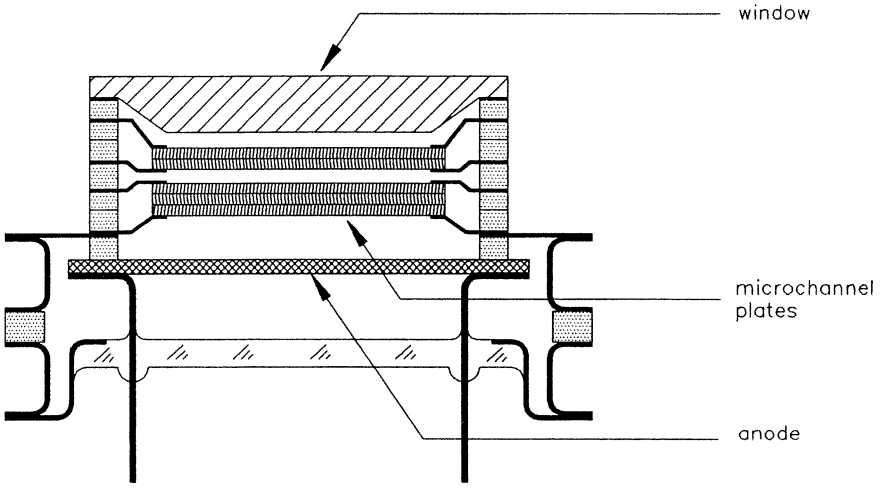


Figure 6.4
The ITT double stack IPD



From Firmani et al

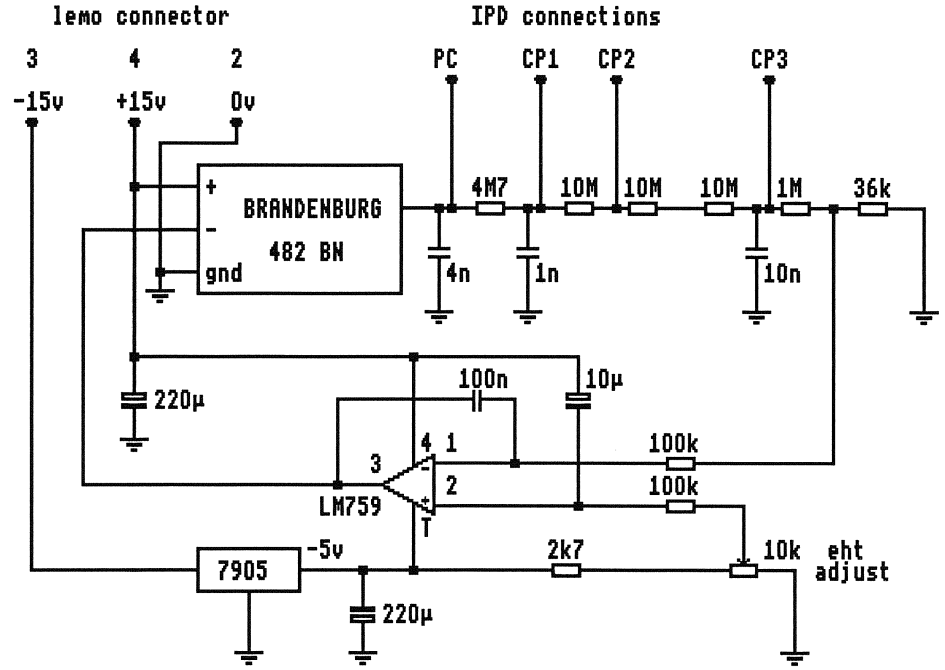


FIGURE 6.5
THE BRANDENBURG HIGH VOLTAGE POWER SUPPLY
FOR THE ITL SINGLE STACK IPD

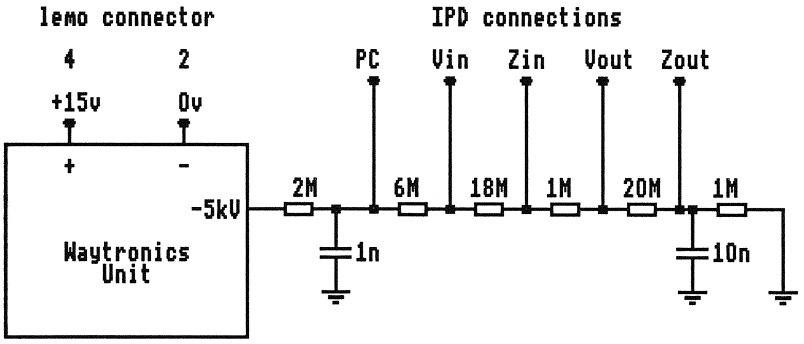


FIGURE 6.6
THE WAYTRONICS HIGH VOLTAGE POWER SUPPLY
FOR THE ITT DOUBLE STACK IPD

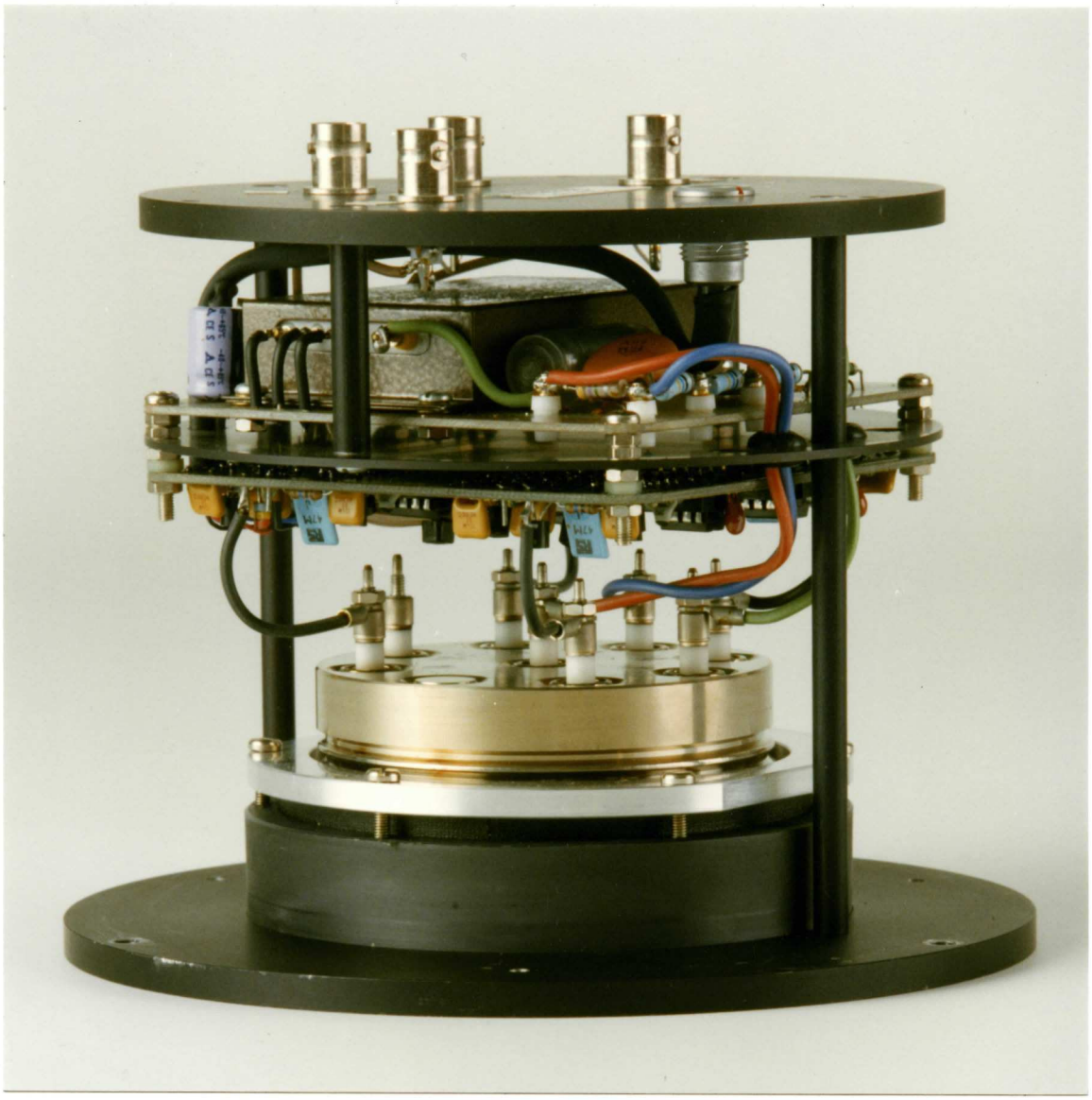


Figure 6.7

The ITL IPD in its housing

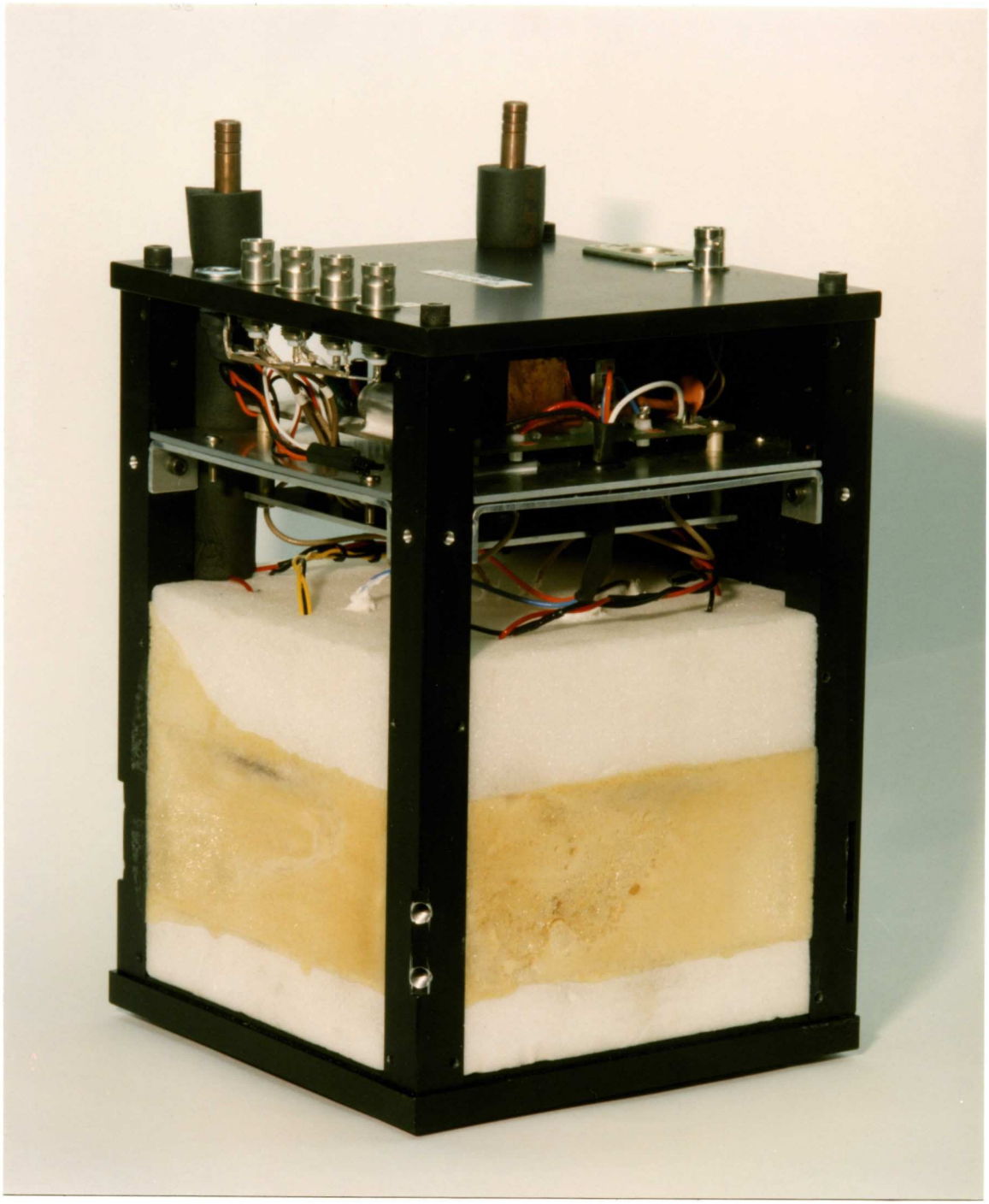


Figure 6.8

The IPD cooled housing

CHAPTER 7 - THE IMAGING PERFORMANCE OF THE RESISTIVE ANODE IPD

7.1 The image quality

The quality of an imaging system may be assessed by addressing the following points:

1. Spatial resolution.
2. Uniformity of sensitivity across the image field.
3. Background noise and blemishes.
4. Geometrical linearity of the image.
5. Linearity of intensity response.
6. Positional stability of the image and stability of the above parameters.

The previous chapters have discussed the mechanisms of image degradation throughout the system. In this chapter the predominant causes are summarised and quantified.

The performance of each IPD is assessed using a test chart comprising black and white bars of known spacing. The chart is projected onto the photocathode at 1:1 magnification. The test pattern is shown in figure 7.1. The dimensions of the bars are given in line pairs per millimetre (lp/mm), a line pair constituting one black bar and one white bar. Figure 7.2 shows a bar chart image taken with a single stack (ITL) detector and figure 7.3 the same bar chart using a double stack (ITT) device. Although the resolution of the software is only 256 x 256 pixels the improved performance of the double stack device is clearly seen. The images were taken using Halo, an imaging and instrument control programme written by Daniel Wade at UCL. The Halo programme is described in more detail in chapter ten.

7.2 Factors affecting spatial resolution

The goal of the IPD system is to determine the exact location of each photoelectron generated by the photocathode of the detector. Positional uncertainties arise at a number of stages between the photocathode and the final stages of the processing electronics. Various mechanisms for this have been discussed in detail in the

previous chapters and the overall imaging performance may be assessed by combining the contributions from each stage. The important stages of image degradation are:

1. The proximity focussed lens.
2. Microchannel plate quantisation.
3. Resistive anode noise.
4. Noise in the amplifiers and the Signal Processing Unit.
5. Computing accuracy of the Signal Processing Unit.

7.3 Assessment of the point spread function

The spatial resolution is mostconveniently quantified by measuring the Full Width at Half Maximum (FWHM) of the point spread function, i.e. the profile of a point source or a narrow slit imaged on the photocathode.

The combined effect of the individual mechanisms can be estimated by adding the squares of all the contributions. Although this is only strictly valid if the profiles are Gaussian, it does provide a reasonable estimate of the expected performance.

Measurements of perturbation and noise are measured in terms of root mean square (RMS) deviation, whilst quantities relating to physical dimensions are usually determined in terms of FWHM. It is a simple matter to convert between the two, as explained in paragraph 3.8.1:

$$\text{FWHM deviation} = 2.3548 \times \text{RMS deviation} \qquad [3.15]$$

7.4 The proximity focussed lens

The mathematics of the proximity focussed lens were discussed in detail in chapter three. It was shown that the RMS deviation is given by equation [3.25]:

$$\sigma = L(2\bar{V}_r/V)^{1/2} \qquad [3.25]$$

It was explained in paragraph 3.8.4 that 500μm is a satisfactory compromise for the size of the gap. Referring to figure 3.5, in which resolution is plotted as a function of applied voltage, it is seen that the RMS value of the point spread function is in the region of

40μm for the 400 volt potential used for single stack ITL devices, and 30μm for 700 volts, used for double stack ITT devices. This assumes an average radial energy, V_r of 200meV. However, there is considerable variation in the accepted values given in the literature, as noted by *Clampin et al (1988)*. This is due not only to the difficulty in measuring V_r but also to the fact that it varies with wavelength and with the precise structure of the photocathode.

7.5 Microchannel plate quantisation

The channels of the MCPs used in the IPD are 12.5μm in diameter and spaced 15μm between centres. This defines the ultimate accuracy to which the position of an incident photoelectron can be determined.

The precise image degradation due to an array of elements such as a microchannel plate is dependent on the exact orientation of the test image with respect to the channel plate matrix. However, *Eberhardt (1981)* has demonstrated that it is reasonable to approximate the FWHM of the point spread function to the diameter of a single channel.

7.6 Resistive anode noise

It was shown in paragraph 4.8 that the mean square charge noise of an anode with two-terminal resistance R_2 is given by:

$$\overline{q_n}^2 = 3.6kt\tau/R_2 \qquad [4.11]$$

where τ is the time constant of the shaping amplifiers.

To determine the effect on resolution it is necessary to express the charge noise in terms of electrons. Assuming a shaping time constant of 0.1 μs and a value of 24k for R_2, the mean square noise is given by:

$$\overline{q_n}^2 = 7.456\times10^{-27}/R_2$$

$$= 3.12\times10^{-31} \text{ coulombs}^2$$

therefore:

$$q_n(\text{RMS}) = 5.58 \times 10^{-16} \text{ coulombs}$$

$$\simeq 3500 \text{ electrons RMS.}$$

The uncertainty in position is thus given by:

$$\text{PSF (FWHM)} = 2.3458 \cdot \text{Anode diameter} \cdot \frac{q_n \text{ (RMS electrons)}}{\text{MCP gain (electrons)}} \qquad [7.1]$$

The anode diameter is 25mm. The same anode is used in both 18mm and 25mm devices.

The anode noise contribution is, therefore:

$PSF_{(anode)}$ = 57μm FWHM for single stack IPDs (MCP gain = 3.6×10^6)

and,

$PSF_{(anode)}$ = 7μm FWHM for double stack IPDs (MCP gain = 3.0×10^7)

It is clear that the anode noise is a dominant form of image degradation for the single stack device. The double stack device offers a very considerable improvement in this area.

7.7 Electron statistics in the charge cloud

The charge cloud incident upon the anode consists of a finite number of electrons, of the order of 10^7. The centroid of the charge is computed according to the "centre of gravity" of the footprint of the cloud on the anode. This is subject to an uncertainty of $\sqrt{N}/N$, where N is the number of electrons. The FWHM of the charge cloud is typically 1mm (*Clampin et al, 1988*), so the uncertainty is less than one micron and may be ignored.

7.8 Charge and shaping amplifier noise

The noise from the charge and shaping amplifier may be measured in terms of electrons RMS referred to the input. Without specialised equipment for noise measurement it is difficult to obtain a precise figure for the amplifier. The original amplifier design by *Alexander (1971)* was quoted as having a noise of 2000 electrons FWHM. Measurements taken using an oscilloscope indicate that the noise from the UCL amplifier is of a similar level.

7.9 Computing accuracy of the Signal Processing Unit

The accuracy with which the signal processing unit can determine the position of an event is subject to a number of uncertainties. Some of the mechanisms are dependent on signal amplitude and position.

Three main effects are observed:

1. Overall positional uncertainty.
2. Degradation towards the edge of the image.
3. Fixed-pattern noise.

The major problem areas are:

1. Electronic noise, resulting in general positional uncertainty.
2. Amplifier non-linearity and slew rate, affecting the amplitude ratios
3. Peak position detector errors, again modifying the amplitude ratios
4. Sample-hold "droop" and post sample-hold noise, resulting in ADC "preferred codes". This is observed as fixed-pattern noise.
5. Accuracy of ADC division, particularly with low amplitude signals.

7.9.1 Overall positional uncertainties

The computing accuracy is significantly reduced at low pulse amplitudes. This is due to the reduced accuracy of the analogue to digital converters at low values of reference voltage. It was seen in chapter five that this input is used as the denominator in the positional algorithm. The successive approximation ADC is very sensitive to noise on the inputs. The signal must be perfectly steady throughout the entire conversion period and any perturbations will result in an inaccurate computation. The electronic noise generated in the Signal Processing Unit should be small compared with the charge amplifier noise. However, careful design is necessary to minimise switching transients from the timing logic, and ground-plane construction is essential.

7.9.2 Edge degradation

There is a slight radial smearing of the image towards the edge. It is most easily visible when a matrix of pinholes is imaged on the device.

It has been suggested (*Durand et al, 1987*) that this image degradation is inherent in the anode, caused by the small amplitude of one or two of the signals when an event is near the edge of the anode. However, this is an incorrect assumption. The noise contribution of the anode depends on the ratio of the anode resistive noise to the magnitude of the denominator signal in the positional algorithm. This is quite independent of signal position. Position computing by the rise-time method will indeed suffer such edge degradation, as noted by *Allington-Smith and Schwartz (1984)*. This is because the uncertainty in deriving the timing information will be greater with small signals.

The most probable cause of impaired edge resolution is amplifier slew rate. Both the shaping amplifiers and the processor input stages have a finite slew rate, that is to say, the rate of change of signal is limited. Slew rate limitations in the charge/shaping amplifiers will affect both the timing and amplitude of the pulse peaks. If the ratios of the pulse amplitudes from each channel are changed, this will obviously cause an incorrect position to be computed. If the time taken for a pulse to reach its peak value varies between channels this will also affect the computed position. The peak amplitudes of the individual pulses are all measured at the same instant in time, which is determined from the peak position of the summed pulse. If the individual pulse peaks occur at slightly different times, then the measured amplitude for each pulse will not be that of the maximum, but the level just before or just after the peak.

This will be most noticeable towards the edges of the image. In this area the pulse amplitudes from the four anode signals are significantly different. Inaccuracies in timing and amplitude linearity have little effect in the centre of the image where all four pulses have approximately the same magnitude.

7.9.3 Fixed-pattern noise

It was explained in chapter five that the ADCs are, in essence, electronic scale pans, where the digital bits are applied in descending order and retained if they fail to "tip" the balance. However, if, for example, a brief noise pulse causes one of the more significant bits to be set incorrectly, the remaining bits will be

discarded (set to zero) in an attempt to correct the imbalance. For the same reasons, great care needs to be taken with the design of the sample-hold circuits as any "droop" or settling of the input signal during the ADC conversion time will have a similar effect.

The ADC output is said to exhibit "preferred codes" under these circumstances. The effect is observed as a fixed-pattern noise on the image, with lines appearing at the transitions between certain bits. In normal operation, fixed-pattern noise can sometimes be observed when the signal level from the charge amplifiers is too low. This occurs when the channel plate gain has dropped over a period of time. It is sometimes seen when uniformly illuminating the detector in order to obtain a "flat field" response. The high illumination level has the effect of reducing the modal gain of the MCPs, due to charge depletion in the MCP channels. When the image is processed it is possible to minimise any fixed-pattern structure by pixel smoothing. Each pixel is reassigned an intensity value which is derived from a weighted average of its own original value combined with that of surrounding pixels.

7.9.4 Measurement of processor PSF

Although it is difficult to quantify precisely the individual contributions of the different mechanisms outlined above, it is easy to assess the overall positional uncertainty of the Signal Processing Unit by applying a suitable test signal. Pulses derived from a signal generator are not satisfactory. They will not have the same pulse height distribution as real detector pulses, the time between pulses will not be random and there will be no pulse pile-up. What is required is a supply of real detector pulses with the positional information removed. The arrangement shown in figure 7.4 can be used to provide this. The four outputs from a standard IPD are added together electronically and then divided in a fixed ratio to give four pulse sources with the required characteristics. By varying the ratio, different positions may be simulated.

Figure 7.5 shows the point spread function obtained with such a test. It is seen that, although there are a few events spreading away from the main peak, the FWHM is in the region of one 25μm pixel.

7.10 Summary of spatial resolution degradation

Single stack MCP (ITL)

FWHM of Point Spread Function for each stage:

Proximity focus, gap = 500μm at 400v (from figure 3.5) (Assuming V_R = 200eV)	40μm
Microchannel plate quantisation	12.5μm
Resistive anode noise for R_2 = 24k, 3500 electrons RMS (MCP gain = 3.6 x 10^6)	57μm
Amplifier noise, 2000 electrons RMS	33μm
Signal Processing Unit uncertainty	25μm
Total positional uncertainty	82μm

Double stack MCP (ITT)

FWHM of Point Spread Function for each stage:

Proximity focus, gap = 500μm at 700v (from figure 3.5) (Assuming V_R = 200eV)	30μm
Microchannel plate quantisation	12.5μm
Resistive anode noise for R_2 = 24k, 3500 electrons RMS (MCP gain = 3.0 x 10^7)	7μm
Amplifier noise, 2000 electrons RMS	4μm
Signal Processing Unit uncertainty	25μm
Total positional uncertainty	42μm

7.11 Measurement of image resolution

In order to measure the resolution more critically it is necessary to store the image at the full 1024 x 1024 resolution available from the Signal Processing Unit. The current imaging software will only record images at 256 x 256 pixel resolution. This is mainly due to memory constraints, both in the computer RAM and in the space required to store images on the computer's disc. In order to achieve the required resolution a switching box was made which used the two most significant bits of each axis to select events from one sixteenth of the overall image. A switch enabled different sections of the image to be chosen and the remaining eight bits from each axis were mapped onto the 256 x 256 Halo format. The selected section of the image was thus displayed at maximum resolution.

Figures 7.6 and 7.7 show a comparison of the point spread functions of the single and double stack devices respectively. These were obtained by imaging a 10μm pinhole on to the photocathode. The full width half maximum of the profile is seen to be in agreement with the calculated values. Figure 7.8 shows a high resolution image of a section of the resolution chart, taken with a single stack device. The sections shown are the 4, 5 and 6.25 lp/mm bars. Figure 7.9 shows a graph of intensity across the 5 lp/mm section.

7.12 Positional stability of the image

For many applications, positional stability is critically important. The measurement of Fabry-Perot interference fringes in the UCL interferometers is a particular example. Any change in image position or size from one image to the next will result in an additional error in the measured Doppler shift. There are two potential problem areas. One is the Signal Processing Unit itself and the other the electron optics of the IPD.

7.12.1 Stability of the Signal Processing Unit

Two of the most critical areas of the processing electronics are the peak position detector and the sample-hold circuit. In the transition between third and fourth generation processing electronics, a prototype unit was built which used the fast pulses of the new high-

speed shaping amplifier with the sample-hold circuits of the third generation processor. The aperture time of the sample-hold was such that the peak position determination was subject to large uncertainties and was dramatically temperature dependent. The problem was largely eliminated by re-designing the sample-hold circuit using a high-speed analogue switch. However, there still remains a small residual temperature dependence, which is attributed to not only the sample-hold circuits, but also to the analogue to digital converters and possibly the summing networks. For optimum image stability it is recommended that the processor should be switched on for an hour before use in order to achieve thermal stability.

7.12.2 Image stability of the IPD

The IPD tube is inherently a very stable device. The electron optics of the proximity focussed lens are such that the image will only be perturbed by a very strong magnetic field. The microchannel plates physically contain the electrons, thus preventing lateral drift. Finally, the centroiding nature of the processing electronics means that the spreading out of the charge cloud in the MCP to anode gap is unimportant as long as it is uniform.

The resistance of the anode, as with any resistor, will change with temperature, but this will not affect position determination by the charge ratio method as the ratios of the pulse heights will not alter. However, it will obviously affect position determination by the rise-time method. It is for this reason that the resistive anode has been dismissed by some (*Allington-Smith & Schwartz, 1984*) in the light of their experience with rise-time electronics.

However, a small but significant change in image size has been observed in the field with a system recording very bright images. The instrument was a Doppler Imaging System, which is described in chapter ten. This produces an image consisting of several concentric rings, and the image is processed to display the size of each ring in radius squared space. An intensity dependent shift of about one radius squared element (out of 256) was observed. This corresponds to a shift in image diameter of about 50μm for a ring near the edge of the detector.

There are a number of possible explanations:

1. Pulse pile-up. In an image with circular symmetry this is a very probable cause.
2. Signal Processing Unit dependence on pulse amplitude. A very slight amplitude dependence has been observed in processors which are not optimally adjusted. Such shifts in image size could occur when the modal gain of the IPD pulses falls with an increase in light intensity.
3. Space charge effects in the IPD. This could be a possibility, in the MCP to anode gap and also in the V-Z gap of double stack devices.

Experiments at UCL have managed to duplicate the effect to some extent, but the results have not revealed the cause. The next version of the Signal Processing Unit will incorporate a pulse pile-up rejector, and particular care will be taken to address the pulse amplitude dependence.

7.13 Uniformity of sensitivity

The uncertainties involved in photocathode processing inevitably lead to non-uniform responsivity over the photocathode surface. The end product is highly dependent on the expertise and equipment employed during manufacture. Variations of up to 50% across the photocathode are not uncommon in the best of devices. It is essential for the majority of applications to obtain a test image with uniform illumination in order to perform a "flat field" correction to the final image.

Slight variations in microchannel plate gain across the diameter of the stack have been observed by measuring the pulse height resolution when selectively illuminating small areas of the photocathode. These variations are not a problem as they are generally small enough to be accommodated by the Signal Processing Unit. Some non-uniformities become less pronounced as the potential of the proximity-focussed gap is increased. These are attributed to variations in the ion barrier transmission, so it is important that the proximity-focussed gap

voltage is high enough to ensure that the barrier is sufficiently transparent over its entire area.

7.14 Thermionic emission

This was discussed in chapter three. Thermionic emission varies considerably, not only from one photocathode type to another, but also between photocathodes of the same type. It is important to bear this in mind when selecting a device for a particular application. Where appropriate, a thermionic emission image should be taken for subtraction from the subsequent images. Although cooling decreases the thermionic emission, the low residual rate must still be taken into account when calibrating the instrumentation.

7.15 Blemishes

These are almost always caused by manufacturing defects, which are outside the control of the end user. The major causes are field emission spots from the photocathode and channel plate noise caused by faulty channels or poor mounting at the plate edges. The use of hard edged plates instead of the early soft edged ones has helped to reduce this problem. Again, the significance of any blemishes will depend on the individual application for which the device is required. Small "hot spots" can easily be subtracted from the final image.

7.16 Positional linearity

The use of the Circular Arc Terminated anode has dramatically reduced the pincushion and barrel distortions inherent in simple resistive position sensors. The electron optics in the IPD are virtually distortion free, and non-linearities in the processing electronics are small compared with the residual distortion in the anode. Some devices, however, have exhibited pronounced barrel-like distortion, sometimes in just one quadrant of the image. This has been attributed to contamination of the anode during processing, and was confirmed in one case by inspection of a device which was dismantled after exhibiting severe image deformity. Furthermore, it has been noted that the distortion usually changes or disappears when anodes are re-processed into new devices. In early ITL devices, some distortion has

been attributed to internal cross-coupling of anode signals due to poor internal geometry of the electrode connections.

The resistive anode has been criticised for having inevitable local non-uniformities (*Allington-Smith & Schwartz, 1984*). However, the anodes provided by Surface Science Laboratories have been of high quality and, provided that the charge cloud is sufficiently spread out between the MCP stack and the anode, any small local variations in resistivity will be unimportant.

7.17 Linearity of intensity response

The linearity of response is subject to two mechanisms: pulse pile-up and MCP channel recovery. Pulse pile-up has been discussed in detail in chapter five. The effects are predictable and are depicted in figures 5.11 and 5.12. However, MCP channel recovery is much more difficult to quantify as it is highly image dependent.

It was explained in paragraph 2.8 that the local channel recovery time is surprisingly long, as much as 250 milliseconds. In a diffuse image the events are spread over many channels and pulse pile-up is a major limitation to count rate. However, for images of small, bright sources, such as those encountered in astronomical imaging, the channel recovery time is dominant.

The maximum count rate for a single MCP channel was estimated to be in the region of four per second. However, this figure was calculated for a single MCP channel and it would be difficult to test this experimentally. In a practical detector, the charge pulse is spread over several channels when it reaches the final plate of the MCP stack.

The most meaningful test is to illuminate the detector with a bright pinhole source and to measure the maximum count rate per pixel in the final image. This has been done, using a 10μm pinhole, for both single and double MCP stack devices. The maximum counts obtainable per 25μm pixel were 7 per second for the single stack device and 30 per second for the double stack device. This corresponds to 11,000 and 40,000 counts/second per mm^2. respectively. These figures are of the same order as the calculated value. The rate is higher for the double stack

device because the charge cloud is spread out over more channels in the gap between the two stacks. Also the double stack device uses low resistance plates for the Z stack, so the re-charge current is higher.

7.18 The Detected Quantum Efficiency of the IPD system

It has been seen that the DQE of the IPD system is related to a number of factors, some of which are highly signal dependent. The factors are listed below, together with an estimate of their contribution to a typical IPD system as used in the Fabry-Perot interferometers.

1. Photocathode RQE: typically 10% for an S25 at 630nm (figure 3.4).
2. MCP efficiency: open area ratio and ion barrier 60% (paragraph 2.5).
3. Local dead time of channels: significant for point sources but can normally be ignored for Fabry-Perot applications (paragraph 2.8).
4. Capture efficiency of the SPU: dependent on pulse height resolution. 80% for a pulse height resolution of 150% FWHM (paragraph 5.9.1).
5. Dead time of the SPU: signal dependent as seen in figure 5.11, but negligible for rates of less than 20,000 photons/second.
6. Data transfer to computer: negligible for modern computers, 80286 and above (paragraph 5.7.5).

The total DQE is, therefore, the product of (1), (2) and (4), or approximately 50% of photocathode RQE. This gives DQE of 5% for an S25 device at 630nm and 10% for S20 in the blue region (450nm) or a gallium arsenide photocathode at 700-900nm. Optimisation of pulse height resolution can improve this figure, as can the use of unfilmed microchannel plates at the expense of reduced device lifetime.

In order to verify the above figures tests were performed using the International Light IL1700 radiometer to calibrate a bandwidth-limited source, and the factor of approximately 50% of photocathode RQE has been confirmed.

7.19 Comparison with other imaging systems

It is important to compare the IPD with other imaging systems. The main alternative to the IPD is the Charge Coupled Device (CCD). The the performance of a CCD may be enhanced by the addition of an image

intensifier. Three devices are considered: Bare CCDs, linear intensified CCDs, and photon-counting intensified CCDs.

7.19.1 The Charge Coupled Device

The Charged Coupled Device can be described simply as a rectangular array of semiconductor light buckets. The device has an array of cells each of which accumulates an electronic charge in response to incident photons. The resolution of the device is equal to the number of cells. The size of the array varies from 256 x 256 cells for "budget" devices to 4096 x 4096 for expensive scientific grade devices. Higher resolutions may be obtained by combining arrays side by side. Recent advances in CCD technology have made them an attractive alternative to the IPD for many applications. However, the IPD still has specific advantages, particularly when compared with CCD systems of similar cost, rather than with state-of-the-art CCDs costing much more.

1. Signal-to-noise ratio

The main advantage of the IPD is that it is a single photon counting device and the readout noise is essentially zero. When each pixel on a CCD is read there is an associated readout noise, amounting to several electrons RMS per pixel. The overall noise is thus proportional to the number of pixels, and, as a result, longer integration times are required for a CCD in order to bring weak signals above the noise. A major advance in CCD technology has been the reduction of this noise. For devices comparable in cost to an IPD system, levels of 10-30 electrons RMS are typical, but for modern devices, using multiple readout techniques, the noise can be reduced to sub-electron level.

2. Detected Quantum Efficiency

To some extent, the readout noise of the CCD is compensated for by its high quantum efficiency. It has been seen that the DQE of an IPD is between 5% and 10%. CCDs typically have efficiencies ranging from 30% to 70%, and figures close to 100% have recently been reported.

3. Time resolution

The CCD is limited by the need to read the entire array, whereas the IPD can record individual photon events to sub-microsecond accuracy. The time resolution of the CCD is limited by the time it takes to read the array. This ranges from 40ms for television video rate readout to several minutes for slow scan multiple readout devices.

4. Positional stability and linearity

It has been seen that the positional stability of the IPD becomes less than perfect at high photon rates, due to pulse pile-up effects. The positional linearity depends largely on the quality of the anode. However, the stability and linearity of the CCD are exact as they are defined by the physical position of the individual cells.

5. Linearity of intensity response

The intensity response of the IPD becomes significantly non-linear above 20,000 photons/second. However, the CCD is linear up to the point where the individual cells saturate, typically 10^5 electrons per cell. The lower limit of the IPD is dictated by the thermionic emission, which is considerably lower than the readout noise of a CCD. CCDs are well suited to imaging point sources as they do not suffer from the local count rate limitations of microchannel plates.

7.19.2 The linear intensified CCD

The signal-to-noise ratio of a CCD at low light levels can be improved by the addition of an image intensifier. Experiments are currently being conducted at APL with a CCD camera fitted with a Gen. III (gallium arsenide) image intensifier. The intensifier consists of a vacuum tube containing a photocathode and a single microchannel plate electron multiplier, with a phosphor on the output window. The intensity of the image on the output window is approximately a linear function of the input illumination. The DQE is rather poor as the efficiency of the microchannel plate in unsaturated mode is low. However, it is expected to be useful for applications which have relatively high signal levels, such as the Doppler Imaging System.

The device was manufactured by EEV Ltd. and was fitted into the UCL cooled housing. This cools both the intensifier and the CCD chip. Early results are encouraging and integration times of several minutes have been achieved. The major problem area is that of low level structured noise in the image.

7.19.3 The photon-counting intensified CCD

This combines features of the IPD and the CCD and has some of the advantages and disadvantages of both devices. Its predecessor is the

original Image Photon Counting System (IPCS) This is well established in astronomical circles as a high quality photon counting imager. It has been described in detail in the literature (*e.g. Boksenberg and Coleman, 1979*). The system consists of an image intensifier, the output scintillations of which are viewed by a raster scan television camera. The output scan lines of the camera are then processed in the time domain to determine the centroid of each event. It has been a very successful system, although more complex than the IPD system and considerably more expensive and cumbersome.

By replacing the conventional camera with a CCD and using a microchannel plate image intensifier, the IPCS has become comparable in size to the IPD (*Fordham et al, 1986*) and modern portable computers can easily handle the more complex signal processing. Similar systems are now available commercially at a cost comparable to that of the IPD.

The main advantages of this technique are:

1. Each photon is recorded as an individual scintillation on the phosphor of the intensifier. The readout noise of the CCD is of secondary importance as it needs only to record the position of the scintillation, which is normally well above the noise level.
2. The resolution is potentially higher than that of a bare CCD as the position of each event can be ascertained by centroiding the scintillations to sub-pixel accuracy. However, it is limited by the proximity-focussed lens of the intensifier.
3. It is less susceptible than the IPD to pulse pile-up problems.
4. The image stability is comparable to that of a CCD.

Some disadvantages are shared with the IPD:

1. The DQE is dictated by the image intensifier, and is similar tho that of the IPD.
2. The local dead-time of the microchannel plates affects local count rates for point sources.

7.20 The role of the IPD in the work of the APL

The IPD has held its own very well over the past decade. Only relatively recently has the CCD been able to compete in terms of signal-to-noise ratio, and even then, at a considerably higher cost.

The main use of the IPD within the APL has been as the imager of the Fabry-Perot interferometers used for measurements of atmospheric emission lines. However, its relatively low cost and modest requirements in terms of supporting equipment has enabled a small pool of IPD systems to be accumulated and used in a number of additional projects for which little or no financial support was available. Its small size and extreme portability has enabled it to be used in many field campaigns where the transportation and operation of large quantities of equipment would have been impractical. It has been employed extensively in the field for narrow-band cometary images (several publications by *Meredith, Wallis and Rees - q.v.*) and for imaging of chemical releases from rockets and satellites (e.g. the AMPTE artificial comet experiment, *Rees et al, 1986*). A list of references relating to the IPD as a stand-alone imaging device is given at the end of this thesis.

An example of the results obtainable with the portable IPD imager is shown in figure 7.11 which is a narrow-band image of Comet Halley, taken by the author and Nigel Meredith during the apparition of 1986 at the Jacobus Kapteyn Telescope, La Palma. The complete imaging system consisted of a 300mm focal length Nikon camera lens attached to the front of an uncooled S20 IPD, with a Compaq portable PC computer to collect the data. This was guided by strapping it to the side of the main telescope, which itself was being used with a second IPD fitted to a more complex instrument, a Doppler Imaging System.

For the APL projects, bare CCDs with satisfactory noise performance and sufficient time resolution are likely to remain prohibitively expensive for the foreseeable future. However, photon-counting intensified CCDs now offer advantages over IPDs at comparable cost.

There is one area where the IPD technology remains superior. This is the ability to provide very high time resolution. Devices with high speed discrete element anodes will be used as the detectors for the ALOMAR atmospheric lidar project. In order to obtain the required height resolution of a back-scattered laser signal, photons will need to be timed to an accuracy of one tenth of a microsecond.

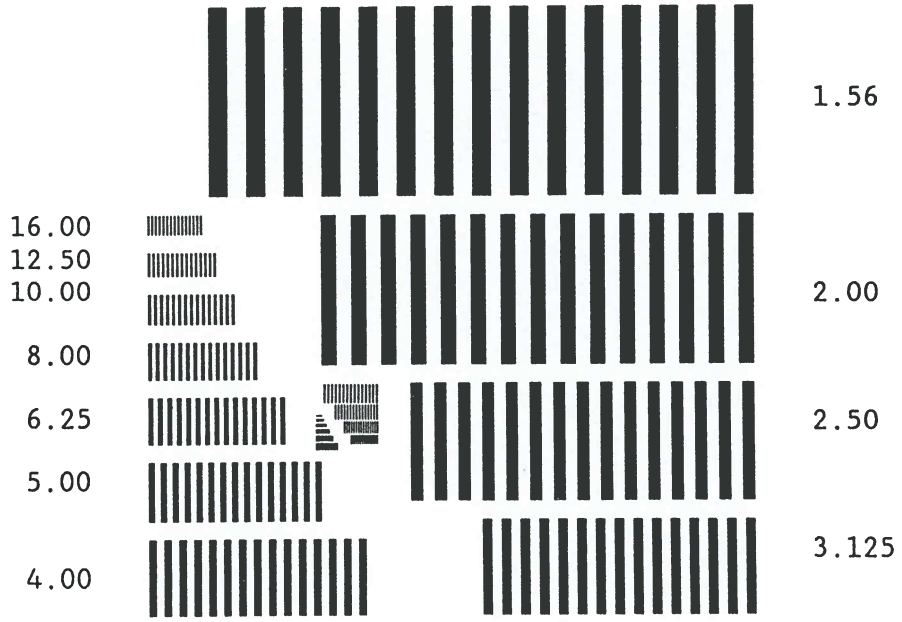
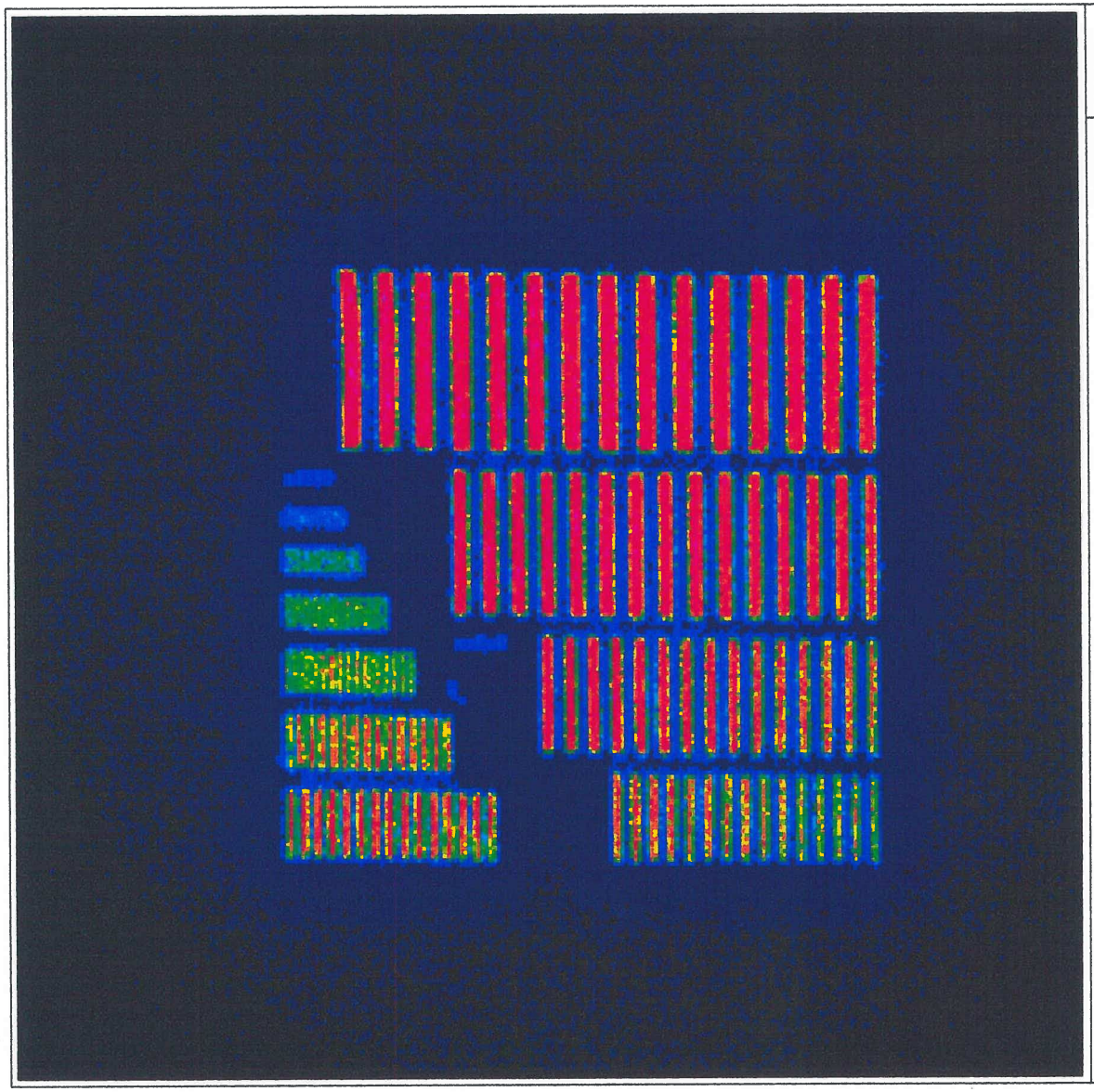


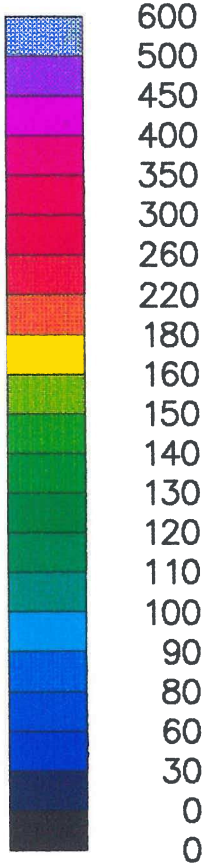
Figure 7.1

The resolution test chart
showing line spacing in line-pairs per millimetre



ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON

Resolution chart
File: TG11607.E08

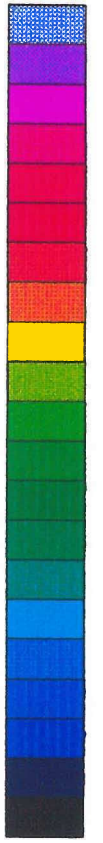


ITL single stack IPD

Figure 7.2

ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON

Resolution chart
File: DFC1341.E07



ITT double stack IPD

Figure 7.3

Figure 7.4

Arrangement for determining the PSF of the Signal Processing Unit

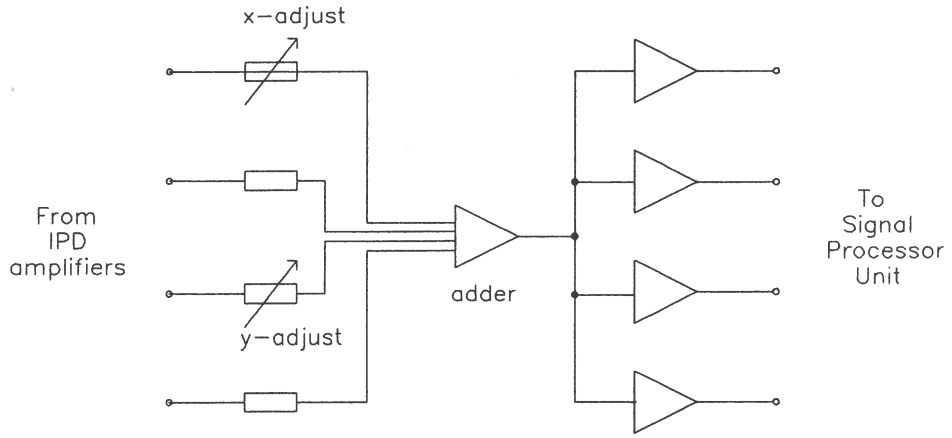


Figure 7.5

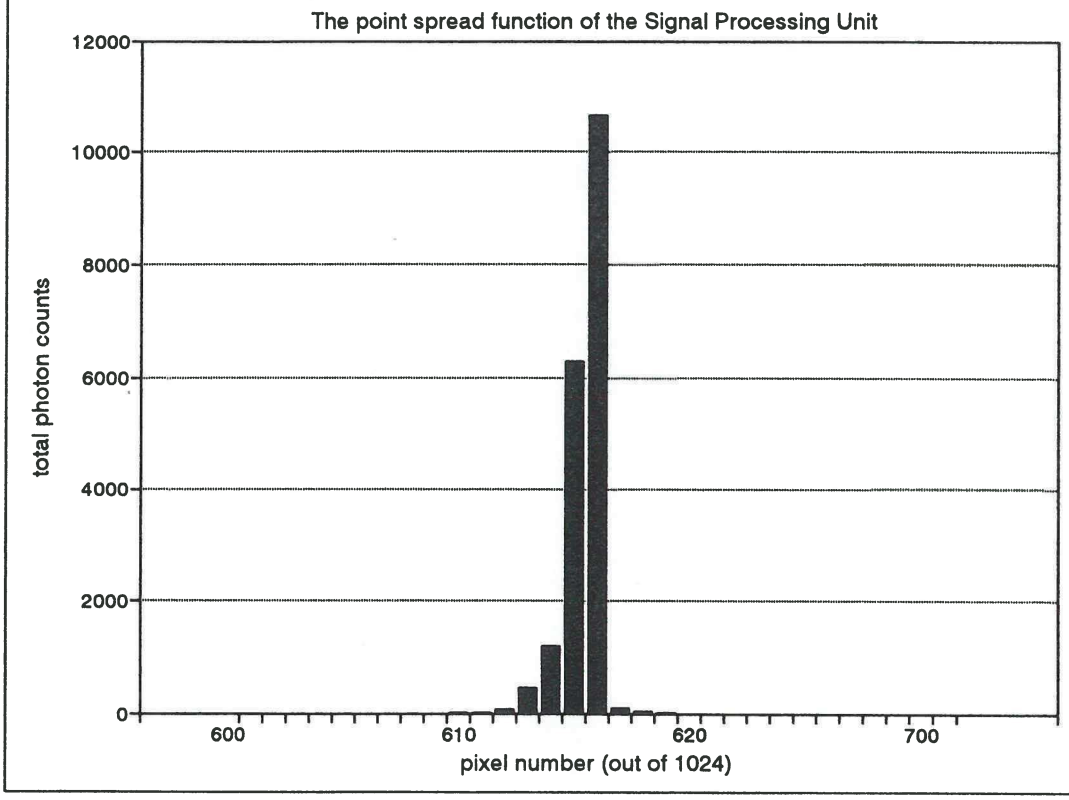


Figure 7.6

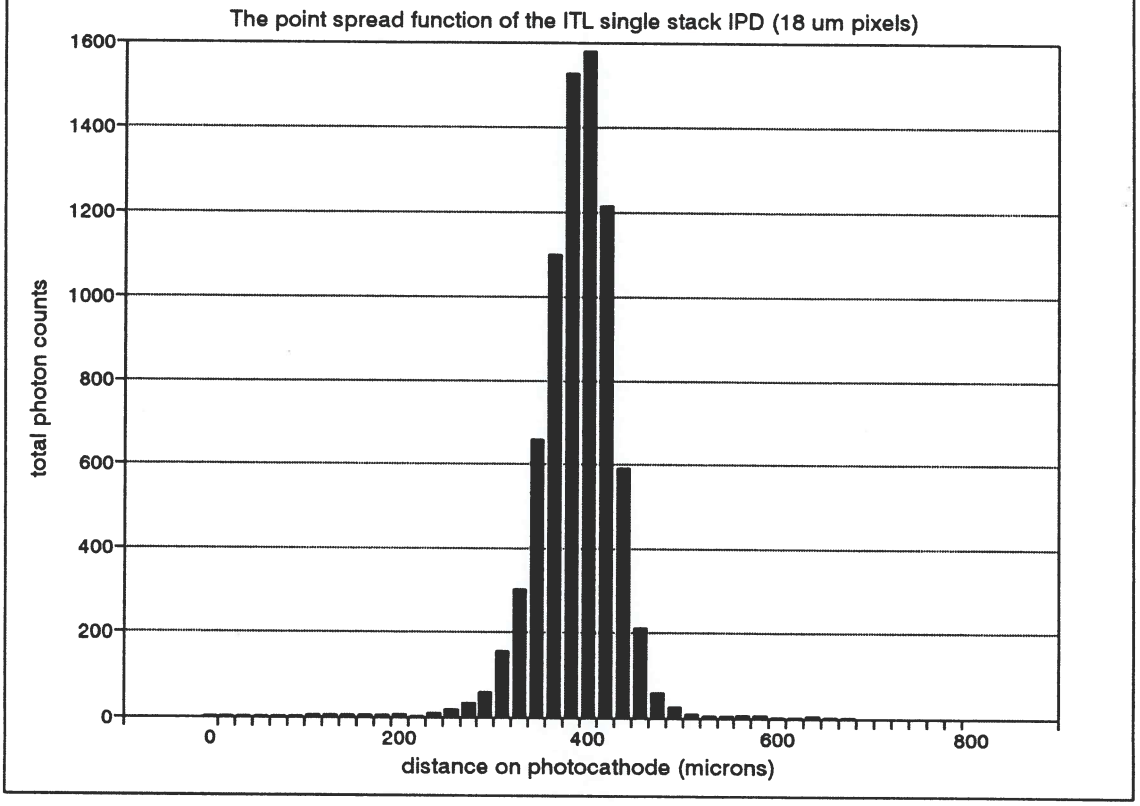
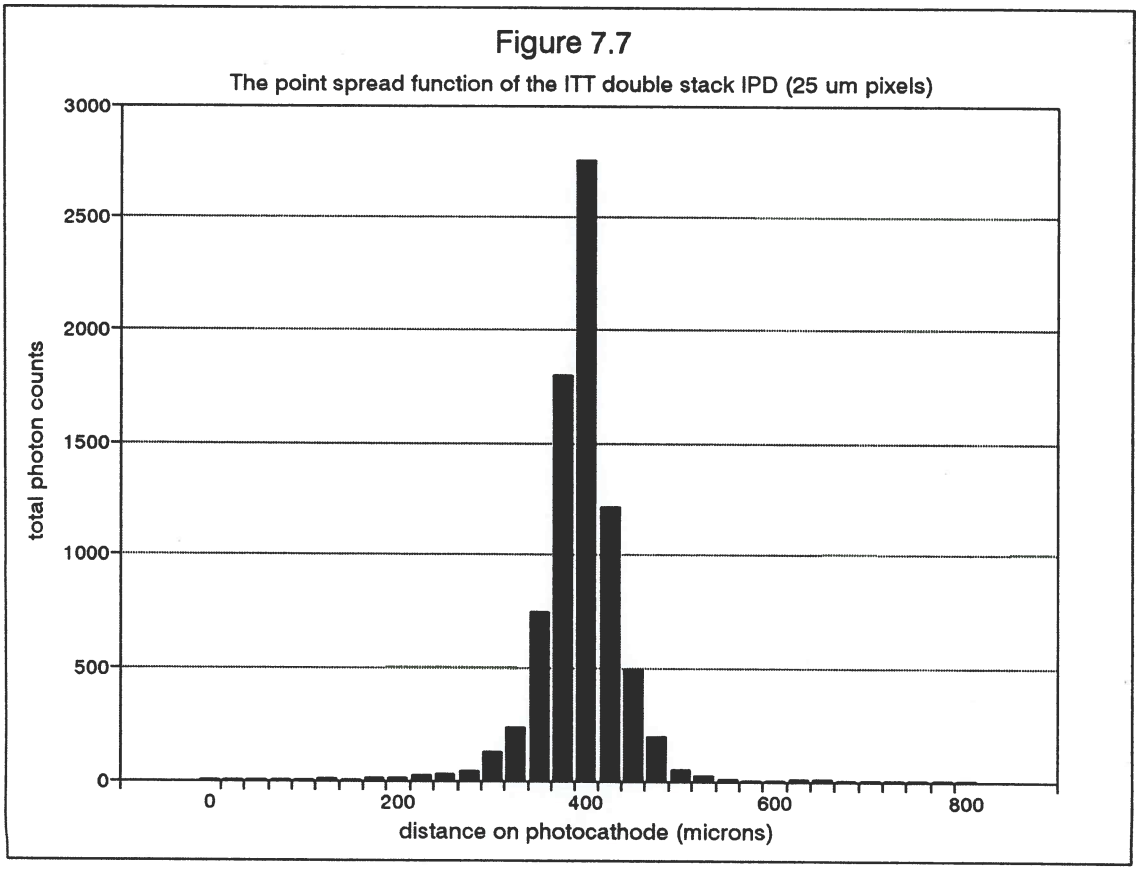
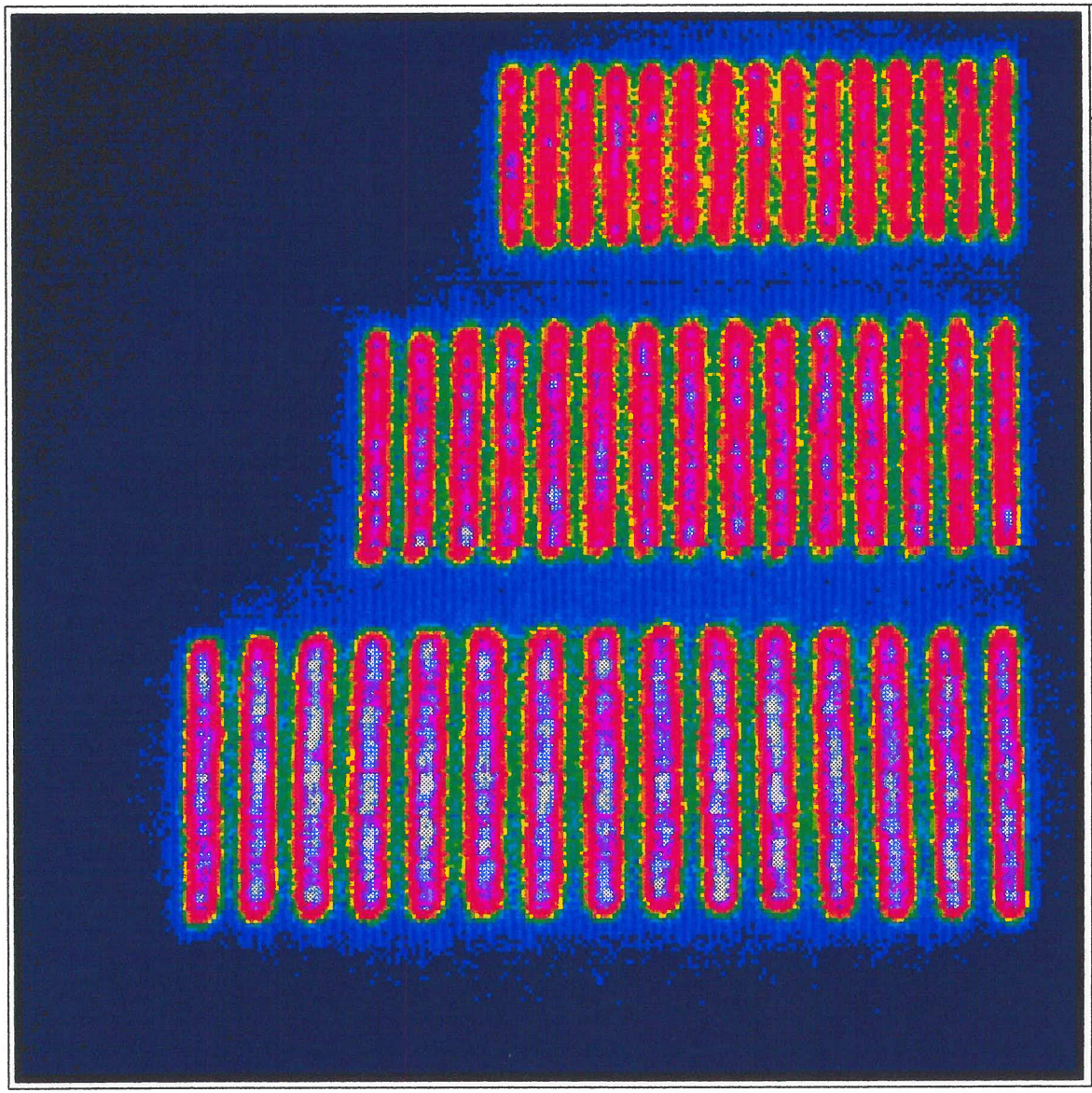


Figure 7.7





ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON

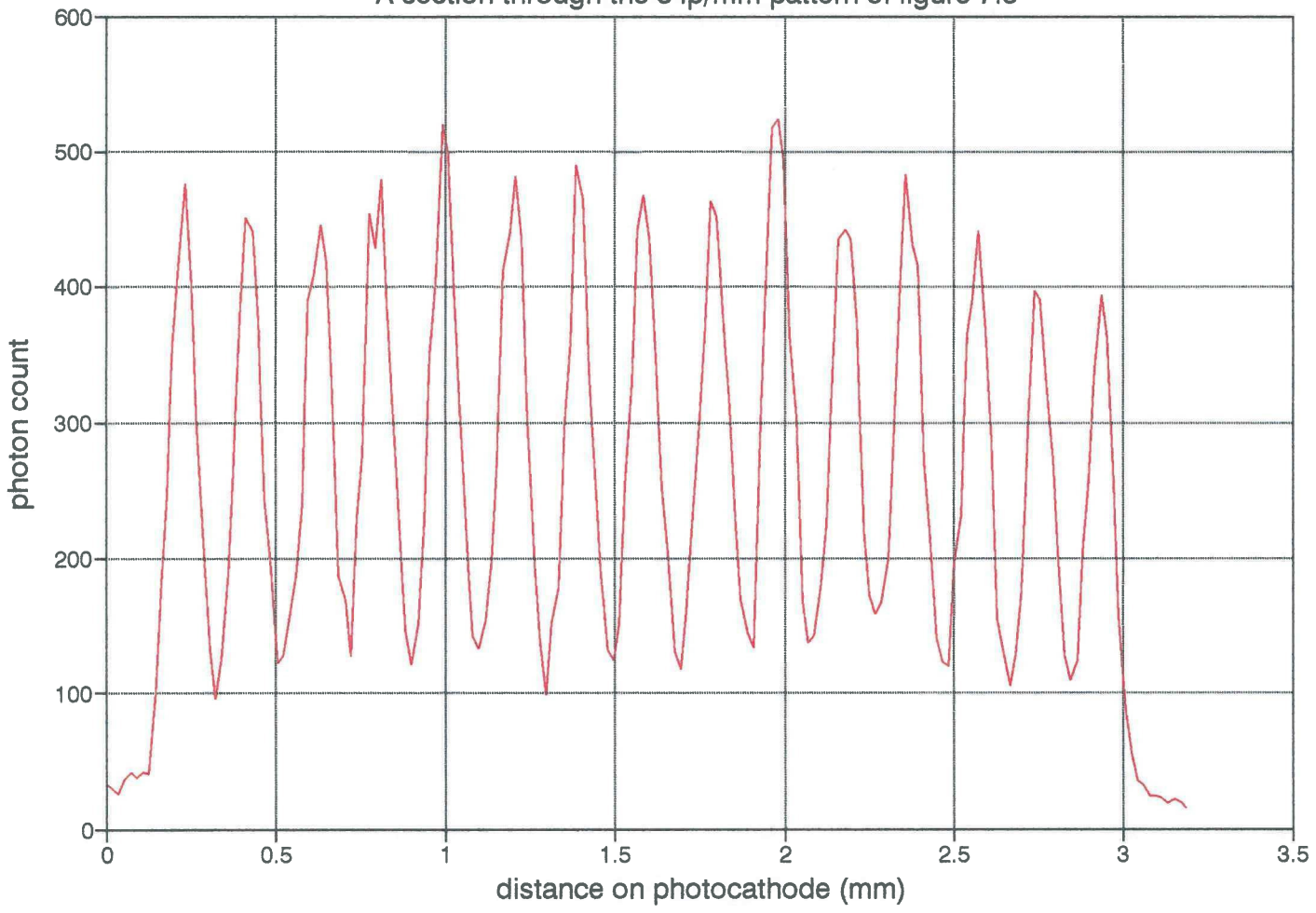
Section at 10 bit resolution
File: TF01702.E31

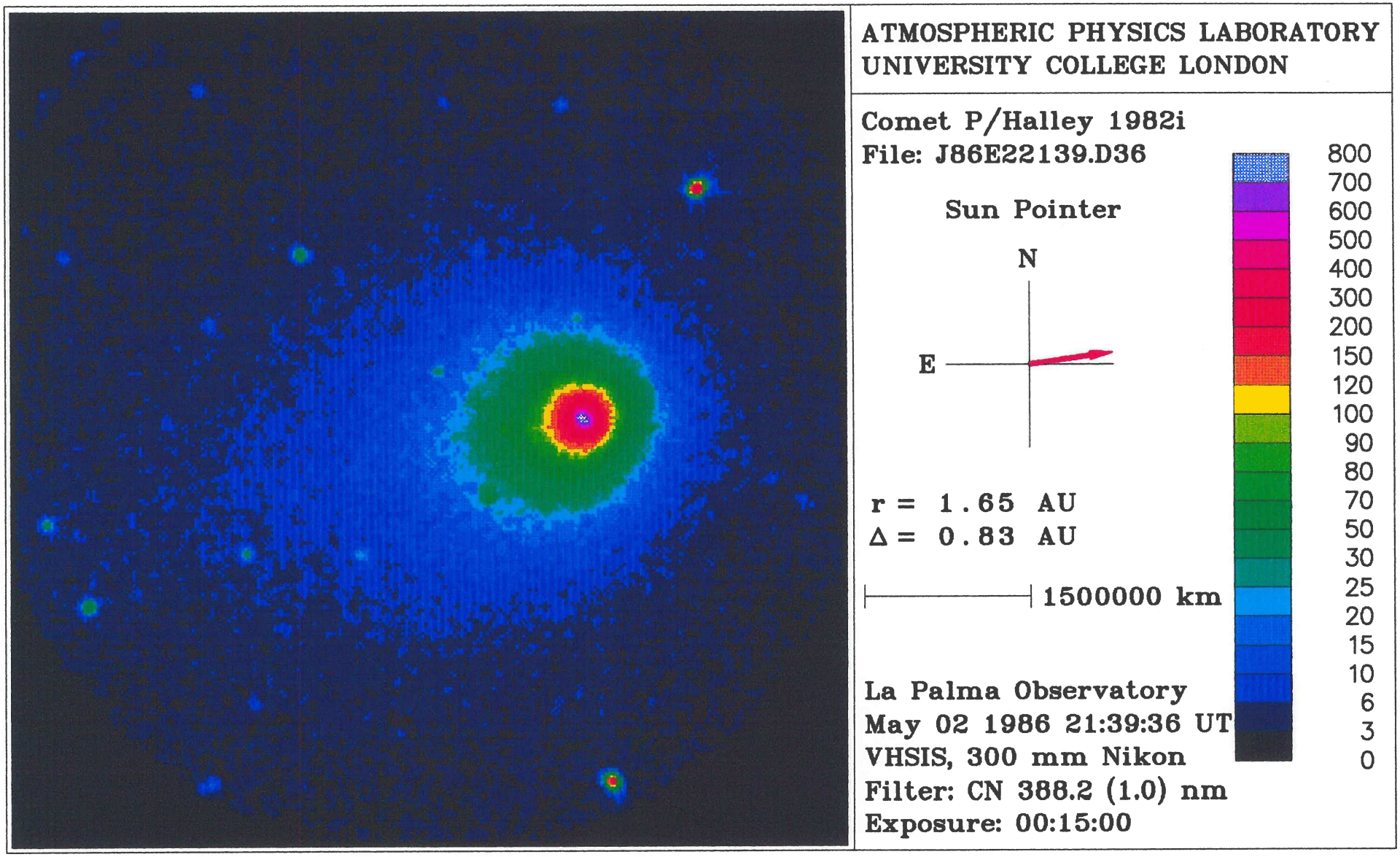
ITL IPD

Figure 7.8

Figure 7.9

A section through the 5 lp/mm pattern of figure 7.8




ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON
Comet P/Halley 1982i
File: J86E22139.D36
Sun Pointer
N
E
r = 1.65 AU
Δ = 0.83 AU
1500000 km
La Palma Observatory
May 02 1986 21:39:36 UT
VHSIS, 300 mm Nikon
Filter: CN 388.2 (1.0) nm
Exposure: 00:15:00
800
700
600
500
400
300
200
150
120
100
90
80
70
50
30
25
20
15
10
6
3
0

CHAPTER 8 - DISCRETE ELEMENT POSITION SENSING ANODES

8.1 Introduction

The majority of the Imaging Photon Detector systems built at UCL have used resistive anodes. However, detectors manufactured for specific applications have used anodes which consist of several discrete elements arranged in a specific pattern.

These elements are fabricated by etching a conductive layer which has been deposited on a ceramic substrate. They do not depend on resistivity for their operation, and the resistance of each element should be as low as possible in order to maximise the rise-time of the anode signal. Electronically, each one is operated as an independent anode. The charge pulse from the microchannel plates illuminates one or more of the elements, from which its position may be deduced.

This is a very simple concept, but one which may be quite extravagant in electronic requirements as the number of signal processing channels is equal to the number of elements in the detector array. This is typically twenty for a practical detector. Each channel may be thought of as a separate photomultiplier. The charge from each electrode is measured using analogue or pulse counting techniques. It may then be displayed as a histogram of intensity versus position in the array.

Detector systems have been developed at UCL using two types of discrete anode: the linear anode and the multi-ring anode. This chapter describes these detectors, together with an outline of their applications.

The linear anode is a simple one dimensional array of elements. The multi-ring anode is a particularly useful configuration, consisting of concentric rings of equal area. This pattern is designed to map directly onto an image of the interference pattern produced by a Fabry-Perot etalon.

The detectors are similar in appearance to the resistive anode devices. The microchannel plate and photocathode requirements are generally the same and there are minor differences in the rear of each device to accommodate the different anodes, which require a large number of electrical connections to be taken through a glass-to-metal seal.

8.2 The linear anode

A linear anode detector was built especially for an auroral imager which was flown on the Aureld-VIP rocket payload from Kiruna, Sweden. The flight was arranged to coincide with overpasses of the Swedish Viking satellite. The instrument is called the "Three Dimensional Auroral Imager" and a full description is presented in *Rees & Steen (1987)*, together with a preliminary analysis of the data obtained. The anode is depicted in figure 8.1. It consists of twenty bars in a linear array. The device may be considered as twenty separate photomultipliers sharing the same microchannel plates and photocathode.

The detector was incorporated into an imaging system which included a narrow-band interference filter (1nm FWHM at 427.8nm). A 100° field of view was imaged onto the photocathode of the detector. The array was aligned with the axis of the rocket so that each anode element gathered light from a different altitude. Each element received light from a 5° field of view in elevation by a 10° field in azimuth, so the complete coverage in elevation was 100°. The spin of the rocket provided a scanning mechanism so that a complete 360° azimuthal scan was obtained with every revolution.

The detector was operated in analogue mode with the anodes feeding a bank of twenty very sensitive logarithmic amplifiers. The circuit diagram of the amplifiers is depicted in figure 8.2. This approach enabled very high time resolution to be obtained over the large range of intensities anticipated for the rocket flight.

8.3 The multi-ring anode

This type of detector was used in the Dynamics Explorer Fabry-Perot Interferometer, described by *Killeen et al (1983)*. It was designed specifically for use with Fabry-Perot interferometers.

Multi-ring anode detectors have been developed at UCL for two specific applications. These are the Doppler Wind Lidar, which requires very high time resolution, and the triple etalon interferometer, which requires very high count rates. The Doppler Wind Lidar is discussed below and the triple etalon interferometer is described in detail in chapter twelve.

The detector elements are arranged as concentric rings of equal area. The Fabry-Perot instruments produce an interference pattern which consists of similar concentric fringes. It will be seen in chapter nine that increments in the square of the radius of the fringes are inversely proportional to wavelength. The anode segments sample equal wavelength intervals on annuli whose width decreases as the inverse square of their diameter. The anode rings thus provide a histogram output which is linear in wavelength space. This offers a considerable advantage over the resistive anode, which requires a large amount of image processing in order to obtain the same histogram. Two types of this anode are shown, in figures 8.3 and 8.4.

The original anode used for the Dynamics Explorer detector had twelve elements. A similar one, using twenty rings, was manufactured by ITL for the initial Doppler Lidar experiments. This is the one shown in figure 8.3. The connection to the rings is achieved by forsaking a "pie slice" of the useful anode surface through which the connections are routed to the perimeter. This segment must be masked off from the incident light. An improved device was manufactured by ITT using the anode illustrated in figure 8.4, which had lead-through connections between the 32 rings and contacts on the rear of the substrate. This detector was used for experiments with the triple etalon interferometer.

In addition to its minimal data processing requirements, the multi-ring anode has two other important advantages over the resistive anode: increased maximum count rate and very high time resolution. The device is operated in the pulse counting mode and, although it is still a sequential device in that simultaneous photon events cannot be resolved, each element of the anode is an independent detector in this respect. Also, the maximum count rate is ultimately limited by the RC time constant of the anode, which is very much shorter than that of the resistive anode.

One problem with the multi-ring anode detector is that it has to be aligned very precisely with the interference pattern. This has been achieved by mounting it on an X-Y table which is adjusted manually whilst continuously monitoring the transmission profile obtained with a calibration source. The adjustment for minimum line width is, in fact, quite easy to achieve. In future instruments this alignment could be done automatically with a computer driven X-Y table.

8.3.1 The electronics for the multi-ring anode detector

The electronic requirements for the two applications are broadly similar. Each channel is equipped with a charge and pulse shaping amplifier. These are similar to those used for the resistive anode IPD, except that the shaping time constants are much shorter. The output of each amplifier is fed into a high speed comparator, producing a digital pulse for each photon event. These pulses are then fed into a multi-channel high speed counter. The counters are controlled and read by a PC computer using the Tecmar parallel interface. The circuit diagram of the amplifier is shown in figure 8.5 and that of the multi-channel counter in figure 8.6.

8.3.2 The Doppler Wind Lidar

The development of the lidar system is presented in *Rees, McWhirter and Wade (1987)*. The instrument was installed on the sodium lidar facility operated by the University of Bonn, Germany. It is situated

at Andoya in Northern Norway, where an advanced lidar facility is currently under construction. This is the ALOMAR project (Arctic Lidar Observatory for Middle Atmosphere Research). It will include a Rayleigh/Mie lidar system, built jointly by teams from Germany, France and the UK (UCL). The UCL instrument will include two multi-ring anode detectors.

The concept of lidar is broadly similar to radar, except that laser light is used instead of microwave radiation. A powerful pulsed laser is fired into the atmosphere and the back-scattered reflections of each pulse are collected by a receiving telescope. The light is passed through a Fabry-Perot interferometer and imaged onto the detector. In order to achieve good height resolution a very high speed pulse counting system is required as the round trip time of flight is 6.7μs per kilometre. The counters need to be read at intervals corresponding to the required height resolution. This is achieved by using either a bank of counters for each height resolution element (termed a range bin), or a single bank of counters with a high speed memory unit to store the counts for each bin. The memory unit may, in fact, be a fast computer.

Rayleigh scattering, at altitudes of up to 85 km, directs a very small fraction of the laser light back to the detector. Only a very few photons may be detected in each range bin for each laser pulse. However, when the photons are accumulated over many laser pulses a complete histogram is built up for each range bin. The back-scatter from low altitudes is many times more intense, and it is necessary to protect the detector with a mechanical shutter. This is synchronised to the laser and opens at a set time after each firing. In practice, this is simply a rotating disc with a hole in it.

A prototype system was tested at Andoya in 1988 using a twenty channel ITL detector. The instrument was fitted to the sodium lidar receiving telescope. It was equipped with three high speed counters to give three range bins. The counters were individually enabled for specific time windows using a three channel clock synchronised to the laser

pulse. The system was aligned and tested using a Helium-Neon laser directed through the complete optical system, whilst obtaining the timing from the firing of the high power sodium laser. The histogram shown in figure 8.7 represents the laser profile obtained by integrating over many laser pulses in a time interval corresponding to a 2km range bin. It may be compared with the etalon profile in radius squared space shown in figure 9.5.

The whole system worked very well, with none of the anticipated problems which can occur when operating sensitive electronics in close proximity to a high power pulsed laser. All that was needed to obtain real data was clear weather. Much to the dismay of all involved, as soon as the sky cleared, the sodium laser sustained a severe malfunction and it was not possible to effect a repair before the end of the visit.

8.3.3 Use with the triple etalon interferometer

The 32 ring anode detector, made by ITT, had a good S25 (extended red) photocathode and so it was necessary to operate it in a Peltier cooled housing. The very high time resolution used for lidar was not required for viewing airglow with the triple etalon instrument. It was, therefore, possible to use amplifiers with a longer pulse shaping time constant, which had the advantage of lower noise and greater stability.

Figure 8.8 shows the detector incorporated into the cooled housing with its amplifiers and high voltage power unit. The detector performed successfully, but it has not been operated for an extended period. This was because of other problems encountered with the interferometer, which are discussed in chapter twelve. Also presented is a sample of the data obtained together with other results from the triple etalon interferometer.

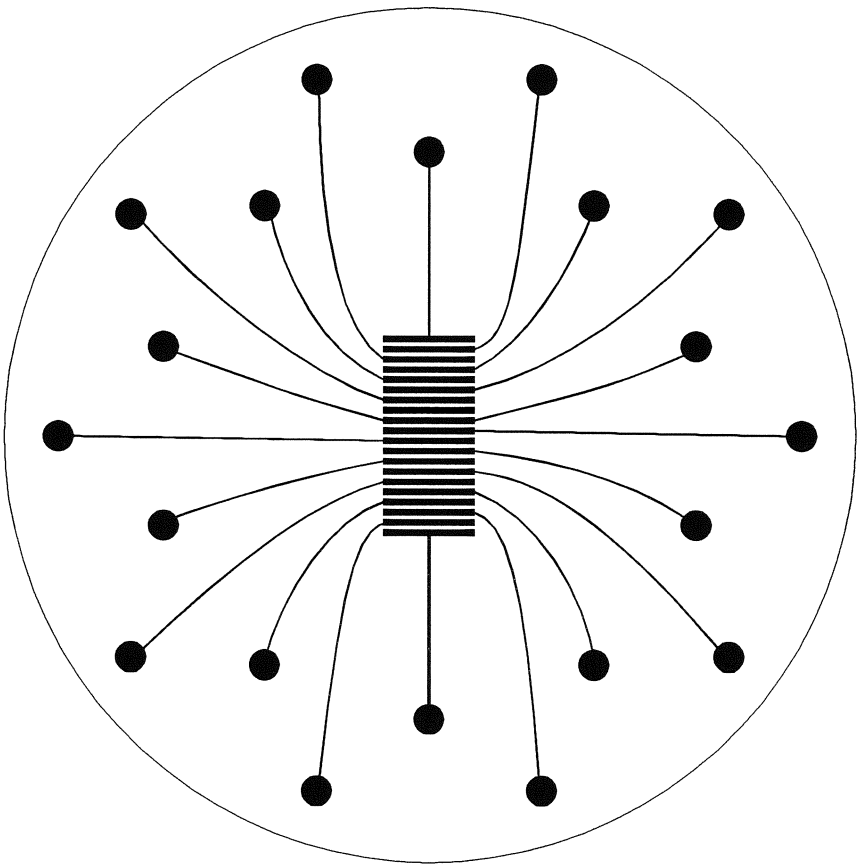


Figure 8.1

The linear anode

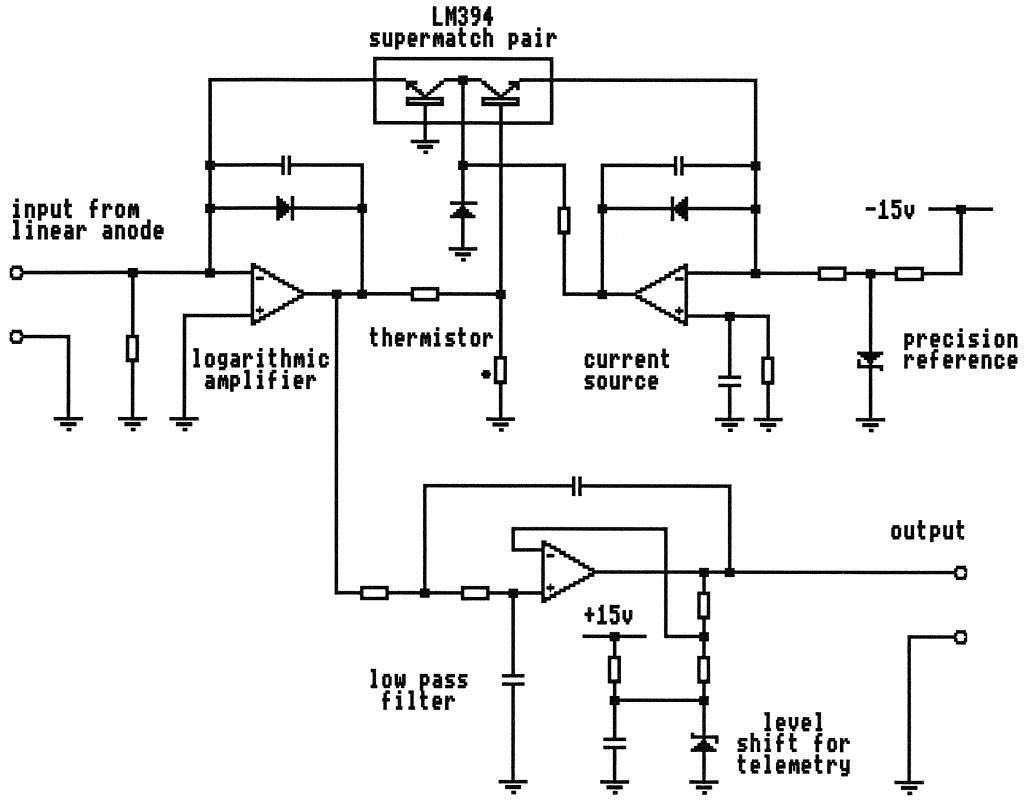


FIGURE 8.2

THE LOGARITHMIC AMPLIFIER
FOR THE ROCKET-BORNE
LINEAR ANODE IPD
(20 channels total)

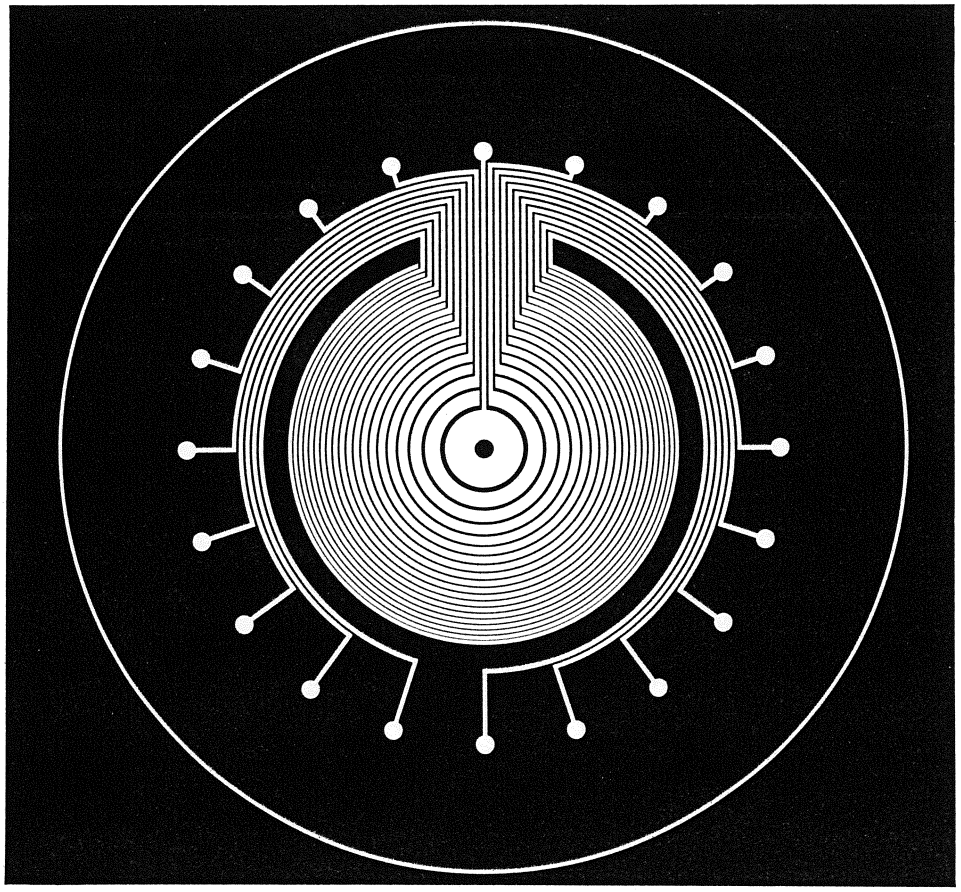


Figure 8.3 The ITL multi-ring anode

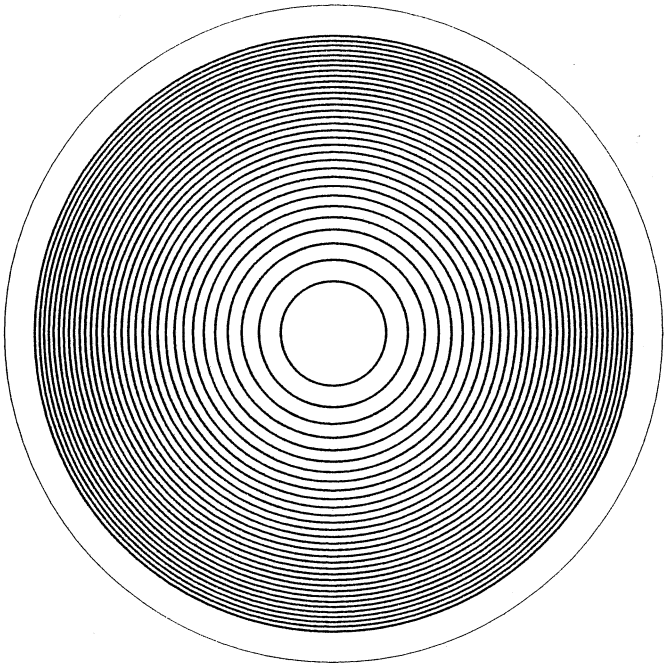


Figure 8.4 The ITT multi-ring anode

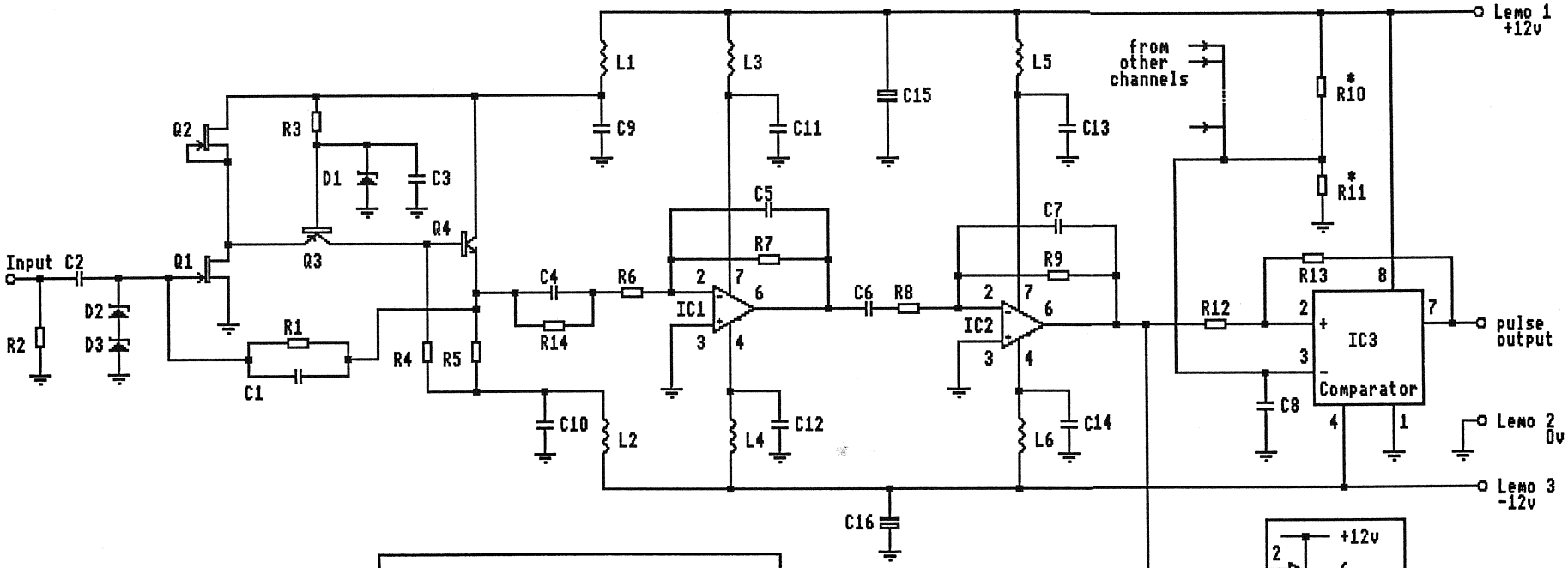


FIGURE 8.5

THE HIGH SPEED PULSE AMPLIFIER FOR THE MULTI-RING ANODE IPD (32 channels total)

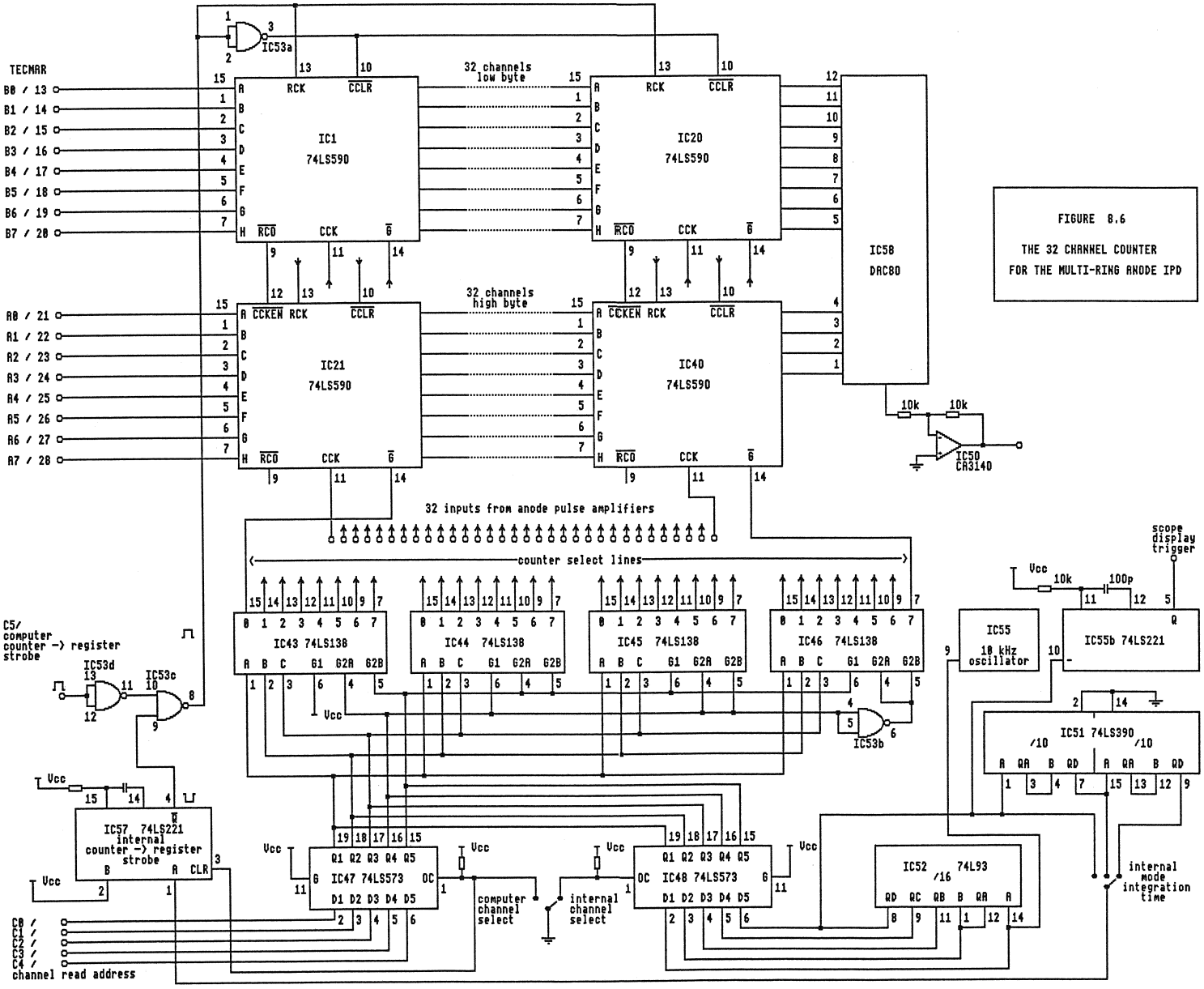


FIGURE 8.6

THE 32 CHANNEL COUNTER
FOR THE MULTI-RING ANODE IPD

Figure 8.7

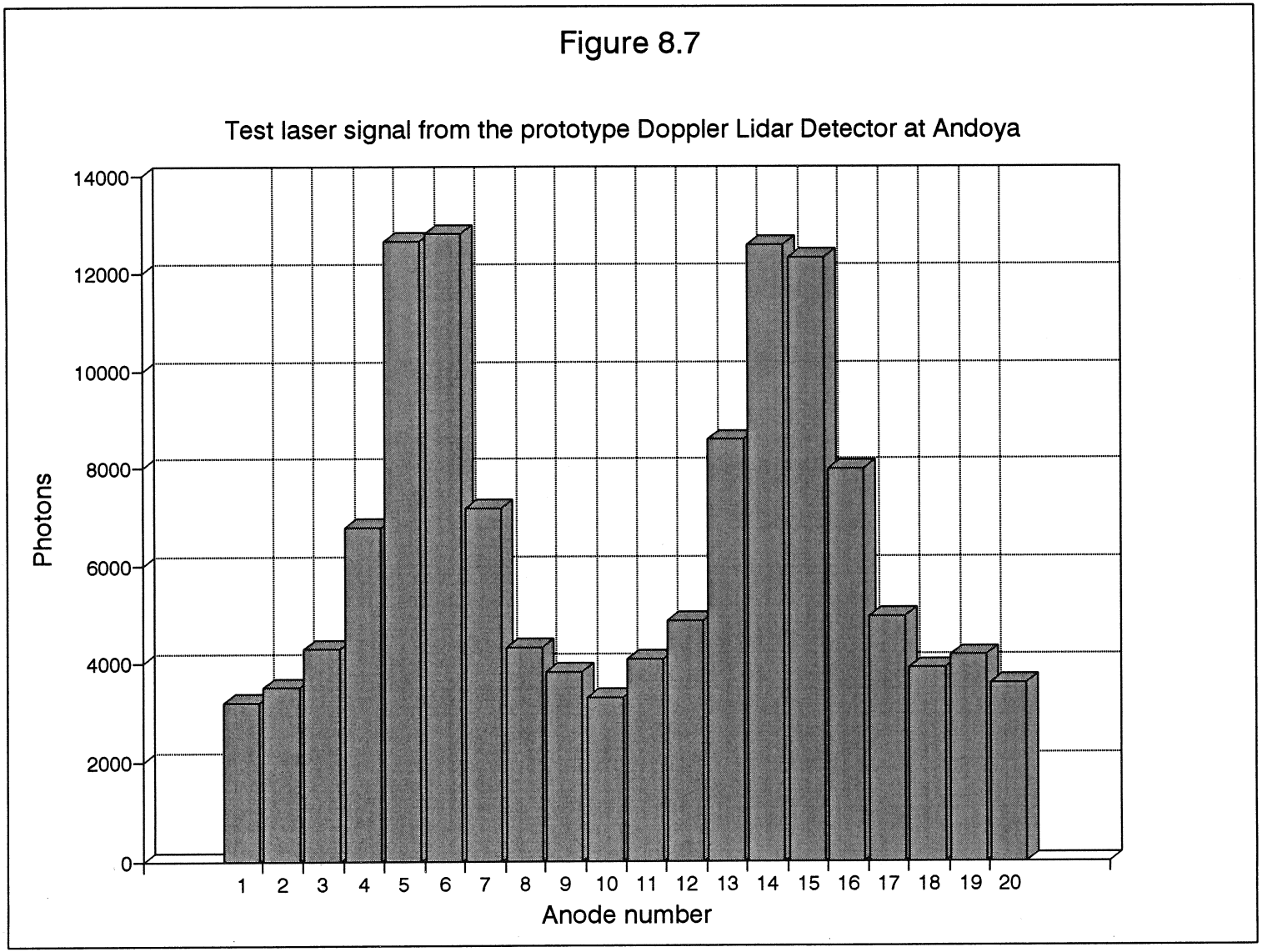

Test laser signal from the prototype Doppler Lidar Detector at Andoya
Photons
14000
12000
10000
8000
6000
4000
2000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Anode number

Figure 8.8

The multi-ring anode IPD in its cooled housing

CHAPTER 9 - THE FABRY-PEROT ETALON

9.1 Introduction

The Fabry-Perot etalon, with its high resolution, is well suited to the determination of Doppler shifts in atmospheric emission lines, and together with the Imaging Photon Detector forms an elegant and robust interferometer. It is capable of accepting light over a relatively large solid angle, Ω. The programme of development at UCL has resulted in the production of high quality large diameter etalons, enabling instruments with a very high sensitivity in terms of $A\Omega$ to be realised.

A full account of the Fabry-Perot etalon may be found in *Hernandez (1986)* and *Vaughan (1989)*. This chapter contains a brief description of the concept of the etalon and resumé of the equations which describe its operation, expressed in a form appropriate to the design of practical interferometers.

9.2 Computer models of the etalon systems

The theoretical responses of the etalon systems described in this thesis have been obtained using a suite of computer models written by the author using the PC-based spreadsheet programme Quattro Pro, by Borland. This has proved to be an extremely versatile tool for assessing the potential performance of etalon systems and assisting with instrument design. It provides a highly interactive environment in which parameters such as the separation and reflectivity of the etalon plates and the details of the imaging optics can be varied at will. The effects of such variations are calculated in seconds to a high degree of accuracy, and can be read directly from the spreadsheet data table or, with a single key press, displayed graphically. The computer models are described in more detail in chapter thirteen.

9.3 The Fabry-Perot etalon

The etalon consists of two flat, partially reflecting mirrors which are perfectly parallel and separated by a fixed distance. The space between the mirrors is termed the cavity, or simply the gap. Light entering the cavity at angles close to the normal of the mirrors will undergo multiple reflections before finally emerging from one side or the other. The resultant emergent beams, both reflected and transmitted, are subject to an intensity modulation. This is caused by mutual interference of the many reflections between the mirrors. The interference is a result of the optical path difference (OPD) between successive reflections. The OPD is a function of the length of the etalon cavity and the precise angle of entry of the beam.

The interference pattern has an inherent circular symmetry which may be observed by viewing an extended monochromatic source through the etalon. A set of bright concentric rings, or fringes, on a dark background will be visible (figure 9.1). The fringes are formed by constructive interference between light rays of equal inclination. It will be shown that the angular diameter of a given fringe is a function of the wavelength of the incident light and the distance between the two surfaces forming the cavity. The Fabry-Perot etalon may, therefore, be thought of as a filter of both wavelength and angle.

9.4 The history of the etalon

Newton observed interference rings between a curved surface and a plane surface - the well known "Newton's rings". Fizeau observed this phenomenon further and the Fizeau interferometer, consisting of two flat plates separated by a wedge-shaped spacer, has been named after him. The fringes obtained with this device are caused by interference resulting from the spatial relationship of the incident light waves, whereas the Fabry-Perot fringes result from the angular distribution of the incident light.

In the early nineteenth century George Biddel Airy, one of the main proponents of the wave theory of light, described mathematically the interference between waves of equal inclination, such as those reflected between the surfaces of a parallel plate. He showed that if light is incident on to a parallel plate of thickness d at an angle of θ, then the transmitted intensity, I, is given by:

$$I = \frac{1}{1 + \frac{4R}{(1-R)^2} \sin^2 \left[\frac{2\pi\mu d}{\lambda_0} \cos\theta \right]} \qquad [9.1]$$

where R is the surface reflectivity, μ the refractive index and λ_0 the wavelength in vacuo. This expression is known as the Airy Function.

It was later realised that when the reflectivity of the surfaces is increased to near unity, the contrast of the interference fringes is dramatically enhanced. Work by Haidinger in the mid-nineteenth century, and subsequently by Lummer and Bouloch, led to a detailed investigation of this phenomenon by Charles Fabry and his colleagues Perot and Buisson. The device as described by Fabry and Perot in the closing years of the nineteenth century eventually became known as the Fabry-Perot etalon. An extensive list of references for the above historical work may be found in the book by *Hernandez (1986)*.

9.5 Conditions for constructive interference

Monochromatic rays of equal inclination entering the etalon cavity interfere either constructively or destructively, depending on the path difference between successive reflections at the inclination in question. The interference pattern appears as a concentric circular fringe pattern centred around the optical axis of the etalon. Because any given point on a fringe is formed by rays of light at a unique angle, the interference pattern is perceived at infinity when viewed through the etalon.

On referring to figure 9.2, which illustrates the multiple reflection undergone by a light ray entering the etalon, it is seen that the path difference, δ, between adjacent transmitted rays is given by:

$$\delta = 2l - p \qquad [9.2]$$

It is easily shown by simple geometry that:

$$\delta = 2d \cos \theta \qquad [9.3]$$

where θ is the angle between the incident ray and the normal to the reflecting surfaces, which are separated by the distance d.

The optical path difference expressed in wavelengths is termed the order of interference, n, and is given by:

$$n = \frac{2d \cos \theta}{\lambda} \qquad [9.4]$$

For constructive interference the path difference must be a whole number of wavelengths, that is to say, n must have an integral value.

9.6 Examination of the Airy function

The phase difference between successive reflections is given by:

$$\Delta\phi = \frac{2\pi \delta}{\lambda} = \frac{4\pi\mu d \cos \theta}{\lambda_0} \qquad [9.5]$$

Thus the Airy function may be expressed more simply as:

$$I = \frac{1}{1 + F \sin^2 \frac{\Delta\phi}{2}} \qquad [9.6]$$

F is termed the coefficient of reflective finesse, given by:

$$F = \frac{4R}{(1-R)^2} \qquad [9.7]$$

It is sometimes referred to as the F parameter (*Hernandez, 1986*) in order to avoid confusion with the quantity N_R. The latter is known as the ideal reflectivity finesse, or simply the reflective finesse, and will be discussed in a later paragraph.

Figure 9.3 is a plot of equation [9.1] and shows the variation of transmitted intensity with phase difference for beams of equal inclination. The expression is plotted for different values of R and is seen to repeat itself at intervals of 2π. It is important to note that the phase difference may be effected by a variation in d, μ, λ or cos θ. The transmitted intensity may be plotted as a function of any one of these variables to give a similar profile.

9.7 The free spectral range

The interdependence of λ and cos θ, as shown in equation [9.1], means that the angular diameter of a given fringe is dependent upon wavelength. The etalon can, therefore, be used to measure small changes in wavelength by measuring the fringe diameter. However, the transmission profile repeats itself at regular intervals. If λ is varied, the fringe pattern eventually passes through a complete cycle and the interference pattern is indistinguishable from that at the starting point.

This repetition interval is known as the free spectral range, or FSR of the etalon, and is a very important parameter. It represents the range of wavelengths which can be displayed without overlap. It is expressed in terms of wavelength shift, $\Delta\lambda_{FSR}$. The FSR can be calculated as a function of the etalon gap by the following reasoning.

If maximum transmission occurs at $2d \cos \theta = n\lambda$, then there will also be a maximum for a wavelength $(\lambda + \Delta\lambda)$ at $2d \cos \theta = (n-1)(\lambda + \Delta\lambda)$.

That is to say:

$$\frac{2d \cos \theta}{\lambda} = \frac{2d \cos \theta}{(\lambda + \Delta\lambda)} + 1$$

which rearranges to give:

$$\frac{\Delta\lambda}{\lambda(\lambda + \Delta\lambda)} = \frac{1}{2d \cos \theta}$$

As $\Delta\lambda << \lambda$ and for near axis rays $\cos \theta = 1$, the equation reduces to:

$$\Delta\lambda_{FSR} = \frac{\lambda^2}{2d} \qquad [9.8]$$

9.8 The etalon as a wavelength measuring device

The etalon may be used to measure a shift in wavelength by exploiting the interdependence of λ, d, μ and $\cos \theta$ shown in equation [9.1]. There are two basic approaches: scanning and imaging.

9.8.1 Scanning methods

Scanning involves measuring the intensity of the light at one particular value of $\cos \theta$, usually on the optical axis of the instrument, ($\cos \theta = 1$), whilst varying either d or μ to step through the transmission profile.

In order to vary the gap, d, adjustable spacers must be used. This presents great problems for maintaining the stability and parallelism of the gap. Such a method is only practical if an accurate means of measuring the gap is used in conjunction with a servo mechanism. This technique is utilised for the capacitance stabilised etalon, described in chapter eleven.

In order to vary the refractive index, μ, it is necessary to vary the pressure of the gas filling the etalon gap. This is known as pressure scanning, and is used extensively. However, it is difficult to scan

the etalon with any speed using this approach. Variation of the pressure is also used with the imaging method in order to adjust the transmission peak to a specific angular diameter.

9.8.2 The imaging method

This exploits the cos θ dependence by measuring the change in angular diameter of the fringes. An easy way of determining the angle, θ, subtended by a given fringe is to form an image of the rings using a converging lens of known focal length. Simple lens theory shows that a circular object at infinity which subtends a half angle of θ will form an image with a radius, r, given by:

$$\tan\theta = \frac{r}{f} \qquad [9.9]$$

where f is the focal length of the lens.

It then remains only to re-write equation [9.4] in terms of tan θ. This is done by using the approximation:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\ldots \qquad [9.10]$$

The angle θ is normally small, so the higher orders can be neglected and θ may be assumed to be approximately equal to tan θ.

Equation [9.4] can, therefore, be expressed as:

$$n\lambda = 2d\left[1 - \frac{r^2}{2f^2}\right] \qquad [9.11]$$

This equation demonstrates that there is a linear relationship between $n\lambda$ and the square of the radius of the observed ring. The adjacent maxima, formed by consecutive integral values of n, form a series of concentric fringes which have equal separation in r^2 space.

A computer model has been written which plots the radial fringe size as a function of etalon gap and the focal length of the imaging lens. Figure 9.4 is a plot of intensity versus image radius and figure 9.5 represents the same data displayed in radius squared space, which clearly shows the linear relationship. The computer simulation takes input data for wavelength, etalon gap, reflectivity, focal length of lens, maximum image radius and radius at the transmission peak. A modified etalon gap is calculated by the programme in order to place the peak of the Airy function at the desired radius, and the function is then plotted. It is, therefore, a very useful tool for interferometer design, where careful matching of etalon gap, optics and detector size is required.

Figure 9.4 may also be compared with the photograph of the fringes shown in figure 9.1. The free spectral range may be determined in terms of radius squared space provided that at least two fringes are displayed in the image. Calculation of λ_{FSR} from equation [9.8] will, therefore, permit precise calibration of the instrument.

The imaging method has two very considerable advantages. Firstly, the incident light can be gathered over a much wider solid angle, Ω. The sensitivity of the instrument in terms of $A\,\Omega$ is, therefore, greatly increased. Secondly, the entire transmission profile is determined in a single measurement, whereas the scanning methods require a measurement to be taken at each step. This means that the imaging method is particularly suited to applications where time resolution is important. However, it does require an expensive two dimensional imaging detector, such as the Imaging Photon Detector or a Charge Coupled Device. Alternatively, if maximum sensitivity is not required, a one dimensional multi-element linear detector, such as a photodiode array, may be positioned to bisect the fringe pattern. The scanning methods have the advantage that they require a much simpler single point detector, such as a photomultiplier, although this is at the expense of providing a scanning mechanism for the etalon.

9.9 The reflective finesse of the etalon

The reflective finesse, N_R, is a measure of the width of the etalon transmission profile. It is the reciprocal of the full width at half maximum of the profile, normalised to the free spectral range. It is an important parameter because it limits the resolving power of the etalon.

Inspection of equation [9.6] shows that the transmission half maxima occur at:

$$F \sin^2 \frac{\Delta\phi}{2} = 1$$

F is normally sufficiently large for the sine to be approximated to the angle. Half maximum points occur at +/- $\Delta\phi$, giving a full width of $2\Delta\phi$. Substituting equation [9.7] for F and recalling that a free spectral range corresponds to a phase change of 2π gives:

$$N_r = \frac{2\pi}{2\Delta\phi} = \pi \sqrt{R}/(1-R) \qquad [9.12]$$

This is an extremely important parameter as it defines the theoretical resolving power of the etalon. However, in practice, the observed finesse is somewhat less.

9.10 The practical finesse of the etalon

The reflective finesse is inevitably degraded because of surface defects and departures from parallelism. The etalon may be considered to be composed of many elemental etalons, each having a particular optical path difference determined by local irregularities. The distribution of the OPDs for these elemental etalons depends on the nature of the defect. For example, if the defects are purely random, the distribution will be Gaussian. The transmission profile of the etalon will then be the convolution of the reflective finesse and the gap distribution.

It is more likely that the major gap variations will not be random, but will be due to plate distortions. One example of this is dishing, which gives a rectangular (top-hat) distribution of OPD. If the etalon is not perfectly parallel, the distribution has a semi-circular profile with a maximum at the mean value of the gap. *Vaughan (1989)* has detailed the form of the distribution for a number of defects. In practice, the nature of any defects can often be determined by visual inspection, whereas the quantitative effect on the transmission profile is best determined by photometric methods.

9.11 Transmission losses

It has been assumed so far that all light undergoing reflection at the etalon surface is either transmitted or reflected, that is to say,

$$T + R = 1 \qquad [9.13]$$

where T and R are the transmittance and reflectance coefficients.

However, a small amount is inevitably lost at each reflection by absorption and scattering. Taking this into account the equation can be rewritten as:

$$T + R + A = 1 \qquad [9.14]$$

Intuitively, one might deduce that a high value of reflectivity will cause the transmitted fringes to be reduced in intensity because the light beams undergo more reflections inside the cavity before they emerge and therefore suffer more attenuation. This is indeed the case, as is shown in the full mathematical analysis given by *Vaughan (1989)*.

The Airy function, equation [9.1], is preceded by a factor $[T/(1-R)]^2$ which, according to equation [9.13], is unity. However, as *Vaughan* points out, if we introduce the attenuation, A, and use equation [9.14] to substitute for T, we have:

$$\left[\frac{T}{(1-R)}\right]^2 = \left[\frac{(1-R-A)}{(1-R)}\right]^2 \qquad [9.15]$$

The transmitted beam is, therefore, decreased by a factor of:

$$\left[1 - \frac{A}{(1-R)} \right]^2 \qquad [9.16]$$

Inspection of this expression shows that for reflectivities approaching unity the transmitted beam can be severely attenuated. Fortunately, modern dielectric coatings, although extremely expensive, are of a very high quality. The absorption is very small and can usually be neglected.

9.12 Practical realisation of the etalon

Etalons are made from cylindrical plates of fused silica which have their surfaces figured and polished flat to a very high degree of accuracy, typically $\lambda/200$. The silica must be of extremely high quality, selected for low defects, uniform refractive index and low thermal expansion. The plates are vacuum coated with a dielectric layer of the required reflectivity.

The reflecting surfaces may be deposited on either side of a single plate. This is termed a solid etalon. However, this approach is normally reserved for etalons with small gaps. The optical path length of the cavity can be made more stable if an air gap is used or, for preference, an evacuated gap.

In the context of etalon type, the general expression "air gap" is taken to imply a non-solid etalon, whether evacuated or filled with either air or an inert gas. When making such an etalon the coatings are deposited on the inner sides of two discs of silica, which are held apart with accurately made spacers. The outer sides are given an anti-reflectance coating to maximise transmission. The plates are each slightly wedge shaped to ensure that any reflections occurring at the outer surfaces are displaced from the optical axis, otherwise each plate would, in itself, form an additional low contrast etalon, which would be most undesirable.

Temperature variations can cause changes in the optical path difference, proportional in magnitude to the cavity length. The primary causes are thermal expansion of the spacers and variations in the refractive index of the cavity medium. The etalon spacers are made of a glass material called Zerodur, which has a very low coefficient of thermal expansion, in the region of $10^{-7}/°C$.

The three Zerodur spacers are mounted just inside the circumference of the plates. They are usually cylindrical in shape. An alternative approach is to use a single spacer in the form of a hollow cylinder with the same outside diameter as the etalon plates. This is a useful technique for the manufacture of small diameter etalons or those with large cavity lengths.

Spacers less than a millimetre long are difficult to fabricate accurately, so small cavity lengths are realised by making a "stepped" etalon. The spacers are made much longer than the required cavity length and a third glass cylinder with a diameter small enough to fit within the spacers is optically contacted to one of the etalon plates. The etalon cavity is then formed between this plate and the second plate. This technique is important for capacitance stabilised etalons, discussed in chapter eleven, where the piezo-electric transducers have a minimum overall length of about 20mm.

The whole assembly is optically contacted together. For very large etalons or for applications where mechanical shock is inevitable, such as in space flight or field campaigns, the strength of the bonds may be increased by adding a ring of ultra-violet setting cement around the base of each spacer where it contacts the etalon plate (*Rees et al, 1982c*).

9.13 Mounting the etalon

It is most important to design the etalon housing correctly in order to ensure maximum stability of the etalon gap. It consists of two

parts: the mount for the etalon itself, and the sealed housing which contains the etalon and its mount.

9.13.1 The kinematic mount

The kinematic mount is the design adopted for use with all the UCL interferometers. It was originally designed by W.H. Hansen at the University of Michigan, for an interferometer on the Dynamics Explorer satellite. The mount consists of two end plates separated by invar pillars. The etalon is clamped at each end by cantilever beam springs at three points which coincide with the positions of the etalon spacers. Adjusting screws enable the pressure on each point to be varied, thereby slightly compressing the etalon spacers and enabling small errors in parallelism to be corrected. It is essential that the pressure applied by these springs remains constant after adjustment. The etalon is restrained laterally by nylon-faced grub screws located in the side pillars of the mount.

The design of the mount is fully described in *Killeen et al (1982)*, the companion paper to *Rees et al (1982c)*. These papers provide a detailed account of the testing carried out at University College London and the University of Michigan, Ann Arbor, as part of the Dynamics Explorer programme. The tests also confirm the suitability of Zerodur as a material for the etalon spacers and demonstrate that there is no significant thermal coupling between the mount and the etalon. An example of the kinematic mount is shown in figure 11.5, which depicts a 140mm diameter capacitance stabilised etalon.

9.13.2 The sealed housing

It is of paramount importance that the optical path length remains constant to within a very small fraction of a wavelength of light. This means that not only the physical path length but also the refractive index of the cavity must be held constant within very close limits. For this reason, air-spaced etalons are routinely operated in sealed, temperature stabilised housings, which are flushed with dry nitrogen and then evacuated. This procedure helps to eliminate

residual moisture and minimise changes in the optical path difference due to pressure changes in the cavity medium. If the fringes are observed during the evacuation process, the etalon will be seen to ramp through several free spectral ranges. This should be considered within the context of the stability requirements for atmospheric wind measurements. It will be seen in the next chapter that this application requires the etalon to be stable to better than one thousandth of an order.

The sealed housing is normally cylindrical and is made from anodised aluminium alloy. It is fitted with windows at either end. The end plates and window clamps are fitted with O-ring seals to ensure that the housing is completely pressure sealed. A tap mounted on the side of the housing allows the etalon to be evacuated and filled with nitrogen.

The O-ring seals must be fitted with very great care as the slightest leak will cause a noticeable drift in the fringe pattern. The windows at either end must be optically flat to avoid distortion of the etalon fringes and should be suitably thick to avoid flexing under vacuum. Good quality anti-reflectance coatings are essential to assure optimum light transmission and minimise further interference fringes. Hard coatings are advisable as the outer surfaces of the windows will be exposed to the outside world and will, therefore, require periodic cleaning. The coatings should be optimised for the range of wavelengths to be observed.

The effect of a pressure leak can be calculated if equation [9.4] is re-written in terms of absolute refractive index. It can be seen that the order, n, is directly proportional to μ:

$$n = \frac{2\mu d\cos\theta}{\lambda_0} \qquad [9.17]$$

The dependence of n on μ can be assessed by plotting Δn, the change in n for one atmosphere change in pressure, as a function of d, the

etalon gap. This is shown for on-axis rays at three wavelengths of interest in figure 9.6. It is seen that Δn increases steeply with etalon gap. The graph illustrates that, as a useful guide, the number of orders per atmosphere is approximately equal to the etalon gap in millimetres. This is exactly so for a wavelength of 580nm.

9.13.3 Temperature control

The temperature of the housing must be controlled with great precision. This is effected by winding a resistive heating coil around the circumference of the housing. The winding should be non-inductive (double wound) so that a magnetic field is not produced. A Pt100 type thermistor is bonded to the housing to measure the temperature. The housing is wrapped in aluminised mylar and insulating foam sheet. The temperature can then be controlled to an accuracy of 0.1°C using a commercial PID (proportional, integral and derivative) temperature controller. When correctly adjusted these units can give very smooth temperature control without the forced oscillation inherent in "on/off" controllers.

It should be emphasised that these controllers must be optimised to suit the thermal characteristics of the system. The temptation is to use them as "black box" controllers and to assume that all is well simply because the set-point and the measured values of temperature are seen to coincide. However, this does not mean that optimum control has been achieved, and small oscillations in temperature can occur which are visible only when analysing the interferograms obtained from the instrument. The effects of less than optimum temperature control are more apparent with large cavity lengths, and were much in evidence during the testing of a 150mm gap etalon, described in paragraph 11.12.

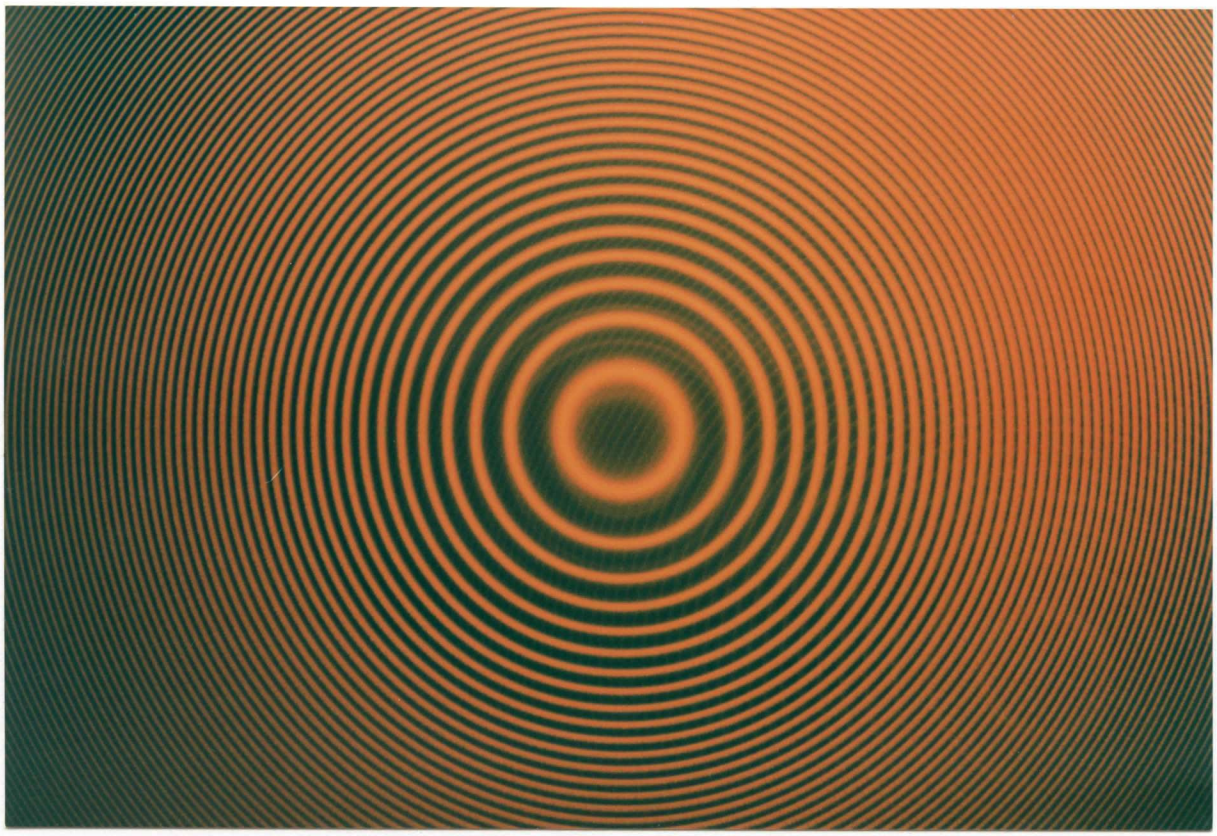


Figure 9.1

Fabry–Perot interference fringes

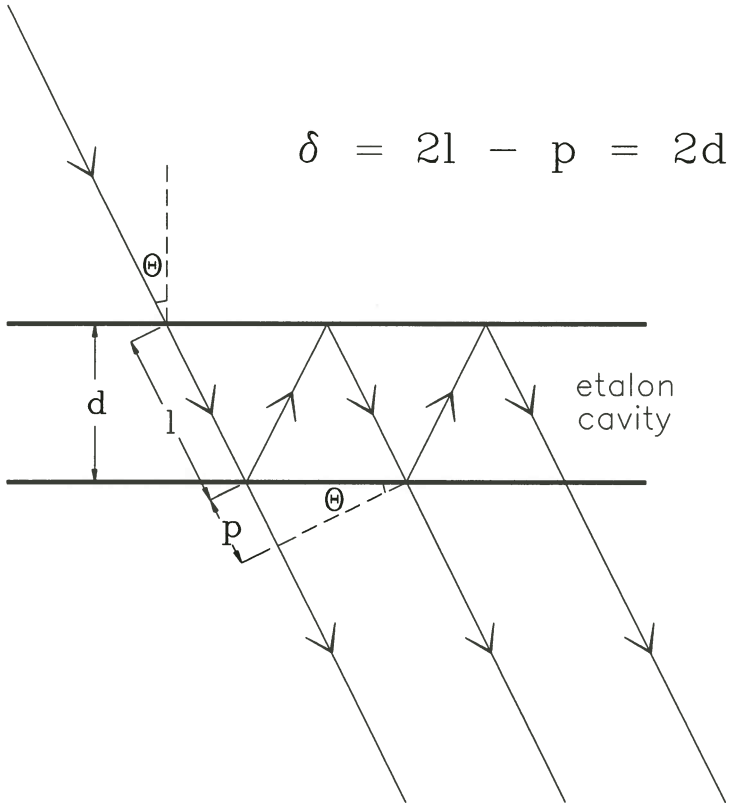


Figure 9.2
Multiple reflection
in an etalon cavity

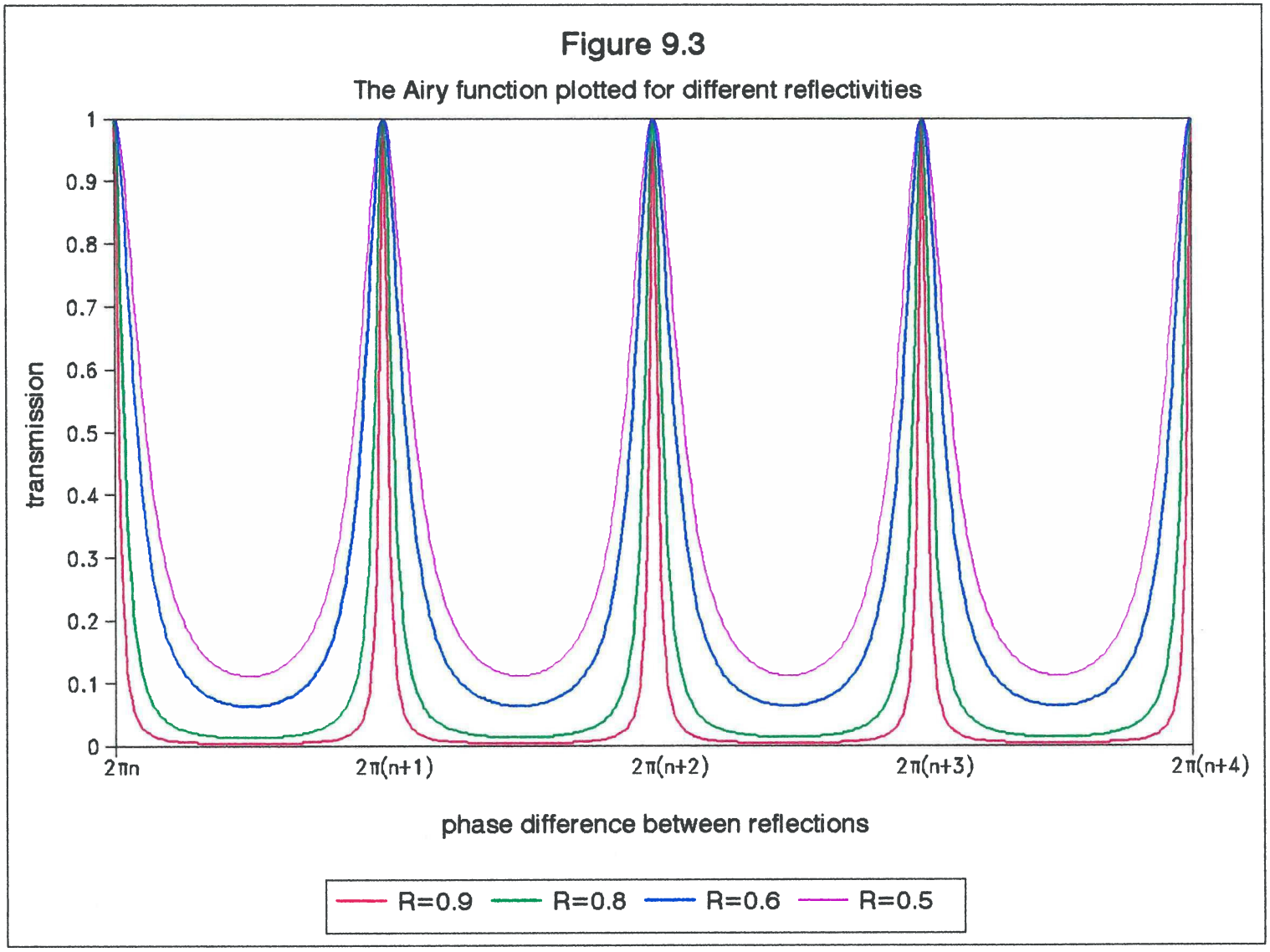


Figure 9.3
The Airy function plotted for different reflectivities
transmission
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2πn
2π(n+1)
2π(n+2)
2π(n+3)
2π(n+4)
phase difference between reflections
R=0.9
R=0.8
R=0.6
R=0.5

Figure 9.4

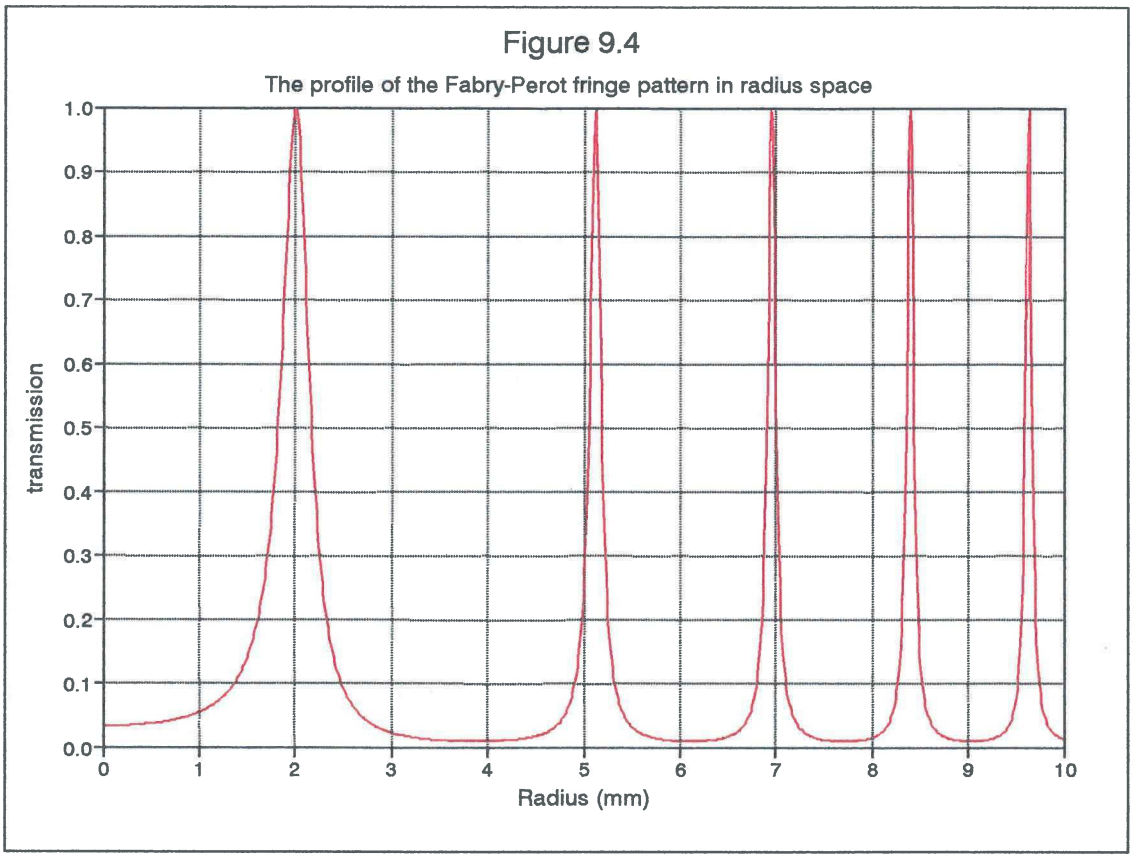


Figure 9.5

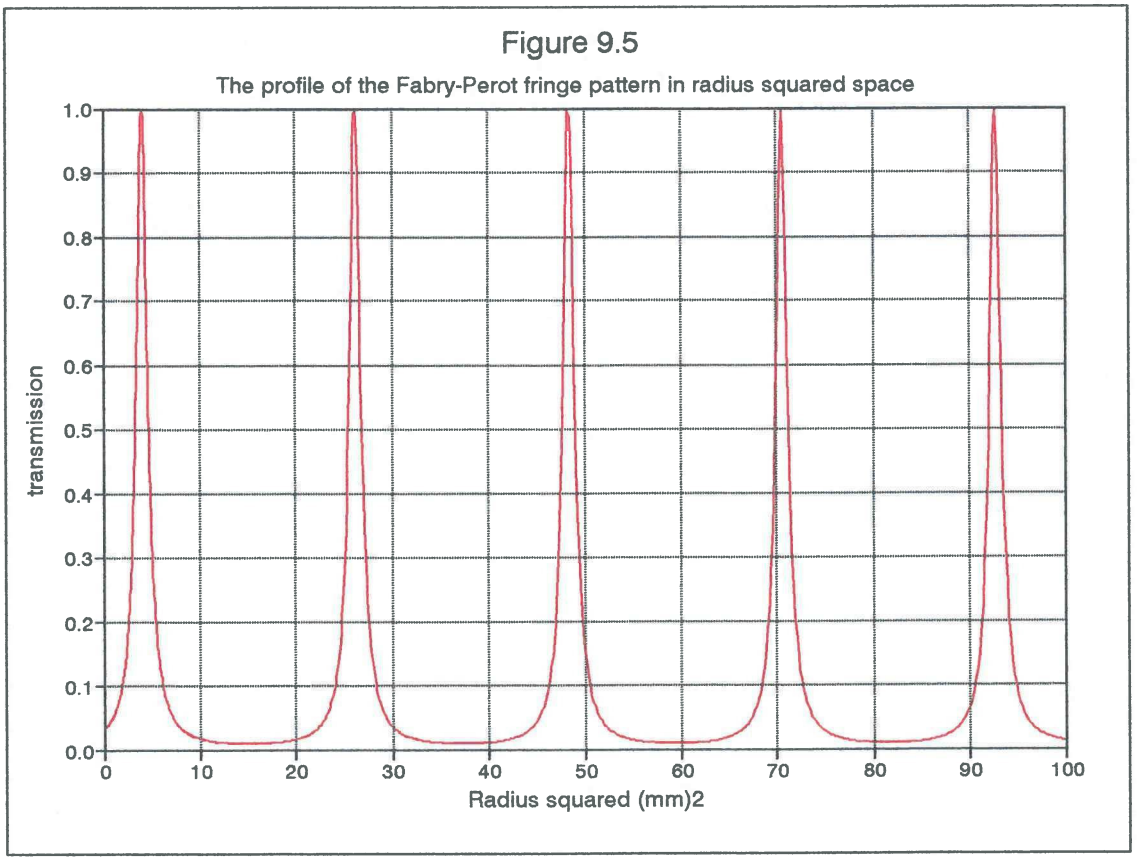


Figure 9.6


Number of orders for one atmosphere pressure change
Orders per atmosphere
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Etalon gap (mm)
580nm
630nm
843nm

CHAPTER 10 - THE FABRY-PEROT INTERFEROMETER

10.1 The UCL interferometers

A series of Fabry-Perot Interferometers (FPIs) have been constructed at the Atmospheric Physics Laboratory at UCL, for operation in Northern Scandinavia, North America and Antarctica. They have been used for routine observations of thermospheric and mesospheric winds. A full description of the observing stations is contained in *Rees, McWhirter, Aruliah and Batten (1992)*, the APL entry in the *World Ionosphere / Thermosphere Study (WITS)* handbook, published by SCOSTEP. Instruments are currently in operation at the following sites:

The Swedish Institute for Space Physics, Kiruna, Northern Sweden.
Kilpisjarvi, Finland.
Longyearbyen, Svalbard.
The British Antarctic Survey Base, Halley Bay, Antarctica.
Bear Lake Observatory, Utah State University, U.S.A.
The University of Saskatchewan, Saskatoon, Canada.
The University of Calgary, Alberta, Canada.
Korhogo, Ivory Coast, Africa.

With the exception of the Canadian and Antarctic instruments, they are all still owned and maintained by the APL. Considerable help has been received from local institutions with their maintenance and operation.

Two classes of instrument have been built: the scanning mirror FPI and the Doppler Imaging System (DIS). This chapter describes their design and implementation.

The scientific goal is to measure wind speeds to an accuracy approaching five metres per second with a time resolution of ten to fifteen minutes for an all-sky measurement. This is achieved by measuring the shift in peak position of a Fabry-Perot fringe which is imaged on to the IPD, using a long focus lens or Cassegrain telescope.

The scanning mirror FPI looks at one portion of the sky at a time, which is selected by the mirror. The field of view is approximately one degree. The time resolution of the final data is limited by the

time that it takes to scan around all mirror positions, and the detector must obtain a low-noise image with an integration time of only one or two minutes. The IPD is well suited to this requirement.

Although the DQE of the IPD is not as high as that obtainable with a CCD imager, the device is virtually noise free, with only a small contribution from thermionic emission. There is no read-out noise to impose a lower limit on integration time. It is, therefore, possible to obtain measurements which are limited only by photon statistics. It will be seen in paragraph 10.7 that the spatial resolution requirements are not demanding, although image stability is paramount.

The Doppler Imager has an "all-sky" lens which maps a large portion of the sky, typically 100 degrees, onto an angle corresponding to several orders of the etalon. Simultaneous measurements of the entire field of view can be made by analysing segments of the image. This instrument is far more critical in terms of image stability and accuracy of analysis than the scanning mirror FPI.

10.2 The scanning mirror Fabry-Perot interferometer

The interferometer is constructed around a rigid optical bench system consisting of four 20mm diameter stainless steel bars. The mounting plates for the various components slide along these on brass bushes. Figure 10.1 shows a diagram of one of these instruments.

The following components are mounted on the optical bench:

1. The Fabry-Perot etalon in its thermally controlled housing.
2. The Imaging Photon Detector or an alternative imaging device.
3. A lens system telescope to focus the fringes on to the IPD.
4. A narrow-band interference filter to select the wavelength of interest. Multiple filter systems have a rotating filter wheel.
5. A scanning mirror collect light from pre-determined directions.

In addition, the following items are required:

1. A spectral calibration lamp, positioned to scatter light from a reflector placed at one of the scanning mirror positions.
2. The IPD Signal Processing Unit.

3. The Interferometer Control Unit, consisting of a thermal control unit and the electronics to control the scanning mirror, calibration lamps, filter wheel and shutter.
4. Cooling equipment for the IPD if required.
5. An IBM compatible computer to control the entire system and collect, process and store the data.

10.3 The etalon

The etalon diameter defines the input aperture of the instrument, and thus its sensitivity. A diameter of 150mm is currently used for the majority of these instruments. The etalon is contained in its kinematic mount inside a sealed housing, as described in chapter nine. It is necessary to adjust the orientation of the etalon so that its optical axis is aligned with the rest of the instrument. This is effected by mounting the etalon with three sets of opposing screws which enable it to be tilted to the correct orientation and securely locked into place.

10.4 Choice of etalon cavity

For atmospheric observations the choice of etalon plate separation is a matter of compromise. The smaller the free spectral range the greater the potential accuracy. However, the free spectral range must be sufficiently large to accommodate the spectral line in question in order to enable accurate peak fitting of the line profile. The line width is largely determined by the temperature of the source because the differential velocities of the gas molecules cause the line to be broadened.

A spacing of 14mm has been found suitable for observations of the oxygen (OI) emissions at 557.7nm and 630 nm. A gap of exactly 20.49mm has proved successful for observations of the hydroxyl (OH) emission at 843.2nm. The emission consists of a doublet, but because the separation of the two peaks is an exact multiple of the free spectral range of the etalon, they are superimposed.

10.5 The imaging optics

The imaging optics must be carefully matched to the etalon cavity length and the diameter of the imaging detector, as explained in chapter nine. A focal length in the region of 1.2 metres is typically required. This is usually realised by a two lens system on a one metre optical bench. The primary lens is the same diameter as the etalon and has a focal length of about 750mm. The secondary, diverging lens has a diameter of 35mm and a focal length of -200mm. Alternatively a Cassegrain telescope will permit an even shorter optical bench to be used.

10.6 The interference filter

The interference filter is essential in order to reduce the amount of continuum in the observed light to a level which will not overload the detector or mask the spectral line which is being observed. It is usually chosen to have a bandwidth of 1nm. The filters used are, in reality, multiple cavity solid Fabry-Perot etalons. The angle of acceptance is in the region of two degrees, so the filter must be positioned at the place in the optical path where the angular size of the light cone is smallest, that is to say, between the final lens of the imaging system and the detector. The full angle from a 150mm diameter etalon and 1200mm imaging lens is approximately seven degrees, so it is necessary to specify the filter pass-band to be centred at a wavelength which is slightly longer than the wavelength of interest. A 630.2nm filter is specified for an interferometer operating at 630.0nm. This places the peak response for the observed wavelength slightly off the optical axis and the whole of the light cone will be accommodated by the filter.

10.7 Stability requirements

The accuracy with which Doppler shifts of atmospheric emissions can be measured is dependent on careful calibration and the inherent stability of the instrument.

The free spectral range of the 14mm gap etalon used for 630nm is

14.2pm. The shift in wavelength of a spectral source is given by:

$$\delta\lambda = \lambda v/c \qquad [10.1]$$

where v is the line of sight velocity and c is the speed of light. A precision of 5 m/s requires the wavelength to be measured to an accuracy of 0.01pm, or 0.07% of a free spectral range, about one eight of a radius-squared element. The highest wind velocities normally encountered when observing thermospheric winds are of the order of 1000 m/s and produce a Doppler shift of 14% of a free spectral range.

The spatial resolution of the detector is not a critical factor as the etalon is designed to produce a line width which is matched to the detector, with a FWHM of approximately 3mm on the image. Because of computer memory limitations the image is restricted to 256 x 256 pixels, each measuring approximately 100μm across. This gives 30 resolution elements FWHM, sufficient to obtain an accurate peak fit. However, an accuracy of 5m/s requires an image stability in the region of 20μm for a 10mm diameter fringe. Larger fringe diameters require even higher stability, as wavelength is proportional to radius squared.

The etalon spacers are subject to a thermal expansion in the region of 10^{-7}/°C, or 1.4nm/°C for a 14mm gap etalon. This will cause a shift of around 0.5% of FSR/°C in the etalon response. The etalon temperature must therefore be held constant to within 0.1°C. This is realisable in practice, using the methods outlined in paragraph 9.13.3.

10.8 Instrument control electronics

The interferometer is controlled by an IBM PC compatible computer fitted with a Scientific Solutions (Tecmar) base board. This is a 96 line input/output card based on four Intel 8255 parallel interface integrated circuits. It provides a versatile arrangement of four ports, each of 24 lines, which may be separately configured as input or output with or without handshaking. One port is used for the Imaging Photon Detector, and one for the Interferometer Control Unit which drives all the ancillary functions.

A typical control station is shown in figure 10.2. This is the 630nm. instrument at Kiruna, Sweden, which was installed by the author. The

photograph shows the IPD Signal Processing Unit, the power supply for the Peltier cooler, the IPD thermometer, the PC - AT computer and the Interferometer Control Unit. The interferometer is suspended in the dome of the adjacent observing room and is shown in figure 10.3.

10.8.1 The Interferometer Control Unit

This contains the necessary circuitry to enable the TTL voltage levels from the computer interface to drive the scanning mirror, filter wheel, calibration lamps and shutter. All functions may be controlled manually using toggle switches on the front panel. The unit also houses the temperature controller for the etalon. The circuit diagram is shown in figure 10.4.

10.8.2 The scanning mirror

The mirror is situated on a rotating table over the vertically mounted optical bench. It is angled to provide a viewing elevation of 30 or 45 degrees above the horizon as required. A simple position encoder is effected by fixing reflective mylar spots to the face of the rotating turntable. These are read by a bank of four optosensors, giving a maximum of fifteen positions. In practice, eight are used for the nominal compass points, plus a parking position, code 15, which is automatically selected when the computer is powered down. One point is reserved for the zenith measurement, which is viewed via a second, fixed mirror, and one point for the calibration lamp, which illuminates a uniformly scattering surface positioned to fill the field of view. These positions are usually chosen to be south-west and south-east. The desired mirror position is output by the computer on four lines of the parallel port, and this is compared with the output from the optosensors. When a match is detected, the power is removed from the motor and a low resistance load applied across the motor windings to effect a brake. The position number is displayed on the front panel of the control unit.

10.8.3 The filter wheel

Electronically this is identical to the scanning mirror. Five interference filters, each 50mm in diameter, are mounted on a

revolving wheel driven by an electric motor. The assembly is mounted on the optical bench directly in front of the imaging detector. The position of the wheel is detected by a bank of three optosensors and displayed on the front panel in the same way as for the scanning mirror. The filter wheel assembly also incorporates a shutter in the form of a rotary solenoid which is actuated by one of the function outputs of the controller. Unused filter wheel positions are fitted with blanks to protect the detector from excessive illumination.

10.8.4 Calibration lamps

Each calibration lamp is driven by a single line from the computer which actuates a relay in the control unit. They are electrodeless lamps which are excited by a radio frequency oscillator. The lamps are sealed quartz drums containing a mixture of noble gases, which will generate emission lines in the regions of interest. The lamp supply operates at a frequency of about 5 MHz, providing an electric field in the region of 400 volts/cm. The lamp is shown in figure 10.5 and its circuit diagram in figure 10.6. A combination of two lamps, neon and a mixture of helium, argon and krypton (HAK) produces a selection of lines suitable for many requirements. In particular, the 630.4nm line of neon provides a very convenient source for the 630.0nm OI emission.

10.9 Control software - Halo

A comprehensive instrument control programme called Halo has been developed at UCL by Daniel Wade. Originally written to take and display images from the IPD, it has been expanded to perform many functions, which can be summarised as follows:

1. Run an almanac to determine the position of the sun and the moon. Start the data collection process when the sun is below a pre-determined elevation.
2. Turn the IPD high voltage on and off as required, not only during daylight hours, but also at night when the moon is in the field of view of the instrument.
3. Take an IPD image for a pre-determined integration time.

4. Reduce the image to a histogram of intensity plotted against the square of the radius of the Fabry-Perot rings.
5. Write the full or reduced image to a data file.
6. Rotate the scanning mirror to the desired position.
7. Turn the calibration lamp on and off when required.
8. Rotate the filter wheel to select a new filter.
9. Operate the shutter.

10.10 Data collection and analysis

The scanning mirror FPI is well proven and a succession of instruments has been collecting data routinely since 1981. This has resulted in numerous publications, e.g. *Rees et al (1982b); Lloyd (1985) and Aruliah (1991)*. Others are listed at the end of this thesis.

A typical image of the 630nm emission line is shown in figure 10.7, and the corresponding Halo radius-squared plot is shown in figure 10.8. The data is stored in the "reduced to radius squared" format, which requires only 0.5 Kbytes of computer disc space per image, compared with the 128 Kbytes needed for a full image. This means that a week's data consisting of two minute integrations can normally be accommodated on two floppy discs, which can easily be copied and sent by post to UCL. Halo has the facility to store a full image at regular intervals, normally once a month, in order to monitor the performance of the IPD. To calculate the reduced image the centre of the ring pattern has to be determined by inspection of the image. The co-ordinates are then used by Halo to generate a look-up table, which is corrected for quantisation effects due to the finite number of pixels. The correction is calculated by assigning each of the pixels in the 256 x 256 image to one of 256 radius squared bins, and dividing the total photon count for that bin by the number of pixels allocated to it.

On arrival at the APL the data is transferred to a DEC workstation where an accurate peak-fitting routine is used to calculate the Doppler shifts from the interferograms.

10.11 The Doppler Imaging System

The Doppler Imaging System differs from the scanning mirror Fabry-Perot interferometer in that it is an all-sky instrument. It samples all directions in azimuth simultaneously, at elevations dictated by the properties of the all-sky lens. The basic optical bench is similar to that of the scanning mirror FPI, but the scanning mirror is replaced with another optical bench which contains a telecentric all-sky lens and a collimating lens. The purpose of this is to map the all-sky image on to several fringes of the interference pattern of the etalon. In other words, a field of view of approximately 100 degrees is mapped on to the field of view of the etalon/lens/detector system. This is approximately two degrees, corresponding to five or six fringes.

The main advantages of the Doppler Imaging System over the scanning mirror FPI are:

1. Simultaneous measurements over a large portion of the sky enable winds to be measured over an area of approximately 800 km.
2. The absence of a scanning mirror means there are no moving parts.
3. The overall instrument sensitivity in terms of the product of the area and solid angle, $A\Omega$, is much greater.

There are, however, some disadvantages:

1. The calibration requirements are more exacting.
2. Stringent stability requirements are placed on the imaging system.
3. The signal-to-noise ratio is inferior because each measurement uses only part of the image.
4. The data analysis is more demanding.

10.11.1 Improvements to the original instrument

The instrument was originally described by *Rees and Greenaway (1983)* and experimental results have been published in *Rees et al (1984)* and

Batten et al (1988). Since these early publications, significant improvements have been incorporated into the instrument.

The major improvement to the optical system was the introduction of an all-sky lens which replaced the original cumbersome mirror assembly. This provided an improved field of view and did not obscure a large area around the zenith. Not only was the optical performance improved, but the alignment and focussing was considerably simplified. The new lens also incorporated a shutter which, when closed, enabled calibration images to be taken without contamination by sky signal. Although this now meant that a moving mechanical component was introduced into the system, the improved calibration facility was considered well worth any slightly increased failure risk.

A calibration source was incorporated into the optical path. This replaced the unsatisfactory method of illuminating a segment of reflecting material placed in the dome. The new method introduced the light by reflecting it from a 45° window placed between the all-sky lens and the etalon. This was made physically possible by the new all-sky configuration.

10.11.2 Analysis of the image

A typical image obtained from the Doppler Imaging System is shown in figure 10.9. It is clear that substantially more information is contained in this image compared to that produced by the scanning FPI. Each ring represents a different angle of elevation and each point on the circumference of the ring pattern corresponds to a different azimuth. In order to extract the Doppler wind data it is necessary to divide the fringe pattern into a number of segments which are then individually reduced to the intensity versus radius squared format in the same way as the scanning FPI data. However, for the Doppler imager the data is reduced into 24 segments, so 24 times as much data is generated per image as with the scanning FPI. A plot of one segment of the image of figure 10.9 is shown in figure 10.10. The stability of the detector is critical as the radius squared bins span very few

pixels of the image. The detailed analysis of the data is described in *Batten et al (1988)*.

10.11.3 Calibration of elevation

It is not possible to calculate accurately the relationship between the fringe diameter in the image and the observed angle of elevation because there are several variables in the optical system. This has to be done empirically. The first step is to ensure that the instrument is perfectly vertical. A point light source is then suspended over the centre of the instrument. This defines the zenith, and the detector is adjusted so that the image of the point falls in the exact centre of the field of the detector. The spectral calibration is then switched on and the etalon is tilted until the Fabry-Perot fringes are concentric around the image of the test source. Once the instrument has been aligned with respect to the zenith it is necessary to provide calibration sources at different angles of elevation. Although this can be done using appropriately positioned point sources, a convenient method is to observe the moon with the instrument and compare its position on the image with its elevation as given by the ephemeris.

10.11.4 The detector

Clearly, the demands placed upon the imaging system are far more stringent than is the case with the scanning FPI. Longer integration times are required in order to obtain a satisfactory signal-to-noise ratio. Image stability is of paramount importance as the outer fringes span only a few pixels of the image. It was seen in chapter seven that the Imaging Photon Detector used for the re-designed instrument was not sufficiently stable in this respect. The instrument is currently operating with a cooled CCD, which is expected to give excellent image stability. However, the data received so far indicate that its sensitivity is insufficient unless a particularly active night is being observed. This was not unexpected, and another detector option to be tested at a future date is that of a cooled intensified CCD.

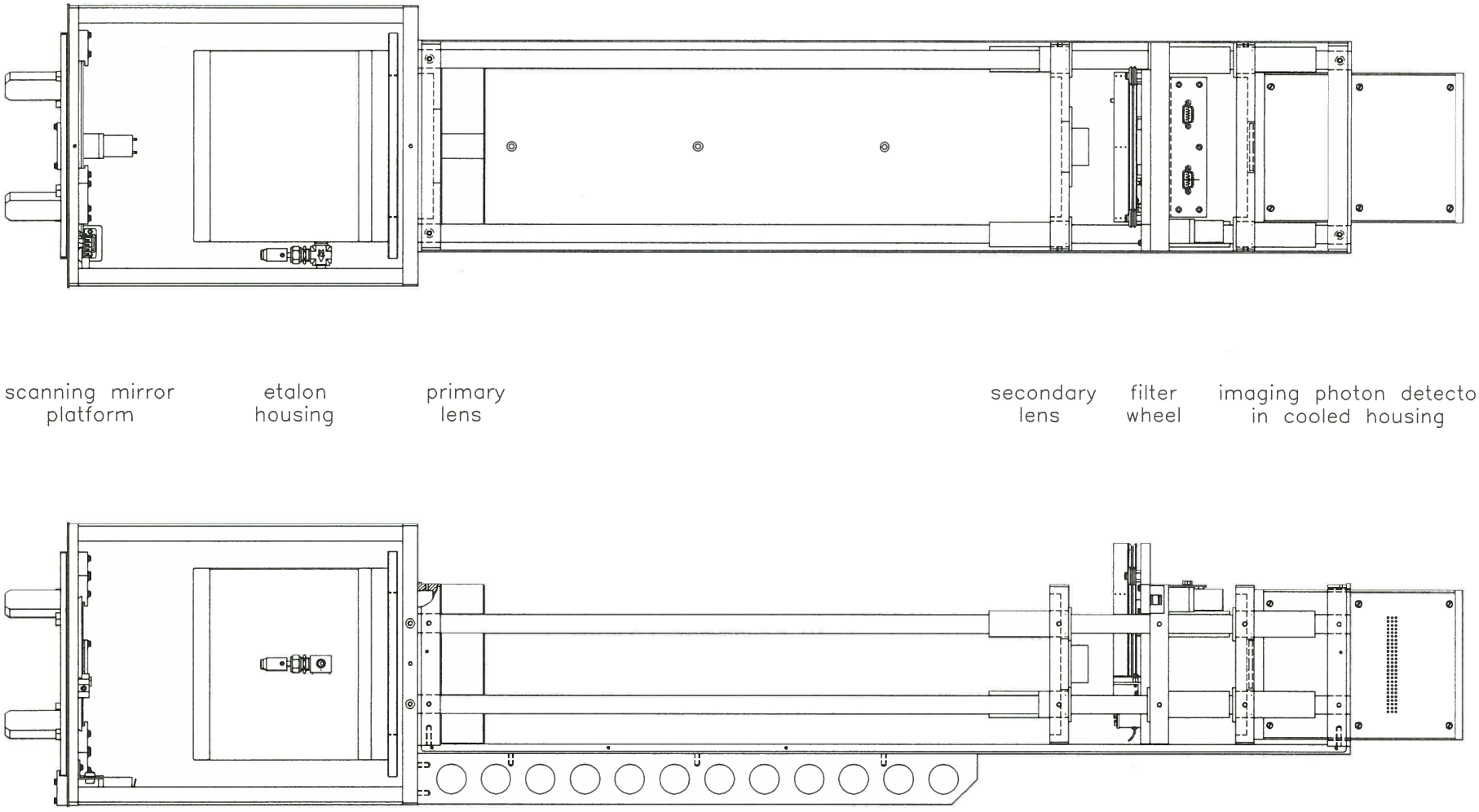


Figure 10.1 The Scanning Mirror Fabry–Perot Interferometer

J. Percival

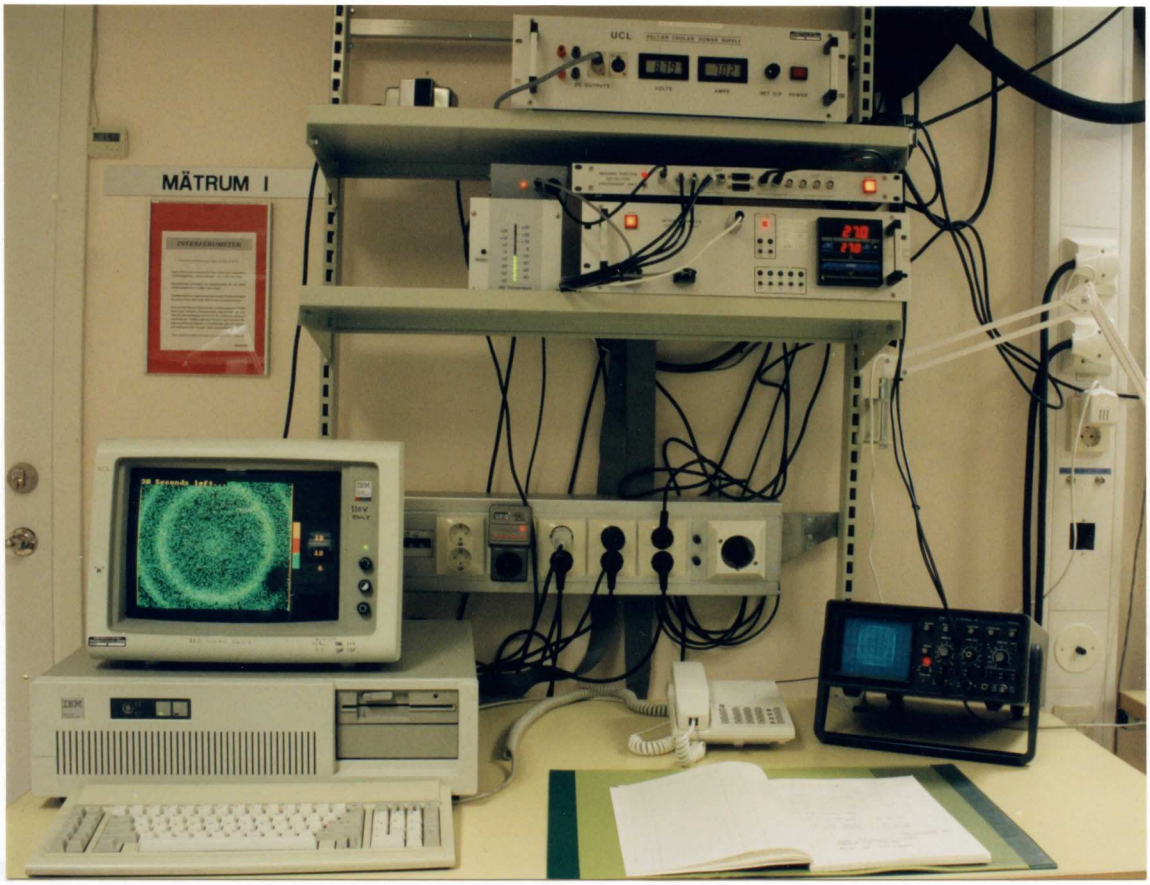


Figure 10.2

The control station for the 630nm FPI at Kiruna

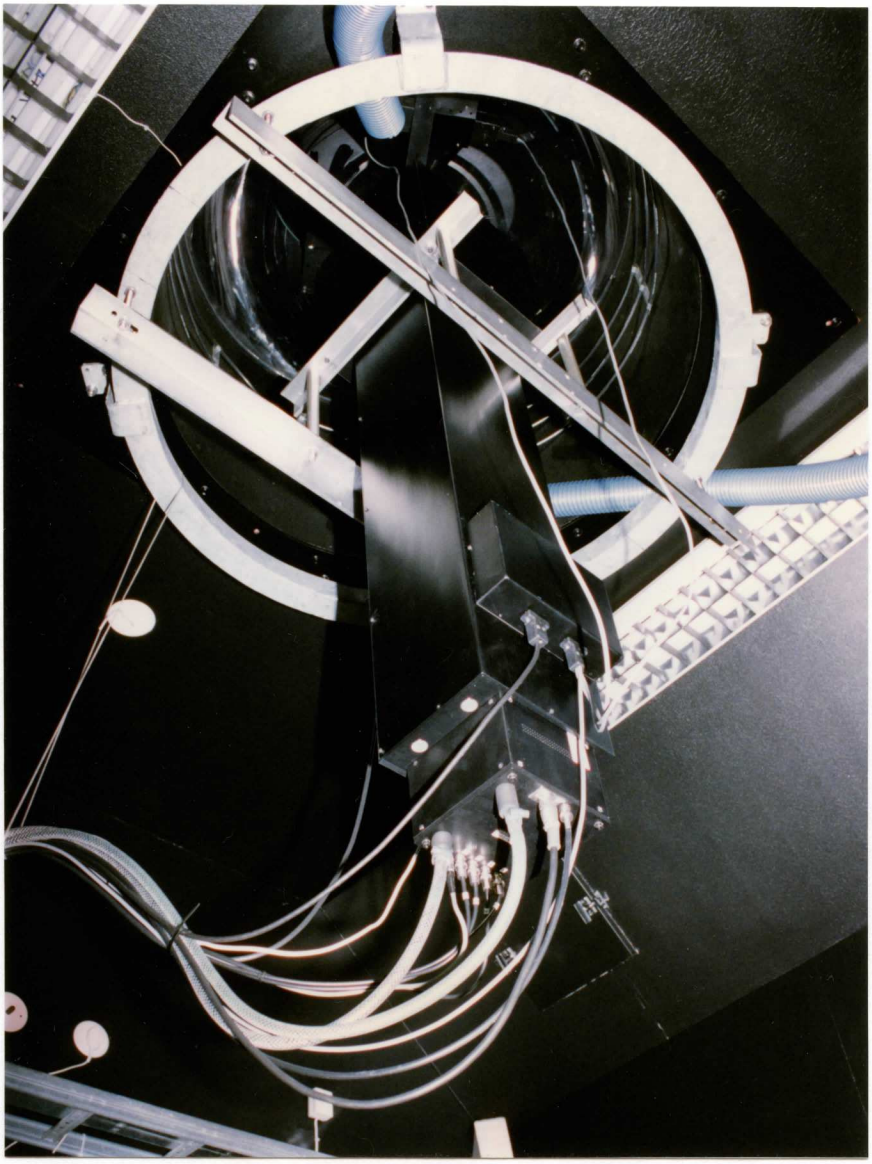


Figure 10.3

The FPI suspended in the observing dome

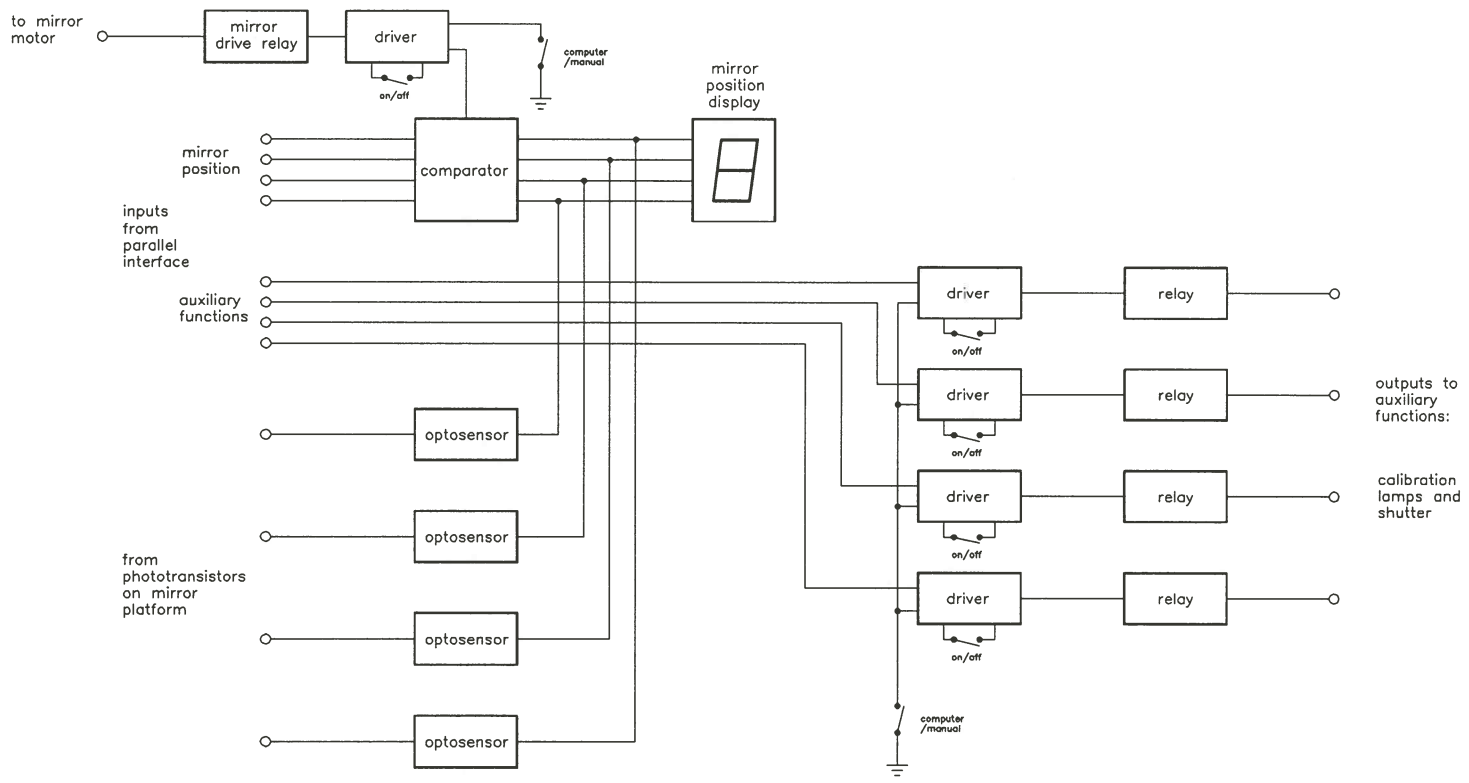


Figure 10.4 The Interferometer Control Unit

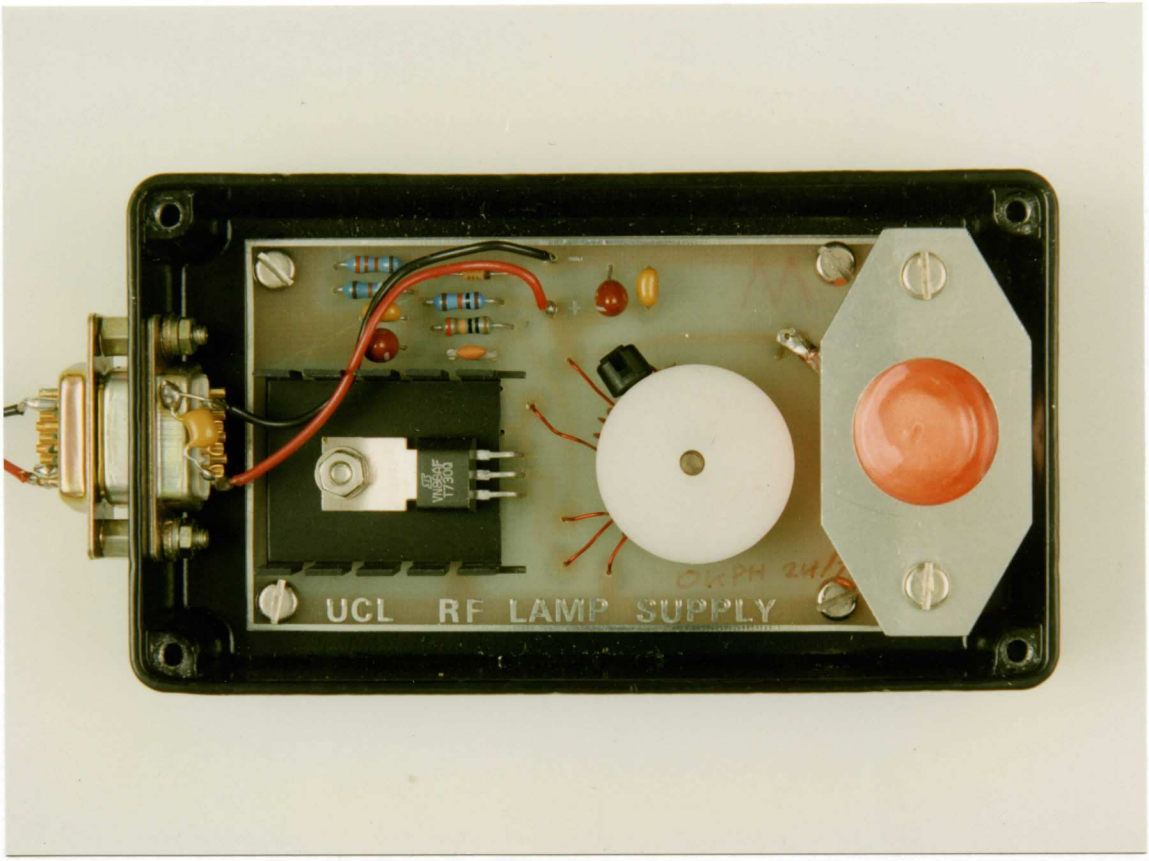


Figure 10.5 The spectral calibration lamp

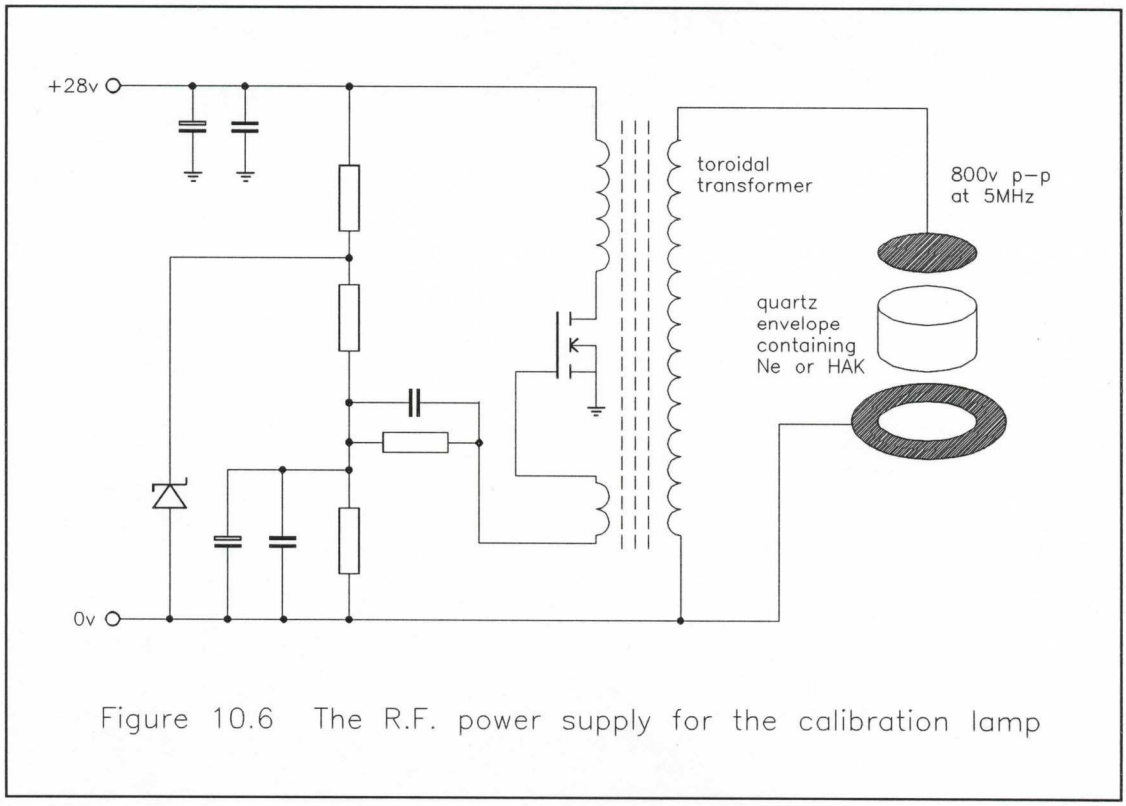
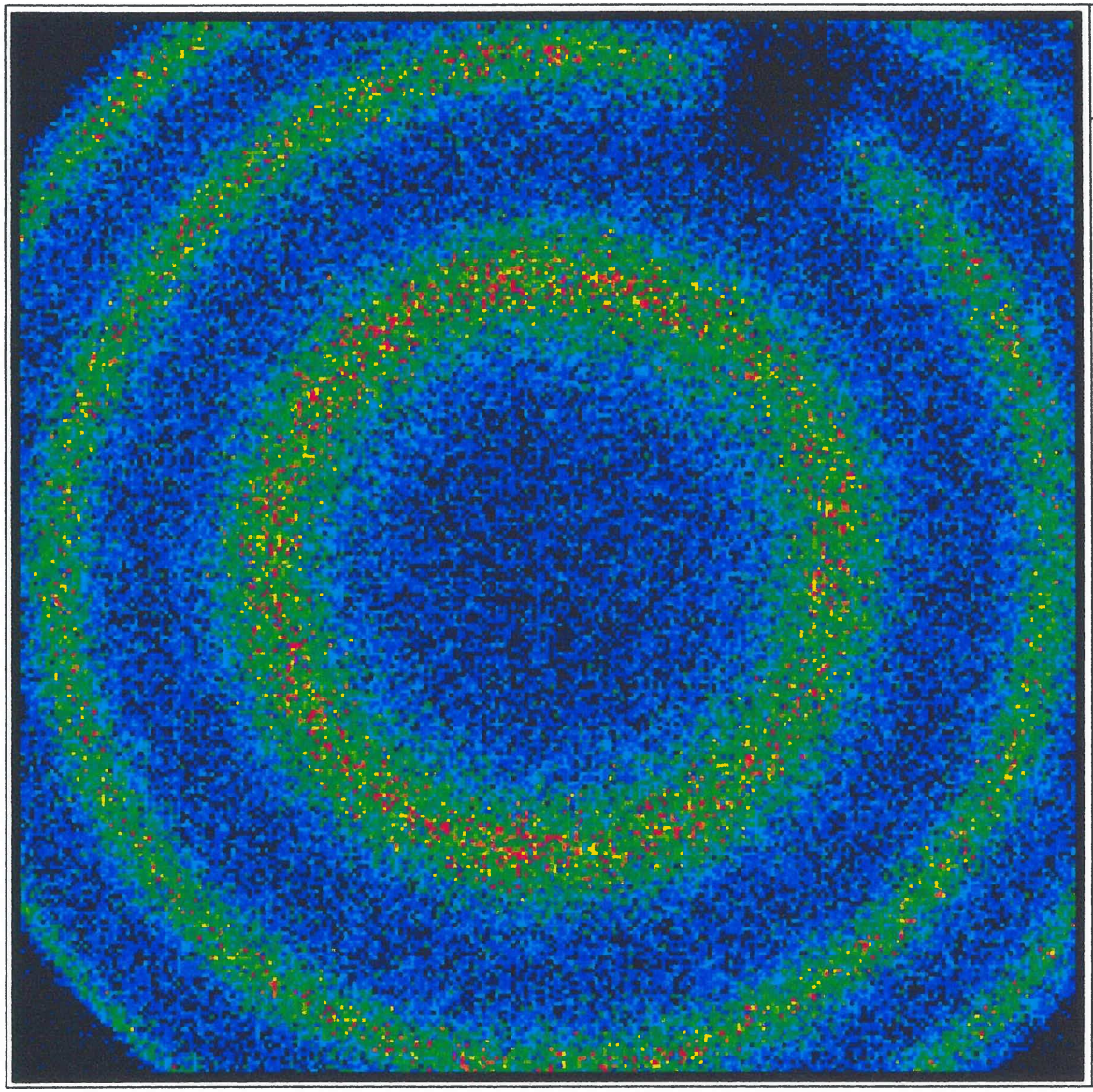
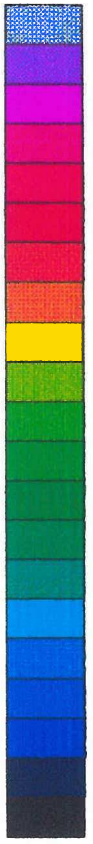


Figure 10.6 The R.F. power supply for the calibration lamp



ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON

Fabry–Perot Interferometer
File: TA20007.G48



630nm OI emission

Figure 10.7

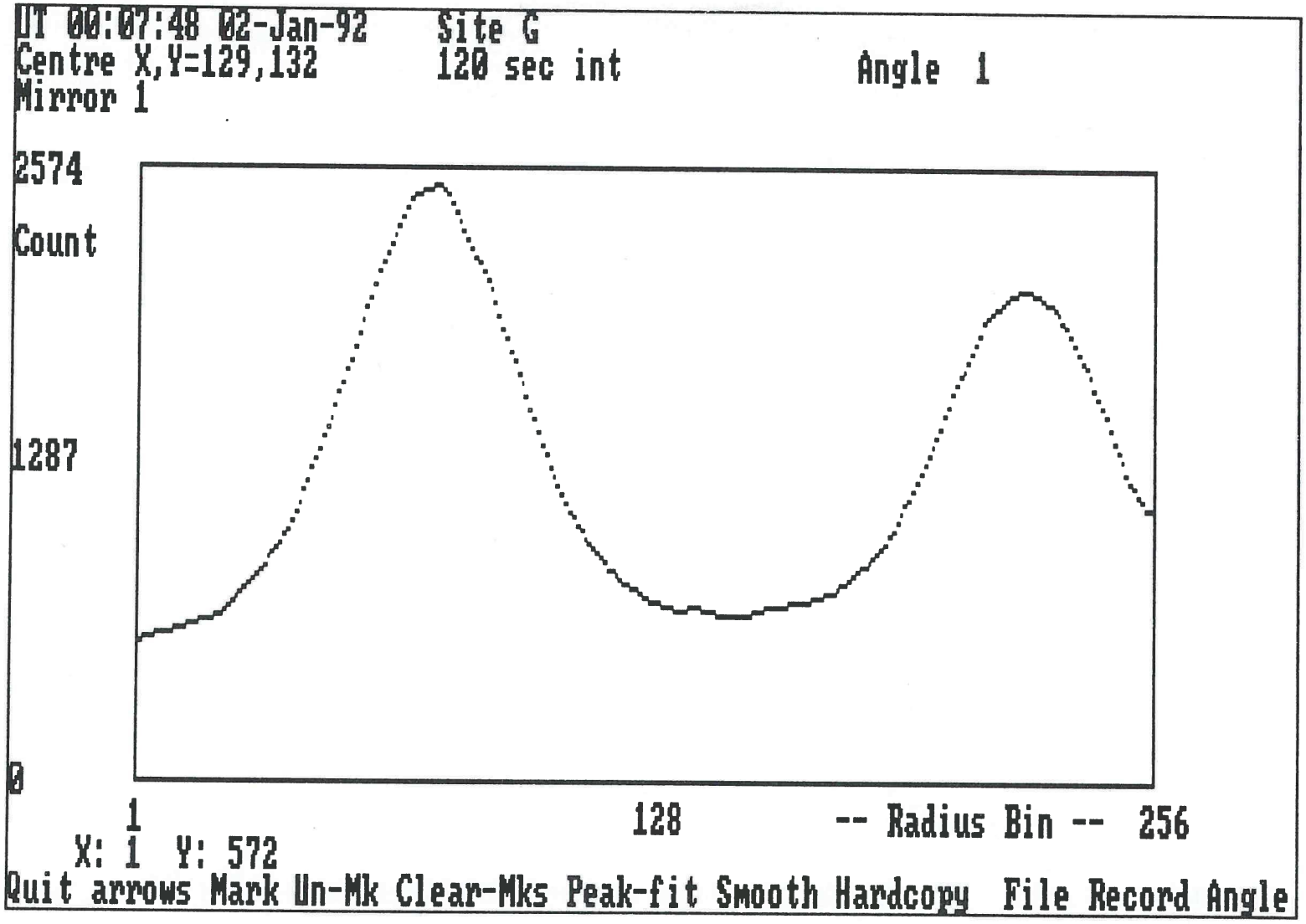
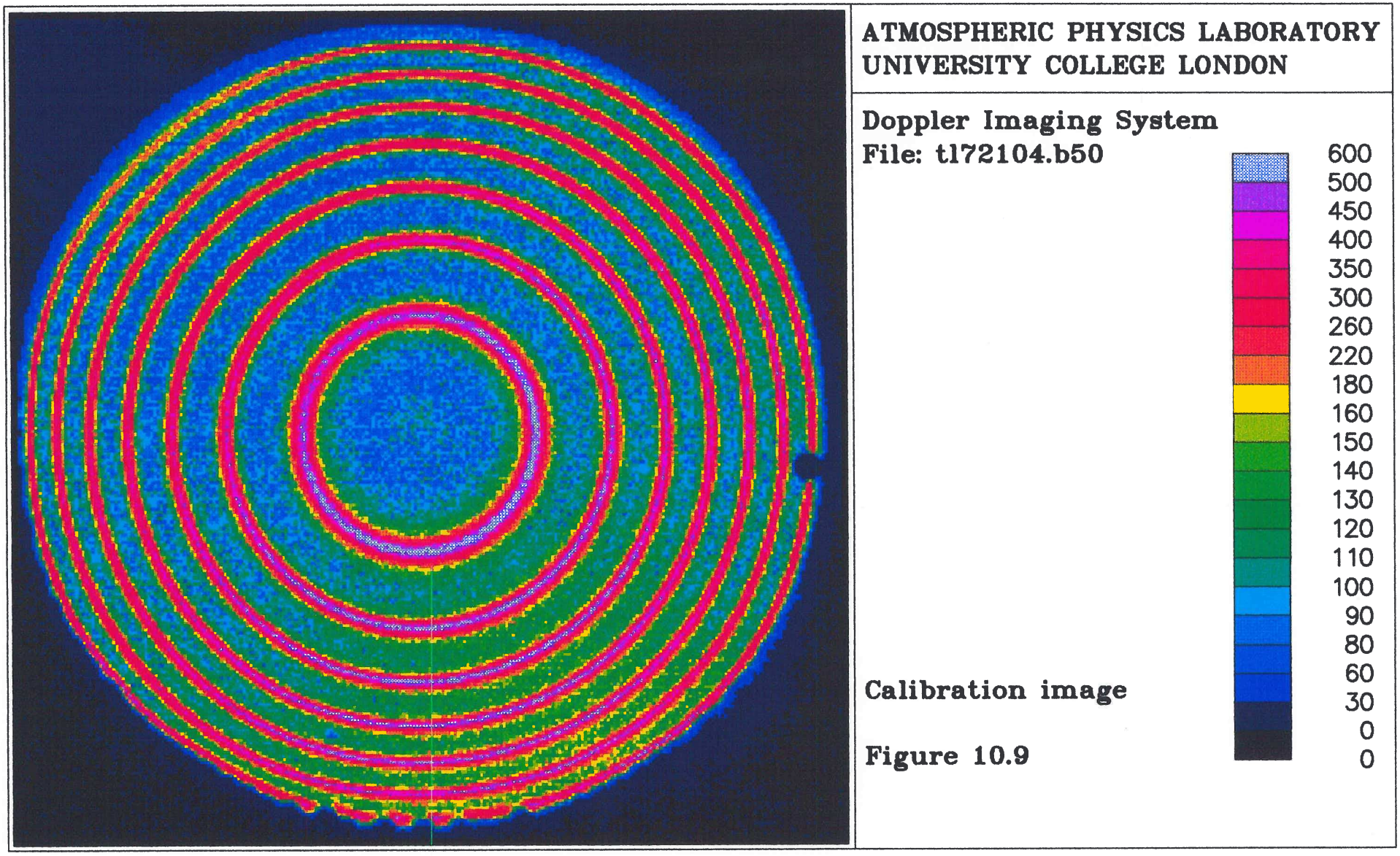


Figure 10.8 Reduced-to radius plot of the FPI image


ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON
Doppler Imaging System
File: t172104.b50
600
500
450
400
350
300
260
220
180
160
150
140
130
120
110
100
90
80
60
30
0
0
Calibration image
Figure 10.9

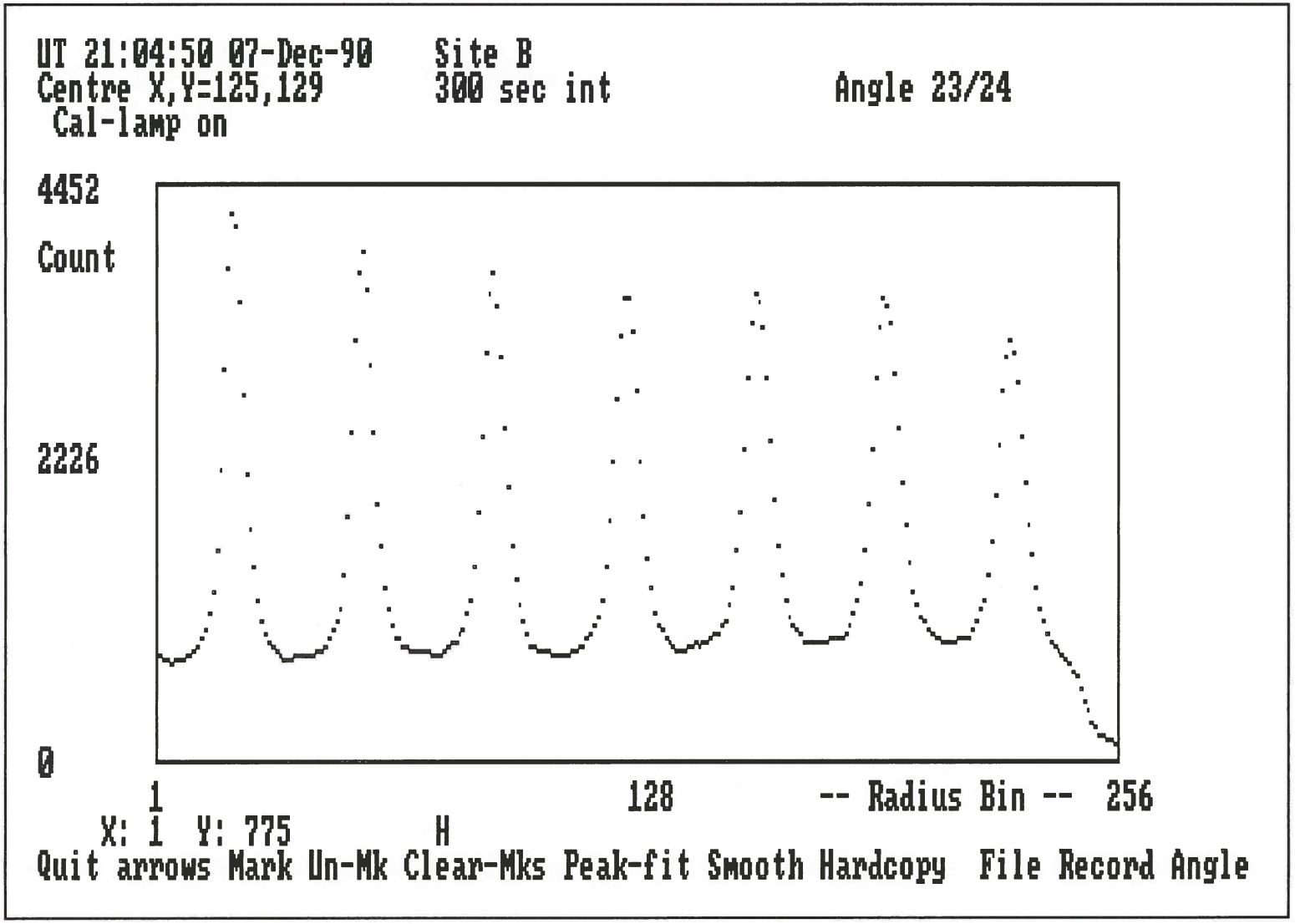


Figure 10.10 Reduced-to-radius plot of the DIS image

CHAPTER 11 - THE CAPACITANCE STABILISED ETALON

11.1 Introduction

This chapter describes the design and development of a series of etalons with variable cavity length, known as capacitance stabilised etalons. They were originally built for the balloon-borne triple etalon interferometer and subsequently used in a number of other instruments, in particular, the ground-based triple etalon interferometer.

It was shown in chapter nine that the angular diameter of the Fabry-Perot interference pattern may be adjusted by tuning the etalon cavity. This entails changing the optical path difference (in terms of wavelengths of light) between the etalon plates. It is achieved either by changing the refractive index in the cavity by varying the gas pressure (pressure scanning), or by changing the physical distance between the plates.

Although pressure scanning is a useful technique it is not generally suitable for rapid or repetitive scanning of the etalon. However, changing the physical length of the etalon cavity in a precise manner is a formidable task when one considers that one free spectral range represents a change in path length of half a wavelength of light, and that an accuracy of one thousandth of a free spectral range is generally required.

A device often used for providing very small amounts of movement is the piezo-electric transducer (PZT). These are made from a material which changes its physical size with an applied voltage. However, such materials are dimensionally unstable. The extension per volt, although predictable within certain limits, is not consistent enough for positioning to the required accuracy and the response exhibits significant hysteresis.

If a method can be devised to determine changes in the etalon gap with sufficient accuracy, the information can be used in an electronic feedback loop to drive the piezo-electric transducers to the required position. This has been done interferometrically by *Hernandez and Roble (1976)* as outlined in *Rees, McWhirter et al (1981b)*, where the difficulties of this approach are discussed. A more elegant solution is provided by the capacitance micrometer. It is possible to fabricate a parallel plate capacitor with one plate attached to each etalon surface. Variations in the etalon gap produce a change in the value of the capacitance which can be measured and used to adjust the piezo-electric voltage. Such a device is termed a capacitance stabilised etalon (CSE).

The early development of this device is described in detail in *Rees, McWhirter et al, 1981b, (A stable, rugged capacitance-stabilised piezo-electric scanned Fabry-Perot etalon)*. The main points of this paper are summarised in this chapter, followed by an account of the substantial improvements effected since its publication.

11.2 The capacitance micrometer

The capacitance micrometer was described by *Jones and Richards (1973)*. Very small changes in length can be determined by measuring corresponding changes in the value of a capacitor. The technique can be adapted to determine the cavity length of a Fabry-Perot etalon.

The micrometer is based on a capacitance bridge, the equivalent circuit of which is shown in figure 11.1. The bridge consists of two sinusoidal voltage sources, e_1 and e_2, and two parallel plate capacitors of nominally equal value, C_1 and C_2. C_2 is a fixed capacitor whilst C_1 is arranged so that the separation between its plates is mechanically variable, in sympathy with the physical length to be determined. The voltage sources are 180 degrees out of phase. The

current, i_b, flowing between the arms of the bridge is given by:

$$i_b = \frac{e_1}{2\pi f C_1} - \frac{e_2}{2\pi f_c 2} \qquad [11.1]$$

where f is the frequency of the voltage sources. This is nominally 16kHz, and is not critical. The selection criteria are that it should be easy to generate in a stable form, it should be high compared with the required response time of the etalons and low enough to allow accurate phase and amplitude information to remain intact when processed by conventional analogue circuitry.

The current in the bridge may be adjusted to zero by making slight adjustments to e_1 or e_2. The value of the capacitor C_1 is given by:

$$C_1 = \frac{A}{\epsilon . d} \qquad [11.2]$$

where A is the area of the plates, d the separation and ϵ the permittivity of the medium separating the plates. A variation in d will, therefore, cause a proportional change in the value of C_1, which will, in turn, create an imbalance in the bridge, causing a current i_b to flow.

This current, i_b, although very small, is straightforward to measure as it is an a.c. signal and may be amplified without d.c. offset and drift problems. Figure 11.2 illustrates the basic micrometer circuit. The bridge current, i_b, is fed into the virtual ground of a current to voltage converter. The resulting voltage signal is then amplified. It remains to convert this to a d.c. voltage which will be proportional to the change in capacitance gap. This is achieved with a phase sensitive detector, also known as a synchronous amplitude modulation detector. It consists of a switched amplifier which changes between inverting and non-inverting modes at alternate half cycles of the input signal. This is achieved with an analogue switch which is synchronised to the oscillator. The output signal has the appearance of a rectified sine wave. It is then filtered to give a smoothly

varying positive voltage equal to the average level of the rectified signal. The signal waveforms are similar to those produced by a simple a.m. detector. However, the important difference is that the output of the simple detector circuit is proportional to the input signal amplitude regardless of phase, whereas the phase sensitive detector will only give its maximum output if the input is in the correct phase. The circuit will automatically reject signals which are caused by resistive offsets in the bridge. These signals are due to minute amounts of surface contamination on the capacitors and surrounding circuitry. The signal resulting from a resistive offset in the bridge is 90 degrees out of phase with the capacitive signal. The phase sensitive detector will, therefore, chop the resistive component of the signal in such a way that the average signal is zero. This makes the micrometer substantially immune to resistive imbalance in the bridge.

The effect is illustrated in figure 11.3, which compares the outputs of the phase sensitive detector for capacitive and resistive offsets. This type of circuit has a very high noise immunity, as random noise will not be synchronous with the clock and, like the resistive offset signal, will be greatly attenuated.

11.3 Application to etalons

A capacitor may be created by metallising the end faces of two pillars which extend towards each other, one from each etalon plate. Changes in the capacitance gap are, therefore, the same as changes in the etalon gap.

This capacitance has a value of typically 15 pF, with a gap in the region of 30 microns. The total capacitance range for scanning one free spectral range of the etalon involves a positional shift of $\lambda/2$, that is, 300 nm for illumination at 600 nm. This gives a total capacitance shift of 0.15 pF. For stability and scanning to one thousandth of a free spectral range the bridge must be sensitive to changes in capacitance of 150 aF.

Tuneable etalon systems using capacitance micrometers have been constructed by *Hicks et al (1974 & 1976)* and *Wells et al (1978)*. These use pairs of capacitance sensors in the X and Y axes to maintain parallelism. A fifth capacitance sensor is compared with a reference capacitor to determine the overall gap.

A different approach, adopted for the UCL etalons, is to use capacitance sensors which are built into the piezo-electric elements themselves. The three transducers are independently adjusted to achieve parallelism and then driven in step to scan the etalon. This relies on precise tracking between the three drive systems, which can be achieved by careful design. This method is neater in construction and cheaper to fabricate. It is also more suitable for the three-point fixing of the space-qualified kinematic mount described in paragraph 9.13.1.

11.4 Transducer designs

The piezo-electric transducers have an extension coefficient of 0.6 nm/volt, which means that 500 volts are required to give an extension of 300nm. This corresponds to one free spectral range at a wavelength of 600nm. As it is desirable to scan the etalon over approximately two free spectral ranges, a potential in excess of 1000 volts must be applied. This is inconveniently high for a linear amplification system and carries an increased risk of electrical breakdown.

It is possible to build a stack of piezo-electric discs of alternating polarity. By connecting sides of the same polarity in parallel, the extension per volt of the stack is increased in proportion to the number of discs. The etalon may then be driven by a supply of around 300 volts.

Three types of transducer/sensor assemblies have been used in the UCL etalons. They are shown in figure 11.4.

The Mk I transducer design is described in *Rees, McWhirter et al (1981b)*. It consists of a co-axial configuration, with a rod of piezo-electric material around which is mounted a hollow cylinder. This cylinder is split half way along its length, so that one half contacts to each etalon plate with a small gap in between. The opposing faces of the two halves are metallised to form the two plates of the capacitor.

The Mk II design was the first attempt at using stacked transducers. The need to make an electrical connection to the individual elements meant that the co-axial approach of the Mk I could not be used. Instead, each piezo-electric stack was flanked by two equal capacitance sensors. This arrangement was, by comparison, somewhat inelegant, and its appearance caused it to be designated the "Stonehenge" design.

The Mk III design was a return to the original concept, except that the capacitance sensor was on the inside and the piezo-electric transducers on the outside. This has proved to be very successful and is currently still in use. An etalon of this type is shown, in its kinematic mount, in figure 11.5. This is one of the latest designs of etalon, with a plate diameter of 140mm and a usable aperture of 104mm.

In addition to the driven capacitors, three reference capacitors were mounted on the etalon next to the transducers. They were similar in construction to the driven capacitors, and were used for the second capacitor in the micrometer capacitance bridge. This meant that any changes in the driven capacitor values due to variations in temperature or pressure would be cancelled out by equal changes in the reference capacitors.

11.5 The balloon-borne electronics system

The system used in the balloon-borne interferometer is described in *Rees, McWhirter et al (1982a)*. The low and medium resolution etalons

were both tuneable. Three capacitance micrometers were used for each etalon to determine the cavity length. These each produced an analogue output ranging from zero to five volts. This voltage was then compared in an error integrator with a command voltage derived from a digital to analogue converter driven by the host computer. This was an 8085 single board computer designed especially for the balloon flight.

In order to minimise interference to the sensitive capacitance bridge input, the circuit was built on an annular circuit board surrounding the etalon. This is shown in figure 11.6. As the etalons used for this version had an outside diameter of only 75mm, the result was a very neat unit which fitted well on the optical bench. However, when the design was adapted for the 132mm diameter etalons used for the 1983 balloon flight, it became necessary to use rectangular boards mounted against the outer sides of the etalon housing.

Two 75mm capacitance stabilised etalons of this design were built at the APL for the University of Michigan as prototypes for the High Resolution Doppler Imager (HRDI) on the Upper Atmosphere Research Satellite (UARS). The Space Shuttle disaster meant that the launch of this satellite was delayed by some years. It finally reached orbit in 1991 and results have just been published (*Hays et al, 1992; Hays et al, 1993*). The spectral response of the HRDI triple etalon system is examined in chapter twelve.

11.6 The ground-based electronics system

After the balloon flights in 1980 and 1983 the design philosophy was changed as the requirement was now for a ground-based instrument. The use of the larger diameter etalons, introduced for the 1983 flight, meant that it was no longer such an attractive solution to place the electronics around the etalon because space was much more restricted. Furthermore, it was desirable to be able to adjust the electronics once the etalon was in its sealed, temperature-stabilised housing. For the balloon flight, the entire instrument, including the electronics, was, of necessity, enclosed in a sealed housing.

It was decided to house all the electronics in a single box, remote from the etalon, with screened cables to carry the capacitor signals and the high voltage for the PZTs. The unit is illustrated in figure 11.7 a & b. This approach proved to be quite satisfactory and the system has been operated with up to four metres of cable separating the etalon and the electronic unit. The introduction of the Mk II and Mk III transducer/capacitor designs meant that the high voltage amplifiers could be re-designed for a much lower maximum output, in the region of 300 volts.

11.6.1 Improvements in servo design

In the balloon system, the error signal was derived by subtracting the output of the capacitance micrometer from the d.c. command voltage, using a simple summing amplifier. However, this approach meant that any variations in the gain of the micrometer amplifiers and phase sensitive detectors would directly affect the etalon gap.

This dependence was dramatically reduced by placing the summing point for the command signal directly at the junction of the driven and reference capacitors. This required the reference signal to be an a.c. signal derived from the master oscillator. It was obtained by using a multiplying digital to analogue converter (MDAC), which is, in effect, a digitally controlled amplifier. A similar approach was originally used by *Hicks et al (1976)* who used a combination of a digital to analogue converter and an analogue multiplier. The computer command for the etalon position was input to the MDAC and the resulting command signal was applied directly as an offset to the bridge. The servo system was arranged so that the high voltage applied to the transducers caused the driven capacitor, C_{etalon}, to change in value until the bridge was once again balanced. By adopting this approach the effect of gain variations in the micrometer circuit was reduced by a factor equal to the loop gain of the system.

The block diagrams shown in figures 11.8 and 11.9 illustrate the difference between the two systems. The full circuit diagram of the ground-based unit is shown in figure 11.10.

11.6.2 Other design improvements

In addition to the revised control philosophy outlined above, a number of substantial improvements were effected in the following areas.

Oscillator and capacitance bridge: the oscillator of the balloon-borne system was based on a hand wound transformer - one for each micrometer. These were very tedious and time consuming to make in any quantity. It was possible to dispense with the transformers by using a Wien bridge oscillator, an RC oscillator noted for its low distortion. The phase reversal, formally achieved with anti-phase transformer tappings, was effected with operational amplifier invertors.

Phase sensitive detector: the phase sensitive detector also used an improved analogue switch and a combined inverting/non-inverting amplifier in order to effect a full wave PSD output. The previous version of the electronics used a half wave detector, which discarded the negative half of each cycle.

Output filter: the two pole Butterworth filter on the output of the PSDs was replaced by a single pole filter with a switchable time constant. This had the advantage that a dominant pole, required for optimum stability, was defined in the servo loop. This is described further in paragraph 11.11. The switchable time constant enabled the system response time to be optimised whilst observing the response to a step in the command signal.

High voltage amplifier: the original high voltage amplifiers for the Mk I transducers incorporated Brandenburg high voltage modules, and were really voltage controlled power supplies rather than amplifiers. The stacked Mk II and Mk III transducers meant that they could be replaced by conventional high voltage amplifiers powered by a single

high voltage d.c.- d.c. converter. These had a much improved noise level and dynamic response. The circuit diagram of the ground-based amplifier is shown in figure 11.11.

Control status indicators: the high voltage amplifier has an output swing of approximately -15 to +300 volts. The electronics are adjusted so that the full excursion of the digital offset command will correspond to transducer voltages ranging from approximately +50 to +250 volts. This provides a safety margin of some tens of volts before the amplifiers limit. However, changes in the transducer characteristics or the mechanical mounting could create a demand for voltages outside the available range. In this case the high voltage amplifier limits, and the servo control is lost. It is essential to be aware of this situation. In the event of a high voltage limit, the output of the phase sensitive detector will assume a value far outside its normal range of values as it tries in vain to direct the high voltage amplifier to a voltage in excess of its capability. The abnormal value is detected and used to illuminate a bi-colour light-emitting diode to provide a warning light indicating that electronic re-adjustment is required. The warning light is red for a high voltage limit and green for a low voltage limit.

11.7 Adjustment of the ground-based unit

As seen in figure 11.7, the front panel has connectors for the etalon signals, comprising a common reference capacitor input, the driven capacitor inputs, and the outputs of the driven and reference capacitor junctions. Three PET 100 connectors provide the high voltage connection to the PZTs. Test outputs are provided for the three PSD outputs, both before and after filtering. Ten multi-turn potentiometers are accessed through holes in the front panel. One is for phase adjustment of the PSD switching and the others are the adjustments for capacitive offset, resistive offset and gain for each channel. The capacitive offset provides an adjustment of approximately ±0.5pF and the resistive offset ±100 MΩ.

A detailed instruction manual has been written for adjustment and operation of the unit. The adjustment procedure consists of balancing the bridge for capacitive and resistive offsets, checking the switching point of the phase sensitive detectors and adjusting the gain of the command offset for each channel.

The adjustment procedure may be briefly summarised as follows:

1. With the high voltage switched off, the capacitive and resistive offsets are adjusted to give zero PSD output on each channel.
2. A pre-set amount of capacitive offset is introduced to ensure that the high voltage amplifier will be approximately 50 volts for zero command input.
3. The phase control is adjusted, if necessary, to ensure that the PSD outputs are switching at the zero voltage crossover.
4. The high voltage is turned on and it is observed that the PSD output falls to near zero.
5. The etalon is illuminated with diffused coherent light and slight adjustments are made to the capacitive offsets whilst observing the fringes to achieve parallelism.
6. A digital offset is applied and the gain controls adjusted to obtain parallelism at the upper end of the command range.
7. When the etalon has been adjusted by eye, the parallelism can be optimised by adjusting the individual leg offsets whilst observing the transmission using an appropriate photometric method.

11.8 Computer control of the ground-based system

The capacitance bridge is offset by a small amount determined by a digital command provided by the computer or the etalon driver unit. There are two methods of control, 8-bit and 10-bit.

The 8-bit control is effected by 24 parallel lines (3 x 8 bits) and uses one 24 line Tecmar port for each etalon. The Halo software reserves Tecmar ports 1 and 3 for controlling two etalons with the 8-bit "Etalegs" programme. This is a very easy method as the data is simply set up on the port output for each leg as required. A hardware

control box has also been built from which the etalons may be driven by push-buttons. The 8-bit mode is, however, very uneconomical in its use of the data lines.

The 10-bit control is a later development and offers not only improved resolution but also uses multiplexing techniques to control up to three etalons using just one 24 line port. To ensure backwards compatibility and to retain the facility for the simple 8-bit option, the main etalon controller board is identical for the 8-bit and 10-bit options. In order to effect 10-bit operation a small multiplexer board is fitted between the mother board and the 40 way connector on the rear panel. This connects the 8-bit inputs in parallel. In addition, it activates the two least significant bits required for 10-bit operation and the chip select lines for the three MDACs.

A jumper on the multiplexer board selects the etalon number (1,2 or 3) and the 10-bit "Decalegs" programme addresses port 1 of the Tecmar board. Four bits are used to select etalon and leg number, and ten bits supply the positional data. A single line is then toggled to latch in the data. For multiple etalon systems the 40 way cable may be daisy-chained to a maximum of three etalon controllers. The circuit diagram of the multiplexer is shown in figure 11.12.

11.9 Mathematical analysis

It is important to determine the way in which the etalon gap varies with respect to the digital command voltage.

Referring to figure 11.9 and defining the input to the reference capacitor, C_{ref} as V_1, the input to the driven capacitor C_{et} is the sum of the following signals:

1. The inverted input signal, $-V_1$.
2. The offset applied to balance the bridge, $K.V_1$.
3. The offset applied by the digital command, $nV_1/1023$, where n is the value of the digital command from 0 to 1023.

The conditions for balance are:

$$V_1C_{ref} - V_1C_{et} - \frac{n}{1023} \cdot V_1C_{et} - KV_1C_{et} = 0 \qquad [11.3]$$

Substituting $A_c\epsilon_0\epsilon_r/d$ for C_{et}, where A_c is the area of the driven capacitor plates and d their separation gives:

$$A_c\ \epsilon_0\ \epsilon_R\ V_1 \left[1 + A + \frac{n}{1023} \right] = d(V_1C_{REF})$$

and:

$$d = \frac{n}{1023} \cdot \frac{\epsilon_0\epsilon_R A_c}{C_{REF}} + (1+K) \cdot \frac{\epsilon_0\epsilon_R A_R}{C_{REF}} \qquad [11.4]$$

This demonstrates that the change in etalon gap is linear with respect to the digital command, an important requirement for many applications.

11.10 Loop gain

The accuracy of a servo system is determined by the loop gain. This is the total gain around the closed loop, given by the product P.S.G, where P is the piezo-electric transducer extension in nanometres per volt, S is the capacitance bridge sensitivity in picoamps/nanometre, and G is the electronic gain in volts per picoamp.

The piezo-electric material is type PZT-5H, manufactured by Vernitron. The extension of a single piece is 0.6 nm/volt, so the six layer stacks used in the Mk III etalons have an extension, P, of 3.6nm/volt.

The capacitors have a value of approximately 15pF, which represents an impedance of 660k at 16kHz. The 16kHz input to the bridge has a peak amplitude of four volts, so the peak current is 6μA. The capacitor gap is nominally 30μm. The current through the capacitors is a linear function of gap, so the bridge sensitivity, S, is 200 pA/nm.

The electronic gain from the bridge amplifier to the PSD output is 125μv/pA (including conversion from peak to average in the demodulator), and the high voltage amplifier gain is 1000. The integrating amplifier has a frequency dependent gain, G_I. The electronic gain is, therefore, $0.125G_I$ volts/pA.

The total closed loop gain, P.S.G is, therefore, 3.6 x 200 x 0.125GI, which comes to $90G_I$. In the d.c. limit, the integrating amplifier has a gain of 470, which gives a total d.c. loop gain of 42,300.

This figure will dictate the repeatibility of the servo loop, that is to say, the ability of the etalon to return to exactly the same position for a particular digital command.

11.11 Dynamic response and servo stability

The etalon plates can be moved to a new position in approximately ten milliseconds. The maximum speed at which a servo mechanism can operate is determined by the gain-phase characteristics of the system. Any time constant in the system will introduce a frequency dependent phase lag in the feedback loop. Negative feedback involves feeding back an error signal which is 180 degrees out of phase with the input. However, this is only achieved below a certain cut-off frequency. Above this point the phase shift around the loop will generally increase. There will be a frequency for which the total phase shift becomes 360 degrees. The feedback will, therefore, be in phase, constituting positive feedback. If the loop gain at this frequency is greater than unity the system will oscillate.

The phase response of the etalon system depends on the mass of the etalon plates, the force applied by the piezo-electric transducers, and the characteristics of the clamping springs on the etalon mount. These parameters are difficult to define and predict, and the phase characteristics will be complex. In order to ensure stability, the electronic feedback loop should contain a dominant pole. This is a simple RC time constant which introduces a 90 degree phase lag above

its cut-off frequency and has a gain-frequency response which rolls off at 6dB per octave, or 20dB (10x) per decade. The time constant is selected so that the loop gain is less than unity at the frequencies for which the feedback is in phase.

The dynamic response may be assessed by observing the filtered outputs of the phase sensitive detectors when responding to a step input in the command signal. However, it should be remembered that the error signal of the servo system is being observed. This is generated by the very small difference between the commanded position and the actual position. The observed response will, therefore, reflect not only the actual position of the etalon plates, but also any slight change in the transfer function of the transducers (in terms of extension per volt) as they settle into the new position. The PSD filtered outputs are, in fact, a direct measure of the applied high voltage because they are used as the inputs to the high voltage amplifiers. Nevertheless, the filtered outputs provide a useful indication of the performance of the servo system. The rise-time may be taken to be a good indication of the response speed of the etalon plates, but the observed settling time will reflect not only the settling of the physical position of the plates, but also any short term variations of the transducer response induced by the sudden movement.

The time constant is introduced by IC10 which is an integrating amplifier. The d.c. gain is equal to the ratio R22/R23, which is 470, and the a.c. gain is defined by the product of R22 and the selected capacitor (C13 to C16). Four RC time constants are available, ranging from 10ms to 0.22s. An analogue switch is used to select the appropriate capacitor. The effect of this time constant on the circuit response time is reduced by the loop gain of the system outside the integrating amplifier which, as seen in the previous paragraph, is calculated to be 91. The time constant is adjusted by experiment for optimum response time, which is observed just prior to the onset of ringing. The loop gain at this point will be just less than unity at the natural frequency of oscillation, and will increase with

decreasing frequency at 20dB per decade until, in the d.c. limit, the maximum gain of 42,300 is reached.

11.12 Temperature and pressure stability

Three areas affect the stability of the CSE system:

1. The thermal response of the physical gap of the etalon. This depends on thermal expansion of the pillars carrying the capacitance sensors or, if a stepped etalon is formed using a third plate, the differential expansion between the sensors and this plate.
2. Variations in the refractive index of the cavity medium due to pressure and temperature change.
3. The thermal stability of the electronics.

A number of capacitance stabilised etalon systems have been built by the author for use by other institutions. One of these was a 150mm gap etalon built for the University of Bonn, Germany. This was supplied as part of a laser wavelength meter, which used a linear photodiode array bisecting the fringe pattern to measure the diameter of the fringes. It was built for monitoring the lasers used for atmospheric lidar (the ALOMAR project).

The production of this instrument coincided with the acquisition of a single longitudinal-mode stabilised laser. This is a Helium-Neon laser which is thermally controlled and frequency stabilised to give a single line output. A normal, unstabilised laser will output a number of closely spaced lines over a bandwidth of about 1.4GHz which drift in frequency within the bandwidth of the laser. The new laser enabled the stability of the system to be measured to a high degree of accuracy. The 150mm cavity length corresponds to approximately 150 orders of interference per atmosphere of pressure change, so this provided a highly critical test of pressure stability. A series of tests was performed (*Meredith and Mugleston 1993, APL internal report*), using the laser wavelength meter software developed at UCL for determining the fringe peak position from the diode array data.

11.12.1 The physical cavity length

The capacitance sensor assembly is made of silica, which is not as satisfactory thermally as Zerodur, from which the fixed cavity etalon spacers are made. However, if the etalon is stepped in construction, with a central element contacted between the outer plates to form the cavity, then this will have nearly the same coefficient of expansion as the capacitor pillars and the system will be to some extent self compensating (*Rees, McWhirter et al, 1981b*). This will not be so for large cavity etalons which do not have the third inner element. In this case, improved performance may be achieved by fabricating the capacitor pillars from Zerodur. For single etalon systems it is difficult to separate the effect of physical cavity length change from the variations in refractive index of the medium. It will be seen in chapter twelve that the effects can become separated for multiple etalon systems.

11.12.2 Stability of the refractive medium

The variation of refractive index of the cavity medium accounts for most of the observed thermal drift. The etalon housing cannot be evacuated because this would cause high voltage breakdown. During the tests the temperature dependence of the refractive index became immediately apparent. Initial measurements were taken with the temperature controller inadvertently operating in on-off mode rather than PID mode. The hysteresis of the controller provided a temperature oscillation of ±1°C and the sawtooth form of the resulting plot of fringe position versus time provided a dramatic demonstration of the thermal response. This shows the importance of ensuring that the temperature controller is properly adjusted and tuned.

The thermal time constant of the heavy aluminium pressure-sealed housing was found to be much shorter than had been expected. The PID controller operates by varying the mark-space ratio of the power switched to the heating coil. The cycle length must be short in comparison with the response time of the can. A length of twenty seconds

was found to be too long and a significant drift in peak position during this period was observed. A two second cycle length was found to be suitable. Once proper temperature control had been established the short term stability of the device was found to be satisfactory. A typical test run is shown in figure 11.13. The peak variation of approximately 0.2 diodes corresponds to 0.3% of a free spectral range.

It was then necessary to assess the integrity of the sealed housing. This is a potential problem area, not only because of the O-ring seals of the housing itself, but also because of the ten bulkhead connectors required for the etalon signals and high voltage leads. The etalon was tested at pressures of 0.5, 1.0 and 1.5 atmospheres, using oxygen-free nitrogen as the refractive medium. The data showed long term drifts of a somewhat haphazard nature. Testing on a helium leak detector failed to locate a leak. It was concluded that the changes were probably due to internal pressure equalisation owing to pockets of gas trapped inside the housing. The most likely place for this to occur is in the co-axial cables carrying the etalon signals, although there could possibly be some porosity of the inner surface of the housing itself. There may also be outgassing from surface contamination or from constituents of the insulation. The testing was curtailed as the instrument had to be shipped to the Alomar team in Germany, who now report excellent performance.

11.12.3 Thermal stability of the electronics

The electronic system has been designed to be substantially immune to temperature change. The design ensures that very few areas of the circuit are gain critical and any drift in such areas is minimised by the use of high stability components. The temperature dependence of the electronics was tested as part of the evaluation of the triple etalon interferometer described in the next chapter. The test was effected by operating the electronic unit in a different environment from that of the optical instrument. No change in etalon peak position was detected over a ten degree change in the ambient temperature surrounding the unit.

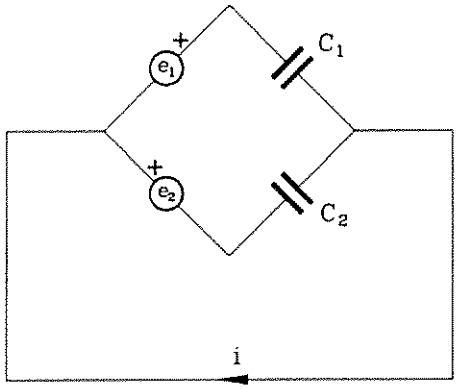


Figure 11.1

The capacitance bridge

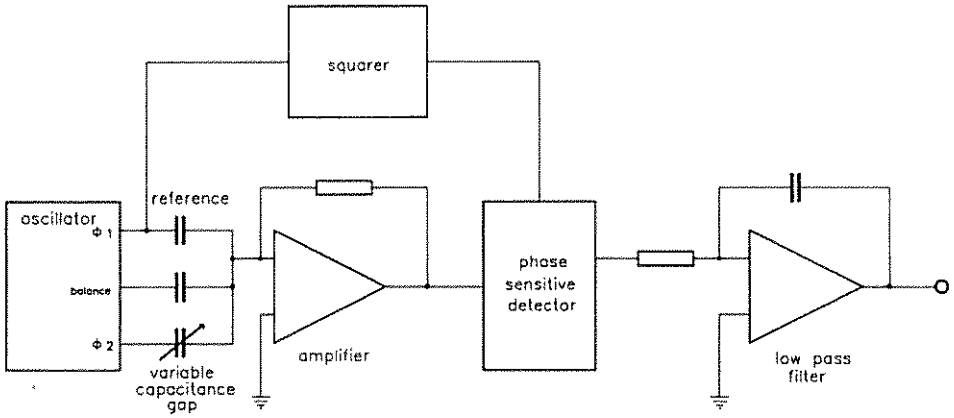


Figure 11.2

The capacitance micrometer

Figure 11.3

Capacitance micrometer waveforms

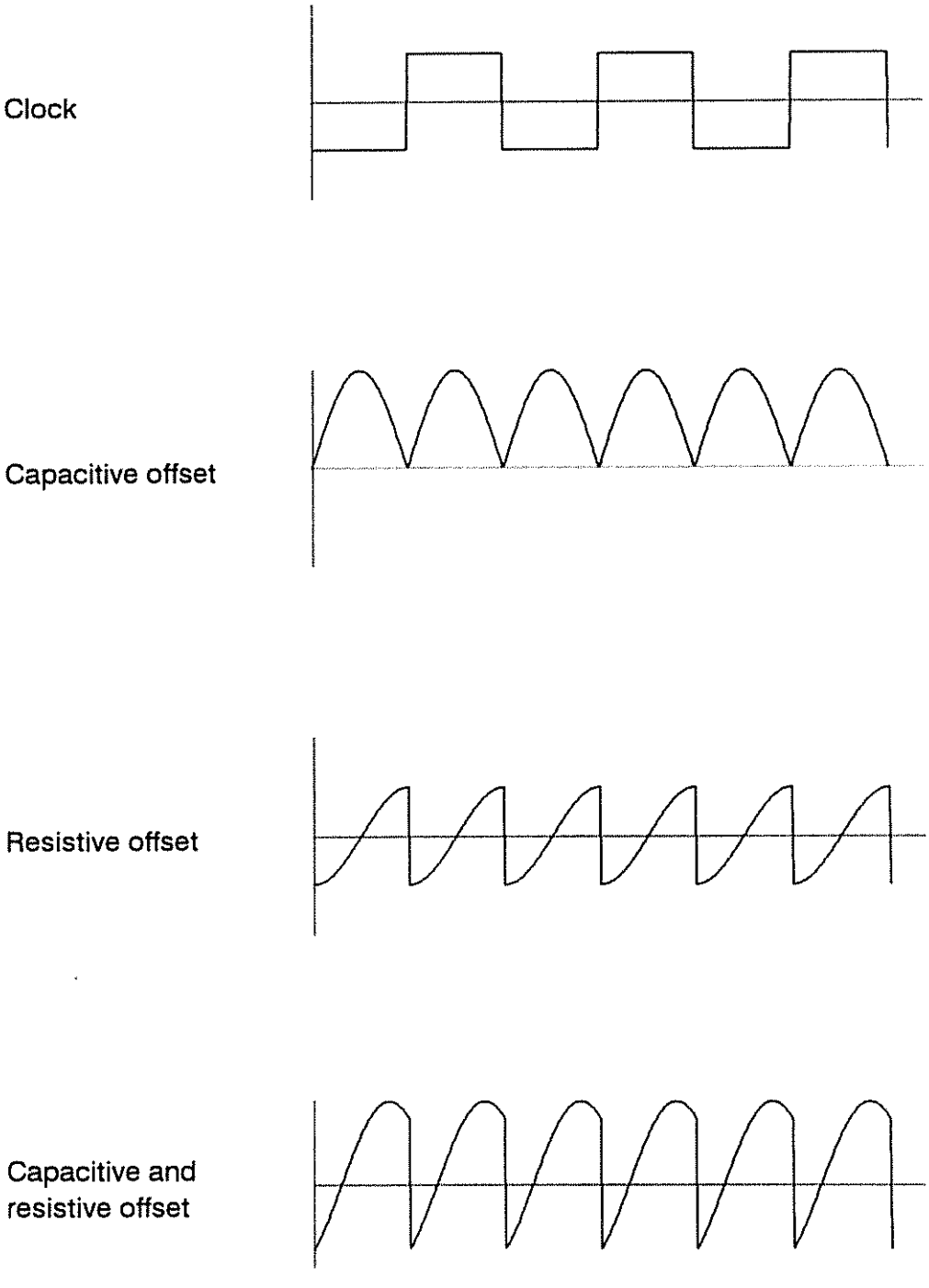
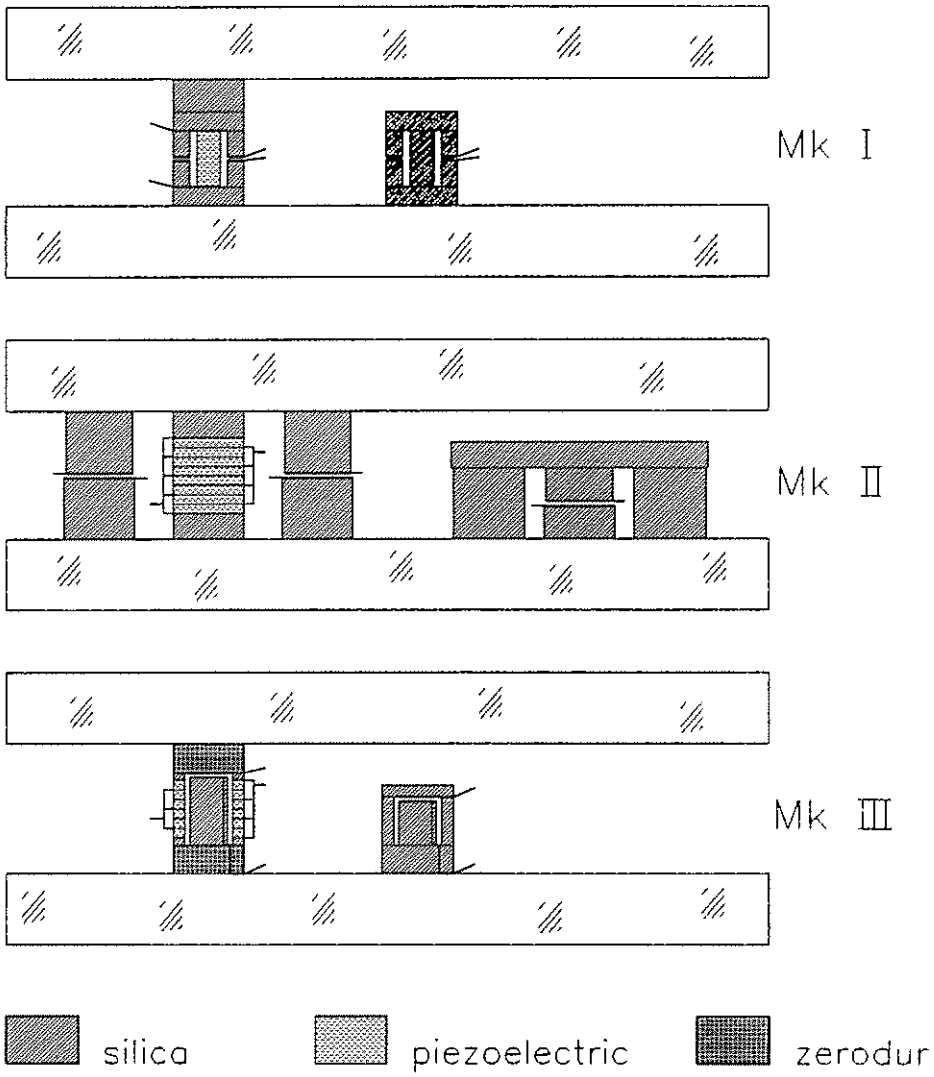


Figure 11.4

Piezoelectric transducer assemblies for the Capacitance Stabilised Etalons



The three types of transducer/capacitor assemblies are shown each with its respective reference capacitor to the right.

The connecting wires to the capacitors and the piezoelectric transducers are shown.

The Mk I and Mk III designs are composed of concentric cylinders and are drawn in section

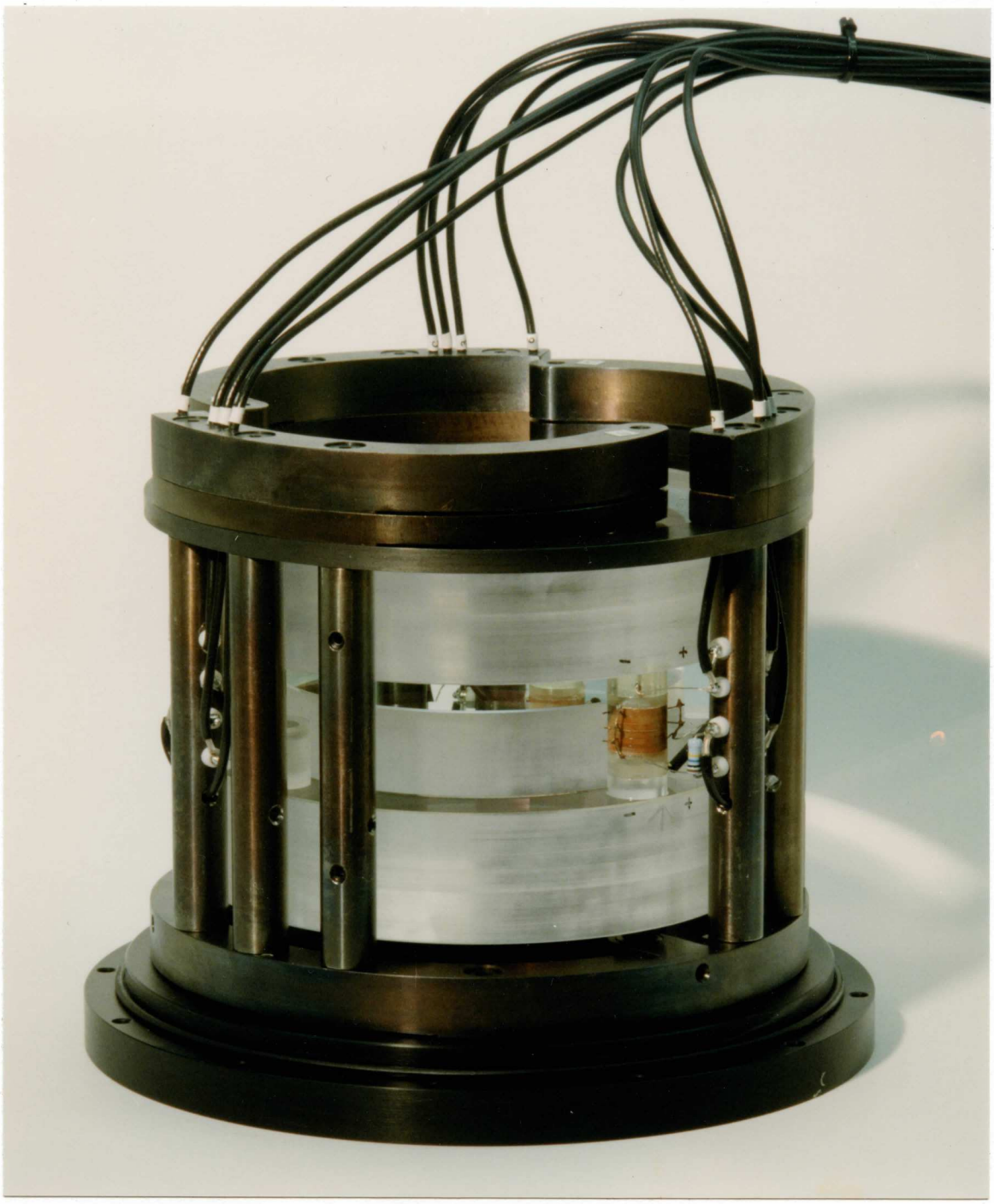


Figure 11.5

A capacitance stabilised etalon in its kinematic mount

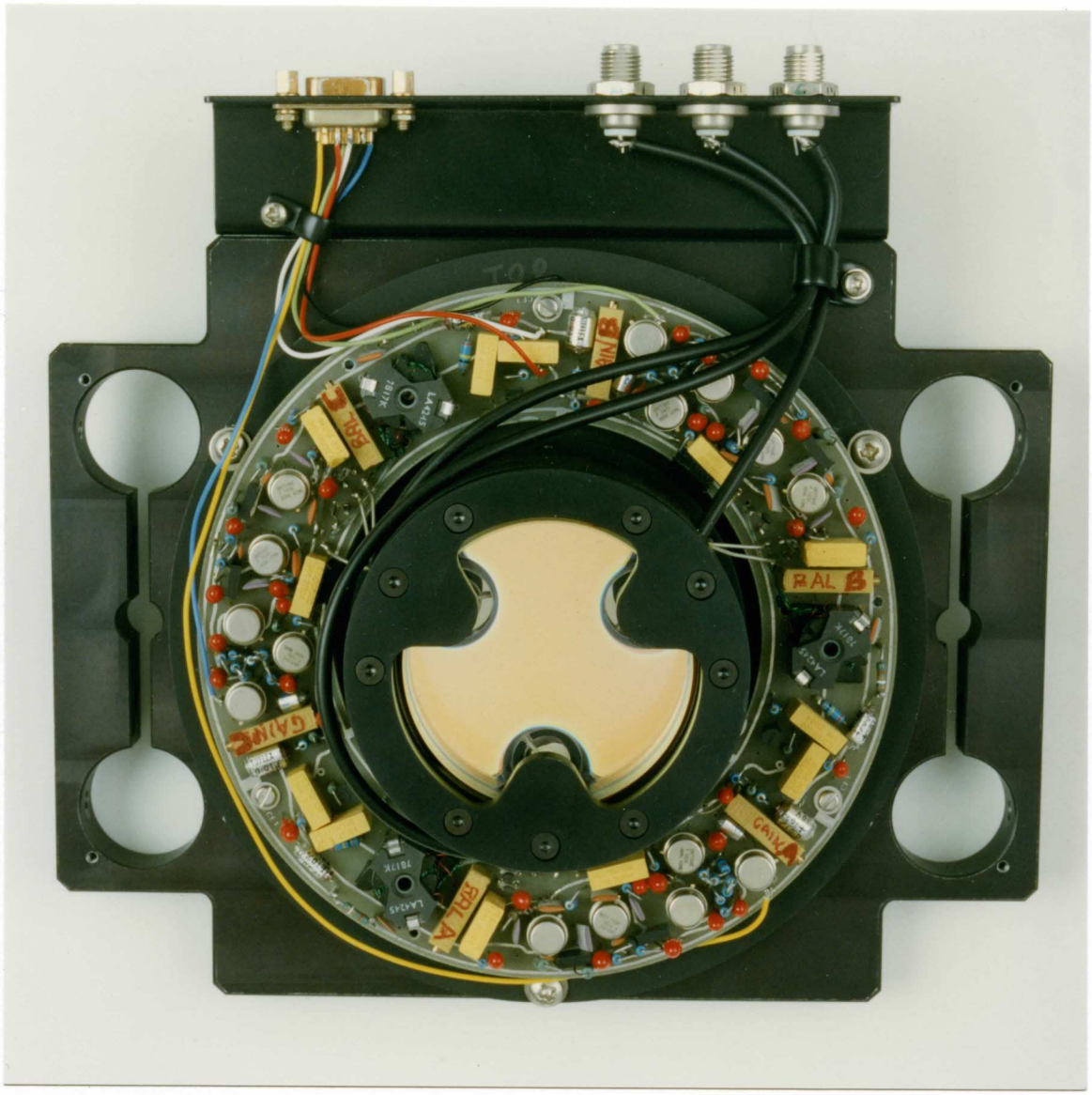


Figure 11.6

The 75mm diameter balloon etalon
with its capacitance micrometers

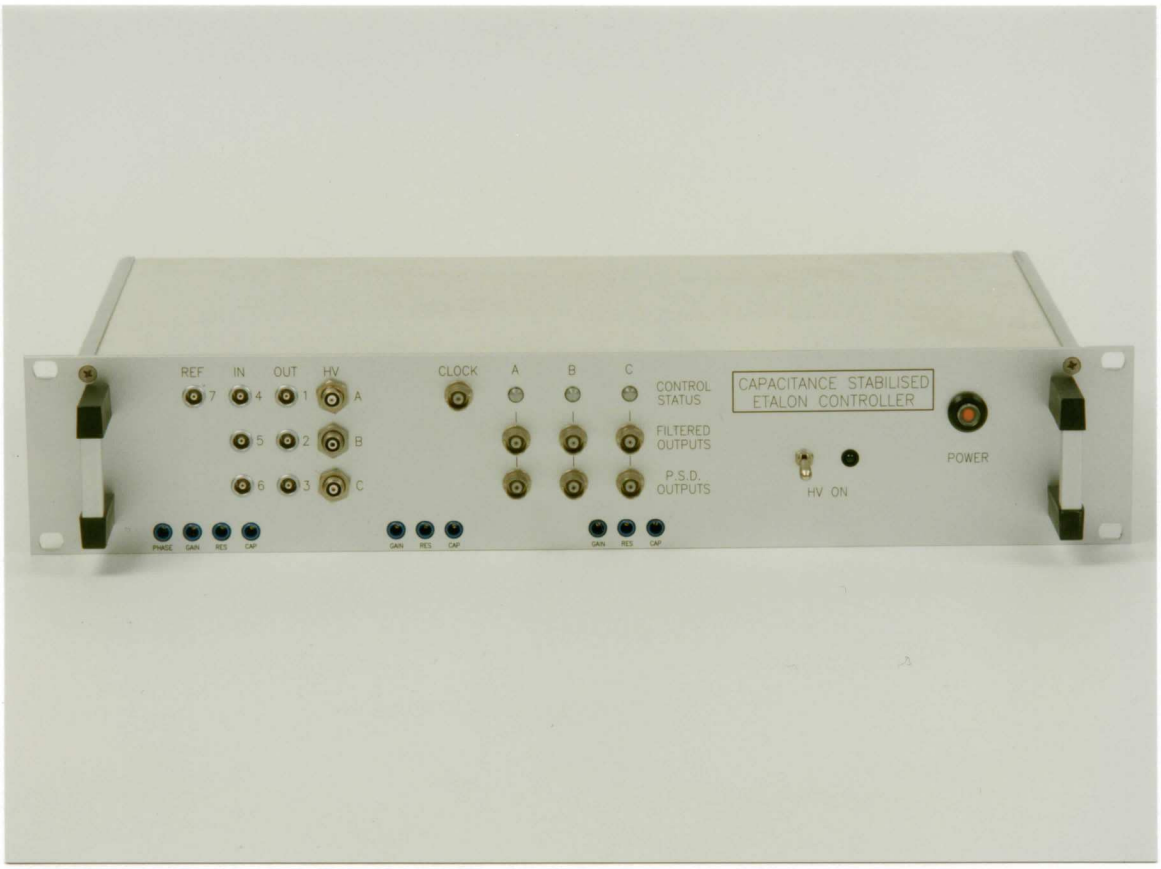


Figure 11.7a The Capacitance Stabilised Etalon Controller

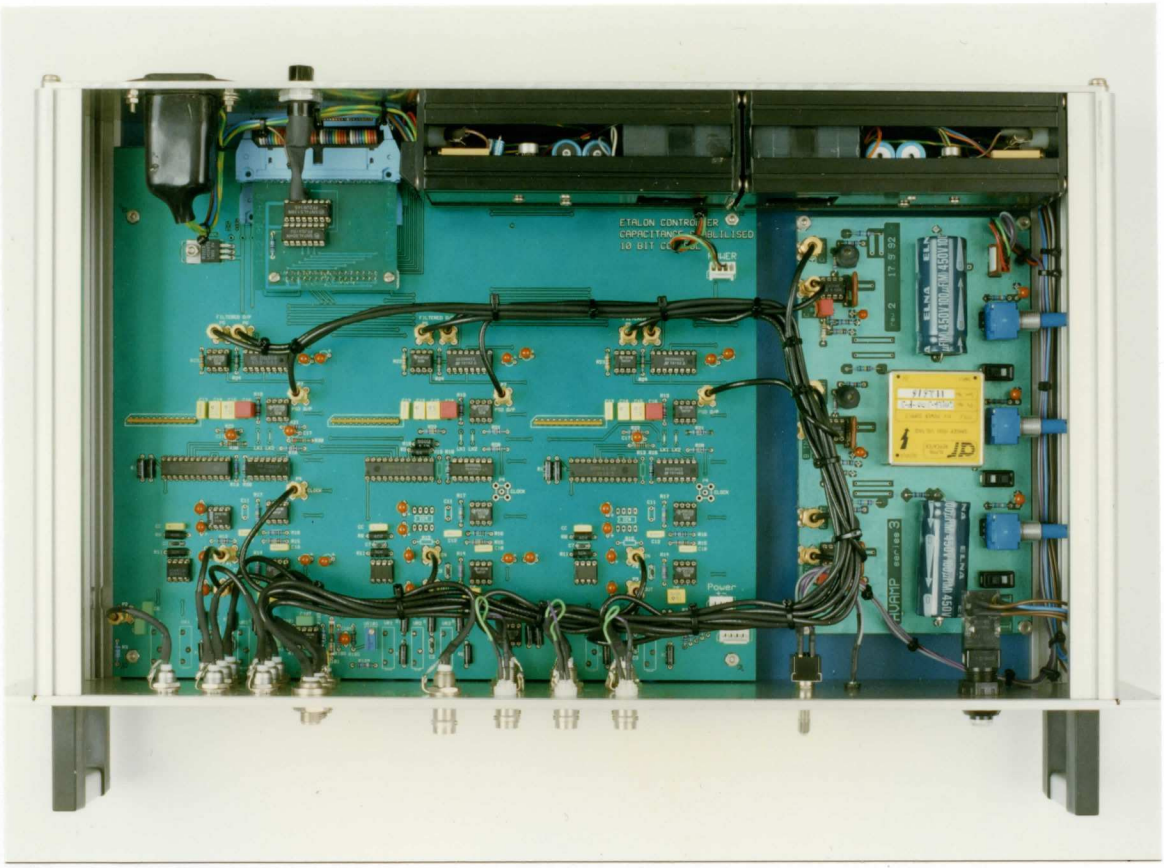


Figure 11.7b The Capacitance Stabilised Etalon Controller

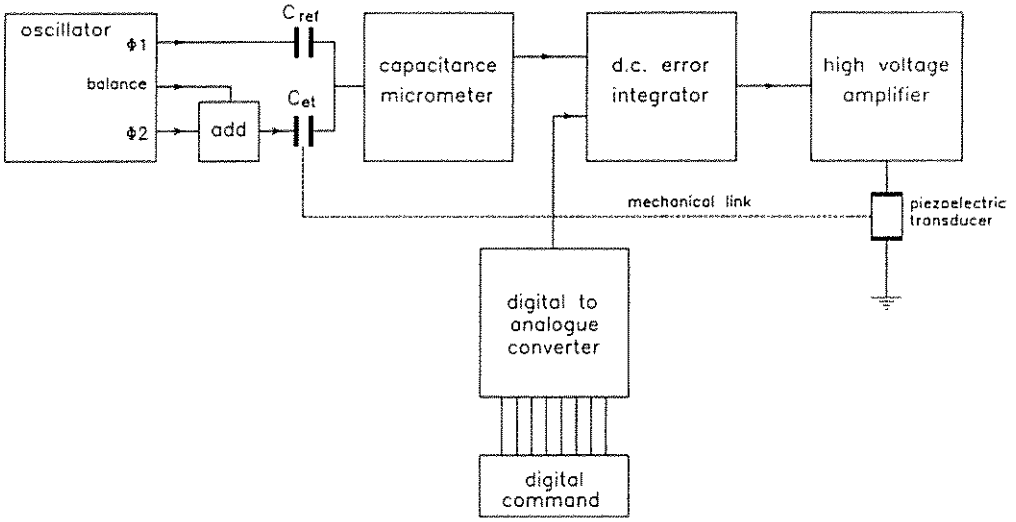


Figure 11.8

CSE electronics – balloon version

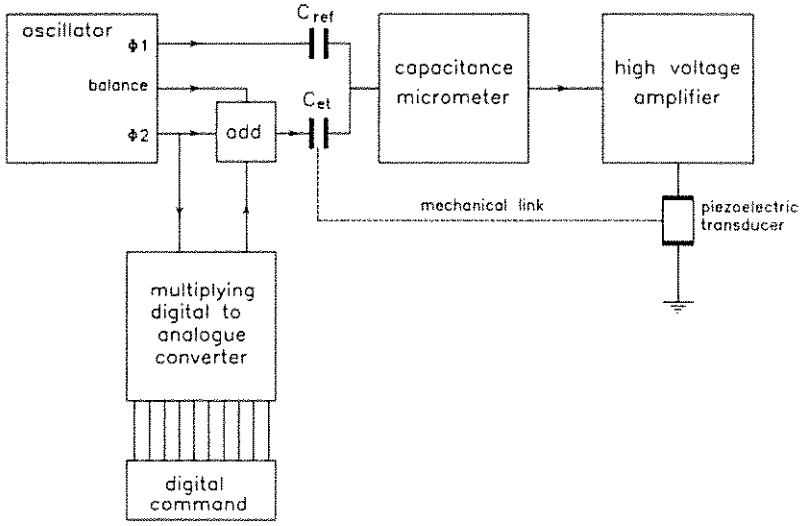


Figure 11.9

CSE electronics – ground based version

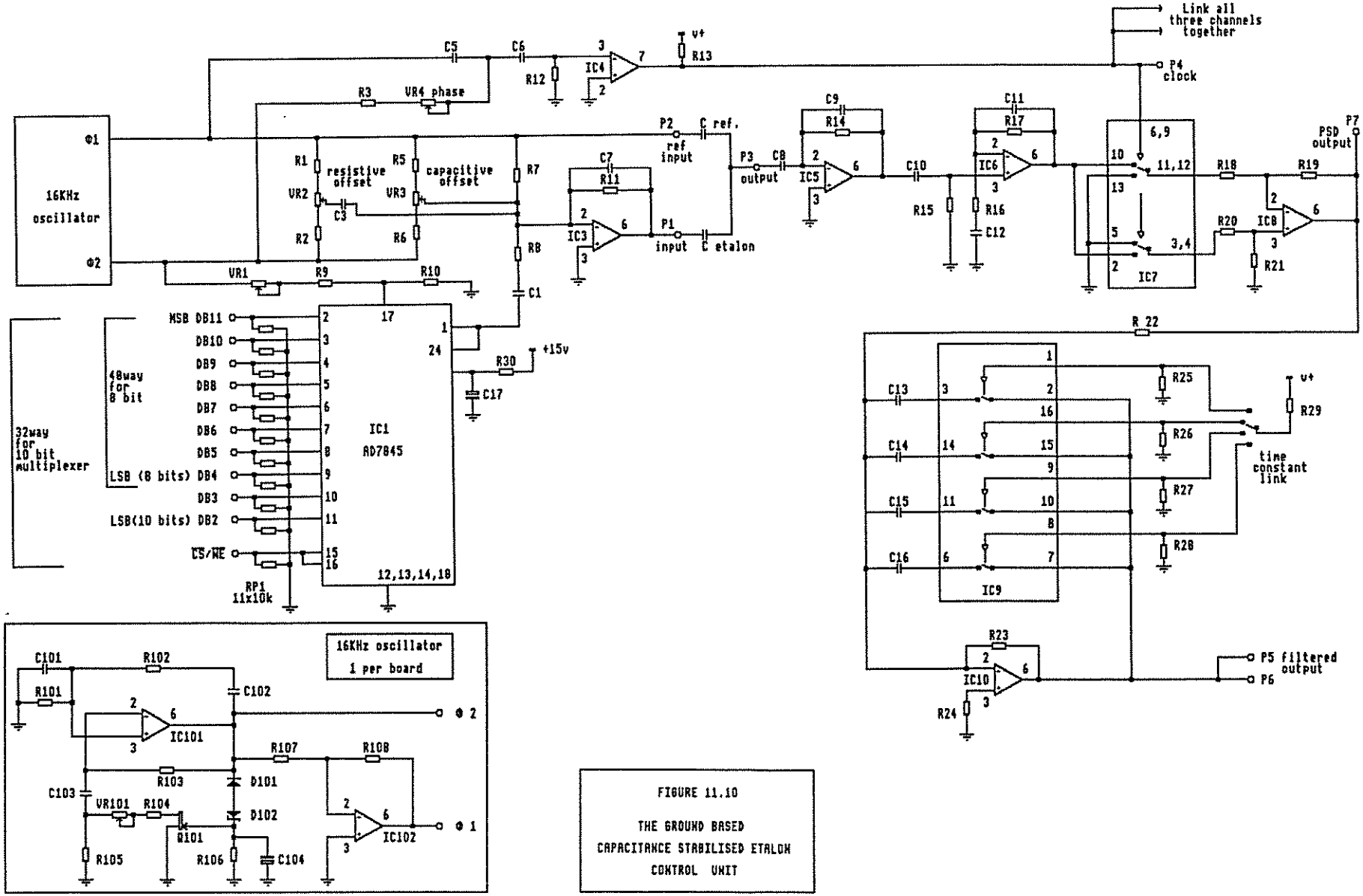


FIGURE 11.10

THE GROUND BASED CAPACITANCE STABILISED ETALON CONTROL UNIT

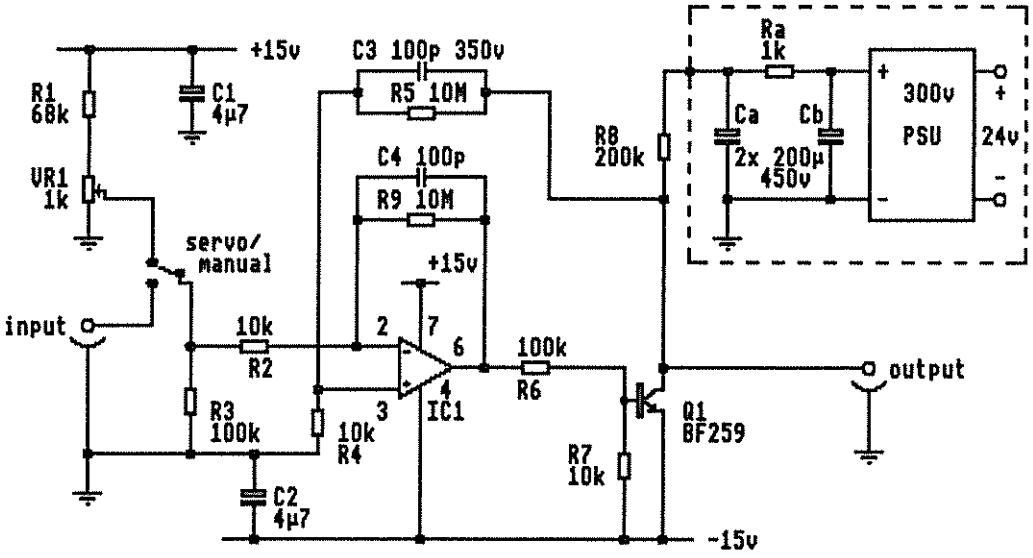


FIGURE 11.11

THE HIGH VOLTAGE AMPLIFIER
FOR THE
CAPACITANCE STABILISED ETALON

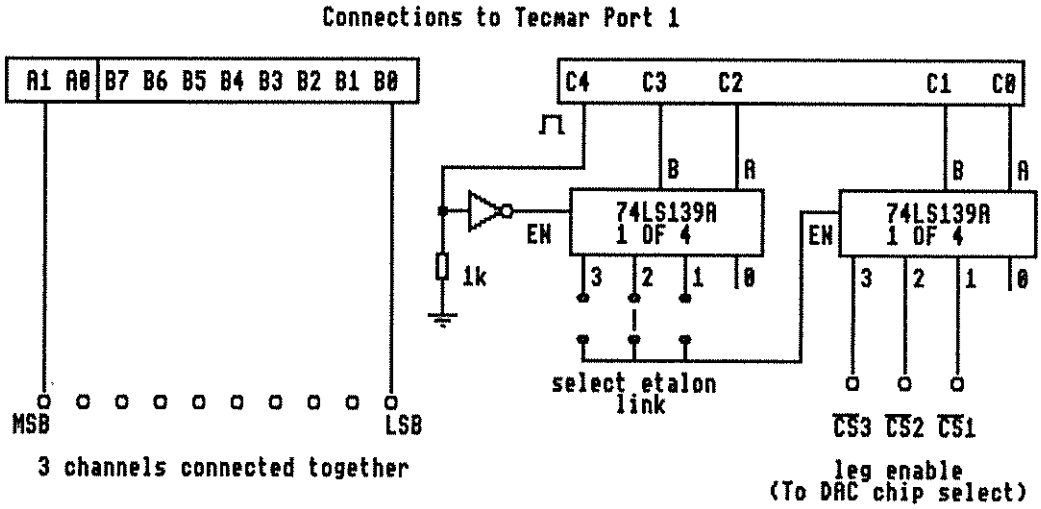
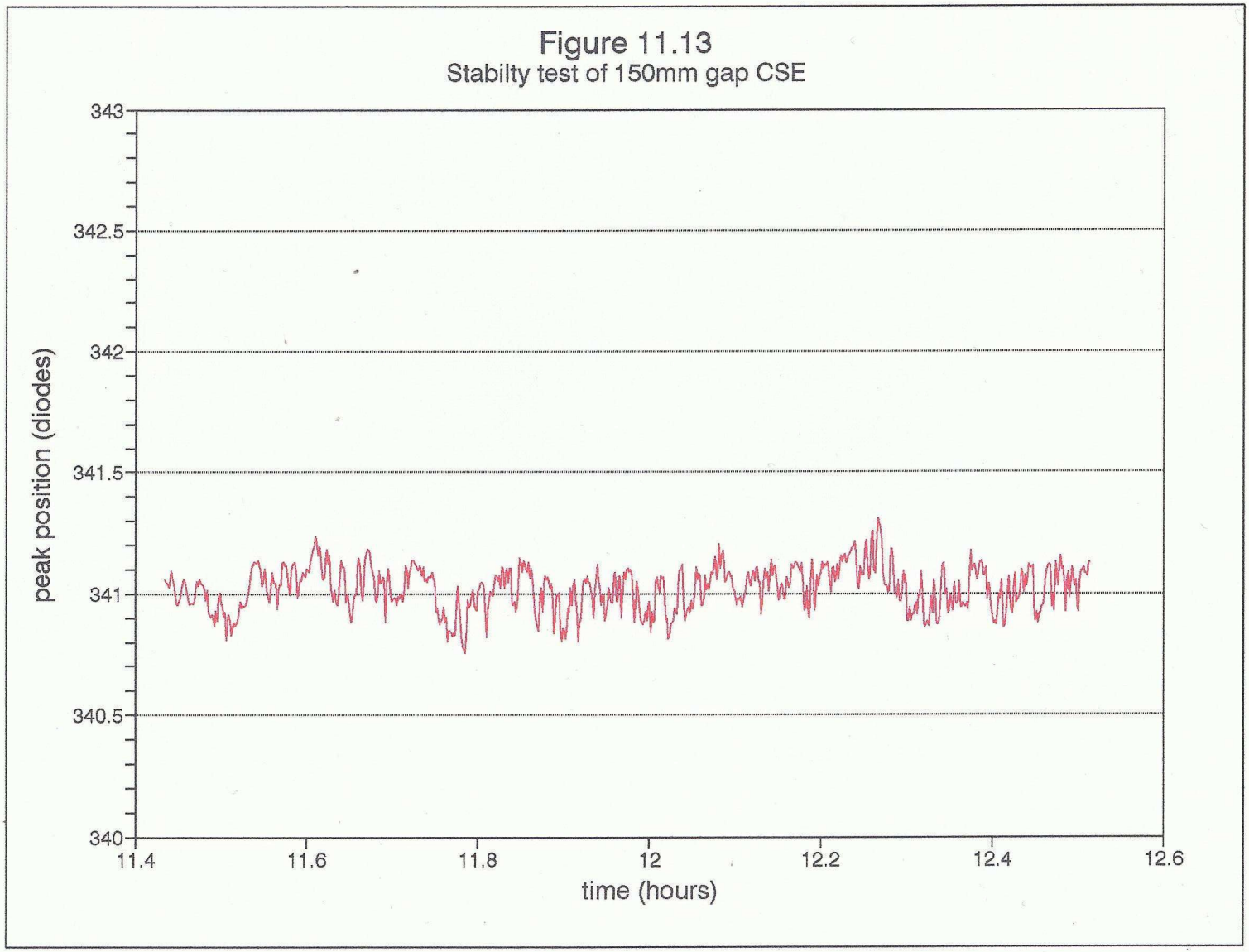


FIGURE 11.12

THE 10-BIT MULTIPLEXER FOR THE
CAPACITANCE STABILISED ETALON

Figure 11.13
Stabilty test of 150mm gap CSE



CHAPTER 12 - THE TRIPLE ETALON INTERFEROMETER

12.1 Introduction

Important work on triple etalon systems was carried out by *Mack et al (1963)* with their PEPSIOS scanning spectrometer. Their instrument consisted of pressure scanned etalons with a pinhole photomultiplier detector. The advent of the capacitance stabilised etalon and the Imaging Photon Detector has now enabled a new class of instrument to be realised.

In this chapter the general concepts of multiple etalon systems are considered. The UCL triple etalon interferometers are described, and the performance limitations are explored quantitatively. The theoretical response and stability criteria are assessed using a computer simulation for comparison with the experimental data. Other instruments discussed are the double etalon interferometers built at UCL for the ALOMAR atmospheric lidar project, and the High Resolution Doppler Imager on the Upper Atmosphere Research Satellite (*Hays et al, 1992 & 1993*), for which the author provided prototype capacitance stabilised etalon systems.

12.2 Multiple etalon systems

The single etalon interferometer is an excellent spectral filter. However, if the observed source is contaminated by spectral lines which differ in wavelength by more than a free spectral range from the region of interest, then the unwanted lines may partly overlap the wanted signal. Also, any continuum not restricted by a pre-filter will be superimposed upon the signal, degrading the signal-to-noise ratio. This applies especially to observations of atmospheric emission lines and absorption lines during daytime. The situation may be improved by combining etalons in series. The overall response thus obtained provides much better rejection of unwanted spectral components, that is to say, better contrast.

The transmission function of an etalon, as seen in chapter nine, is given by an Airy function, and may be thought of as a comb filter. The requirement for combining etalons is that transmission maxima of one order from each etalon should coincide to give the instrument pass-band, and the remaining maxima should not coincide at all, thereby giving high attenuation at all other frequencies. This requires very careful selection of the etalon gap ratios. It is clear that if integer ratios are chosen then the pass-bands will coincide at any order which has those integers as factors.

In order to align the transmission functions so that the transmission maxima coincide, it is necessary to tune the etalon cavities as described in chapter eleven. If the experimental configuration permits, one etalon may remain fixed whilst the remainder are varied. The low amplitude fringes formed by partial superposition of the individual etalon peaks are known as parasitic fringes (*Hernandez 1986*). The aim of multiple etalon system design is to minimise the effect of such fringes.

12.3 Double etalon systems

There are two approaches which may be used when selecting the gap ratios of a double etalon system: a low-high resolution arrangement where the free spectral range of one etalon is several times that of the other, and a "vernier" configuration, where the gaps are similar, yet sufficiently different to ensure that the transmission peaks do not coincide except at the chosen order. In practice, the results obtained are comparable. The vernier pair has the advantage that the overall finesse is greater because the transmission profile is the product of two similar high resolution etalons, whereas the finesse of the low-high pair is approximately that of the high resolution etalon alone. However, the vernier pair has parasitic fringes which are very close to the main peak, and the overall number of parasitic fringes is greater. Also, the alignment and stability requirements are more stringent as both etalons have a small angle of acceptance. For these reasons it is very difficult to achieve good results when using an

imaging detector with vernier ratios and wide ratios are used in preference.

A double etalon interferometer with vernier ratios has been built at APL for use at the University of Bonn, Germany, in connection with the ALOMAR lidar project. ALOMAR is an international atmospheric lidar facility currently under construction at Andoya, Norway. Instruments will be provided by Britain (UCL), Germany and France. A photograph of the Bonn double etalon system is shown in figure 12.1 and a computer simulation of the response in figure 12.2. This may be compared with figure 12.3, which shows the predicted response of a large ratio pair currently being built for the UCL ALOMAR instrument.

12.4 Triple etalon systems

The parasitic peaks of the double etalon configuration can be substantially reduced with a triple etalon approach using non-comensurate gaps, that is to say that at no point other than at the chosen order do the three transmission peaks coincide. The gap ratios obviously need to be chosen with care. The optimum ratios for the UCL and HRDI triple etalon interferometers have been determined by simulations (*Skinner et al, 1987;* also *Rees and Hays, private communications*). The etalons are referred to as the low, medium and high resolution etalons (LRE, MRE, and HRE). The response of a triple etalon system is fundamentally the product of the Airy functions for the three etalons. A computer simulation of the response of each etalon, together with the overall response, is shown in figure 12.4, and on an expanded scale in figure 12.5. A logarithmic response is shown in figure 12.6, where the improvement over the double etalon interferometer is clearly seen. It will be shown later in this chapter that several factors conspire to degrade the theoretical performance.

12.5 The UCL triple etalon interferometers

A triple etalon interferometer was developed at UCL initially for observations from a balloon-borne platform launched at Palestine, Texas, in Spring 1980. The instrument was further refined for a second

flight in September 1983 and then re-built as a ground-based instrument for operation in Northern Sweden.

12.5.1 The balloon instrument

The original balloon instrument is described in detail in *Rees, McWhirter et al (1982a), "Observations of atmospheric absorption lines from a stabilised balloon platform and measurements of stratospheric winds"*, together with an account of the scientific objectives of the flight and a presentation of the results. Briefly, the main objective of the flight was to detect Doppler shifts in several species of atmospheric absorption lines. These lines are too broad to be measured when observed from the ground. A photograph of the instrument is shown in figure 12.7.

12.5.2 The ground-based instrument

The ground-based instrument is essentially identical in concept to the scanning mirror FPI described in chapter ten, except that a large sealed housing containing the three etalons replaces the single etalon unit. The improvements made to the capacitance stabilised etalon system, detailed in chapter eleven, enabled the electronics to be stripped away from the optical bench so that the sealed housing now contains only the etalons. The three etalons mounted on their optical bench are shown in figure 12.8 and a drawing of the complete instrument in figure 12.9. An image obtained with a single-frequency stabilised laser is shown in figure 12.10.

The main objective of the ground-based triple etalon interferometer is to observe emission lines, such as 630.0nm OI, which are also measured by the single etalon interferometers. However, the rejection of unwanted light is very much greater than with the single etalon system and measurements can be made during twilight. The goal is to be able to measure winds during full daylight. This is particularly important for high latitude sites where the sun is above the critical -6° elevation for most of the summer months.

The instrument was operated at the Swedish Institute for Space Physics at Kiruna between 1988 and 1990. However, unattended long term operation was not possible because of a number of unresolved problems with the instrument.

The re-location of the Atmospheric Physics Laboratory from the Gower Street site to Riding House Street provided a very considerable improvement in the available facilities for long term instrument testing, so it was decided to return the triple etalon instrument to UCL for rigourous analysis. A comprehensive test programme carried out over several months has provided considerable insight into the nature of the problems.

12.6 The measured response of the ground-based instrument

The performance of the instrument will be degraded by a number of factors. The important parameters to be considered when characterising the triple etalon interferometer are:

1. Filtrage, defined as the ratio of wanted light to total light detected.
2. Contrast or extinction, defined as the ratio of the transmission maximum to the transmission minimum for each etalon.
3. The finesse of the individual etalons.
4. Alignment of the maxima of the individual transmission functions, achieved by adjusting the etalon gaps.
5. Mechanical alignment of the optical axes.

In order to assess the performance of the instrument it is necessary to measure the responses of the individual etalons. The measurement procedure is outlined below and the test results are presented. Subsequent paragraphs then discuss in greater detail the factors affecting instrument performance.

12.6.1 The responses of the LRE and MRE

When the transmission profiles of all three etalons have been aligned as accurately as possible, the individual responses of the LRE and MRE can be obtained. This is done by scanning each etalon in turn through a free spectral range whilst monitoring the transmitted intensity. The measured intensity is then plotted as a function of etalon gap (in terms of CSE steps). This determines the free spectral range of each etalon in terms of CSE command and the FWHM of the profile enables the finesse of each etalon to be calculated. The measured intensity at the mid-point between the peak positions gives the contrast of each etalon.

The measured responses of the LRE and MRE are shown in figures 12.11 and 12.12 respectively, and finesse has been calulated to be 11.5 and 16.5 for the LRE and MRE respectively.

12.6.2 The response of the HRE

The free spectral range of the HRE can only be measured directly if the angular field of view imaged on the detector is sufficient to accommodate two fringes of the HRE. The LRE and MRE must be adjusted to illuminate each fringe in turn. The extinction of the LRE and MRE prevent viewing of the two fringes simultaneously, although it may be possible to do this if the LRE and MRE are deliberately set out of parallel. The HRE free spectral range is expressed in terms of radius squared bins. It is dependent on the focal length of the imaging optics and the diameter of the detector, as explained in paragraph 9.8.2.

Unfortunately, with the optical arrangement available at UCL, namely an 18mm diameter detector and a Cassegrain telescope of approximately 1200mm focal length, it is not possible to display two complete fringes. However, by offsetting the image on the detector, it was possible to obtain an image of half of the ring pattern, in which the three inner rings were visible. By measuring the ring spacing in pixels and converting the values to r^2 space, an approximate value for

the free spectral range of 350 radius squared bins was obtained. The maximum radius that the reduction to radius software can handle is 128 pixels, so the outer rings were out of range of the software routine even though they were visible.

The HRE profile can be assumed to be that of the complete instrument. It is seen in the computer simulation shown in figure 12.5 that the overall instrument profile is only slightly narrower. The measurement of the HRE finesse was greatly simplified by the acquisition of a longitudinal-mode stabilised laser. This is a Helium-Neon laser which is controlled to give a single line output. A normal, unstabilised laser will output a number of closely spaced lines over a bandwidth of about 1.4GHz. The stabilised laser may be assumed to be a single wavelength, so it will not be necessary to de-convolve the spectral bandwidth of the source from the measured instrument response.

The detector used for the tests exhibited a slight elliptical distortion which resulted in a broadening of the "reduced to radius" profile. In order to minimise this broadening effect and obtain the true profile of the etalon it was necessary to examine individual radial segments of the image. This was done by setting the software routine to reduce to 24 sectors. The measured profile of the fringe seen in figure 12.10 is shown in figure 12.13, from which the finesse of the HRE was calculated to be 11.0.

12.6.3 Summary of the etalon characteristics

The parameters of the three etalons are summarised as:

	LRE	MRE	HRE
Usable aperture (mm)	90	90	90
Reflectivity (%)	88	89	83
Finesse	11.5	16.5	11.0
Contrast	8.3	50	-
Free spectral range	644/1024 CSE steps	806/1024 CSE steps	350 r^2 bins

12.7 Examination of the parameters affecting instrument performance

The two most noticeable features of the above table are the poor contrast of the LRE and the poor finesse of the LRE and HRE. The factors affecting these parameters are discussed in the following paragraphs.

12.7.1 Filtrage and contrast

The important advantage of the triple etalon interferometer over a single etalon system is that it has much better filtrage. This is the ratio of wanted light to the total light detected by the entire system. It depends on the spectral distribution of the incident light, the transfer function of the optical system and the spectral response of the detector. The definition of "wanted light" is somewhat arbitrary. It is often defined as the light contained within the half maxima of the main transmission peak.

The bandwidth of the incident light is defined by the 1nm interference filter. The spectral response of the detector may be taken as uniform within this narrow region, but it must be remembered that filters may exhibit some residual transparency at regions far removed from their design wavelength.

The unwanted light within the region accepted by the filter is removed by the etalon system. The efficiency of this is determined not only by the choice of the etalon gap ratios but also by the extinction factors of the etalons both individually and collectively.

The contrast of an ideal etalon is defined as the ratio of the transmission maximum to the transmission minimum, and is given by:

$$\text{extinction} = \frac{(1+R)^2}{(1-R)^2} \qquad [12.1]$$

where R is the plate reflectivity. For a reflectivity of 0.86, as used in the UCL instruments, the expected extinction is in the region of 150. In practice, this value is rarely obtained and a realistic target is for each etalon to attenuate to around 2% at minimum transmission, corresponding to an extinction factor of 50. However, the attenuation of a multiple etalon system is not simply the product of the individual values for each etalon. One important mechanism which degrades the contrast is scattering.

12.7.2 Scattering between the etalons

In order to maximise the contrast of the system it is critically important to minimise scattering between etalons. The etalon is an angular filter and scattering creates angular dispersion. The contrast of the final fringes can be markedly reduced as scattering in the space between adjacent etalons will effectively "fill in" the transmission minima of preceding stages.

Problems were encountered with the UCL triple etalon instrument at Kiruna because the observed contrast of the LRE was unsatisfactory. The transmitted intensity when the LRE was placed at its transmission minimum was 12% of the maximum obtained when all three etalons were aligned. However, the MRE would attenuate to the expected 2%. This is clearly seen in the measured profiles of the LRE and MRE shown in figures 12.11 and 12.12.

It was assumed that the problem was with the LRE itself and considerable effort was applied to solving this in the field. The baffling of the instrument was improved and eventually a new LRE was fitted, but the results were always the same.

Eventually the new accommodation at the APL meant that it was feasible to return the instrument to UCL for detailed testing. Whilst observing the passage of diffused laser light through the instrument, it became clear that the problem was, in fact, a result of poor coatings on the MRE. The MRE had suffered from condensation damage when in Texas

during preparation for one of the balloon flights and the coatings had become crazed. Whilst it still performed acceptably as an etalon in its own right, the crazed surface was scattering the light from the LRE and filling in the transmission minima of the latter. The surfaces of the HRE were relatively undamaged, so the MRE profile was not being degraded in this way. The instrument, therefore, requires a new MRE as soon as funds permit. The poor finesse of the HRE suggests that it should also be replaced.

12.7.3 Etalon cross-coupling

To say that the overall transmission function is the product of the three etalons is generally true to a first approximation, but in reality the situation is more complex. Multiple beam interference can also occur between plates of the different etalons. When the system is in perfect mechanical alignment, further etalon cavities are formed between the reflecting surfaces of adjacent etalons.

The interference fringes from these "extra" etalons are clearly visible, especially if the inter-etalon distance has been kept to a minimum. They are a useful indication that exact alignment has been obtained. However, the coupling of energy from the main peak into these subsidiary fringes can significantly degrade the contrast of the instrument.

Mack et al (1963) have derived an expression for the transmission of a triple etalon system, taking into account all the inter-etalon reflections. They calculate that inter-etalon coupling degrades the filtrage of such a system by a factor of four. The effect can be drastically reduced by "spoiling" the inter-etalon cavities. One method of doing this is to introduce a small amount of attenuation between adjacent etalons, thereby absorbing multiple reflections, and *Mack et al* have shown that a transmission factor of 0.85 between each etalon is sufficient to return the filtrage to the value predicted when neglecting inter-etalon effects. They have suggested omitting the anti-reflectance coating on the glass surfaces between the etalons to

provide the attenuation. Experiments at UCL indicate an improvement in performance using neutral density filters, but the results have been masked by the inter-etalon scattering which, with the present etalons, is a more serious problem.

Another way of overcoming the coupling effect is to tilt the central etalon slightly off-axis. If this technique is employed it is best to place the lowest resolution etalon in the centre as it has the largest angle of acceptance. However, this method is not suitable with two-dimensional imaging systems, such as the UCL triple, because an uneven illumination of the imaged ring will result.

A full account of the available methods of minimising the inter-etalon coupling has been presented by *Schwider (1965)*. In addition to attenuation and tilting, he discusses polarisers, lens systems and quarter wavelength couplers. However, these methods are largely inappropriate to the UCL instrument.

12.7.4 Reflections between the outer surfaces of the etalon plates

The etalon plates are normally wedged, as explained in paragraph 9.12, so reflections from the outer surfaces are thrown off the optical axis and do not cause problems. An anti-reflectance coating has been used on the outer surfaces of the etalon plates in order to minimise any spurious reflections. However, as mentioned in the previous paragraphs, the coatings on the inter-etalon surfaces can be omitted to advantage, although this may give problems with ghost images in the system.

12.7.5 Surface irregularities and deviation from parallelism

It was explained in paragraph 9.9 that, with a single etalon interferometer, the finesse of the instrument is degraded if the surfaces are not perfectly flat and parallel. However, the effect is compounded when etalons are combined - the so-called automasking effect (*Mack 1963; Hernandez 1986*). In simple terms, the transmission peaks will not coincide over the full area of the etalons. The overall

response may be considered as the sum of many elemental areas. The throughput of each one is degraded by the effect of the local surface defects on the matching areas throughout the system. It is, therefore, essential that the plates are extremely flat. A flatness of $\lambda/200$ is considered acceptable. To obtain this accuracy the plates must be very precisely figured and allowances made for any distortions introduced by the coatings. It is also essential that the parallelism is carefully adjusted and maintained.

The finesse of the three etalons has been obtained from the measurements described in paragraph 12.6. The values are 11.5 for the LRE, 16.5 for the MRE and 11.0 for the HRE. The MRE finesse is just acceptable, but the performance of the other two etalons is rather poor. However, the quality of such large diameter etalons has increased considerably since these particular etalons were made. One problem with the older etalons was that they exhibited significant plate distortion near the edges. The triple etalon interferometer uses a Cassegrain telescope to focus the image onto the detector, so the secondary mirror partially obscures the flatter central area of each etalon and the image is formed from the more distorted outer regions. A lens system would be more satisfactory in this respect, but the overall length of the instrument would be significantly increased. One effect of the degraded finesse is to increase the width of the main transmission peak. A more serious consequence is the increase in amplitude of the parasitic peaks. Figure 12.14 shows the response of the instrument re-calculated using the measured finesse values. The increased amplitude of the parasitic peaks compared to those shown in ideal etalon response of figure 12.6 is clearly shown.

12.7.6 Transmission function alignment: stability of etalon gaps

The etalon gaps are adjusted so that the transmission maxima of the three etalons coincide. If the peaks do not coincide exactly then a loss of transmission will result. However, more importantly, a change in the peak position of any one of the etalon profiles will cause the peak position of the overall response to be shifted. This will result

in an error in the measured wind speed. The effect of such misalignment has been predicted with the computer simulation. The MRE is much more critical than the LRE in this respect as the steeper slope of its transmission profile has a more pronounced effect on the overall peak position if the peak is offset.

Figure 12.15 shows a simulation of the overall response for a 1% of FSR offset of the LRE, and figure 12.16 for a 1% offset of the MRE. The displacement of the overall peak is barely discernible, amounting to a few femtometres for the LRE and a few tens of femtometres for the MRE. However, when the equivalent wind speed errors are calculated they are seen to amount to tens of metres per second. This is shown in figures 12.17 and 12.18 which are plots of transmission peak offset versus wind error for the LRE and MRE respectively.

The simulation assumes ideal etalons, so if the finesse of the LRE and MRE are impaired then the peak pulling effect will be less. Conversely, if the HRE response is broader than the ideal response, its ability to dominate the overall peak position will be impaired.

12.7.7 Mechanical alignment of etalon axes

If an etalon is illuminated with an extended monochromatic source, then an observer looking through the etalon towards the source will see the alternating transmission maxima and minima as a concentric ring pattern at infinity. The rays of the observed transmission peak can, therefore, be considered to lie along the surface of a cone, the apex of which is at the observing point. In a multiple etalon system, the tuning of the cavities ensures that the apex angle is identical for the chosen maximum of each etalon. That is to say, the angular diameter of the fringe from each etalon is the same and a single bright ring is observed. However, if the longitudinal axes of the cones for each etalon are not aligned precisely, the conical surfaces will not coincide at all points and the observer will not see a uniformly illuminated circle. Instead, the intensity of the observed

ring will vary around its circumference, resulting in one or more bright arcs.

In order to effect the angular alignment of the transmission functions, the etalons must be mechanically tilted with respect to each other so that all the reflecting surfaces are precisely normal to the optical axis. In practice, the HRE is aligned to give a fringe pattern which is positioned centrally on the detector and further angular adjustment is carried out on the MRE and LRE only. In the first balloon instrument this was achieved with locking screws which were adjusted whilst examining the fringe pattern by eye and then securely clamped. Although this method worked, it was a very difficult and painstaking procedure, and the alignment could not be optimised once the etalons had been sealed in their housing.

An improved method was incorporated into the second balloon instrument and is currently used in the ground-based instrument. This is a motorised alignment system. The etalons are mounted on platforms which can be tilted in two axes, 120° apart. This is done using electric motors which rotate a fine screw thread. The pitch of the screw thread is 0.5mm and each axis has a pivot length of 160mm. This gives an adjustment of 3.125 milliradians per complete turn of the screw. The motors are mounted on a rigid platform attached to the optical bench and the screw threads adjust the spacing between this and the etalon platform. The motors are driven for fixed time intervals. This gives some measure of repeatability to the adjustment. The intervals chosen were 3 seconds for a coarse adjustment and 0.1 seconds for a fine adjustment. It is an unsophisticated system, as dictated by the restrictions of the balloon command system, but one which has proved to be successful. The motorised mounts can be seen in figures 12.1 and 12.8.

In order to assess the required accuracy of the mechanical alignment, it is instructive to calculate the angle of acceptance for each etalon. This has been done using the Quattro etalon simulation. The transmission intensity as a function of angle has been calculated for

the MRE and the LRE. Simulations were performed for on-axis alignment of the etalon profiles and also for alignment at the first off-axis order of the HRE. The latter occurs at an angle of eight milliradians to the optical axis. This is the largest off-axis angle that will be observed and the one which requires the most precise mechanical alignment.

The results of the simulations are shown in figures 12.19 and 12.20. Inspection of the spreadsheet table shows that for the on-axis case the transmission drops to 90% of the maximum for a displacement of 1.73 milliradians for the MRE and 4.05 milliradians for the LRE. However, for the more critical off-axis case the values fall to 0.17 and 0.90 milliradians respectively. The MRE must, therefore, be aligned to an accuracy approaching one twentieth of a revolution of the screw thread.

The initial alignment of the etalons is most easily performed by examination of the fringes by eye whilst observing a diffused coherent source. Initially, light is injected into a diffuser placed between two of the etalons so that the MRE can be aligned to the HRE and then the LRE to the MRE. Final adjustment can be achieved photometrically by observing the image produced by the complete instrument whilst stepping the etalon gaps.

Figure 12.21 shows the uneven illumination of the ring that results from mechanical misalignment of the etalons.

12.7.8 The effect of image distortion

It was mentioned in paragraph 12.6.2 that distortion in the detector image results in a broadening of the measured peak profile. This is because the "reduce to radius" routine produces a plot of intensity versus radius squared, calculated by summing all points on the image, measured from a pre-determined centre.

If the light intensity around the fringe is not uniform, then a weighting effect in favour of the brighter segments of the fringe will result. If this brightness distribution varies with time and the image is in any way distorted, then the calculated radius of the ring will vary accordingly. It was explained in paragraph 12.7.7 that any slight mechanical misalignment of the etalons will result in intensity variations around the fringe when the etalons are scanned through the peak. Such variations are inevitable as it is impossible to obtain absolutely perfect mechanical alignment of the etalons.

This means that any drift in one of the capacitance stabilised etalons will result in a change in the measured peak position, resulting in an apparent Doppler shift. This is additional to the peak pulling described in 12.7.6 which occurs when the transmission profiles are slightly offset.

The effect was observed during testing at UCL, when attempting to verify the simulations of the peak pulling effect shown in figures 12.15 to 12.18. The Imaging Photon Detector available for the tests produced a slightly elliptical image. The nature of the slight intensity variations due to mechanical misalignment were such that the measured ring diameter decreased when the etalons were placed off tune, regardless of the direction in which the offset was applied.

The reduce to radius routine was changed to reduce to 24 sectors in order to observe the variations in measured radius around the ring. These were found to span a number of radius squared bins. Furthermore, the FWHM of each profile was approximately 70% of that obtained by summing around the complete ring.

Although the detector used for these tests represented a "worst case", it is clear that a distortion free detector is critically important for multiple etalon systems.

12.8 Choice of detector

The resistive anode IPD is a suitable detector for the triple etalon FPI for measurements during hours of darkness. It is most useful for the precise photometric alignment of the etalon axes using the motorised alignment. However, as seen in the previous paragraph, even slight image distortion can put great demands on etalon stability.

During twilight the transmitted light intensity becomes too great for the resistive anode IPD. The ring anode detector described in chapter eight is a more appropriate device as it has a much higher photon rate capability. The 32 ring ITT device shown in figure 8.8 has been fitted to the instrument and an example of the data obtained is shown in figure 12.22.

Another candidate for the detector is an intensified CCD. Hopefully this would have the required sensitivity together with a distortion free image. This option will be explored when the problems with the instrument contrast have been corrected.

12.9 Selection of etalon gap ratios

The selection of gap ratios for multiple etalon systems must be done with great care to ensure optimum performance.

The gap ratios of the UCL triple are 10 : 1.82 : 0.331, that is to say, a ratio of 5.5 : 1 between each etalon. The interferometer is bandwidth limited by an interference filter. The most critical region with regard to the overall response of the etalons is the region within the pass-band of the interference filter, which is normally chosen to be 1nm. In this respect, the danger points occur at the first three off-axis peaks of the MRE.

Inspection of figure 12.6 shows that the first and third peaks coincide with HRE minima. The second peak, however, coincides with the eleventh off-axis peak of the HRE, so the attenuation at this point is determined by the LRE alone, the response of which is close to its

minimum. The poor contrast of the LRE is, therefore, a very serious problem.

12.10 Comparison with the UARS HRDI interferometer

Results obtained with the The High Resolution Doppler Imager (HRDI) on the Upper Atmosphere Research Satellite (UARS) have recently been published. (*Hays et al, 1992; Hays et al, 1993*). The ratios of the etalon gaps differ slightly from those of the UCL instrument. They are: 10.007mm, 1.861mm and 0.241mm (5.38 : 1 and 7.72 : 1). The theoretical response has been determined with the Quattro model and is shown in figure 12.23. The simulation was performed using the published finesse values of 10.6, 14.5 and 14.4 for the HRE, MRE and LRE respectively. The low value of the HRE finesse is partly a result of the inclusion of the aperture broadening effect of the multi-ring anode detector. The response has been plotted in terms of wave number in order to compare directly with the published data which include a measurement of the instrument response using a scanning dye laser. This is reproduced in figure 12.24. The computer model is seen to be in close agreement with the measured response.

At this wavelength (760nm) the etalon gap ratios are such that the third peak of the MRE nearly coincides with the sixteenth peak of the HRE. It is interesting to compare this response with the 630nm response of the UCL triple:

1. The response at the first MRE maximum no longer coincides with the HRE minimum and as a result it is more susceptible to the effects of poor finesse, becoming slightly larger than the equivalent peak in the UCL triple.

2. The response peak corresponding to the second MRE maximum is somewhat lower than its UCL counterpart, the latter having coincident HRE and MRE peaks at that point.

3. The response at the third MRE peak is larger, because of the coincidence with the HRE peak.

12.11 Temperature and pressure stability

One of the major problems encountered in operating the triple etalon interferometer in the field has been that of small drifts in the instrument response due to temperature instability and pressure changes inside the sealed housing. Such problems are only apparent when testing over a period of several days, and it is not generally feasible to isolate the specific cause of any observed drift during a field trip of approximately two weeks. The return of the instrument to the APL enabled a detailed examination of the stability to be made (*Harris, 1993*).

The tests in the field had concentrated on looking for a pressure leak through the bulkhead O-rings and connector seals. However, the performance of the 150mm CSE (paragraph 11.12.2) indicated that reservoirs of trapped air and general outgassing are a more likely explanation for the variations seen. In order to minimise these effects the instrument was operated at near atmospheric pressure.

The temperature dependence of a triple etalon system is more complex than that of a single etalon interferometer. Not only is the stability of the ring diameter to be considered, but the fringes of the three etalons must stay completely aligned with each other. In assessing this one must differentiate between changes in refractive index (caused by temperature and pressure variations), and the thermal characteristics of the physical gap lengths.

If the refractive index of the cavity medium changes, provided that the environment is uniform throughout the housing, the fringes should vary in step with each other and the peak intensity will be unaffected, although the fringe diameter will change. If this happens on a long enough time scale it can be compensated for by frequent calibrations.

However, if the individual etalons are subjected to the same temperature variation, the physical gap length will inevitably change

in a different way for each etalon. The HRE gap is dependent on the stability of the etalon spacers and housing. The LRE and MRE gaps are dependent on the capacitance sensor geometry. As these are both stepped etalons the capacitance sensor pillar will expand by approximately the same amount as the third etalon plate, so the system should be self-compensating. However, there will be a small residual temperature dependence which will not be precisely the same between the two. It is therefore essential to minimise the temperature dependence of the physical gap by careful design. In the case of the CSEs it is possible to optimise the matching between the capacitance sensors and the etalon plates by making the sensors from a combination of Zerodur and silica.

The most important result of the testing so far is the finding that the heavy cylindrical housing does not provide the temperature equalisation throughout the instrument that had been assumed by virtue of its undoubtedly large thermal capacity. It was concluded that the internal temperature differential between the middle and the ends of the housing was as much as three or four degrees. This is sufficient to cause a significant difference in the refractive index of the cavity medium for the three etalons. This theory was substantiated by the fact that the LRE and HRE, at opposite ends of the housing, were moving in step with each other, whereas the MRE had to be re-adjusted in order to obtain alignment of the peaks. The thermal insulation has now been improved substantially in preparation for another series of tests and a small fan may be added in order to minimise thermal stratification. Another important result arising from the testing programme was the failure to detect any observed temperature dependence in the capacitance stabilised etalon electronics.

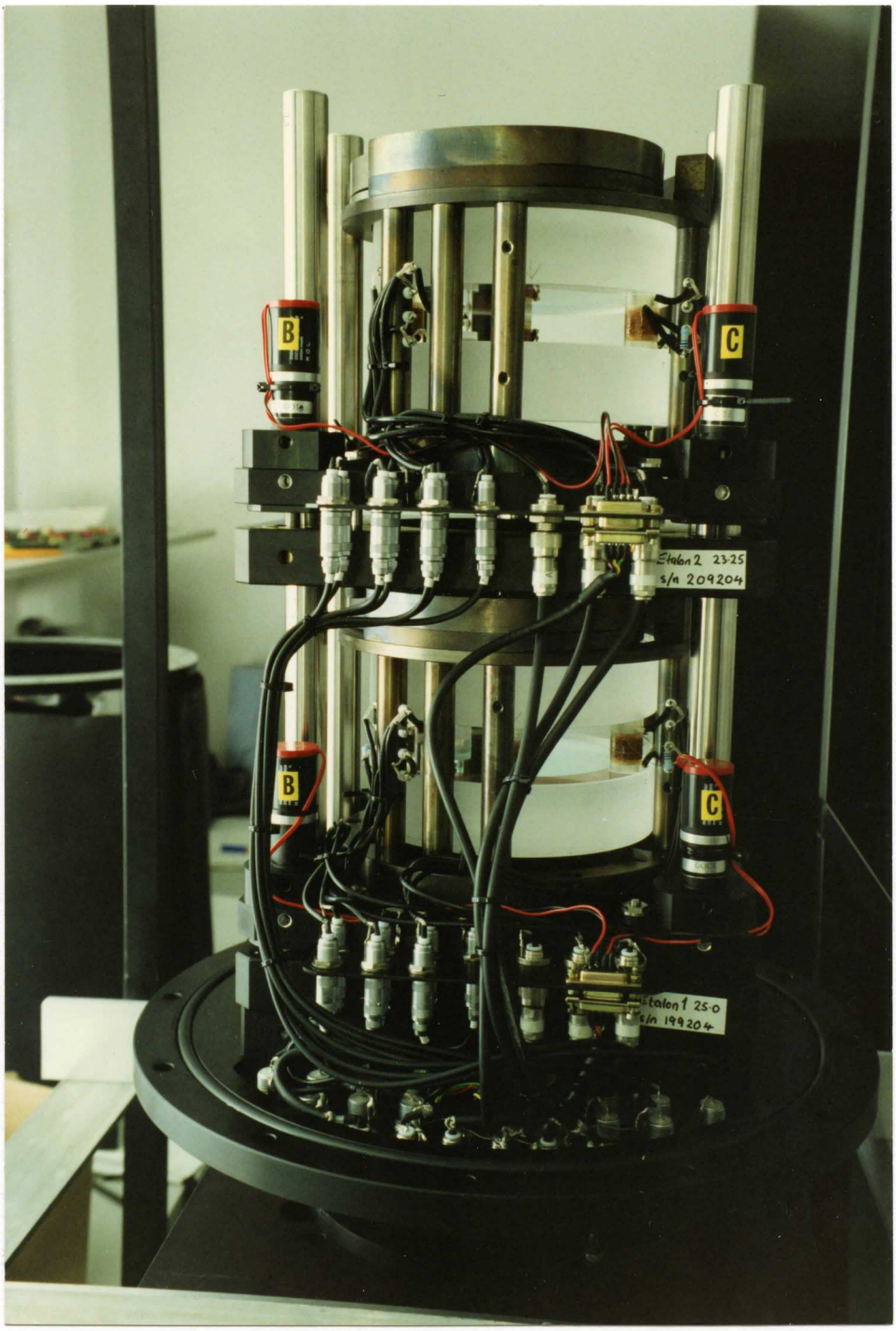


Figure 12.1

The Bonn double etalon interferometer

Figure 12.2

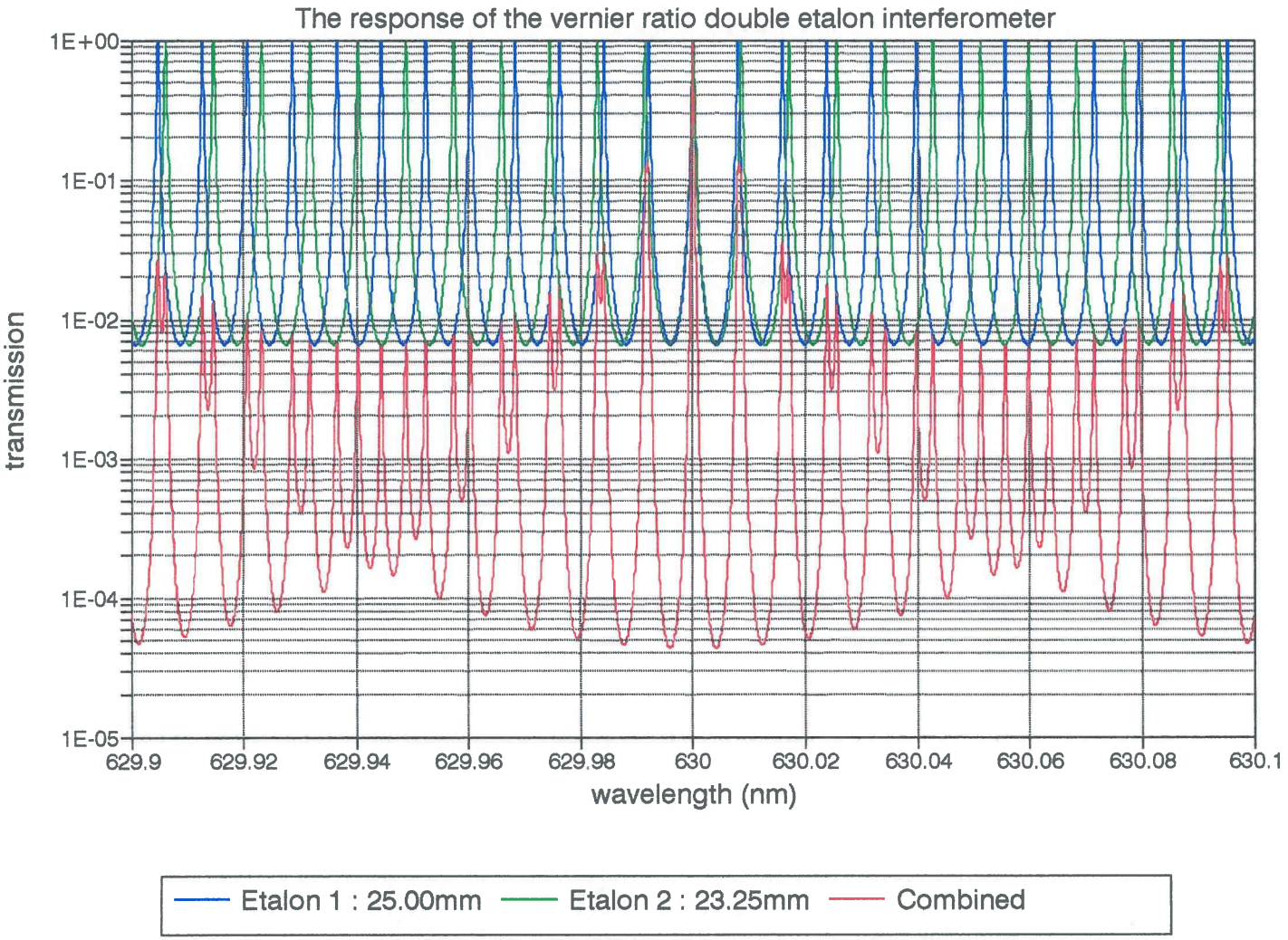

The response of the vernier ratio double etalon interferometer
1E+00
1E-01
1E-02
1E-03
1E-04
1E-05
transmission
629.9
629.92
629.94
629.96
629.98
630
630.02
630.04
630.06
630.08
630.1
wavelength (nm)
Etalon 1 : 25.00mm
Etalon 2 : 23.25mm
Combined

Figure 12.3

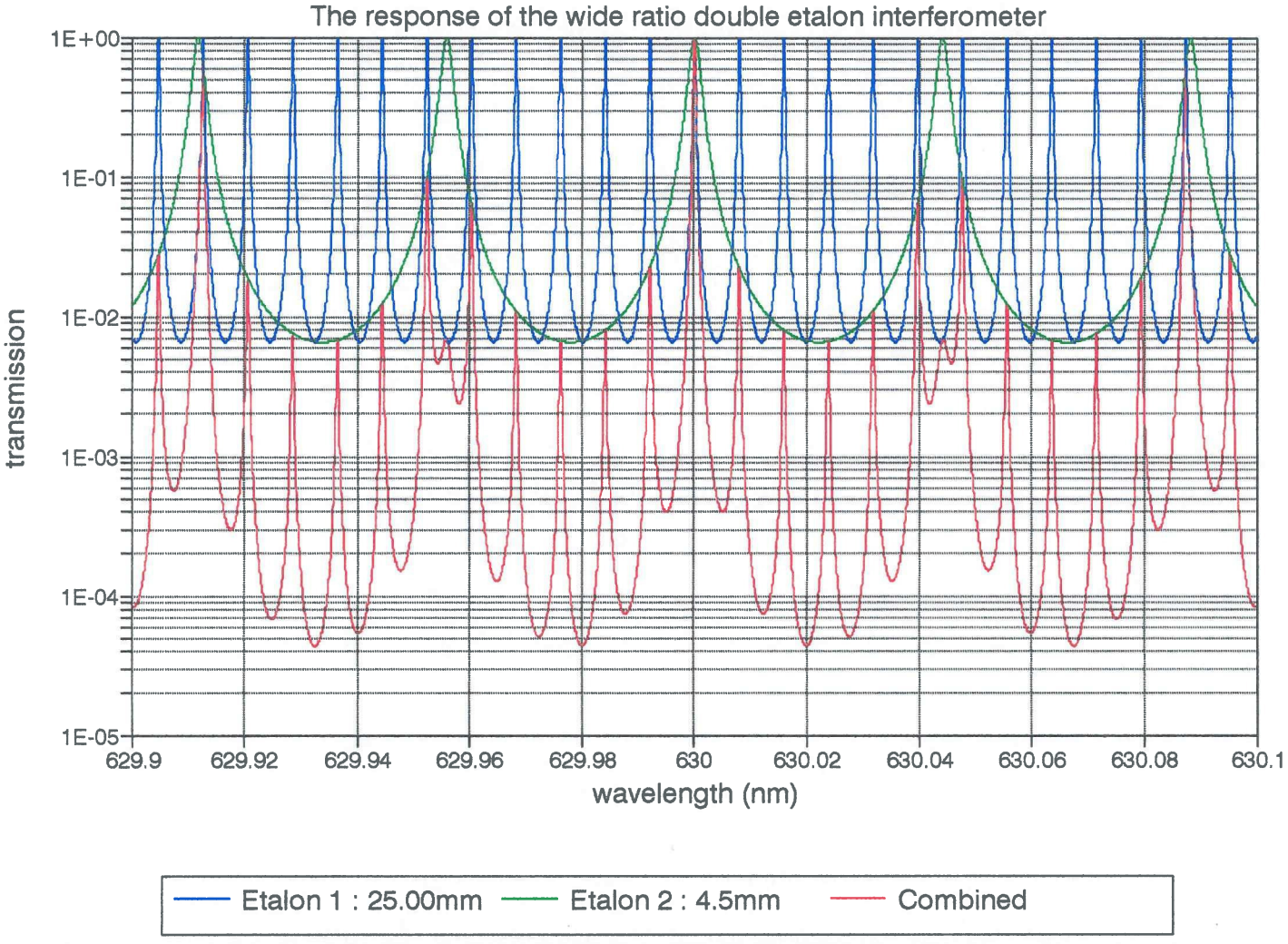

The response of the wide ratio double etalon interferometer
1E+00
1E-01
1E-02
1E-03
1E-04
1E-05
transmission
629.9
629.92
629.94
629.96
629.98
630
630.02
630.04
630.06
630.08
630.1
wavelength (nm)
Etalon 1 : 25.00mm
Etalon 2 : 4.5mm
Combined

Figure 12.4

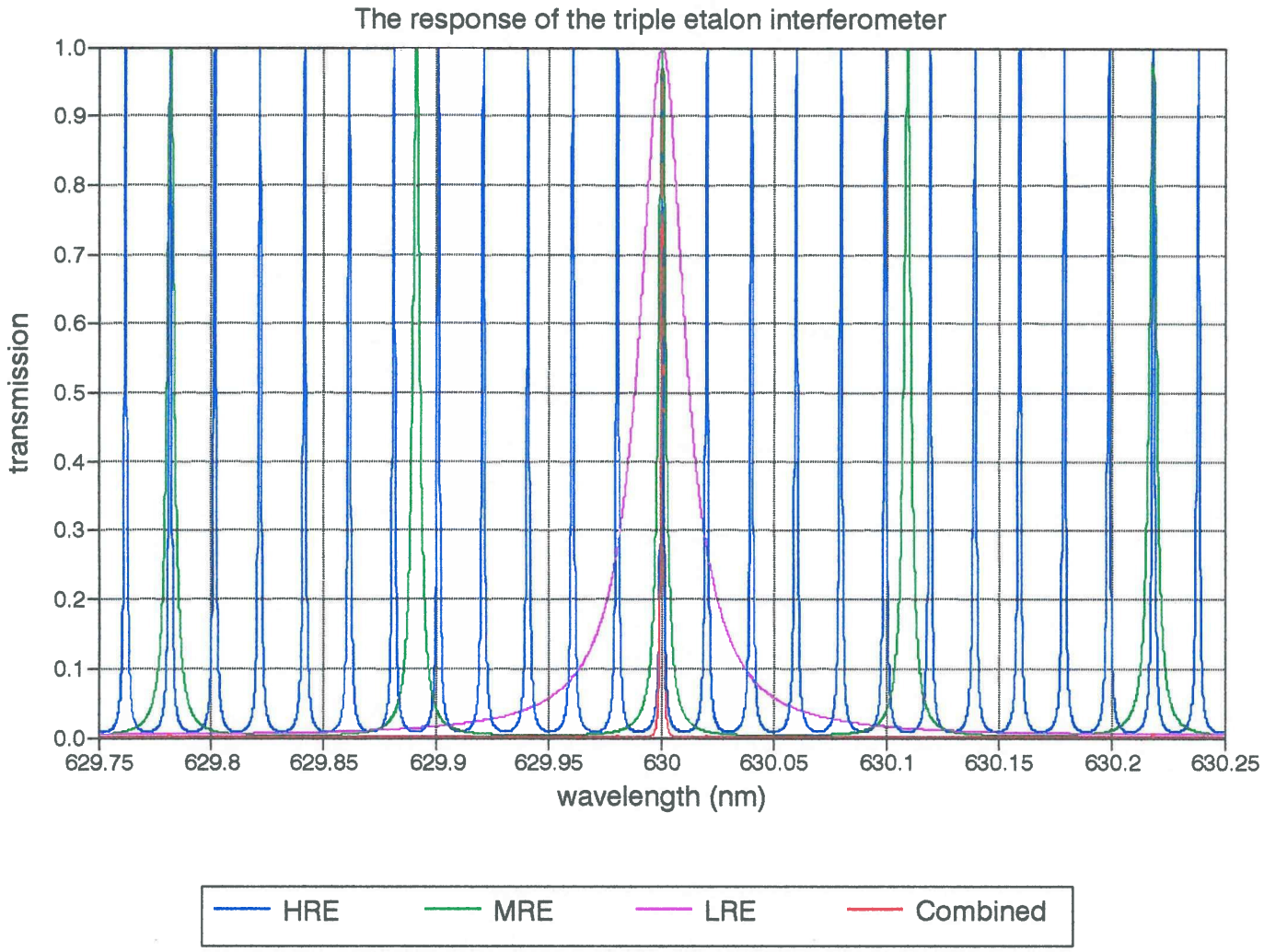

The response of the triple etalon interferometer
transmission
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
629.75
629.8
629.85
629.9
629.95
630
630.05
630.1
630.15
630.2
630.25
wavelength (nm)
HRE
MRE
LRE
Combined

Figure 12.5

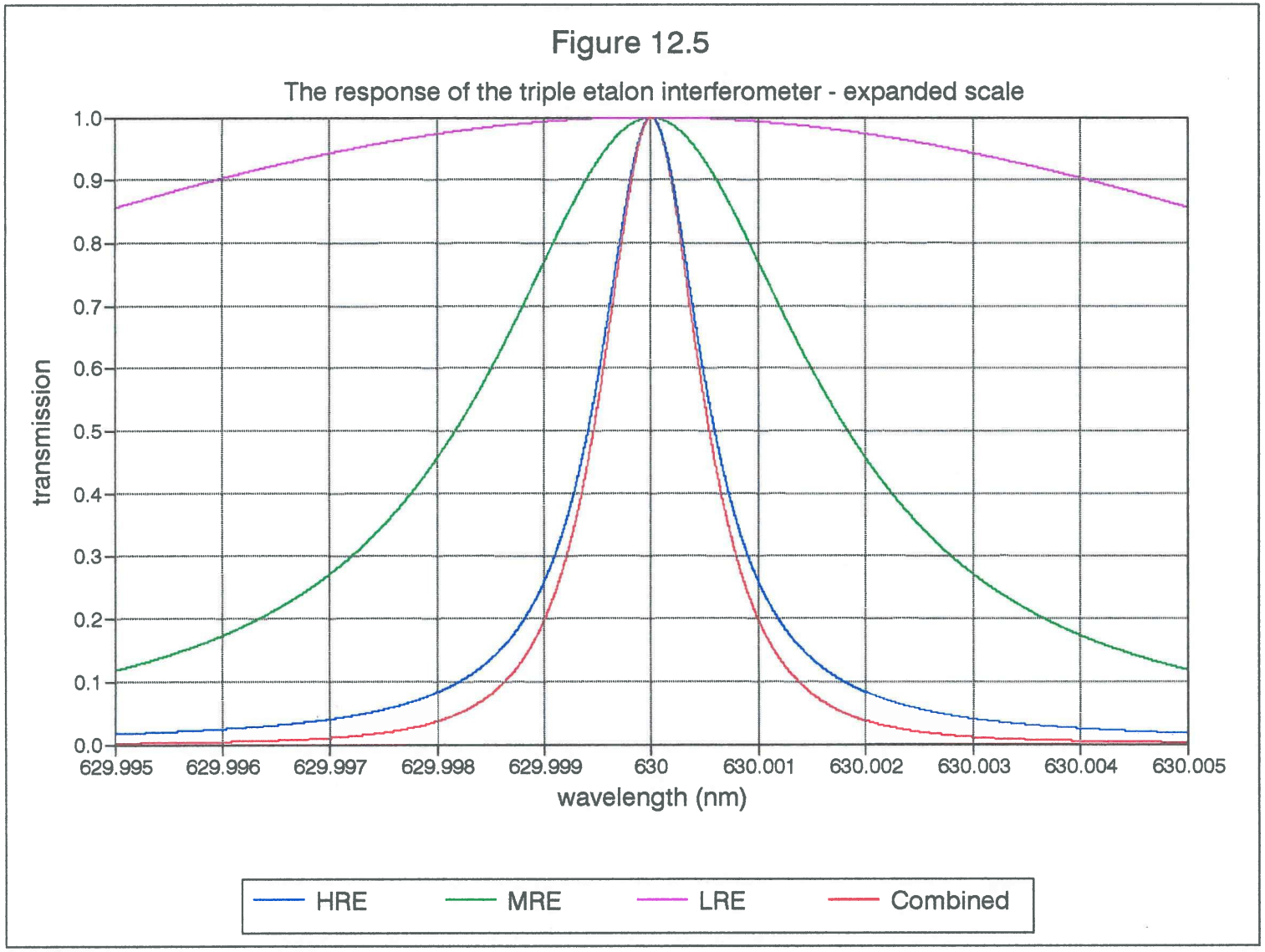
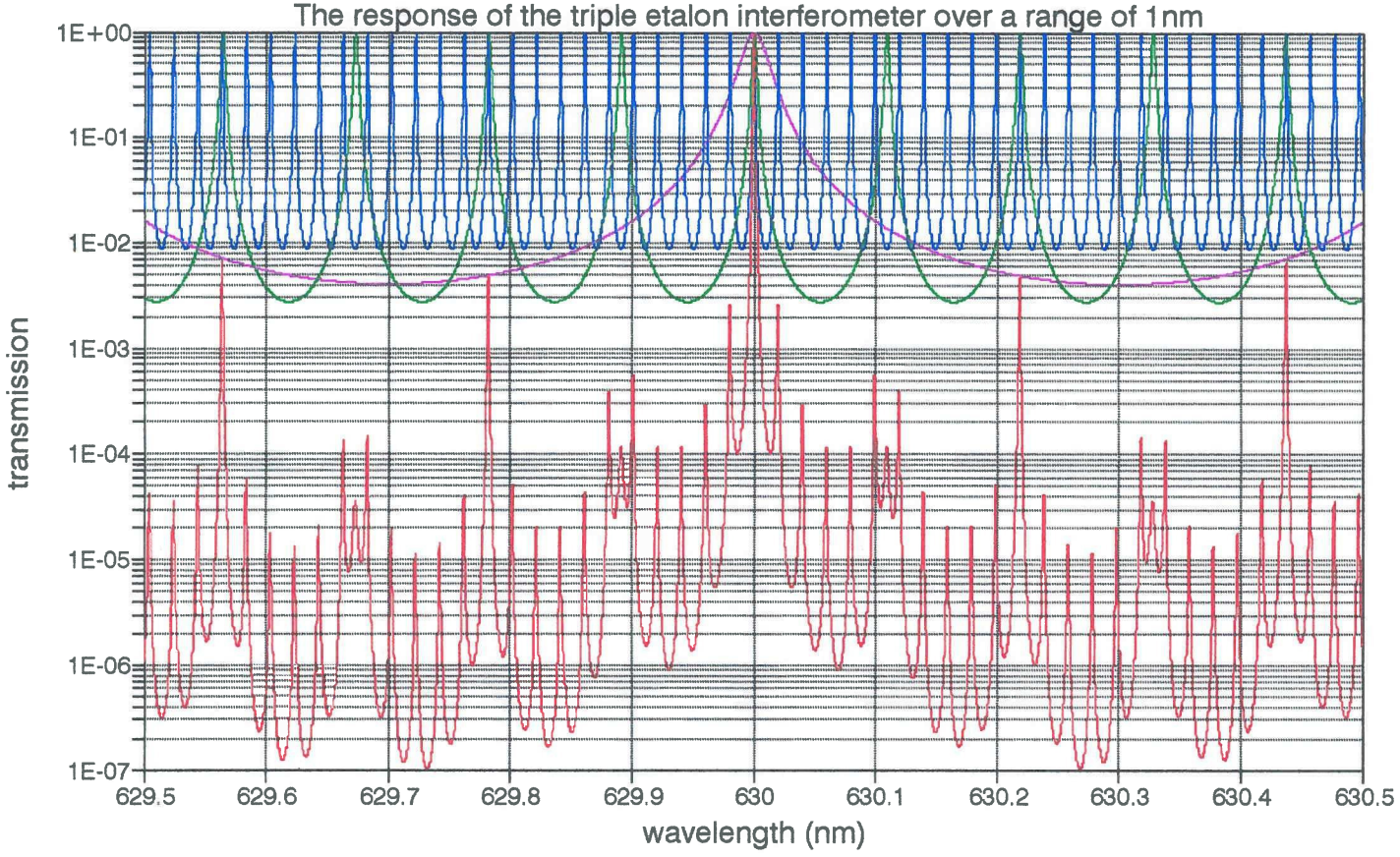

The response of the triple etalon interferometer - expanded scale
transmission
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
629.995
629.996
629.997
629.998
629.999
630
630.001
630.002
630.003
630.004
630.005
wavelength (nm)
HRE
MRE
LRE
Combined

Figure 12.6


The response of the triple etalon interferometer over a range of 1nm
transmission
1E+00
1E-01
1E-02
1E-03
1E-04
1E-05
1E-06
1E-07
629.5
629.6
629.7
629.8
629.9
630
630.1
630.2
630.3
630.4
630.5
wavelength (nm)
HRE
MRE
LRE
Combined

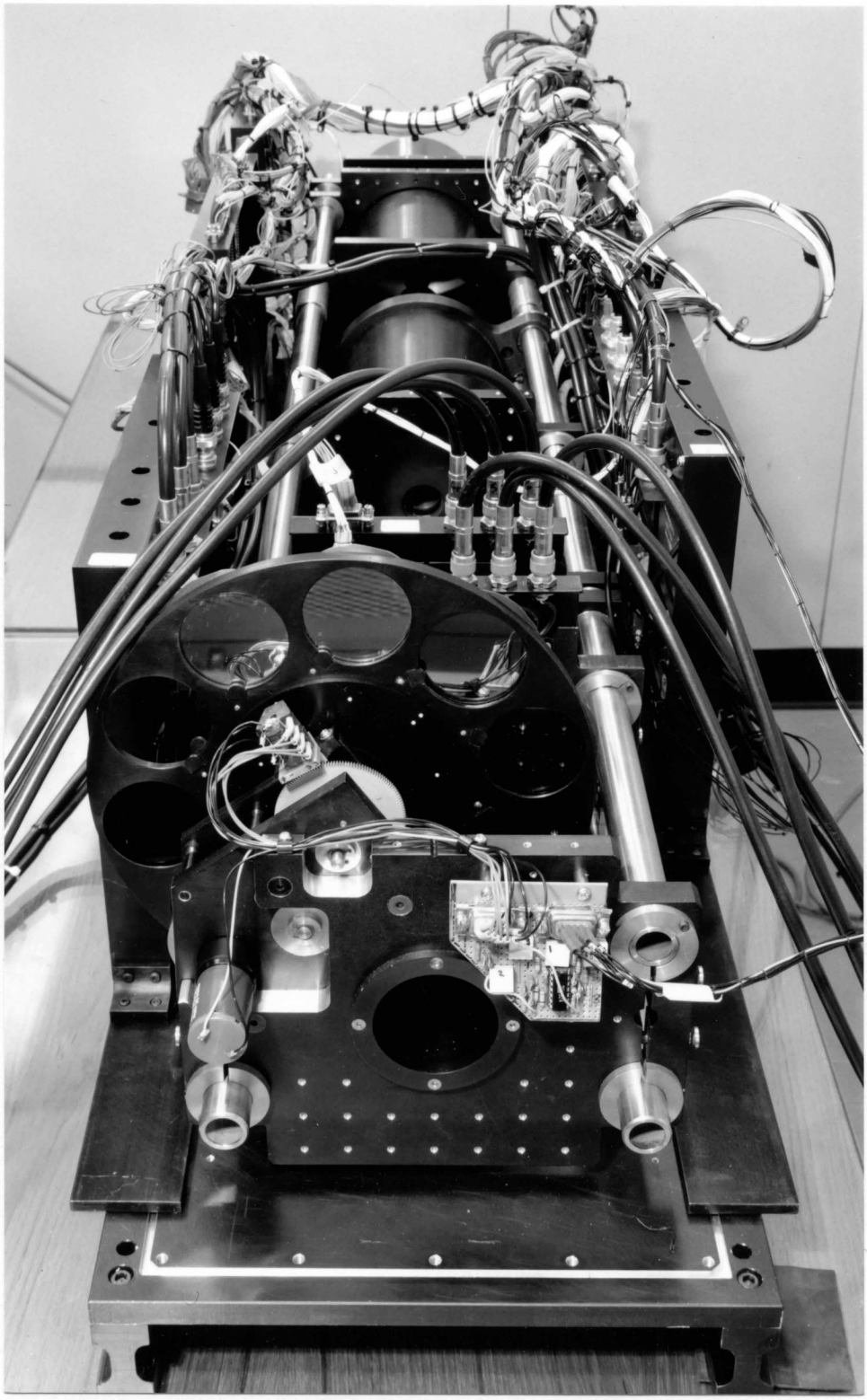


Figure 12.7

The balloon-borne triple etalon interferometer

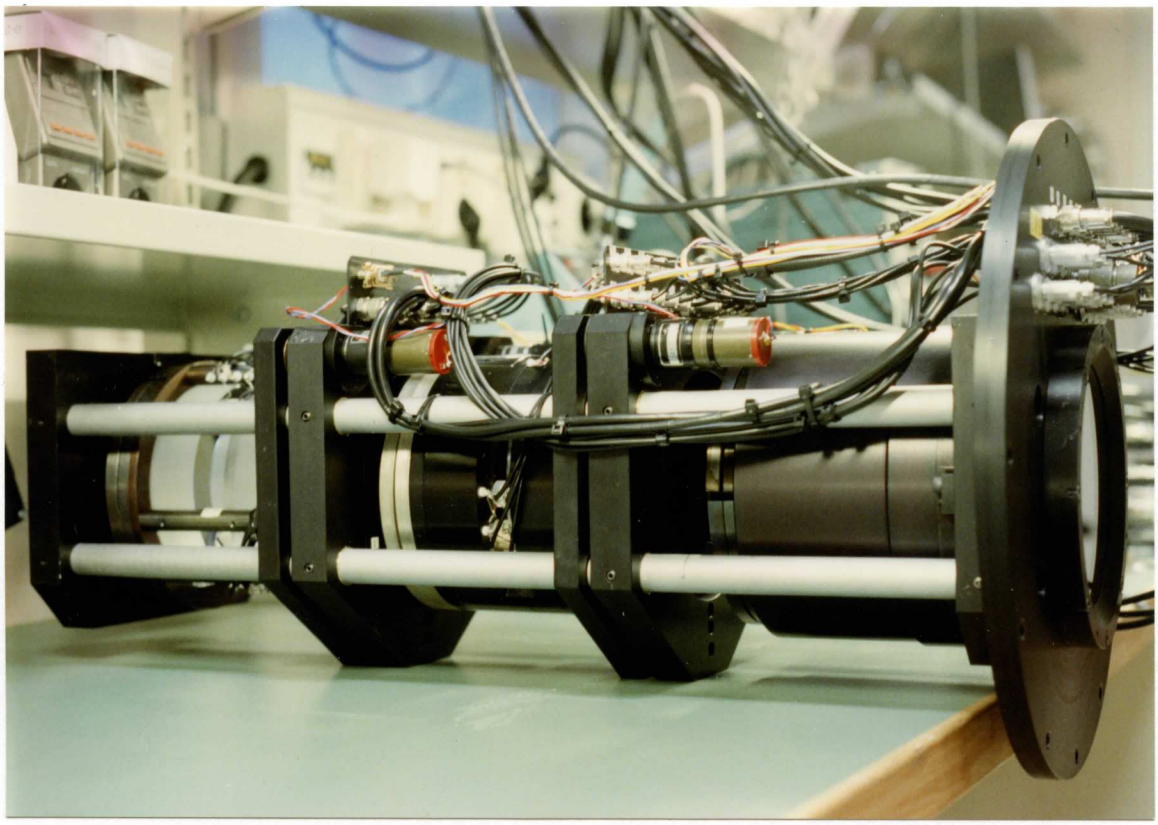


Figure 12.8

The ground-based triple etalon optical bench

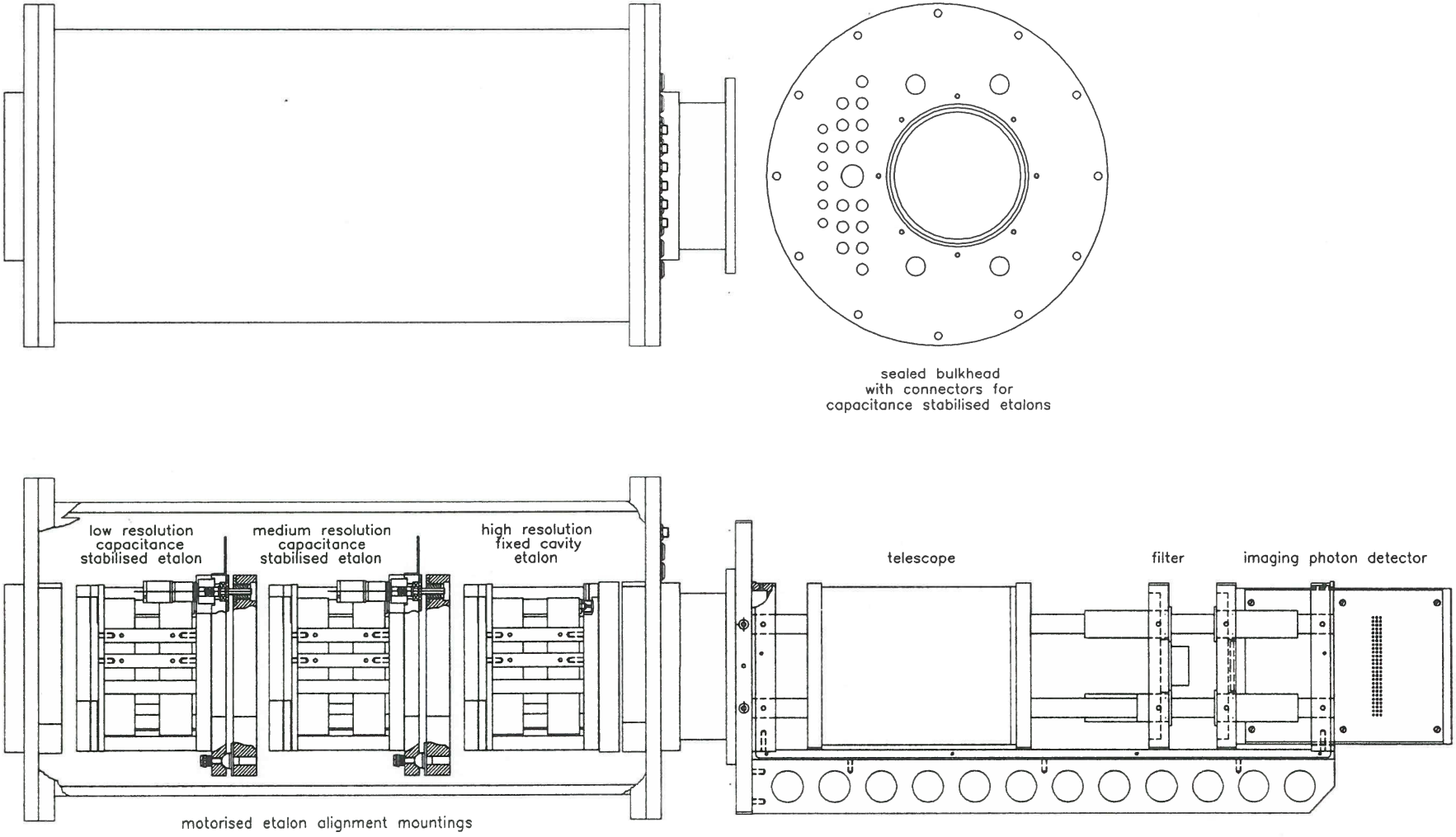
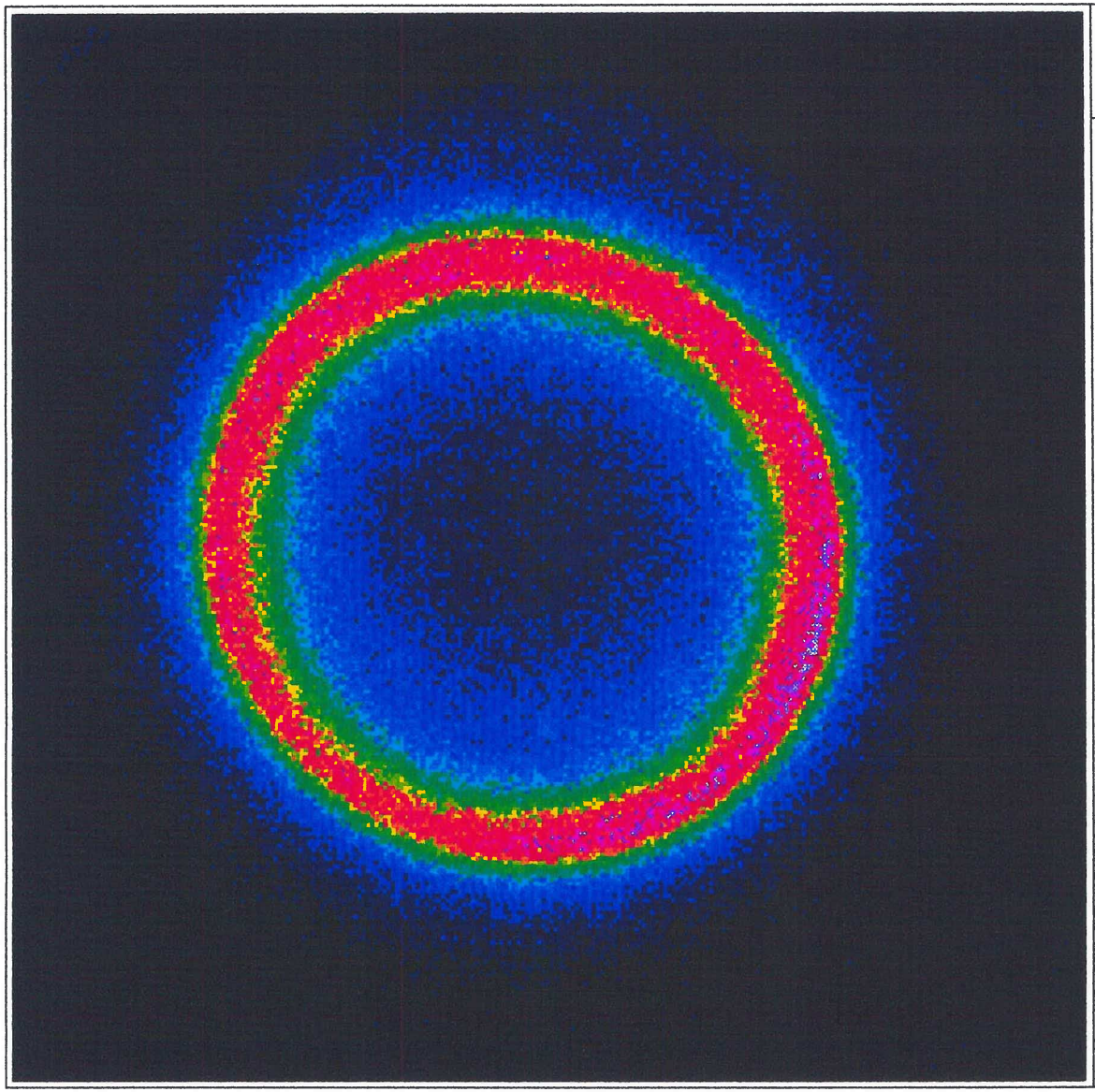


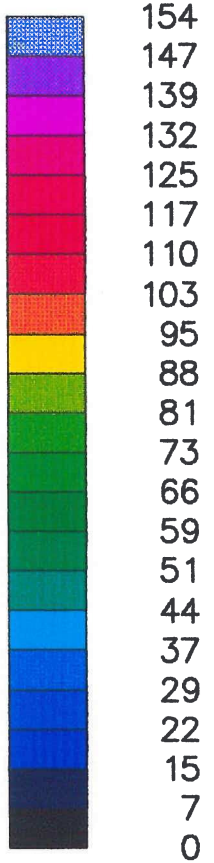
Figure 12.9 The Triple Etalon Interferometer

J. Percival



ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON

Triple etalon interferometer
File: TEQ1213.E01



Stabilised laser image

Figure 12.10

Figure 12.11

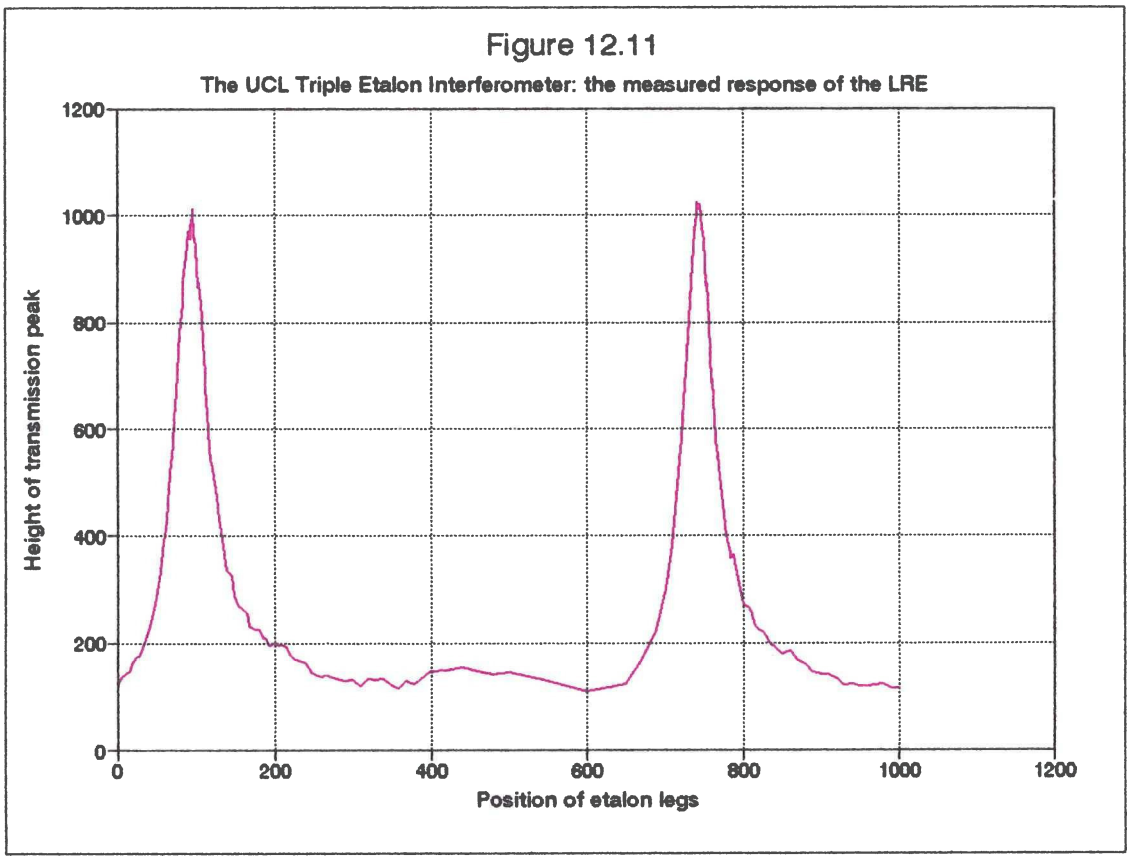
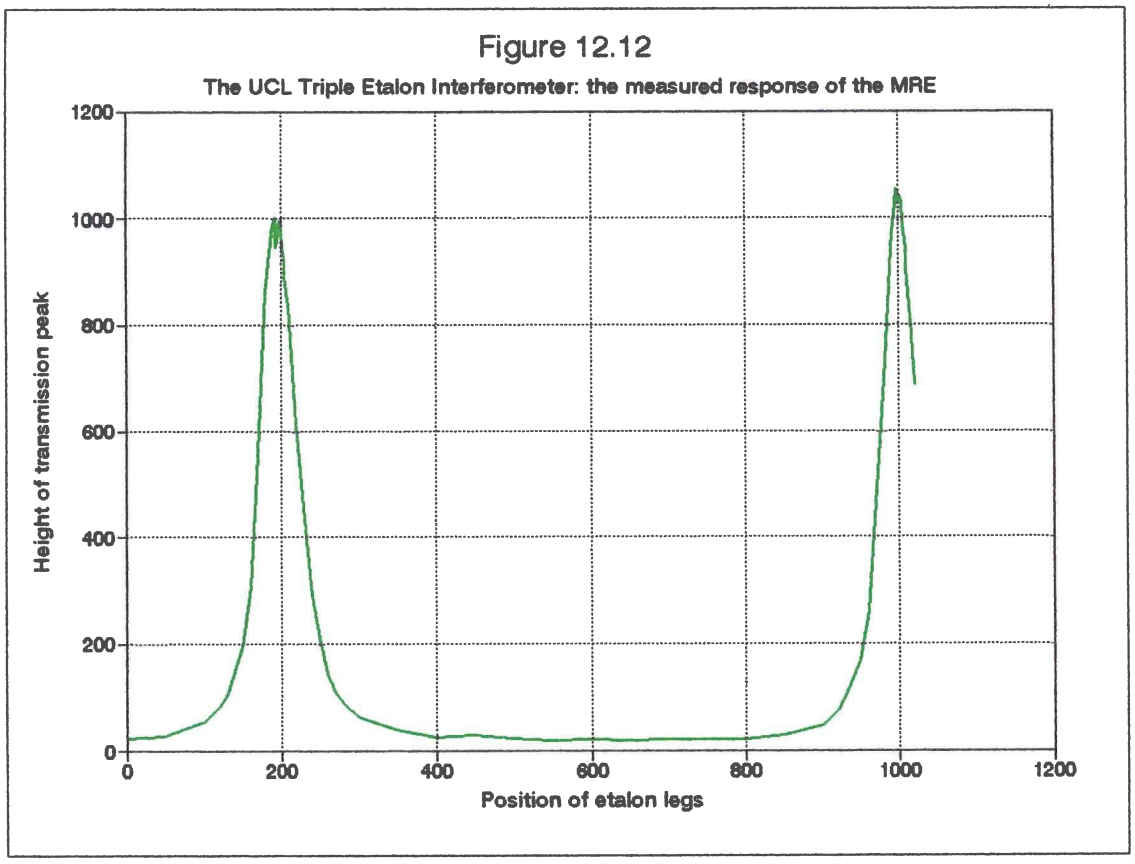


Figure 12.12



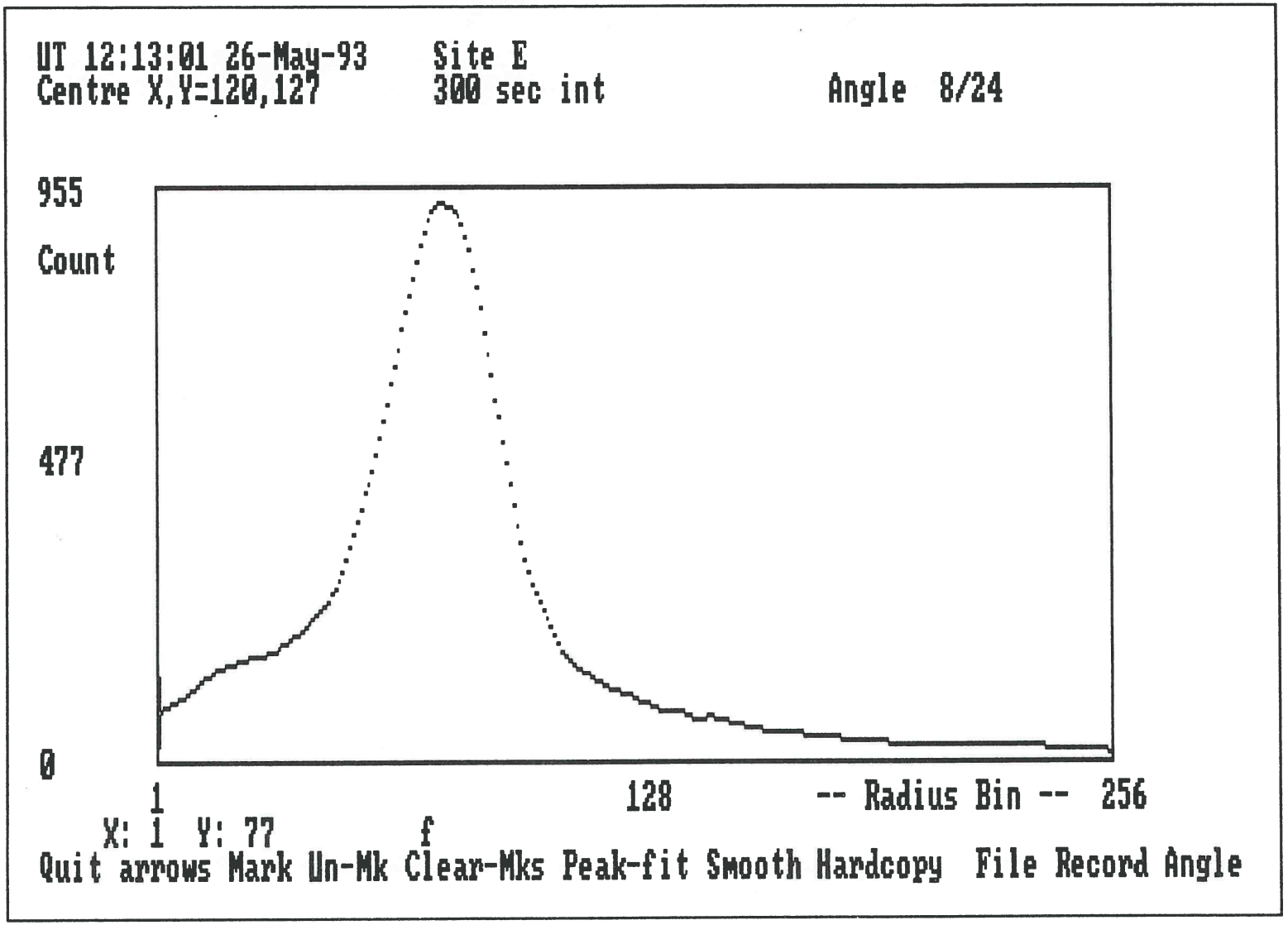


Figure 12.13 Reduced-to-radius plot of the Triple Etalon Interferometer image

Figure 12.14

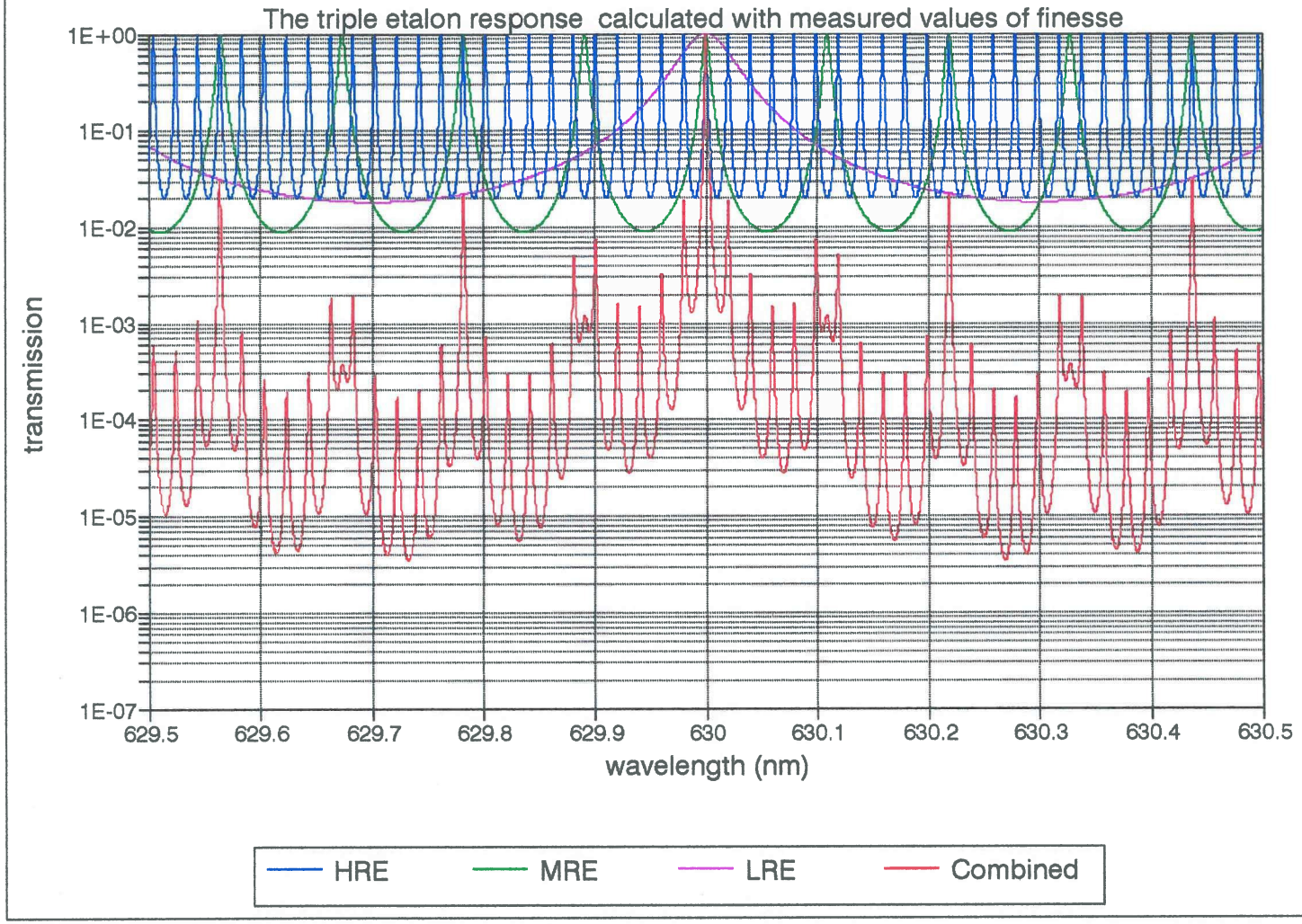

The triple etalon response calculated with measured values of finesse
transmission
1E+00
1E-01
1E-02
1E-03
1E-04
1E-05
1E-06
1E-07
629.5
629.6
629.7
629.8
629.9
630
630.1
630.2
630.3
630.4
630.5
wavelength (nm)
HRE
MRE
LRE
Combined

Figure 12.15

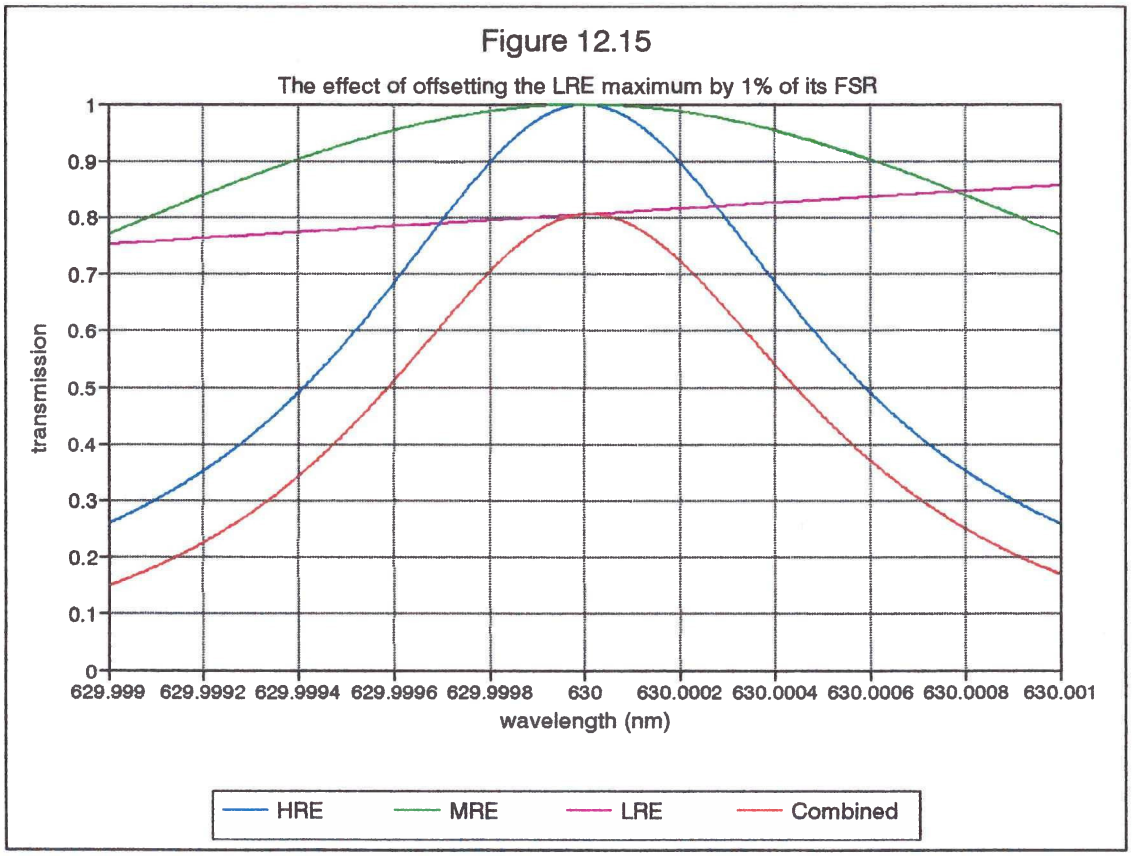


Figure 12.16

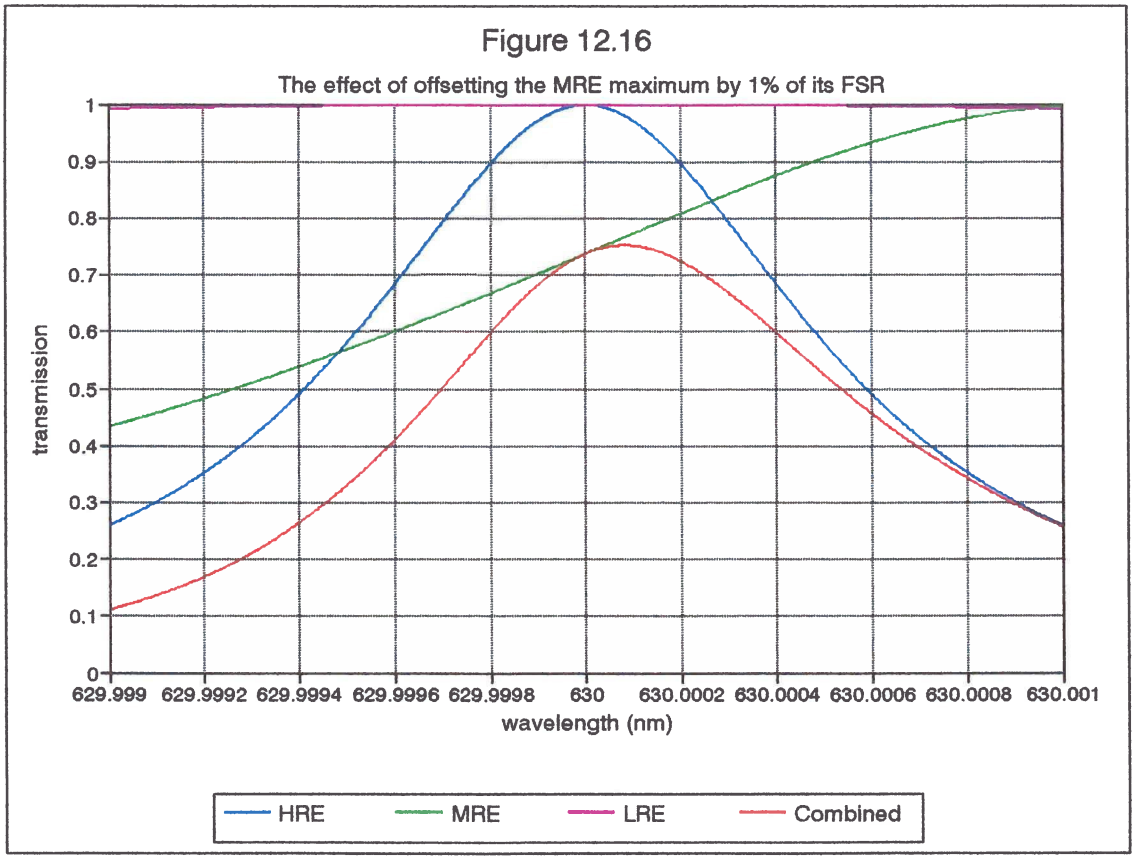


Figure 12.17

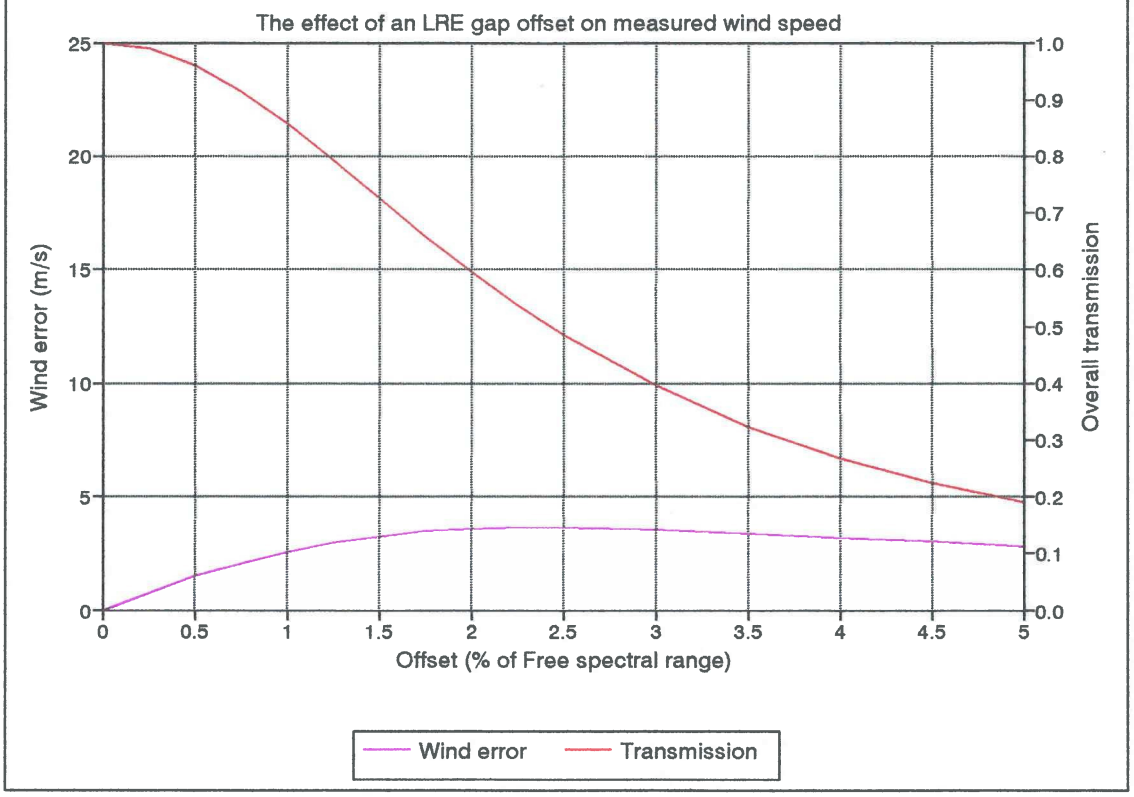


Figure 12.18

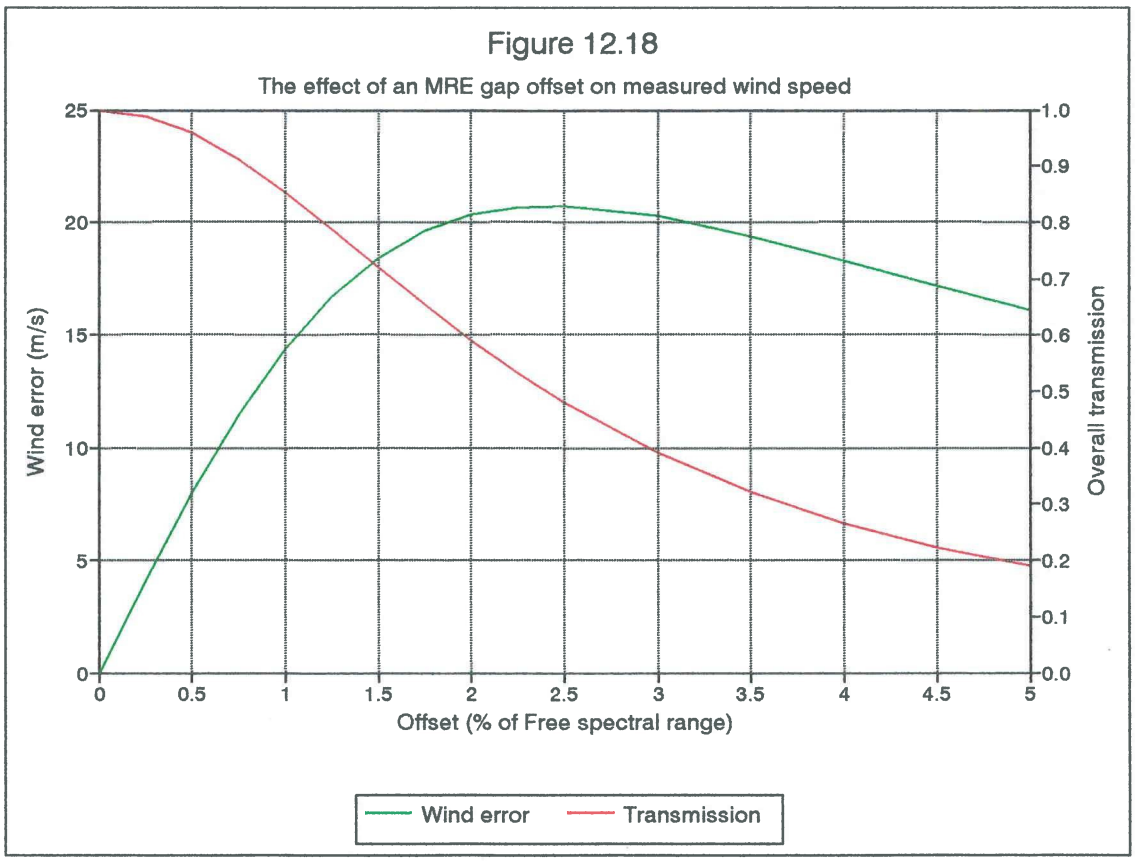


Figure 12.19

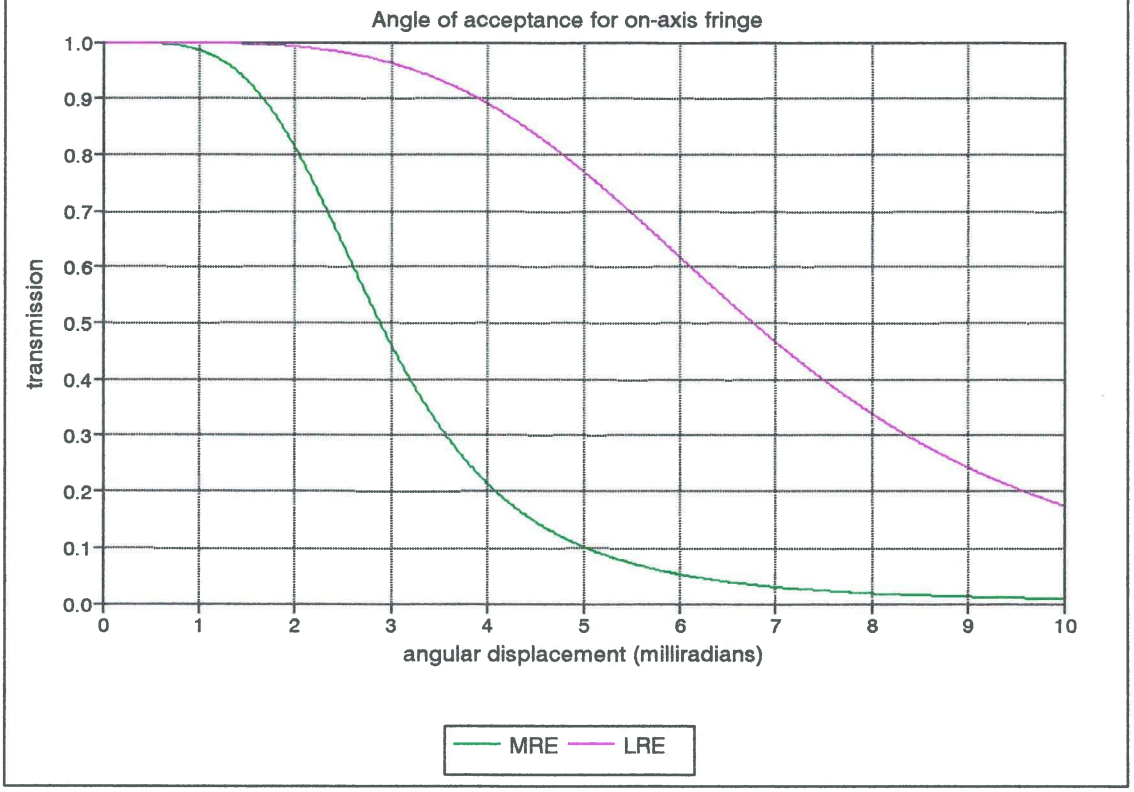
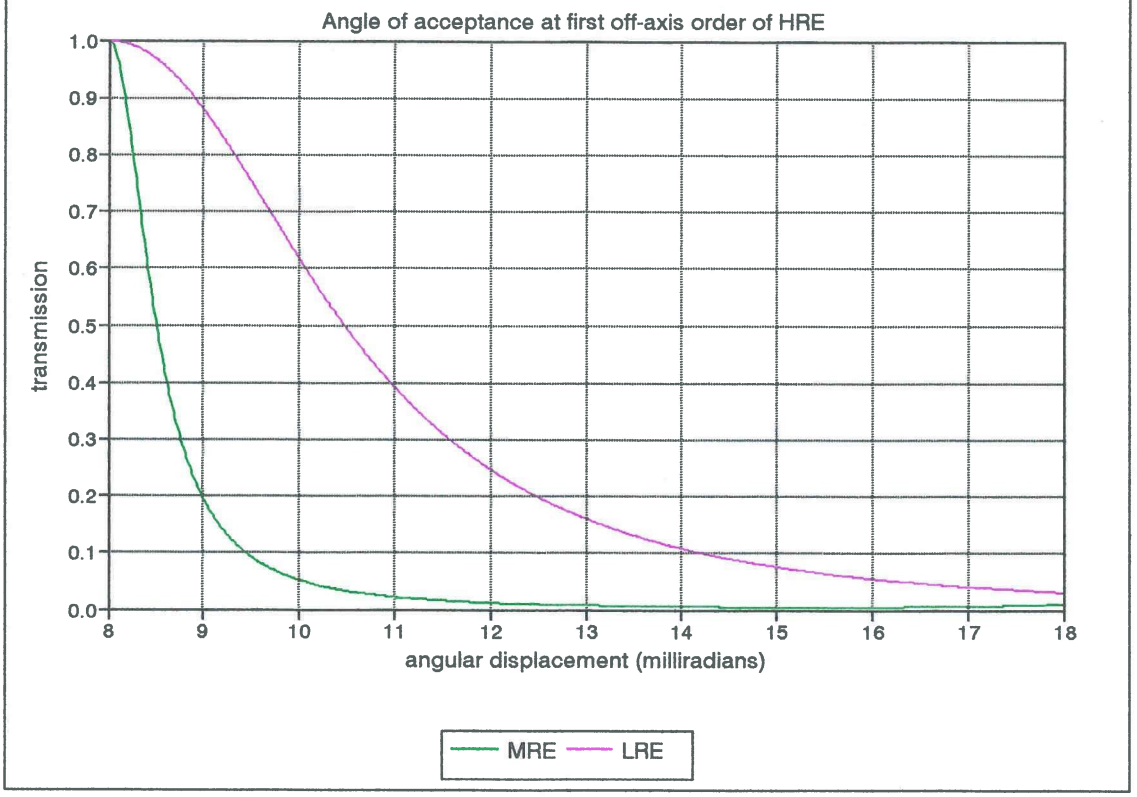
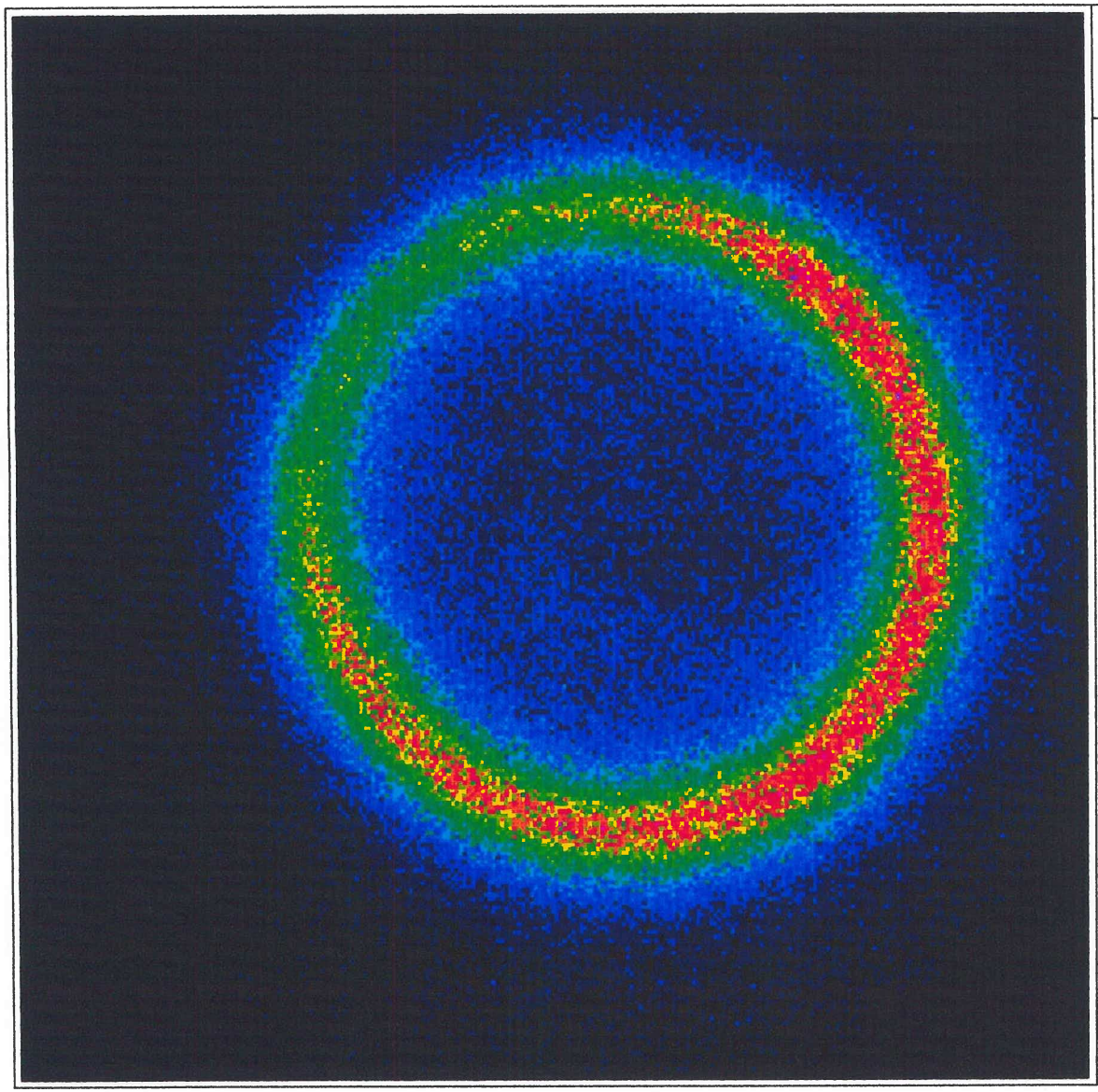


Figure 12.20





ATMOSPHERIC PHYSICS LABORATORY
UNIVERSITY COLLEGE LONDON

Triple etalon interferometer
File: TKK1058.E38

Misaligned axes

Figure 12.21

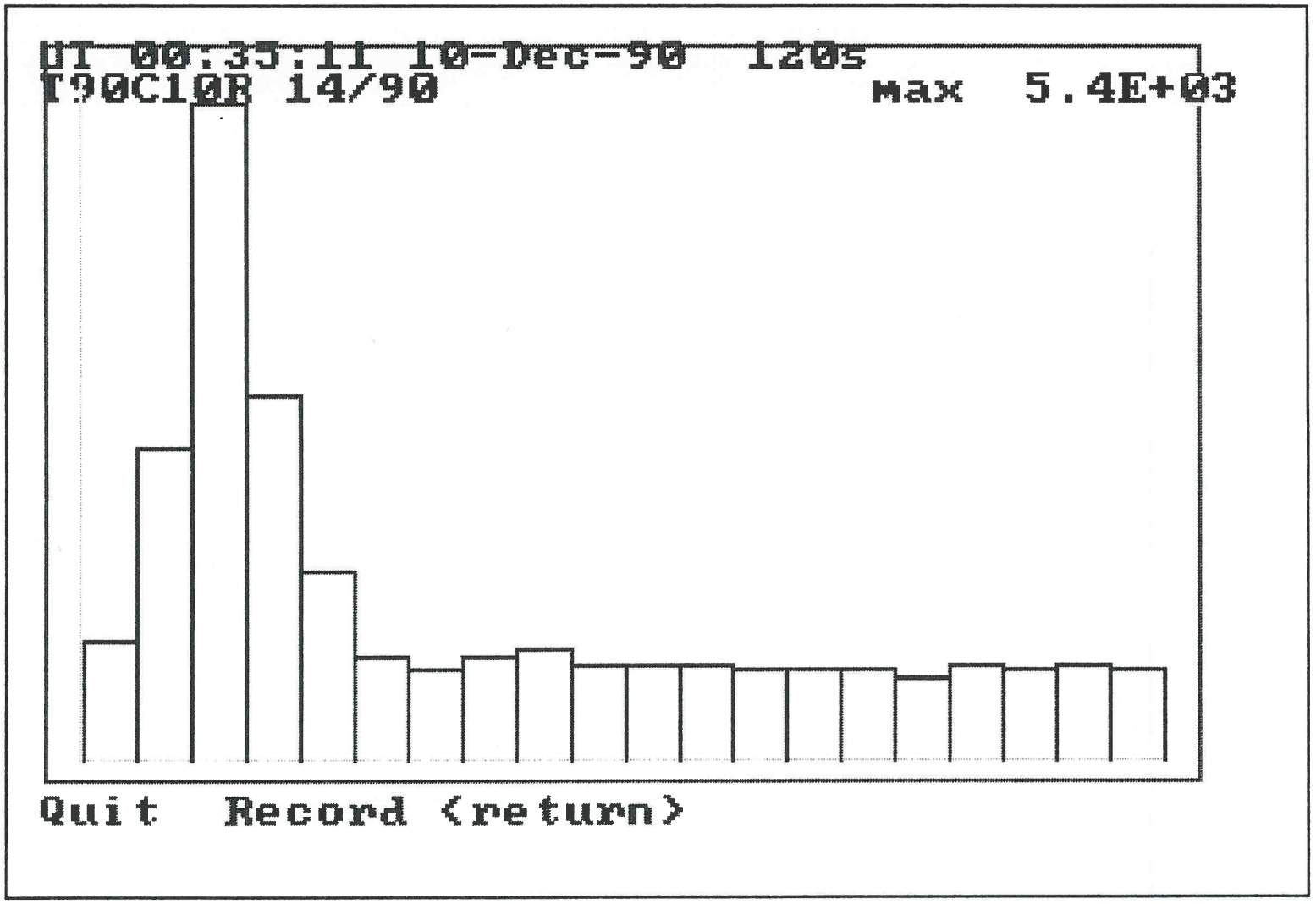


Figure 12.22 Ring Anode image from the Triple Etalon Interferometer

Figure 12.23

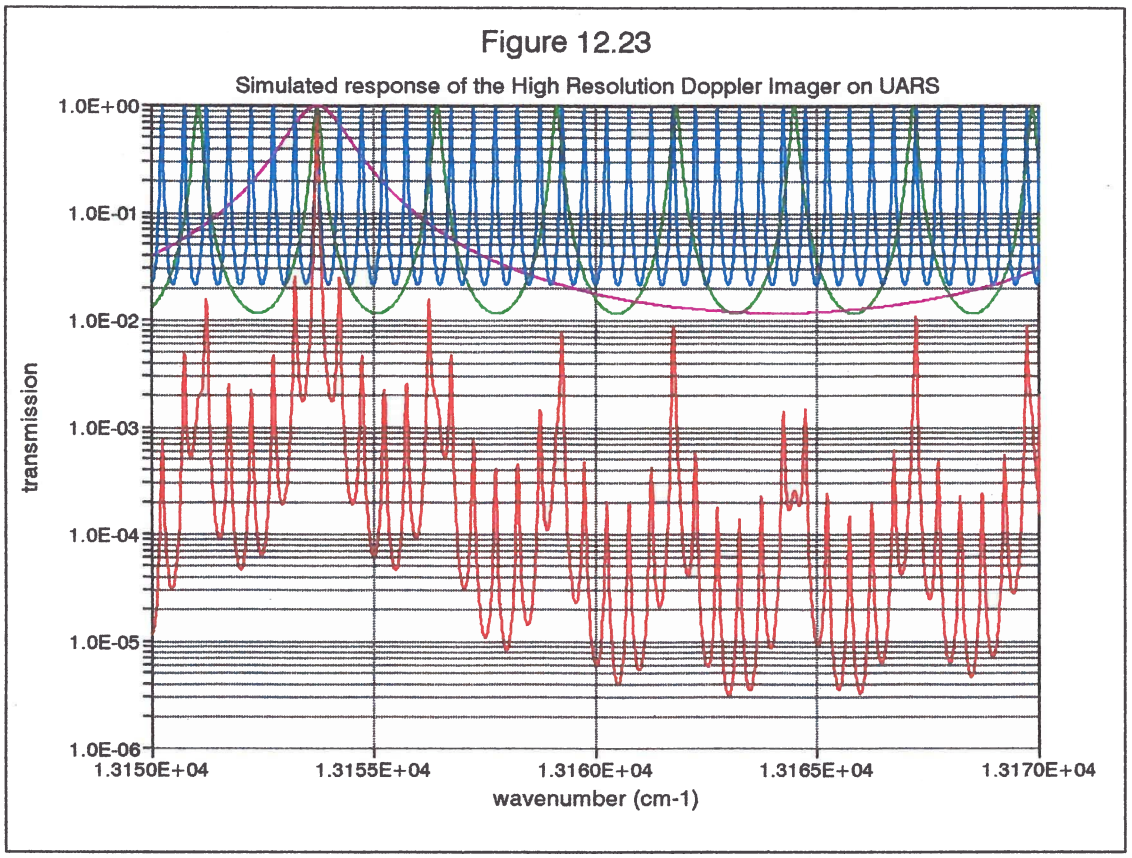
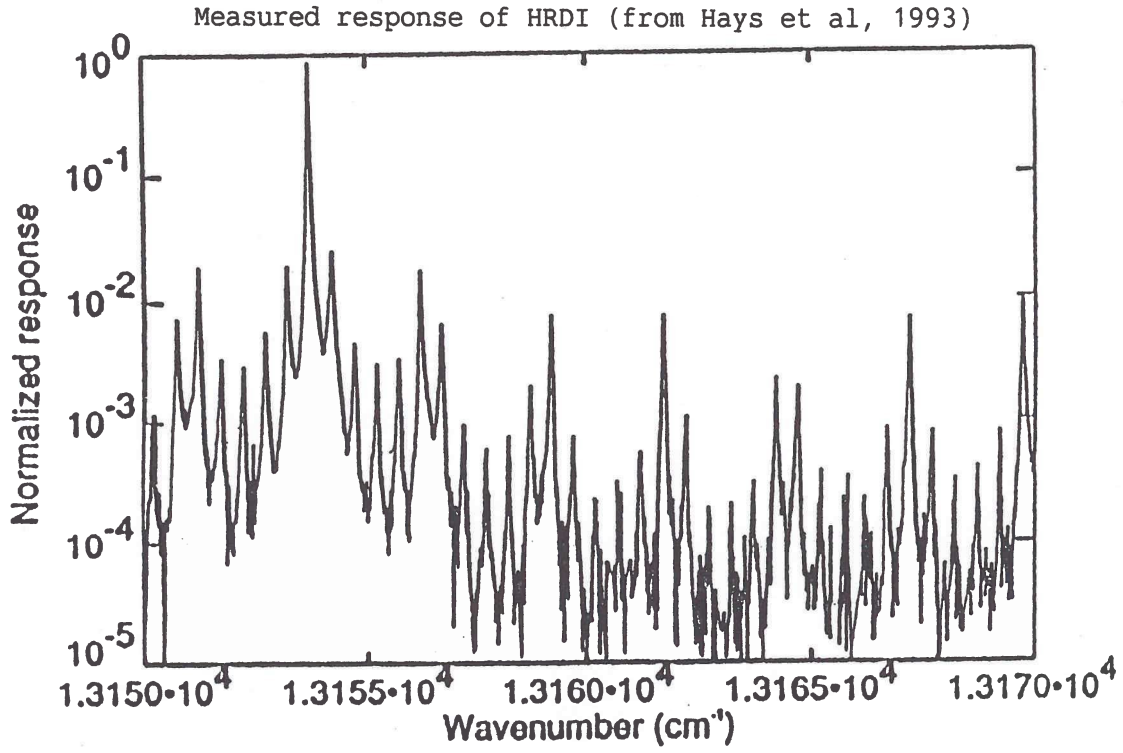


Figure 12.24



13.1 The Quattro Pro spreadsheet

Simulations of etalon responses have been performed at the APL by writing programmes in Fortran on the VAX main frame computers and the DEC workstations. However, the author has found that a commercial spreadsheet programme available for IBM-compatible personal computers, Quattro Pro, provides an extremely powerful and versatile method of performing the calculations involved and presenting them in both tabular and graphical forms.

The working environment is considerably more interactive than that provided by the mainframe computers, and it is well suited to investigations into the effects of varying critical parameters. The etalon simulations constitute an important part of this thesis, and all the graphs were produced using Quattro Pro, so it is appropriate to provide a short description of the method used.

The spreadsheet consists of "cells" which are referenced in terms of rows and columns. Some of the cells are used to enter the instrument parameters. The remaining cells each contain the appropriate formulae to calculate further instrument parameters or performance data, which they achieve by referring to the contents of other cells. The result of each calculation is displayed in the corresponding cell. The data can then be displayed in graphical form if desired. The copy facility means that each formula has to be entered only once. References to cell addresses may either be incremented when copying to other cells or they may remain absolute.

Up to 8000 rows of calculation may be performed. This is more than enough for examining the fine structure of etalon responses, and 1000 points are usually adequate. Regions of interest may be homed in on by changing plot width and plot centre variables each contained in a single cell. Peak positions may be determined to a very high degree of

accuracy by running the calculation with a very small plot width and inspecting the data table.

13.2 Computer simulation of the triple etalon response

The general method may be illustrated by referring to the spreadsheet written for the triple etalon interferometer. The other simulations in this thesis were performed along similar lines. The spreadsheet header is shown in figure 13.1, together with the first few lines of the resulting data table. 4000 rows of data were calculated in order to resolve the fine structure of the LRE when plotted over a wide range of wavelengths.

The following parameters are input to the appropriate cells:

1. Central wavelength. This is the wavelength for which the maxima of the three transmission profiles are aligned, giving the combined transmission maximum.
2. Plot width. This defines the width in nanometres of the total plot.
3. Plot Centre. The Airy functions are evaluated and plotted for 2000 points either side of this value. The plot centre may be the same as the central wavelength or it may be offset to examine a region of interest at some distance from the overall transmission maximum.
4. Nominal Gap and Reflectivity. These are entered for each etalon.
5. Offset. The transmission peak for each etalon may be offset from the central wavelength by a specified amount to simulate the effects of inaccuracies in alignment.

The following parameters are then automatically calculated:

1. Plotting increment. This is the wavelength interval between successive points, equal to the plot width divided by the number of plotting points.
2. The order of interference, for each etalon.
3. A modified gap. The order for each etalon is truncated to its integer value and the gap re-calculated. This ensures that the

transmission maxima coincide for each etalon at the central wavelength (apart from the above offset).

4. The free spectral range of each etalon in terms of wavelength and frequency.
5. The peak shift corresponding to the Doppler shift obtained from a 5 metre/second wind variation - a realistic target for instrument accuracy. It is useful to have this reminder of the practical implications of peak position shifts.

A data table is defined, with a column assigned to each etalon. Each cell in the column calculates the Airy function for its respective etalon and wavelength increment. The combined response of the three etalons is calculated in a fourth column. The information is displayed in the cells in tabular form and, at the press of a key, in graphical form. Changes in the input parameters are re-calculated in seconds to give the new response. Peak position shifts resulting from applied offsets can then be determined by inspection of the table.

13.3 Other simulations

The triple etalon spreadsheet may be used to assess the position and magnitude of the parasitic fringes and to explore the effects of different gap ratios. Other spreadsheets have been written to simulate many aspects of etalon behaviour.

One in particular has proved very useful for optimising the size of the fringe pattern image for specific instruments. It takes as its input parameters the focal length of the imaging lens, the etalon cavity length and the detector diameter. This produced the plots shown in figures 9.4 and 9.5.

The Quattro package may also be utilised for simple plotting of experimental results, as data files in ASCII format can easily be assimilated.

Figure 13.1

Spreadsheet for triple etalon simulation

Peak wavelength (nm)	Plot width (nm)	Plotting increment (nm)	Centre of plot (nm)	Doppler shift for 5 m/s (pm)
630	0.010	0.0000100	630.00	0.0105

	HRE gap	MRE gap	LRE gap
Nominal gap (mm)	10.00	1.82	0.331
Reflectivity	0.830	0.920	0.850
Refl. finesse	16.84	37.67	19.31
Order	31746.03	5777.78	1050.79
Modified gap(mm)	10.000305	1.820070	0.331065
FSR (pm)	19.85	109.04	599.55
FSR (GHz)	14.99	82.42	453.17
Offset (% FSR)	0.00	0.00	0.00

Wavelength	HRE	MRE	LRE	Combined
629.9950000	0.01690685	0.07781053	0.90600621	0.00119188
629.9950100	0.01695899	0.07809633	0.90634656	0.00120040
629.9950200	0.01701146	0.07838368	0.90668648	0.00120899
629.9950300	0.01706425	0.07867259	0.90702598	0.00121767
629.9950400	0.01711737	0.07896307	0.90736505	0.00122643
629.9950500	0.01717083	0.07925512	0.90770368	0.00123527
629.9950600	0.01722462	0.07954876	0.90804189	0.00124420
629.9950700	0.01727874	0.07984400	0.90837967	0.00125320
629.9950800	0.01733321	0.08014085	0.90871701	0.00126230
629.9950900	0.01738802	0.08043932	0.90905392	0.00127148
629.9951000	0.01744317	0.08073943	0.90939039	0.00128074
629.9951100	0.01749867	0.08104118	0.90972642	0.00129009
629.9951200	0.01755452	0.08134459	0.91006202	0.00129954
629.9951300	0.01761072	0.08164967	0.91039717	0.00130907
629.9951400	0.01766728	0.08195643	0.91073189	0.00131869
629.9951500	0.01772420	0.08226488	0.91106616	0.00132841
629.9951600	0.01778148	0.08257504	0.91140000	0.00133821
629.9951700	0.01783913	0.08288691	0.91173338	0.00134812
629.9951800	0.01789714	0.08320051	0.91206632	0.00135811
629.9951900	0.01795552	0.08351586	0.91239882	0.00136821
629.9952000	0.01801427	0.08383296	0.91273086	0.00137840
629.9952100	0.01807340	0.08415182	0.91306246	0.00138869
629.9952200	0.01813291	0.08447247	0.91339360	0.00139907
629.9952300	0.01819281	0.08479491	0.91372430	0.00140956
629.9952400	0.01825308	0.08511915	0.91405454	0.00142015

CHAPTER 14 - FUTURE DEVELOPMENTS

14.1 Overview of future work

There are two aspects to the future work of the APL: maintenance and improvement of existing field stations and participation in new projects. Both these areas rely on the refinement of existing designs and the development of new technology.

14.2 Field station improvements

The scanning mirror Fabry-Perot interferometers have been operating routinely for many years and are now largely free of problems. The main goal must be the similar routine operation of the Doppler Imaging System and the triple etalon interferometer.

14.2.1 The Doppler Imaging System

The major improvements in the DIS at Kiruna have not resulted in a proportionate increase in quality data, largely because of detector problems. The IPD assigned to the DIS after its major upgrade during 1990 and 1991 suffered from an intermittent high thermionic emission which was signal induced. It was an ITT device with an S25 photocathode, and as such had a very high sensitivity. Unfortunately, the resulting high photon counts induced a degree of dimensional instability of the image, as described in chapter seven. These two factors made precise data analysis very difficult. A cooled CCD was fitted during 1992, but the results received so far indicate that the sensitivity is well below the requirement. The cooled intensified CCD system currently under development at the APL may prove to be a suitable option, or alternatively, an up-graded IPD Signal Processing Unit with pulse pile-up rejection may overcome the image size variation.

Another problem area with this instrument is the large amount of data generated - in the region of one floppy disc per night. The present system depends on the help of local personnel in copying the data to

disc. Optical storage devices are now a more affordable option for transporting large quantities of data and this may be explored. Also, a single optical drive could possibly be shared by the other instruments at the Kiruna site.

14.2.2 The Triple Etalon Interferometer

Recent work at the APL has done much to increase our understanding of the problems with this instrument. A substantial financial investment is necessary as it is clear that the required performance will not be achieved without at least one and preferably three new etalons, made to the latest specification, with a 104mm usable aperture. The provision of a suitable detector is also a problem. The multi-ring anode will handle the high photon rates but the light intensity is likely to degrade the device quite quickly. It is clear from the experiments with the DIS that a bare CCD lacks the sensitivity for these instruments. One possibility is to use the EEV cooled intensified CCD, possibly with an external light sensor to reduce the gain of the intensifier during daylight.

14.3 The ALOMAR project

Without a doubt, this will be the major area of effort for the APL in the coming months and years. The Arctic Lidar Observatory for Middle Atmosphere Research is expected to provide a wealth of new information on the dynamics of the middle and upper atmosphere. Funding has been provided for the APL to proceed directly with the more complex option of multi-ring anode detectors, rather than starting with a simpler photomultiplier alternative. The ALOMAR receivers will consist of these detectors coupled to a double etalon interferometer. The APL has already built a successful double etalon interferometer and two capacitance stabilised laser wavelength meters for the participating German groups.

The use of the multi-ring anode detector for lidar was described in chapter eight. The demise of the ITT facility has left the APL with only one good multi-ring anode detector, and negotiations are

currently underway with a U.K. company, Photek Ltd. who, it is to be hoped, will provide a suitable device. This is largely dependent on the fabrication of a satisfactory anode. A high speed counter and store unit is currently being designed at the APL, and the aspects of interfacing with the data acquisition computers and the pulsed laser will be very challenging.

14.4 The TIDI project

TIDI, the TIMED Doppler Interferometer is part of the TIMED satellite mission (Thermospheric Ionospheric Mesospheric Energetics and Dynamics) and consists of a double etalon interferometer with a CCD detector. The low resolution etalon will be capacitance stabilised. The APL role will mainly be concerned with the provision of the etalons. As a satellite project, this presents an interesting addition to the main body of work.

ACKNOWLEDGEMENTS

During my time at the Atmospheric Physics Laboratory I have seen many people come and go. Rather than single out any individuals, I would like to congratulate all members of the group, past and present, who have contributed towards making these instruments so successful.

Outside the group, we have received much assistance from institutions throughout the world wherever we have taken our instrumentation. I would particularly like to thank Åke Steen and Urban Brändström at the Swedish Institute for Space Physics at Kiruna. They have not only given us considerable help and support with the installation and operation of our instruments, but they have also been most generous with their laboratory facilities at times when the equipment disgorged from our crates must have appeared as nothing less than a takeover bid.

Special thanks go to David Rooks for providing the photographs, both for this thesis and on many other occasions, and to Terry Dines of I.C. Optical Systems, for painstakingly meeting our ever more stringent demands for etalons. I would also like to thank Martyn Wells for many useful discussions concerning capacitance stabilised etalons during his brief stay at UCL.

Most of all, my love and thanks go to my wife Mary, for her unflagging support during the trials and tribulations of working for the APL, for sharing my enthusiasm in the good times and for picking up the pieces during the bad. This thesis has been much enhanced by her tireless proof reading, which sadly, or perhaps fortunately, has not awakened a latent passion for physics.

In conclusion, I would like to remember Arthur Dennington, who studied physics under Appleton before dedicating his life to music. It was he who started it all those years ago by bringing to my attention a small advertisement in the Daily Telegraph for what looked like quite an interesting job.....

REFERENCES

Adams J. and Manley B.W. (1966) *The mechanism of channel electron multiplication*. IEEE Transactions on Nuclear Science NS-13 p.88.

Alexander R.N. (1971) *A laboratory quality charge sensitive amplifier suitable for space applications*. IEEE Transactions on Nuclear Science NS-18 p.264.

Allington-Smith J.R. and Schwartz H.E. (1984) *Imaging Photon Detectors for optical astronomy*. Journal of the Royal Astronomical Society 25 p.267.

Armstrong E.B. (1956) *The observation of line profiles in the airglow and aurora with a photoelectric Fabry-Perot interferometer*. The Airglow and Aurora - Journal of Atmospheric and Terrestrial Physics Special supplement 5 p.366.

Aruliah A.L. (1991) *The synoptic variability of thermospheric and mesospheric winds observed using a Fabry-Perot interferometer*. Ph.D. Thesis, University of London.

Baldinger E. and Franzen W. (1956) *Amplitude and time measurement in Nuclear Physics*. Advances in Electronics and Electron Physics Vol.8 p.255.

Batten S., Rees D., Wade D. and Steen Å. (1988) *Observations of thermospheric neutral winds by the UCL Doppler imaging system at Kiruna in Northern Scandinavia*. Journal of Atmospheric and Terrestrial Physics Vol.50 No.10/11 p.861.

Boksenberg A. and Coleman C.I. (1979) *Image Photon Counting Detectors for Spaceborne Applications*. Advances in Electronics and Electron Physics Vol.52 p.355.

Bronwell A. (1953) *Advanced Mathematics in Physics and Engineering*. McGraw-Hill.

Bryant D.A. and Johnstone A.D. (1965) Review of Scientific Instruments 36 p.1662.

Carslaw H.S. and Jaeger J.C. (1959) *Conduction of heat in solids*. Clarendon Press, Oxford.

Clampin M. and Edwin R.P. (1987) *Large format imaging photon detector for astronomical spectroscopy*. Review of Scientific Instruments 58 (2) p.167.

Clampin M., Crocker J., Paresce F. and Rafal M. (1988) *Optical Ranicon detectors for photon counting imaging*. Review of Scientific Instruments 59 (8) p.1272.

Clampin M. and Paresce F. (1989) *Spatial resolution of a GaAs-photocathode RANICON*. Review of Scientific Instruments 60 (6) p.1092.

Csorba I.P. (1979) *Recent advances in the field of image intensification: the generation 3 wafer tube*. Applied Optics 18 No.14 p.2440.

Csorba I.P. (1985) *Image Tubes*. Published by Howard W.Sams.

Dance J.B. (1969) *Photoelectronic devices*. Iliffe Books.

Dolizy P. and Legoux R. (1969) *A new technology for transferring photocathodes*. Advances in Electronics and Electron Physics 28A p.367.

Durand D., Hardy E. and Couture J. (1987) *A panoramic photon-counting detector system*. Publications of the Astronomical Society of the Pacific July p.686.

Eberhardt E. (1977) *Image transfer properties of proximity focussed tubes*. Applied Optics 16 No.8 p.2127.

Eberhardt E. (1979) *A Gain model for microchannel plates*. Applied Optics 18 No.9 p.1418.

Eberhardt E. (1981) *An operational model for microchannel plate devices*. IEEE Transactions on Nuclear Science NS-28 (1) p.712.

Feller W.B., Cook L.M., Fraser G.W., Pearson J.F., Murray S.S. and Garcia M.R. (1989) *Low noise microchannel plates*. Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.1072 p.138.

Firmani C., Ruiz E., Carlson C.W., Lampton M. and Paresce F. (1982) *High resolution imaging with a two-dimensional resistive anode photon counter*. Review of Scientific Instruments 53(5) p.570.

Floryan R.F. and Johnson C.B. (1989) *Resistive anode photomultiplier tube optimum operating conditions for photon correlation experiments.* Review of Scientific Instruments 60(3) p.339.

Fordham J.L.A., Bone D.A. and Jordan A.R. (1986) *The UCL CCD-based Image Photon Counting System.* Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.627.

Fraser G.W. and Mathieson E. (1981a) *Signal location by uniform resistive anodes. (A) Square anodes in the DC limit.* Nuclear Instruments and Methods 179 p.591.

Fraser G.W. and Mathieson E. (1981b) *Signal location by uniform resistive anodes. (B) Circular anodes in the DC limit.* Nuclear Instruments and Methods 184 p.537.

Fraser G.W. Mathieson E., Lewis M. and Barstow M. (1981c) *Signal location by uniform resistive anodes. (C) Time dependent performance.* Nuclear Instruments and Methods 190 p.53.

Fraser G.W. and Mathieson E. (1981d) *Signal location using anodes of Gears design.* Nuclear Instruments and Methods 180 p.597.

Gear C.W. (1969) Proceedings for the Skytop Conference on Computer Systems in Experimental Nuclear Physics. USAEC Conf-670301 p.552.

Hamamatsu Technical Manual RES-0795 (1982) *Characteristics and applications of microchannel plates.*

Harris M. (1993) *The stability characteristics of the Atmospheric Physics Laboratory's Fabry-Perot Triple Etalon Interferometer.* UCL Astronomy III project.

Hays P.B. and the HRDI Science Team. (1992) *Remote sensing of Mesospheric winds with the High Resolution Doppler Imager.* Planetary and Space Science Vol.40 (12) p.1599.

Hays P.B., Abreu V.J., Dobbs M.E., Gell D.A., Grassl H.J. and Skinner W.R. (1993) *The High Resolution Doppler Imager on the Upper Atmosphere Research Satellite.* Journal of Geophysical Research 98 D6 p.10,713.

Hernandez G. (1986) *Fabry-Perot Interferomters.* Cambridge University Press.

Hernandez G. and Roble R.G. (1976) *Measurements of night-time thermospheric winds and temperatures. 1: Seasonal variations during geomagnetic quiet periods*. Journal of Geophysical Research 81 p.2065.

Hicks T.R., Reay N.K. and Scadden R.J. (1974) *A servo-controlled Fabry-Perot interferometer using capacitance micrometers for error detection*. Journal of Physics E: Scientific Instruments 7 p.27.

Hicks T.R., Reay N.K. and Stephens C.L. (1976) *A servo-controlled Fabry-Perot interferometer with On-line Computer Control*. Astronomy and Astrophysics 51 p.367.

Jones R.V. and Richards J.C.S. (1973) *The design and some applications of sensitive capacitance micrometers*. Journal of Physics E: Scientific Instruments 6 p.589.

Kalbitzer S. and Melzer W. (1967) *On the charge dividing mechanism in position sensitive detectors*. Nuclear Instruments and Methods 56 p.301.

Killeen T.L., Hays P.B., Kennedy B.C. and Rees D. (1982) *Stable and rugged etalon for the Dynamics Explorer Fabry-Perot interferometer*. Applied Optics Vol.21 No.21 p.3903.

Killeen T.L., Kennedy B.C., Hays P.B., Symanov D.A. and Ceckowski D.H. (1983) *Image plane detector for the Dynamics Explorer Fabry-Perot interferometer*. Applied Optics Vol.22 No.22 p.3503.

Kroller L.R. (1929) *Some characteristics of photoelectric tubes* Journal of the Optical Society of America 19 p.135.

Lampton M. and Paresce F. (1974) *The Ranicon: A resistive anode image converter*. Review of Scientific Instruments 45 (9) (1974) p.1098.

Lampton M. and Carlson C.W. (1979) *Low-distortion anodes for two-dimensional position-sensitive anodes*. Review of Scientific Instruments 50 (9) p.1093.

Laprade B.N. and Cortez J. (1983) *Curved Microchannel Plates*. Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.427 p.48.

Laprade B.N. (1989) *The high output technology microchannel plate.* Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.1072 p.102.

Laprade B.N. and Wheeler M. (1990a) *Dynamic range optimisation in microchannel plates.* Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.1243 p.132.

Laprade B.N., Reinhart S.T. and Wheeler M. (1990b) *A low noise figure microchannel plate optimised for GenIII image intensification systems.* Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.1243 p.162.

Lloyd N.D. (1985) *Measurement of neutral thermospheric winds using ground based Fabry-Perot interferometers.* Ph.D. Thesis, University of London.

Lo C.C., Lecomte P. and Lescover B. (1977) *Performance studies of prototype microchannel plate photomultipliers.* IEEE Transactions on Nuclear Science NS-24 p.302.

Mack J.E., McNutt D.P., Roesler F.L. and Chabbal R. (1963) *The PEPSIOS Purely Interferometric High-Resolution Scanning Spectrometer. 1. The Pilot Model.* Applied Optics Vol.2 No.9 p.873.

Martin C., Jelinsky P., Lampton M., Malina R.F. and Anger H.O. (1981) *Wedge and Strip anodes for centroid-finding position sensitive photon and particle detectors.* Review of Scientific Instruments 52(7) p.1067.

McCarroll W.H., Paff R.J. and Sommer A.H. (1970) *The role of Cs in the (Cs)Na_2KSb (S20) multialkali photocathode.* Journal of Applied Physics 42(2) p.569.

McNutt D.P. (1965) *PEPSIOS Purely Interferometric High-resolution Scanning Spectrometer. 2. Theory of spacer ratios.* Journal of the Optical Society of America 55 p.288.

McWhirter I., Rees D. and Greenaway A.H. (1982) *Miniature imaging photon detectors: III. An assessment of the performance of the resistive anode IPD.* Journal of Physics E: Scientific Instruments 15 p.145.

Meredith N.P. and Mugleston S.B. (1993) *The High Resolution Spectrum Analyser*. UCL Atmospheric Physics Laboratory internal report.

Owen R.B. and Awcock M.L. (1968) *One and two dimensional position sensing semiconductor devices*. IEEE Transactions on Nuclear Science NS-15 p.290.

Parkes W., Evans K.D. and Mathieson E. (1974) *High resolution position-sensitive detectors using microchannel plates*. Nuclear Instruments and Methods 121 p.151.

Radeka V. (1974) *Signal, noise and resolution in position-sensitive detectors*. IEEE Transactions on Nuclear Science NS-21 p.51.

Radeka V. and Rehak P. (1979) *Charge dividing mechanism on resistive electrode in position-sensitive detectors*. IEEE Transactions on Nuclear Science NS-26 (1) p.226.

Read P.D., Powell J.R,, van Breda I.G., Lyons A. and Ridley N.R. (1986) *Production and testing of microchannel plate intensifier*. Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.627 p.645.

Rees D., McWhirter I., Rounce P.A., Barlow F.E. and Kellock S.J. (1980) *Miniature Imaging Photon Detectors*. Journal of Physics E: Scientific Instruments 13 p.763.

Rees D., McWhirter I., Rounce P.A. and Barlow F.E. (1981a). *Miniature Imaging Photon Detectors. II: Devices with Transparent Photocathodes*. Journal of Physics E: Scientific Instruments 14 p.229.

Rees D., McWhirter I., Hays P.B. and Dines T. (1981b). *A Stable, Rugged, Capacitance-stabilized Piezoelectric Scanned Fabry-Perot Etalon*. Journal of Physics E: Scientific Instruments 14 p.1320.

Rees D., Rounce P.A., McWhirter I., Scott A.F.D., Greenaway A.H. and Towlson W. (1982a). *Observations of Atmospheric Absorption Lines from a Stabilized Balloon Platform and Measurements of Stratospheric Winds*. Journal of Physics E: Scientific Instruments 15 p.191.

Rees D., Rounce P.A., Charleton P., Fuller-Rowell T.J., McWhirter I. and Smith K. (1982b). *Thermospheric Winds during the Energy Budget Campaign: Ground-based Fabry-Perot Observations Supported by Dynamical Simulations with a Three-dimensional, Time-dependent Thermospheric Model*. Journal of Geophysical Research 50 p.202.

Rees D., Fuller-Rowell T.J., Lyons A., Killeen T.L., and Hays P. (1982c) *Stable and rugged etalon for the Dynamics Explorer Fabry-Perot interferometer. 1. Design and construction*. Applied Optics Vol.21 No.21 p.3896.

Rees D. and Greenaway A.H. (1983) *Doppler Imaging system; an opticals device for measuring vector winds. 1: General principles*. Applied Optics Vol.22 No.7 p.1078.

Rees D., Greenaway A.H., Gordon R., McWhirter I., Charlton P.J. and Steen Å. (1984). *The Doppler Imaging System: initial observations of the auroral thermosphere*. Journal of Planetary and Space Science Vol.32 No.3 p.273.

Rees D., McWhirter I. and Wade D. (1987) *Development of a Doppler Wind Lidar* System for atmospheric wind measurements. Proceedings of 8th. ESA symposium Sunne, Sweden. May 1987.

Rees D. and Steen Å. (1987) *A new technique for rocket and space-borne investigations of time-dependent auroral morphology*. Proceedings of 8th. ESA symposium Sunne, Sweden. May 1987.

Rees D., McWhirter I., Aruliah A. and Batten S. (1992) *Upper Atmospheric Wind and Temperature Measurements using Imaging Fabry-Perot Interferometers*. World Ionosphere / Thermosphere Study (WITS) handbook, edited by C.H.Liu, published by SCOSTEP,

Rouget H. and Baud C. (1979) *Negative electron affinity photoemitters*. Advances in Electronics & Electron Physics 48 p.1.

Schmidt K.C. and Hendee C.F. (1966) *Continuous channel electron multiplier operated in the pulse saturated mode*. IEEE Transactions on Nuclear Science NS-13 p.100.

Schwartz H.E. and Lapington J.S. (1985). *Optimisation of wedge and strip anodes*. IEEE Transactions on Nuclear Science NS-32 (1) p.433.

Schwider J. (1965) *Entkopplungsmöglichkeiten von Fabry-Perot-Interferometern*. Optica Acta 12 p.65.

Siegmund O.H.W., Clothier S., Thornton J., Leman J., Harper R., Mason I.M. and Culhane J.L. (1983) *Application of the Wedge and Strip Anode to position sensing with microchannel plates and proportional counters*. IEEE trans. Nuclear Science NS-30 1 p.503.

Skinner W.R., Hays P.B. and Abreau V.J. (1987) *Optimization of a triple etalon interferometer*. Applied Optics Vol.26 No.14 p.2817.

Soul P.B (1971) *Operational properties of channel-plate electron multipliers*. Nuclear Instruments and Methods 97 p.555.

Stümpel J.W., Sanford P.W. and Goddard H.F. (1973). *A position sensitive proportional counter with high spatial resolution*. Journal of Physics E: Scientific Instruments Vol.6 p.397.

Timothy J.G. (1981) *Curved microchannel plate arrays*. Review of Scientific Instruments 52 (8) p.1131.

Vaughan J.M. (1989) *The Fabry-Perot Interferometer: history, theory, practice and applications*. Adam Hilger series on Optics and Optoelectronics.

Wells M., Atherton P.D., Hicks T.R. and Reay M.K. (1978) *Multi-etalon servo-controlled Fabry-Perot interferometers*. Proceedings of 4th International Colloquium on Astrophysics, Trieste. p.586.

Wiza J.L. (1979) *Microchannel Plate detectors*. Nuclear Instruments and Methods 162 p.587.

PAPERS CO-WRITTEN BY THE AUTHOR

Rees D., McWhirter I., Rounce P.A., Barlow F.E. and Kellock S.J. (1980) *Miniature Imaging Photon Detectors.* Journal of Physics E: Scientific Instruments 13 p.763.

Rees D., Rounce P.A. and McWhirter I. (1980) *Middle Atmosphere winds from a Balloon-Borne Fabry-Perot Interferometer.* Proceedings of the Vth ESA-PAC Symposium on European Rocket & Balloon Programmes & Related Research - Bournemouth (UK), 14-18 April 1980 p.89.

Rees D., McWhirter I., Hays P.B. and Dines T. (1981) *A Stable, Rugged, Capacitance-stabilised Piezoelectric Scanned Fabry-Perot Etalon.* Journal of Physics E: Scientific Instruments 14 p.1320.

Rees D., McWhirter I., Rounce P.A. and Barlow F.E. (1981) *Miniature Imaging Photon Detectors. II. Devices with Transparent Photocathodes.* Journal of Physics E: Scientific Instruments 14 p.229.

Rees D., Rounce P.A., McWhirter I., Scott A.F.D., Greenaway A.H. and Towlson W. (1981) *Observations of Atmospheric Absorption Lines from a Stabilised Balloon Platform and Measurements of Stratospheric Winds.* Journal of Physics E: Scientific Instruments 15 p.191.

Greenaway A.H., Lyons A., McWhirter I., Rees D. and Cochrane A. (1982) *Miniature imaging photon detector.* Proceedings of the Society of Photo-Optical Instrumentation Engineers Vol.331 - Instrumentation in Astronomy IV p.365.

McWhirter I., Rees D. and Greenaway A.H. (1982) *Miniature imaging photon detectors III. An assessment of the performance of the resistive anode IPD.* Journal of Physics E: Scientific Instruments 15 p.145.

Rees D., Rounce P.A., Charleton P., Fuller-Rowell T.J., McWhirter I. and Smith K. (1982) *Thermospheric Winds during the Energy Budget Campaign: Ground-based Fabry-Perot Observations Supported by Dynamical Simulations with a Three-dimensional, Time-dependent Thermospheric Model.* Journal of Geophysical Research 50 p.202.

Rees D., Greenaway A.H., McWhirter I. and Lyons A. (1983) *Imaging Photon Detectors for One and Two-dimensional Applications*. Quantum Electronics and Electro-Optics. Ed. P.L. Knight. John Wiley and Sons p.419.

Rees D., McWhirter I. and Cochrane A. (1983) *Time-resolved Imaging and Spectroscopy using Imaging Photon Techniques*. Quantum Electronics and Electro-Optics. Ed. P.L. Knight, John Wiley and Sons p.415.

Rees D., Greenaway A.H., Gordon R., McWhirter I., Charleton P.J. and Steen Å. (1984) *The Doppler Imaging System: Initial observations of the auroral thermosphere*. Planetary and Space Science 32 p.273.

Rees D., McWhirter I. and Wade D. (1987) *Development of a Doppler wind lidar system for atmospheric wind measurements*. Proceedings of the 8th ESA Symposium on European Rocket and Balloon Programmes and Related Research sponsored by ESA/SSC. Sunne, Sweden. 17-23 May 1987, SP-270 p.99.

Rees D., McWhirter I. and Klein V. (1987) *Development of a detector system for incoherent Doppler lidar wind measurements in the lower and middle atmosphere*. Proceedings of the 14th Annual meeting on Atmospheric studies by Optical Methods 17-22 August 1986.

Rees D., McWhirter I., Aruliah A. and Batten S. (1989) *Upper Atmospheric Wind and Temperature Measurements using Imaging Fabry-Perot Interferometers*. World Ionospheric/Thermospheric Study (WITS) Handbook (SCOSTEP). Ed. C.H. Liu. Vol.2 pp.188-223

OTHER PUBLICATIONS FOR WHICH THE WORK IN THIS THESIS PROVIDED THE ENABLING TECHNOLOGY

1. FABRY-PEROT INTERFEROMETERS

Rees D., Charleton P.J., Lloyd N.D., Steen Å. and Witt G. (1983) *Interferometric and Doppler imaging studies of the auroral thermosphere from Kiruna Geophysical Institute.* Proceedings from Sixth ESA Symposium on European rocket & balloon programmes.

Rees D. and Greenaway A.H. (1983) *The Doppler Imaging System: An Optical Device for Measuring Vector Winds. I. General Principles.* Applied Optics 22 pp.1078-1083.

Rees D., Philbrick C.R., Carlson M., Charleton P.J. and Fuller-Rowell T.J., (1983) *In Situ and Remote Sensing of Thermospheric Winds during the Energy Budget Campaign.* "Advances in Space Research" 2 No.10 pp.129-132.

Rees D., Lloyd N.D., Charleton P.J., Carlson M., Murdin J. and Häggström I. (1984) *Comparison of Plasma Flow and Thermospheric Circulation over Northern Scandinavia using Eiscat and a Fabry-Perot Interferometer.* Journal of Atmospheric Research 46 pp.545-564.

Lloyd N.D. (1985) *Measurement of neutral thermospheric winds using ground-based Fabry-Perot interferometers.* Ph.D. Thesis, University of London.

Rees D., Charleton P.J., Carlson M. and Rounce P. (1985) *High latitude thermospheric circulation during the Energy Budget Campaign.* Journal of Atmospheric and Terrestrial Physics 47 pp.195-232.

Schmidlin F.J., Carlson M., Rees D., Offerman D., Philbrick C.R. and Widdel H.U. (1985) *Wind structure and variability in the middle atmosphere during the November 1980 Energy Budget Campaign.* Journal of Atmospheric and Terrestrial Physics 47 pp.183-193.

Stewart R.D., Smith R.W., Rees D., Dudeney J.R. and Rodger A.S. (1985) *First measurements of thermospheric winds in Antarctica by an optical ground-based method.* Nature 317 pp.45-47.

Rees D. and Wells M. (1986) *A precision wavelength meter for a pulsed laser*. Journal of Physics E: Scientific Instruments 19 pp.301-308.

Smith R.W., Rees D., McCormac F.G. and Charleton P.J. (1986) *Two-station observations of thermospheric winds in the auroral zone and the polar cap*. Journal of Atmospheric and Terrestrial Physics Vol.48 No.1 pp.97-105.

Stewart R.D., Dudeney J.R., Rodger A.S., Smith R.W. and Rees D. (1986) *Neutral wind measurements by Fabry-Perot interferometry in Antarctica*. Proceedings of the 13th Annual Meeting on Upper Atmosphere Studies by Optical Methods, Oslo, Norway, 1985. Ed. Karl Maseide. pp.229-239.

Rees D. and Fuller-Rowell T.J. (1987) Comparison of theoretical models and observations of the thermosphere and ionosphere during extremely disturbed geomagnetic conditions during the last solar cycle. Advances in Space Research Vol.7 No.8 (8)27-(8)38.

Rees D., Lloyd N.D., Fuller-Rowell T.J. and Steen Å. (1987) *Observations of the variations of thermospheric winds in Northern Scandinavia between 1980 and 1986: a study of geomagnetic activity effects during the last solar cycle*. Advances in Space Research Vol.7 No.8 pp.(8)39-(8)47.

Batten S., Rees D., Wade D. and Steen Å. (1988) *Observations of thermospheric neutral winds by the UCL Doppler imaging system at Kiruna in Northern Scandinavia*. Journal of Atmospheric and Terrestrial Physics Vol.50 No.10/11 pp.861-888.

Batten S.M. (1989) *The measurement of neutral winds in the thermosphere using a Doppler imaging system*. Ph.D. Thesis, University of London.

Aruliah A., Rees D. and Steen Å. (1990) *The Seasonal Behaviour of High-latitude Thermospheric Winds over one Solar Cycle*. Quarterly Journal of the Royal Astronomical Society 31 pp.405-410.

Rees D., Aruliah A., Fuller-Rowell T.J., Wickwar V.B. and Sica R.J. (1990) *Winds in the Upper Mesosphere at Mid-Latitude: First Results using an Imaging Fabry-Perot Interferometer*. Geophysical Research Letters Vol.17 No.9 pp.1259-1262.

Aruliah A. (1991) *The Synoptic Variability of Thermospheric and Mesospheric winds observed using a Fabry-Perot Interferometer*. Ph.D. Thesis, University of London.

Aruliah A.L., Rees D. and Fuller-Rowell T.J. (1991) *The Combined effect of Solar and Geomagnetic Activity on High Latitude Thermospheric Neutral Winds*. Journal of Atmospheric and Terrestrial Physics 53 (6/7) pp.467-483.

Aruliah A., Rees D. and Steen Å. (1991) *Seasonal and Solar Cycle Variations in High Latitude Thermospheric Winds*. Geophysical Research Letters Vol.18 No.11 pp.1983-1986.

Aruliah A.L., Farmer A.D., Fuller-Rowell T.J., Wild M.N. and Rees D. (1992) *An Equinoctial Asymmetry in the High-Latitude Thermosphere and Ionosphere*. Submitted to Geophysical Research Letters.

Aruliah A.L., Farmer A.D., Rees D. and Steen Å. (1992) *The Seasonal Behaviour of High-Latitude Thermospheric Winds and Plasma Velocities Observed Over One Solar Cycle*. Submitted to Geophysical Research Letters.

Aruliah A.L., Rees D., Steen Å., Pellinen R., Kyro E. and Henricksen K. (1992) *Simultaneous Observations of Thermospheric Winds from Three Stations in Northern Scandinavia*. To be submitted to Planetary and Space Science.

Greenaway A.H. (1982) *Diffraction-limited pictures from single turbulence-degraded images in astronomy*. Optical Communications 42 pp.157-161.

Rees D., Conboy J., Heinz W., Witt G., Stegman J. and Murtagh D.P. (1986) *Auroral and Airglow observations using a grating spectrometer with an imaging photon detector*. Journal of Physics E: Scientific Instruments 19.

Rees D., Hallinan T.J., Stenbaek-Nielsen H.C., Mendillo M., Baumgardner J. and Barnard P.A. (1986) *Optical observations of the AMPTE 'Artificial Comet' release from the northern hemisphere stations*. Nature 320 pp.740-748.

Rees D., Meredith N. and Wallis M. (1986) *Optical observations of ion tail structures and velocity flows in comet Giacobini-Zinner near the time of the I.C.E. encounter*. Advances in Space Research Vol.5 No.12 pp.267-273.

Rees D., Meredith N. and Wallis M. (1986) *Optical observations of ion tail structures and velocity flows in comet Halley during the period of the five spacecraft encounters*. Advances in Space Research Vol.5 No.12 pp.255-261.

Rees D., Meredith N. and Wallis M. (1986) *Optical Observations of the neutral and ion coma of comet Halley*. ESA SP-250 No.1 pp.493-498.

Rees D. (1987) *The March 21, 1985 Ampte Barium Magnetotail release: imaging and Doppler imaging observations from the NASA airborne observatory*. Monograph on "Magnetotail Physics. Ed. Anthony Ty Lui, published by the John Hopkins University Press, pp.353-361.

Rees D. (1987) *New technique for rocket and balloon-borne instrumentation for investigations of the structure, dynamics and electrodynamics of the middle atmosphere*. Proceedings of the 8th ESA Symposium on European Rocket and Balloon Programmes and Related Research sponsored by ESA/SSC. Sunne, Sweden 17-23 May 1987, SP-270 pp.121-130.

Rees D., Meredith N. and Wallis M. (1987) *Comparison of Ion Structures in Comet Halley and Giacobini-Zinner*. Planetary and Space Science 35 pp.299-311.

Rees D. and Steen Å. (1987) *A new technique for rocket and space-borne investigations of time-dependent auroral morphology*. Proceedings of the 8th ESA Symposium on European Rocket and Balloon Programmes and Related Research sponsored by ESA/SSC. Sunne, Sweden 17-23 May 1987, SP-270 pp.294-254.

Wallis M., Meredith N. and Rees D. (1987) *Gas coma of comet Giacobini-Zinner: emission from grains 1 & 2*. Papers presented at the 'International Symposium on the diversity and similarity of comets' Brussels, Belgium 6-9 April 1987. ESA Special Publication SP-278.

Meredith N.P., Wallis M.K. and Rees D. (1989) *The Anomalous Behaviour of C2 in P/Borrelly 1987p*. Monthly Notes of the Royal Astronomical Society 240 pp.647-655.

Wallis M.K., Meredith N. and Rees D. (1989) *Gas coma of Comet Giacobini-Zinner: emission from grains*. Advances in Space Research Vol.9 No.3 pp.213-216.

Meredith N.P. (1990) *Narrow-band Imaging and Doppler Imaging of Natural and Artificial Gas and Plasma Clouds in the Interplanetary Medium and in the Earth's Atmosphere*. Ph.D. Thesis, University of London.

Meredith N.P., Wallis M.K. and Rees D. (1992) *Narrow-band IPD images of cometary CN and C2: the effect of the solar activity on coma scales*. Monthly Notes of the Royal Astronomical Society 254 pp.693-704.