INFRARED PHOTOMETRY OF INTERACTING GALAXIES

by

Stephen Alan Morris B.Sc., A.R.C.S.

A Thesis submitted for the Degree

of

Doctor of Philosophy of the University of London

and for the

Diploma of Imperial College

March 1984

Astronomy Group
Blackett Laboratory
Imperial College
London SW7 2AZ

Abstract

Bursts of star formation in galactic nuclei have been put forward as an explanation for anomalous UBV colours in certain peculiar galaxies, and as a possible energy source for some types of active galaxies. The triggering mechanism for this phenomenom is not known, but gravitational interaction between neighbouring galaxies is thought to be a likely candidate. Dust is often present in such regions, and so 10 μm observations of interacting galaxies ought to provide evidence as to whether interaction will initiate such bursts of star formation in the nuclei of galaxies.

To carry out such an observing program, a 10 μm photometer was designed and constructed, based around an Si:As photoconductor. The physics of this device is analysed to obtain the maximum sensitivity attainable under the observing conditions expected, and the results of laboratory tests on the photometer reported. A microcomputer-based data acquisition and reduction system to be used with the photometer was built and programmed, and this system, together with its software, is described.

Two trips were made to the 1.5 m infrared flux collector at Tenerife, in autumn 1980, and autumn 1981. For the second trip, the photometer was redesigned to make use of an image intensifier guiding system gained during the year. Bad weather on both trips prevented any useful observations being done, but an estimate of the photometer sensitivity was obtained.

Finally, 10 μm and 20 μm observations of NGC 3227 (one of the Arp 94 pair) made by R.D. Joseph and collaborators are analysed and discussed in light of the above theory.

To My Parents

Acknowledgements

I would first of all like to thank my supervisor, Bob Joseph, for the guidance he has given me over the period of this Ph.D. His enthusiasm, keen insight, and ability to get instantly to the heart of a problem have been a source of constant inspiration.

I would also like to thank Peter Meikle, Martin Kessler, Gillian Wright, Simon Chase, James Graham, and Norna Robertson for their help in the IRPC observations, and for much else besides. In matters electronic, Jim Allen, Roger Sidey and Mick Bartholomew provided much wisdom, as did Jack Crabtree, Bill Stannard, and the rest of the workshop staff in matters mechanical. Mike Selby must also be mentioned, for the loan of much equipment.

Practically all of the writing, and most of the numerical analysis was done while I was holding the post of Applications Programmer/Starlink Node Manager at the Institute of Astronomy, Cambridge. The use of their superb library was much appreciated, as were Martin Ward's helpful comments on Chapter 6, and the loan of drawing materials by Richard Sword. I also acknowledge use of the Starlink VAX 11/780 for the numerical analysis, and for the preparation of this thesis.

I would like to thank the S.E.R.C. for the award of a postgraduate grant, and Imperial Chemical Industries, who provided a much-needed source of funds in the shape of a postgraduate scholarship.

And finally, I would like to thank Gill, my wife-to-be, for all of her love, support, and encouragement during this endeavour.

CONTENTS

CHAPTER 1

Star Formation in Interacting Galaxies

1.1) Introduction

Recently, there has been much work done on the role that star formation plays in the energetics of galaxies. Studies have indicated that many galaxies are experiencing rapid bursts of star formation, and that this process may be responsible for the large infrared luminosities observed in some of them. An unanswered question though, is what is the cause (or causes) of the star formation burst? One promising candidate is interaction with nearby galaxies. Models of galactic interaction have concluded that there would be mass transfer between the components, and it is possible that this could be enough to initiate the star formation. A definite correlation between galaxy interaction and star formation bursts would be strong evidence for this theory.

This chapter examines the evidence for the proposition that star formation can be caused by galactic interaction. The importance of infrared observations in resolving this question is highlighted, and a programme of observations suggested.

1.2) Star Formation in Normal Galaxies

In order for an interstellar cloud to undergo collapse and form stars, the self gravity of the cloud must overcome the internal pressure and temperature (the Jeans criterion). For a given temperature, this implies a minimum density for a given mass of gas. Under typical interstellar conditions, a cloud of mass $10^3 M_\odot$ or

greater could collapse under self gravity (Larson 1978). Most interstellar clouds are smaller than this however, and so "spontaneous" star formation seems to be an inefficient method of producing stars.

Despite this, ongoing star formation activity seems to be a fairly common event. Numerical models by Tinsley (1968) showed that the colours of galaxies could be attributed to their star formation history. Searle et al. (1973) found that the star formation rate in galaxies is falling exponentially with time, but the rate of decline is very different for galaxies of the same morphological type. Also, the star formation rate is highly irregular over regions ~few hundred parsecs in diameter, and over time scales of 10^8 years.

Radio measurements of spiral galaxies by Bierman (1976) have shown that many galaxies are too blue to be consistent with a monotonic decline in the rate of star formation, and that bursts of star formation are needed to match theoretical colours with observed ones. Radio observations of bright spiral galaxies by Condon (1980) and Condon et al. (1982) have had to invoke bursts of star formation in order to explain the observed luminosity, and this mechanism was needed by Rieke and Lebofsky (1978) to explain their low value of M/L (~0.2) for a selection of spiral galaxies.

Given this selection of results, it does seem that star formation must be stimulated in some manner, i.e. that the interstellar matter is induced to collapse by some external event. Such a process occurs when an interstellar gas cloud is in collision with other material. The resulting shock will cause the collapse of the cloud (or part of it) by enhancing the density of the material. For example, with a relative velocity of 10 km s^{-1}, and assuming a post shock temperature of 10 K, a cloud of initial density $10^7 m^{-3}$

could be compressed to a density of ~3 x 10^{10} m^{-3}, sufficient for a 1 $M_\odot$ protostar to collapse (Larson 1978). Certain conditions have to be met, however. Assuming there is sufficient gas for star formation, the most important is that the shock must exist long enough for gravitational collapse to become significant, or else the cloud will return to its original state. The second is the orientation of the magnetic field of the cloud; magnetic support against collapse is reduced in that part of the shock propogating parallel to the field (Elmegreen 1978). The numerical simulations of Woodward (1976) have provided support for such a theoretical model. He has found that as a cloud passes through a shock, it will flatten, and fragment into smaller sections, which will then collapse.

Assuming then that shocks can induce star formation, what mechanisms can produce the shocks? The following processes have been suggested (Blitz 1980, Madore 1980): radiation from OB stars, cloud-cloud collisions, supernovae, stellar winds, spiral density waves, and galaxy-galaxy interactions. These will be considered in the following sections, with interacting galaxies being deferred to section 1.3.6.

1.2.1) Radiation from OB Stars

It is well known that early-type stars are formed in groups. Blaauw (1964), on examining some of these groups within the Milky Way, found subgroups within these groups, in different evolutionary stages. He noted that the evolutionary stage was a direct function of position within the group, the youngest stars being at one end, and the oldest stars at the other. Such a sequence can be explained by assuming that one subgroup of OB stars brings about the formation of an adjacent subgroup.

A model along these lines has been advanced by Elmegreen and Lada (1977), and Lada et al. (1978). They suggest that the HII region around a group of newly formed OB stars will expand into the associated molecular cloud, forming a shock front as it does so. The shock front will eventually separate from the ionization front, and material will accrete between the two. The enhanced density of this material leads to gravitational collapse, and the formation of another subgroup. The predicted separation of 10 to 20 pc. in distance, and $2 - 4 \times 10^5$ years in time agrees well with observations.

1.2.2) Cloud - Cloud Collisions

Shock formation due to the random direct collision of two gas clouds is another triggering mechanism. There is uncertainty though, over both the efficiency of this process (cloud magnetic fields may inhibit shock formation), and of the frequency of such collisions (Elmegreen 1978). At the moment, only NGC 1333 has been put forward as a candidate for this method (Loren 1976).

1.2.3) Supernovae

Another gas compression mechanism, suggested by Opik (1953), is that of supernova blast waves. The attractive feature of this is that it can generate quite high star formation rates, a fact noted by Loose and Fricke (1981). It has been used by Harwit and Pacini (1975) to explain the strong infrared emission of some galaxies, a rate of one supernova per century triggering enough star formation to give an infrared luminosity of $\sim 10^{37}$ W. Closer to home, Herbst and Assousa (1978) suggest that supernovae may be the main cause of star formation in the Milky Way, at least at the distance of the sun from the galactic centre.

1.2.4) Stellar Winds

In an O star, the stellar wind can have a terminal velocity of greater than 3000 km s^{-1}. With a mass loss rate of $10^{-6}M_{\odot}yr^{-1}$, over the lifetime of the star the total energy output will be some $10^{44}J$, which is comparable to the energy output of a type II supernova. This wind may sweep up material and initiate star formation. As yet, there is no hard evidence to support this idea, but the young object CRL-961 in the Mon OB1 region may have formed this way (Blitz 1980).

1.2.5) Spiral Density Waves

One mechanism that may be important in spiral galaxies is that of compression by a spiral density wave. Put forward by Lin and Shu (1964) as a way of explaining the large scale structure of galaxies, it is a travelling enhancement in the local density. Shock fronts can develop in the wave (Roberts 1969, Jones and Tremaine 1979) which can trigger the collapse of material, initiating subsequent star formation. Material leaving the density wave is decompressed, causing star formation to cease. This model explains the incidence of HII regions in the spiral arms of galaxies, and there does seem to be a correlation between the number of HII regions and the degree of compression (Lynds 1980). For our galaxy at least, the spiral density wave may initiate (either directly, or through subsequent supernovae) up to 60% of all star births (Kaufman 1979).

1.2.6) Implications for Galactic Structure

That star formation can be induced by external events has consequences for galactic structure. Numerical simulations by Mueller and Arnett (1976) produced small scale spiral structure in galaxies on

this assumption. Their model required that the galaxy be divided into cells, and that star formation in one cell initiated star formation in adjacent cells by means of supernova blast waves. Differential rotation then produced the spiral pattern. To prevent the whole galaxy going supernova at one time, they introduced a "regeneration" time for the gas in a cell, during which it cooled, and no star formation could take place. A rate of three supernova per century produced a suitable pattern.

This model was taken up and expanded by Gerola and Seiden (1978) who introduced a random star formation probability. Their model, christened the "Stochastic Self Propagating Star Formation" model (SSPSF) was able to produce large scale spiral structure in galaxies, with a similar supernova rate to that of Mueller and Arnett (one per century) being required. A link between SSPSF and the spiral density wave was suggested by Seiden et al. (1982), who noted that if the regeneration time were long enough, their test galaxies had two armed symmetry for about 40% of the time. A density wave initiated by the SSPSF may restart star formation in recovering regions, instead of these regions having to wait for some spontaneous event.

Further support for SSPSF comes in the verification of the prediction that spiral type will be correlated with rotational velocity - most star formation will occur in galaxies that rotate fastest, as fresh material is brought adjacent to current star formation regions (Seiden and Gerola 1979).

1.3) Interacting Galaxies

It was thought at one time that galactic collisions were rare events, this assumption being based on the visible sizes of galaxies and their relative separations. Recent observations however (Faber and

Gallagher (1979), and references therein), have revealed that many galaxies have much of their mass as non-luminous matter. Spiral galaxies are thought to possess halos that may extend as far out as several hundred kiloparsecs (Einasto et al. 1974, Ostriker et al. 1974). Given this larger cross-section, the probability of collision and interaction become much greater. Indeed, it is possible that our galaxy has a massive halo (Hawkins 1983), and may be interacting with the Large Magellanic Cloud (Spight and Grayzeck 1979).

Certainly, the preferred interpretation of peculiar structure in galaxies is one of interaction rather than explosion, e.g. NGC 520 (Krienke 1975, Barbieri and di Tullio 1979, Stockton and Bertola 1980), NGC 1275 (Adams 1976, Metik and Pronik 1979), M82 (Gottesman and Weliachew 1977, Solinger et al. 1977). In addition, signs of interaction in the distribution of HI around galaxies are detected by 21 cm observations, even where there is no obvious evidence in the morphology, as in the case of NGC 4631 / NGC 4656 (Weliachew et al. 1978).

There are two ways that galaxies can interact, by gravitational attraction, or by the interaction of their magnetic fields. If magnetic fields did play a significant role in the interaction, then one would expect to see non-thermal emission from energetic charged particles within them. Radio observations (Allen et al. 1973, Burke and Miley 1973, Allen and Sullivan 1973, von Kapp-Herr et al. 1977) have failed to find emission from the bridges and tails of interacting galaxies, and so it can be concluded that the observed interaction is by means of gravity alone.

1.3.1) Dynamics of Galaxy Interactions

Some of the first work on predicting the behaviour of

galaxies interacting under their mutual gravitational attraction was carried out by Holmberg in 1941. Simulating two galaxies by 74 lamps (37 lamps per galaxy), he used a photocell to find the luminous flux on each lamp due to the rest, and then used this as a measure of the relative force on each galactic component. Following these lamps through a number of interactions, he concluded that spiral arms developed in the galaxies, and that the loss of energy from the centre of mass motion was large enough for mergers to occur.

Later analytical work, based on the impulsive approximation of Spitzer (1958), has tended to confirm this latter conclusion. In essence, this approximation makes the assumption that the time over which the interaction occurs is short enough for the internal structures of the galaxies not to change appreciably, although the velocities of the constituent stars do. Alladin (1965) found that for significant interpenetration of the two galaxies, the increase in internal energy could be of the order of its initial value. As this can be about the same as the centre of mass energy, the energy given up to the random internal motions of the stars can be enough to cause a merger. This conclusion has been verified by further work by Lauberts (1974) and Sastry and Alladin (1979).

1.3.2) Computer Simulation Techniques

The advent of digital computers has led to an increase in the scale of the problems tackled. In practically all cases, galaxies are simulated by a large number of massive points, and the system allowed to develop in appropriate time-steps. Much work has gone into devising suitable algorithms to carry out these simulations, since the direct approach (each particle interacting with every other) becomes expensive in computer time for relatively small numbers of particles.

Early investigators (e.g. Toomre and Toomre 1972) used the restricted three body approach, where galaxies are simulated by a central mass and massless stars (massless rings in the case of Toomre 1974). The drawback with this method is that in terms of galaxy merging, it is a two body problem, and so mergers cannot occur. Later work has tended to be divided between the direct method of Aarseth (e.g. Aarseth and Fall 1980) and the mesh method (e.g. Miller and Smith 1980). In the former, each particle interacts with its nearest neighbours, using a time-step appropriate to its needs. The field due to the rest of the system is calculated at appropriate intervals, and interpolated as needed. The mesh method depends upon the calculation of the gravitational potential on a mesh of points by an algorithm based on the fast Fourier transform. The particles are then advanced by a fixed time-step at each iteration.

1.3.3) Results of Computer Simulations

The results of the restricted three body simulation (Toomre and Toomre 1972, Wright 1972, Eneev et al. 1973) showed that gravitational interaction was a strong candidate as the cause of the formation of tails and filaments; the simulation by Toomre and Toomre of the "Antennae" galaxies (NGC 4038/9) is particularly impressive. These experiments also showed the importance of galactic rotation in the interaction: for maximum effect, the galaxies must pass round each other with the angular momentum vector of the passage in the same direction as the angular momentum vector of their spin. This has been confirmed by White (1979a).

Later simulations with models having distributed mass have shown that mergers will occur. In the words of Aarseth and Fall (1980), "...merging is a natural and non-negligible consequence of the

gravitational instability mechanism". The product of the merger seems to nearly always be an elliptical galaxy (White 1978, 1979a, 1979b, Aarseth and Fall 1980, Gerhard 1981), although it is unclear as to whether such a merger could provide the extended envelope of a cD galaxy (Gallagher and Ostriker 1972, White 1978, Tremaine 1980, Carnevali et al. 1981, Villumsen 1982).

As practically all of these authors have noted, for the rotation of the merger products to agree with observational data on elliptical galaxies, the pre-merger components must either be in orbit around one another, or intersect on a low angular momentum orbit. Some angular momentum (~20%) is carried away by escapees from the system (White 1980, Gerhard 1981), but not enough to account for the observed difference with the product of an initially unbound interaction. This was one of four objections raised by Ostriker (1980) against the suggestion that ellipticals are the merged remnants of spiral galaxies. The other three were:- (1) The scale length for ellipticals is hundreds of parsecs, whereas for spirals it is kiloparsecs; and that mergers have a more extensive envelope than either of their progenitors (White 1978). (2) That the best merging environment is in small groups, the worst being in centrally condensed fields (the largest fraction of elliptical galaxies is in centrally condensed fields), and (3) the colour-luminosity and metal-luminosity relationship of elliptical galaxies is not consistent with that of spirals.

The scale length argument seems to have been overcome with the discovery by Malin and Carter (1980) of giant shells around several normal elliptical galaxies, with sizes ~180 kpc. Point two has been countered by the simulations of Aarseth and Fall (1980), whose simulations have agreed qualitatively (and sometimes quantitatively)

with the clustering of elliptical galaxies. In addition, simulations of cluster collapse by Roos and Norman (1979) have showed that after the collapse, 20% of the resultant galaxies were the products of mergers, and that there tended to be an increase in these towards the collapse centre. This last point fits in well with the data of Hickson et al. (1977), who find a depletion of galaxies with HII regions and dust (features not usually associated with elliptical galaxies) in groups.

The final objection, that of the colours of elliptical galaxies, fails to take into account star formation induced by the merger. This will be discussed in section 1.3.6.

1.3.4) Ring Galaxies

Whilst considering the dynamics of interacting galaxies, it is worthwhile to digress upon ring galaxies. These comprise 0.7% of all galaxies (Dostal and Metlov 1977). They are unstable and, after $\sim 10^8$ yr, the ring tends to break up into smaller galaxies (Theys and Spiegel 1977). Ring galaxies are usually found with a companion, and it is this fact that argues powerfully for a collisional formation. Both analytical models (e.g. Chaterjee 1981, Sunder and Chaterjee 1982) and numerical methods (Lynds and Toomre 1976, Theys and Spiegel 1977) predict the formation of a ring if a small galaxy passes through a disk galaxy normal to the plane of the latter. As the latter rebounds, a transient ring-like density wave develops. This picture is consistent with observations of ring galaxies, e.g. AM 064-741 (Few et al. 1982), II Hz 4 (Lynds and Toomre 1976).

It has been argued that ring galaxies may be produced by internal activity (Dostal and Metlov 1977, 1979) as the probability of collision is low, but the evidence for massive halos seems to

undermine this line of reasoning. It is quite possible that ring-like structures in galaxies may be due to nuclear activity (Kolesnik et al. 1979); a survey of Seyfert galaxies by Hodge (1968) showed five out of sixteen with rings.

1.3.5) Possible Links with Active Galaxies

Observations by Karachentsev and Karachentsev (1975) showed that a higher proportion of Markarian galaxies with active nuclei are found amongst double galaxies than as isolated objects (20% of all Markarian galaxies are paired objects (Vardanyan and Melik-Alverdyan 1975)). Similarly, Adams (1977), in a study of 80 Seyfert galaxies, found a surplus of Seyfert nuclei in disturbed and interacting systems, and indicated that greater than 10% of all Seyfert galaxies were disturbed in some way. This tends to confirm earlier work by Khachikyan and Weedman (1971), who noted a connection between the nuclear activity and the appearance of type 2 Seyfert galaxies. In addition, examination of the cluster membership of Markarian and Seyfert galaxies by Smirnov and Komberg (1980) also hinted as a relationship between the pair/group membership and the properties of such objects.

If such a relationship were to be unambiguously established, it would go far towards providing an explanation of the energy source of such objects. Type 1 Seyfert galaxies are generally conceded to be powered by a small non-thermal source in the nucleus. For example, NGC 4151, one of the original galaxies studied by Seyfert (1943), is variable in the near-infrared (Penston et al. 1974), optical (Fitch et al. 1967, Penston et al. 1974) and X-ray (Elvis et al. 1978) over periods of months. The energy source of type 2 Seyferts, on the other hand, is still open to discussion. Although a non-thermal energy

source is not ruled out, their spectra are consistent with the re-radiation of energy by dust (Rieke 1978), and observations by Balick and Heckman (1979) of the Balmer absorption lines in NGC 1068, the archetype type 2 Seyfert, showed the presence of early-type stars. Far-infrared observations of the same object by Telesco et al. (1976) found a nuclear $M/L \leq 0.003$, consistent with star formation, the stars following the Salpeter initial mass function. Further 20 μm observations by Telesco et al. (1980) implied that a number of local sources heat the dust rather than one single compact one. However, if early-type stars do produce the observed luminosity in NGC 1068, then a prohibitively large nuclear mass ($\sim 10^{11}$ $M_\odot$) would accrue over the estimated lifetime of the Seyfert stage ($\sim 2 \times 10^8$ yr) (Telesco and Harper 1975). This would seem to argue for bursts of star formation, rather than continuous star formation, as the energy source.

1.3.6) Star Formation in Interacting Galaxies

With the evidence that star formation can be induced by the interaction of large clouds of interstellar gas, it is perhaps natural to expect enhanced star formation in galaxies showing signs of disturbed structure. Larson and Tinsley (1978) (hereafter LT) compared the colours of a sample of morphologically peculiar galaxies (from the Arp (1966) atlas) with those of a sample of normal galaxies (from the Hubble atlas (Sandage 1961)). They found that both sets tended to occupy a well defined locus in the (U - B, B - V) diagram, but that the peculiar galaxies had a large amount of scatter. Noting that the position of a galaxy on this locus is almost uniquely determined by the star formation rate per unit mass averaged over the past 10^8 years, LT attributed the scatter in the peculiar galaxy plot to recent bursts of star formation.

Further analysis of the scatter revealed that most of it was due to galaxies showing signs of tidal interaction, and that of these, there was a small difference between those systems with tails, and those without. This difference, whilst not statistically significant, did seem to indicate that interacting galaxies have colours consistent with bursts of star formation, followed by subsequent decay, the decay time being roughly the same as the age of the interaction estimated by dynamical means. From this, L/T argued that galactic interaction could induce bursts of star formation, probably by direct collision of the gas clouds associated with the galaxies, but also by gas, deprived of angular momentum by a grazing passage, falling into the nucleus (Bierman et al. 1979).

Further evidence to support this theory has come from a variety of sources. Bothun and Schommer (1982) made HI and optical measurements of peculiar galaxies contained in clusters, and found essentially the same results as L/T. Stocke (1978), in a survey of 600 pairs of galaxies, noted an inverse correlation between detectability of the galaxies, and their physical separation. This was attributed to mass transfer between the galaxies, and led to the possibility that the radiation may be due to rapid star formation, or accretion onto some central object. Smirnov and Tsvetkov (1981) compared the supernovae rates in normal and interacting galaxies, and found that the frequency of supernovae in interacting galaxies was more than twice that in field galaxies. Since supernovae are the explosions of massive stars, they were led to conclude that either a higher proportion of massive stars are formed in interacting galaxies, or that the star formation rate was higher. A radio survey of 33 spiral galaxies by Condon et al. (1982) also linked the supernova rate with

the presence of a companion, and a survey of interacting galaxies by Bergvall (1981), and Bergvall et al. (1981) revealed extensive regions of ionized gas attributable to star formation. In the near-infrared, JHKL photometry by Joseph et al. (1983) of interacting galaxies from the Arp (1966) atlas revealed that most had colours consistent with a recent burst of star formation.

Observations of particular objects also support the link between star formation and galaxy-galaxy interaction: NGC 4038/9 (the "Antennae", and one of the systems modelled by Toomre and Toomre (1972)) (Schweizer 1976), NGC 6052 (Alloin and Duflot 1979), NGC 1510/2 (Van Woeden et al. 1979), NGC 4485/90 (Viallefond et al. 1980), II Zw 40 (Baldwin et al. 1982), NGC 3690 (Fischer et al. 1983) have all been interpreted as interacting galaxies where star formation is proceeding as a result of the interaction. The high infrared flux of M82 (Rieke et al. 1980) has also been put down to star formation, due to an infall of gas from an interaction with M81, a view shared by Gottesman and Weliachew (1977) on the basis of HI studies, but Rieke et al. point out that to agree with observed luminosities, formation of stars at the lower end of the mass spectrum ($M_* < 3M_\odot$) has somehow to be supressed.

A further ramification of this theory is that it is able to explain the elliptical galaxy colour problem mentioned in section 1.3.3. Tinsley and Larson (1979) have suggested that the formation of elliptical galaxies could have proceeded via a heirarchical series of mergers, with stars being formed by bursts induced by collisions. The mass/metallicity correlation of elliptical galaxies is explained if star formation is a function of mass, and star formation ceases before all the gas is used up; either by the ending of the merger series, by gas being unable to cool between collisions or, possibly, gas being

blown out of the region by a "galactic wind" driven by the starburst (Mathews and Baker 1971). Galaxy disks may then form by subsequent accretion. A prediction of this theory is that the spherical components of spiral and S0 galaxies will be identical to elliptical galaxies.

1.4) The Need for Infrared Observations

Building on the work of LT, Struck-Marcell and Tinsley (1978) found that the (U - B, V - K) colour index was more sensitive to star formation rates than (U - B, B - V). In general, blue V - K colours indicate either a young age, or a substantial burst of star formation. The V - K colours can, however, be affected by the presence of dust, in which case the M/L ratio of the system becomes a more reliable indicator of star formation. Considering the bolometric luminosity, they advanced the empirical relationship,

$$\Psi \simeq 0.2\ L_{bol} M_{\odot} Gyr^{-1} L_{\odot}^{-1} \quad (\text{if } M/L_{bol} \leq 0.5). \qquad (1)$$

where Ψ is the star formation rate ($M_{\odot}$/Gyr). They suggested that this result should hold to within a factor of 4. The condition that $M/L_{bol} \leq 0.5$ occurs because models show that monotonically decreasing star formation rates have $M/L_{bol} > 0.5$. Observations of star formation regions in our own galaxy have indeed shown that dust is a major component. Figure 1 shows the energy distribution of W3, a typical HII region (Wynn-Williams and Becklin 1974). The spectrum consists of essentially two components. The solid line is the spectrum produced by the thermal bremsstrahlung of electrons within the HII region itself. This has $F_{\nu} \propto \nu^{-0.1}$, turning over at the low frequency end at the point at which the source becomes optically thick to the radiation (this point will be a function of the electron density within the region)

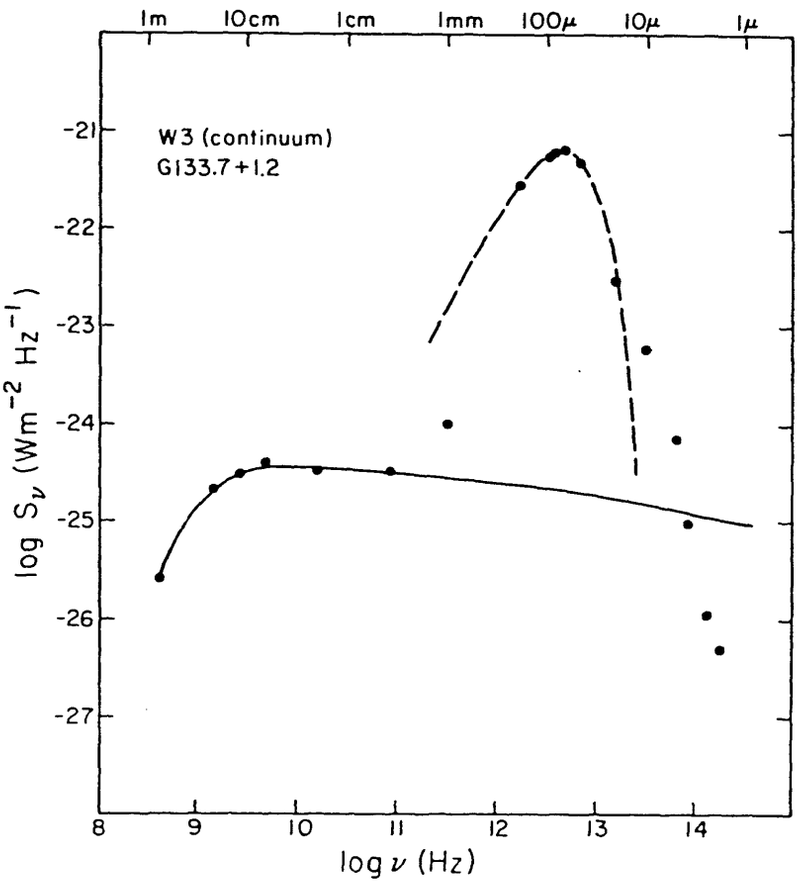


Figure 1. The radio and infrared spectrum of the HII region W3 (from Wynn-Williams and Becklin 1974). The solid line is the emission expected from transitions in the HII plasma, and the dashed line is a 70 K black-body curve

(Fazio 1978). The second component (dashed line) is interpreted as dust emission. In Figure 1, this is the spectrum of a 70 K black body, but the slow fall off towards short wavelengths indicates a range of dust temperatures, while the sharp fall at long wavelengths is compatible with particle sizes less than the wavelength (Wynn-Williams and Becklin 1974, Beckman and Moorwood 1979).

Owing to the poor resolution of far-infrared observations, the structure of the dust in HII regions is not known for certain, but the spatial correlation of the infrared and radio fluxes (e.g. Panagia 1974, Lebofsky et al. 1978) implies that the emitting dust is mixed well with the ionized gas. The cooler dust is probably in the neutral hydrogen region surrounding the ionized regions. Observations of the 10 μm silicate absorption feature indicate that the dust can cause visual extinctions in excess of 10 magnitudes (Beckman and Moorwood 1979). This ties in well with the observation by Becklin et al. (1976) of a sharp increase in dust density at the HII/HI boundary of the Orion nebula (one of the best studied HII regions), and the prediction by Panagia (1974) that a relatively thin dust shell could absorb practically all the stellar radiation. Whatever the distribution of dust in such regions, it is clear from Figure 1 that, to a good approximation,

$$L_{\lambda>3.5\mu} \simeq L_{bol} \qquad (2)$$

and this seems to hold for a large number of such regions (Harper 1975).

Combining this equation with equation (1), it can be seen that measurements of the total infrared luminosity can be used to estimate the star formation rate in galaxies, an approach used by Gehrz et al. (1983) for the "star burst" galaxy NGC 3690. Measurement

of the total infrared luminosity is not easily accomplished though, as most of the infrared region of the spectrum is absorbed by the earth's atmosphere. Ideally, one would want to observe in the "broad 100 μm" window (40 - 300 μm), where the luminosity of HII regions is estimated to be ~75% of the total (Wynn-Williams and Becklin 1974). Unfortunately, this is where the atmospheric absorption is the strongest. The longest convenient wavelength for ground-based observations is the 8 - 13 μm band, as the windows at 20 μm and 34 μm have poor transmission characteristics. Although 10 μm measurements cannot give exact results, measurements of a heterogeneous selection of galaxies by Rieke and Low (1975) gave, to within a factor of 2 or 3,

$$L_{8-13\mu} \simeq 0.07\ L_{bol} \qquad (3)$$

(This is also within the bounds set be Figure 1 for HII regions). We can thus summarize this section by saying that broad band 10 μm observations of galaxies where star formation is occurring will give an estimate of the bolometric luminosity of the system, and this will, in turn, lead to an estimate of the star formation rate in those galaxies.

1.5) Summary

In this chapter, we have examined the processes of star formation, and have noted that "contagious" star formation can be a very efficient process. As this involves the interaction of large masses of interstellar material, it has been natural to look at interacting galaxies (where large quantities of matter are being disturbed) for signs of star formation. This has been done by L/T, and evidence of a correlation between the degree of interaction and the

star formation rate has been found.

The processes of galaxy interaction have been looked at, and it has been seen from simulations that a possible consequence of galaxy interaction is a merger, and that this merger nearly always produces an elliptical galaxy. Possible links between interacting galaxies and Seyfert galaxies have been explored.

Finally, it was noted that in cases where the star formation regions are obscured by dust, broad band 10 μm observations could help towards estimating the star formation rate in such a galaxy.

Given the above, it is possible to speculate along the following lines: Interaction of galaxies leads to the formation of elliptical galaxies, and during these processes, bursts of star formation are induced. In some circumstances, these bursts are able to fuel Seyfert (type 2) activity, but otherwise they lead to the abnormal colours observed by LT. During the time scale of the interaction, most of the available gas in the galaxies is used up, and this accounts for the low rate of star formation observed in elliptical galaxies (and also neatly ties in with the observation by Adams (1977), that there are few Seyfert nuclei in elliptical galaxies).

In view of the crucial importance of star formation in the above sequence of events, it is proposed to observe interacting galaxies at 10 μm, with a view to estimating their bolometric luminosity and hence the rate of star formation within them.

CHAPTER 2

The Physics of Photoconductive Infrared Detectors

2.1) Introduction

As with practically all physical phenomena, photoconductivity can be viewed on two levels, the microscopic and the macroscopic. In this chapter, the former is considered first, with a description of its various manifestations. One type in particular, extrinsic photoconductivity is concentrated upon, and the microscopic viewpoint is related to the bulk properties.

After discussion of the noise processes that operate within an extrinsic photoconductor, the operational conditions for such a device to be used in a detection system, together with the appropriate circuitry are examined. Finally, a survey of photoconductive detectors for 10 μm astronomical observations is presented, and a choice of detector made.

2.2) Microscopic Picture of Photoconductivity

2.2.1) Electrical Conductivity

Before discussing photoconductivity in any depth, a general picture of electrical conductivity is needed. For a substance to be able to conduct an electric current, the electrons must be able to move when an electric field is applied. In terms of k-space (or momentum space), this is equivalent to saying that the Fermi sphere must be able to be displaced by an amount

$$\delta \underline{k} \simeq - \frac{e\tau_c}{\hbar} \underline{E} \ , \qquad (4)$$

where E is the applied electric field, e is the electronic charge, and τ_c is the time between collisions of the electrons and the lattice.

In an insulator, the electrons cannot acquire this momentum. The valence band is full, the gap between the valence and conduction bands being large enough so that thermal excitation of electrons from the valence band is negligible, and to give each electron an added momentum $\hbar\delta k$ requires those at the edge of the Brillouin zone to be given enough energy to jump the gap (Figure 2a). This will only happen when the electric field is large enough to disrupt the electronic structure of the material.

Semiconductors are similar to insulators, but here the band gap is small enough for there to be a significant number of electrons thermally excited into the conduction band (Figure 2b). When an electric field is applied, the conduction band electrons can move into free states, as can the corresponding holes in the valence band. Conduction is thus a two carrier process in this class of materials, the holes carrying a charge of +e.

Finally, in metals, the conductions band is only partially full at all temperatures (Figure 2c). When an electric field is applied, there are always free states available to the electrons, and so conduction occurs. Unlike semiconductors, conduction in this class of materials is a one-carrier process.

2.2.2) Types of Photoconductivity

A material is said to be photoconductive if, on the absorption of radiation, the conductivity of the material changes. Strictly speaking, only processes where energy is given to the electrons are called photoconductive, processes where the energy is absorbed by the lattice being referred to as bolometric. Using this

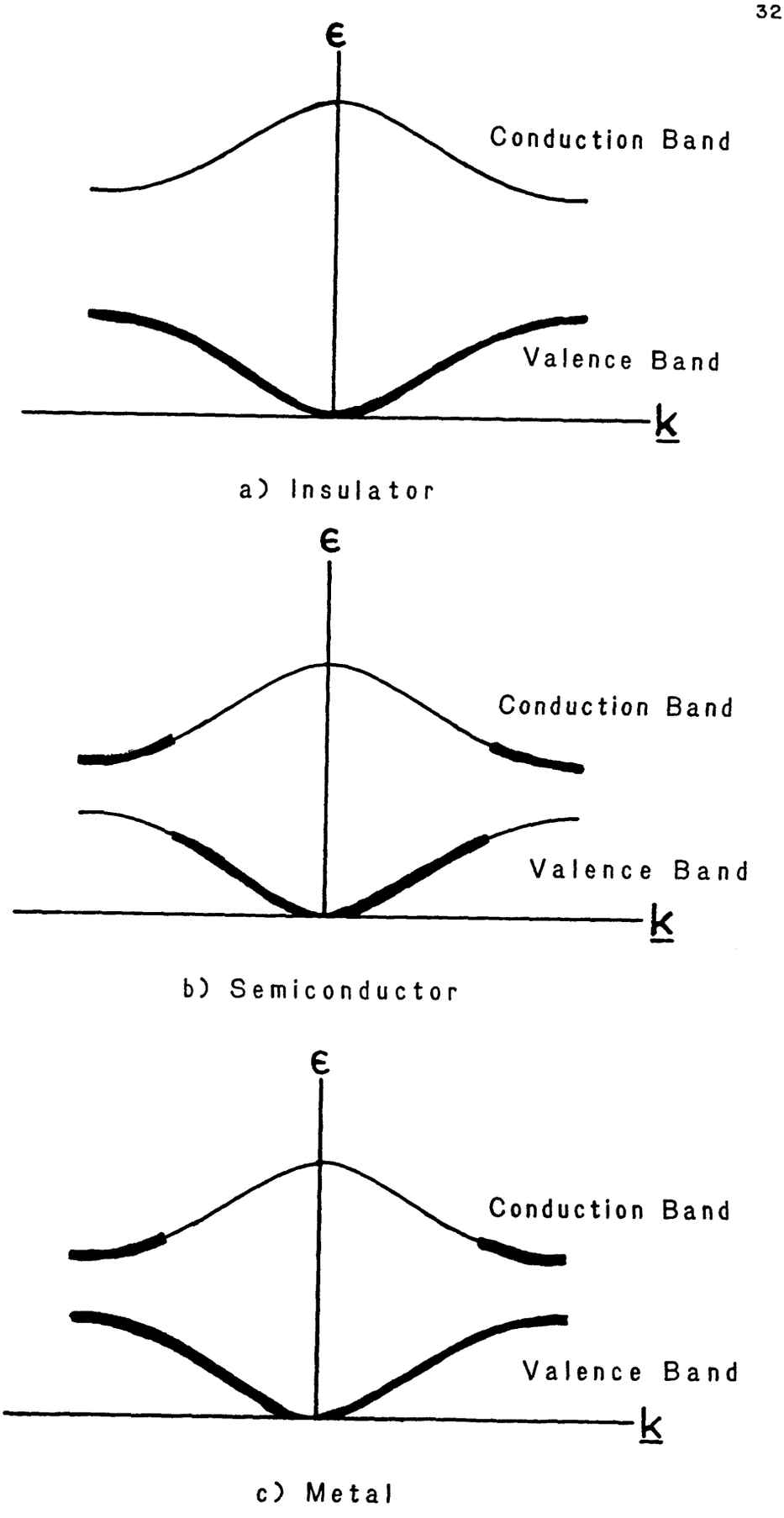


Figure 2. Occupied states and band structures of a) an insulator, b) a semiconductor, and c) a metal. Bold lines indicate the limit of occupancy. (Adapted from Kittel 1971).

definition, photoconductivity can be divided into three classes, intrinsic photoconductivity, extrinsic (impurity) photoconductivity, and free-electron (hot-electron) photoconductivity.

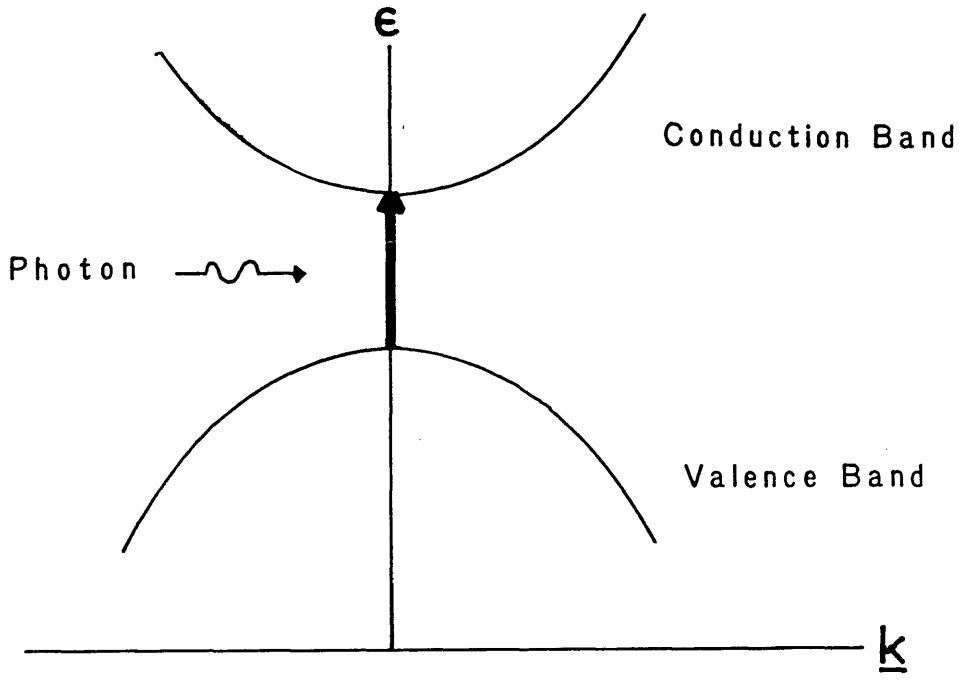
2.2.3) Intrinsic Photoconductivity

In intrinsic photoconductivity, absorption of a photon causes excitation of an electron from the valence band to the conduction band. If this transition is unaided by phonons, it is a direct transition, and corresponds to Figure 3a. The photon contributes essentially nothing to the momentum of the electron, and the transition is between states with the same value of k. Figure 3a shows the maximum in the energy of the conduction band at the same point in k-space as the minimum in energy of the conduction band. Often, this is not the case and the threshold energy for transition is greater than the true band gap. To excite an electronic transition with a threshold energy photon requires the absorption or ejection of a phonon (Figure 3b). The phonon carries an energy $\hbar\Omega$, where Ω is the phonon angular frequency, and so this will lower the threshold energy to $\epsilon_g - \hbar\Omega$ (ϵ_g being the band gap energy). In general though, $\hbar\Omega << \epsilon_g$, and so the threshold energy for transition is the band gap energy.

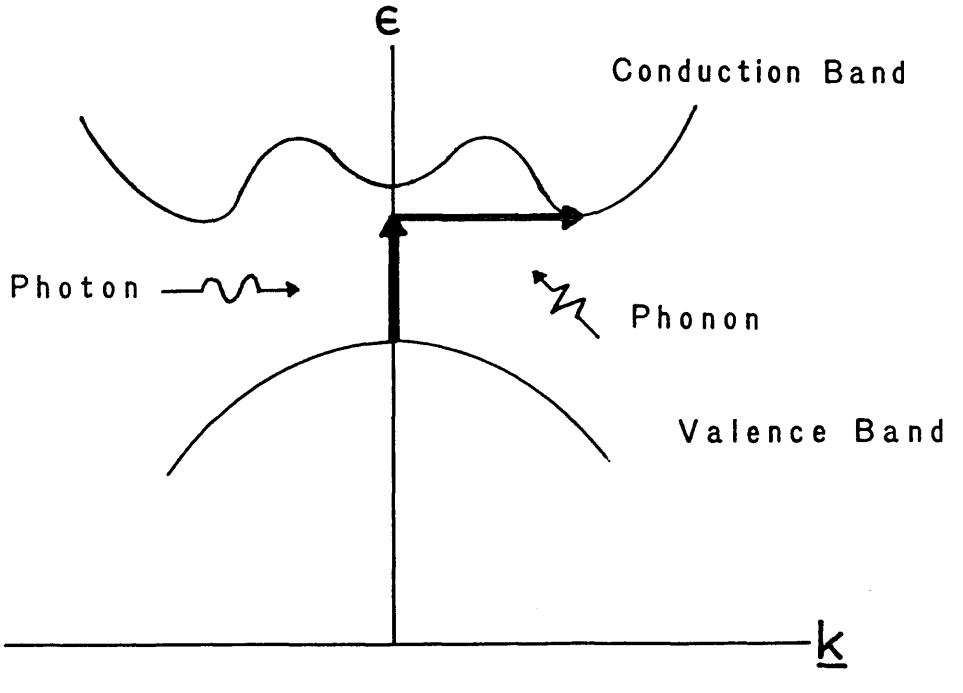
2.2.4) Extrinsic Photoconductivity

Impurities and imperfections in a semiconductor can drastically alter the electrical characteristics of these materials. Of considerable importance are the group IV elements (e.g. Si, Ge) doped with either group III or group V elements, and these are the ones that will be examined in detail.

Doping a group IV element with an impurity of the type



a) Direct Transition



b) Indirect Transition

Figure 3. Direct and indirect transitions across a band gap. (Adapted from Kittel 1971)

mentioned above has the effect of replacing some of the group IV atoms in the crystal lattice with impurity atoms. If the impurity is a group III element, the impurity atoms lack one bonding electron and hence have an excess positive charge. These are p-type impurities. Replacement of an atom by a group V atom leads to an excess electron and so the name n-type impurity. The effect that these impurities have on the energy level structure of a semiconductor is shown in Figure 4. In the following discussion, only n-type impurities will be considered. The argument applies equally to p-type impurities, but with holes substituted electrons, and valence band for conduction band.

The impurity level in an n-type semiconductor lies just below the conduction band (e.g. in arsenic-doped silicon, this gap is 0.054 eV, compared with a band gap of 1.166 eV (Bratt 1977)). At room temperature, the number of electrons thermally excited into the conduction band from this impurity level will be many times the number excited from the valence band; in the example above, this factor is $\sim 2 \times 10^9$. The great disparity in numbers is one factor in determining the degree of doping in the semiconductor. Another consideration is that too high a concentration of dopant, in addition to giving the semiconductor metallic-like properties over practically all temperatures, will cause the impurity level to spread out into a band, which may overlap the conduction band.

Assuming that the impurity concentration is sufficiently low, absorption of a photon will cause an excitation of an electron from the donor level into the conduction band. Under the influence of an electric field, this electron will be free to move in the practically empty conduction band, and so be able to conduct a current. Unlike intrinsic photoconductivity though, the hole is not free to move, and

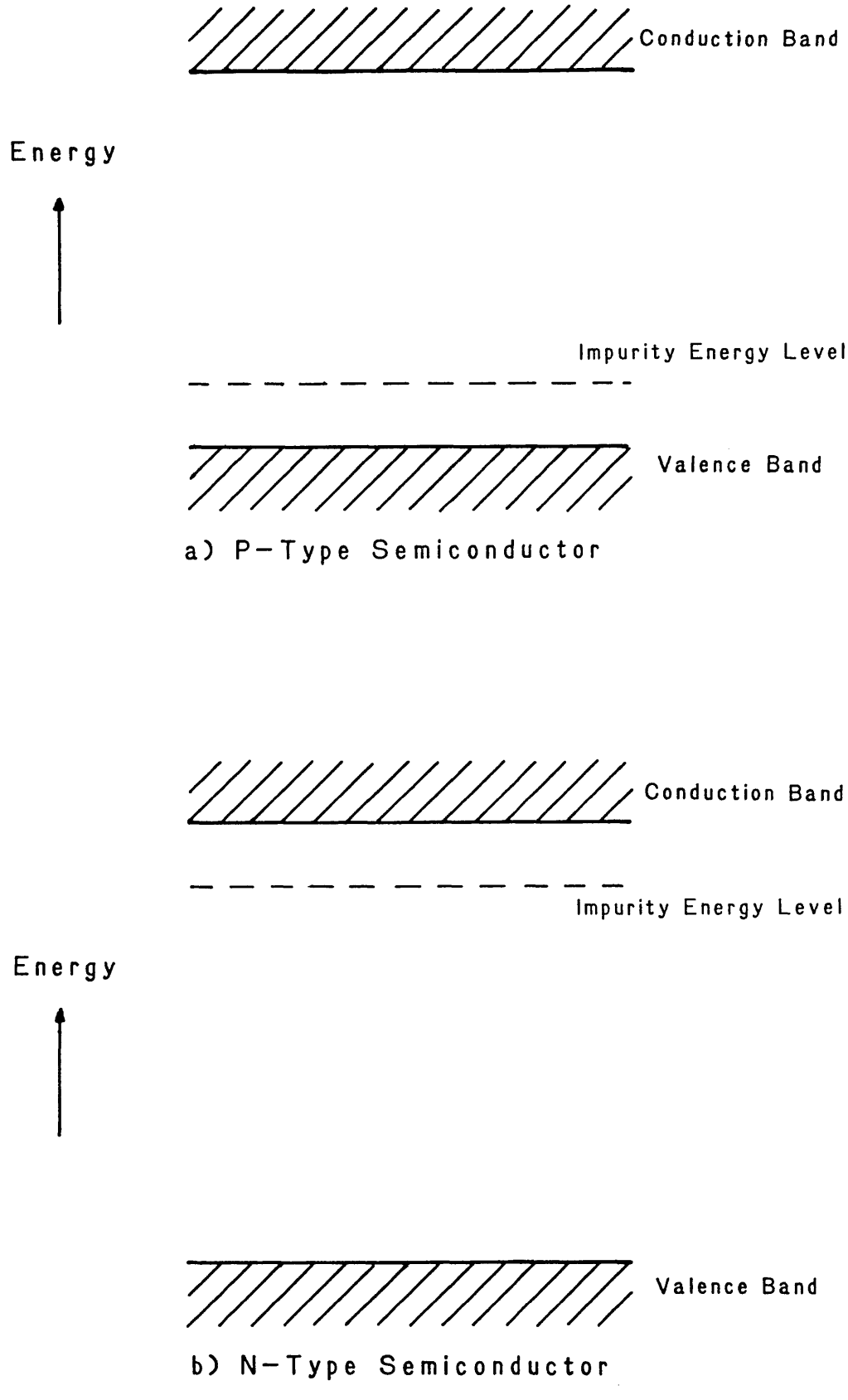


Figure 4. Energy levels in a) a p-type and b) an n-type semiconductor.

remains bound to the lattice. There will, of course, be some hole conductivity due to the thermal excitation of electrons from the valence band, but this will be many orders of magnitude below the electron conductivity.

2.2.5) Free Electron Photoconductivity

The process described in this section is the principle of the "hot-electron" bolometer. By the definition of photoconductivity given above, this process is photoconductive rather than bolometric and so is included for completeness.

In thermal equilibrium, electrons in the conduction band have energy ~kT, and any energy gained in photon absorption is soon lost, either by re-radiation or by collision with phonons. At low temperatures, the coupling between the electrons and the lattice is weak, and a steady state can exist where the effective temperature of the electrons is different from the lattice temperature. Absorption of photons will increase this effective temperature. The effect on the electron mobility is the same as if the temperature of the lattice were to have been raised (in that the mobility will decrease) and so the conductivity is lowered (Putley 1964).

2.3) Macroscopic Picture of Photoconductivity

2.3.1) Basic Assumptions

Before an expression is derived relating the electrical response of a photoconductor to the flux of incident radiation, four simplifying assumptions have to be made. These are: 1) only one carrier is generated for each photon absorbed, 2) there is a uniform distribution of impurities throughout the photoconductor, 3) the

photoconductor has ohmic contacts, and finally, 4) there are no trapping effects in the photoconductor. These conditions are found to be fairly well obeyed in most photoconductors, but a justification for them will be given.

The first condition, that of one carrier per photon allows us to drop consideration of the case where an excited electron picks up enough energy from the bias field to collisionally excite another electron. Such a process would be complex and, to a first approximation at least, we want to ignore this. Conditions (2) and (3) have been made to avoid electric fields (other than the bias field) being present in the detector: a non-uniform impurity concentration would lead to the diffusion of carriers not being due to the bias field alone, whereas non-ohmic contacts have a potential barrier between them and the material, and so distort the bias field. Finally, absence of trapping effects - caused by impurities or crystal dislocations - means that only a single time constant for recombination need be considered.

2.3.2) Elementary Case

The following derivation of photoconductive response has been taken from a number of authors, notably Petritz (1959), Putley (1964), Bratt (1977) and Kruse (1977).

An extrinsic photoconductor, thickness L, surface area A is illustrated in Figure 5. A photon flux density of ϕ photons $m^{-2}s^{-1}$, coupled with an electric field E within the detector gives rise to the short circuit photocurrent I_s. The drift velocity v_d, of electrons through the detector is given by

$$v_d = \mu E \ , \qquad (5)$$

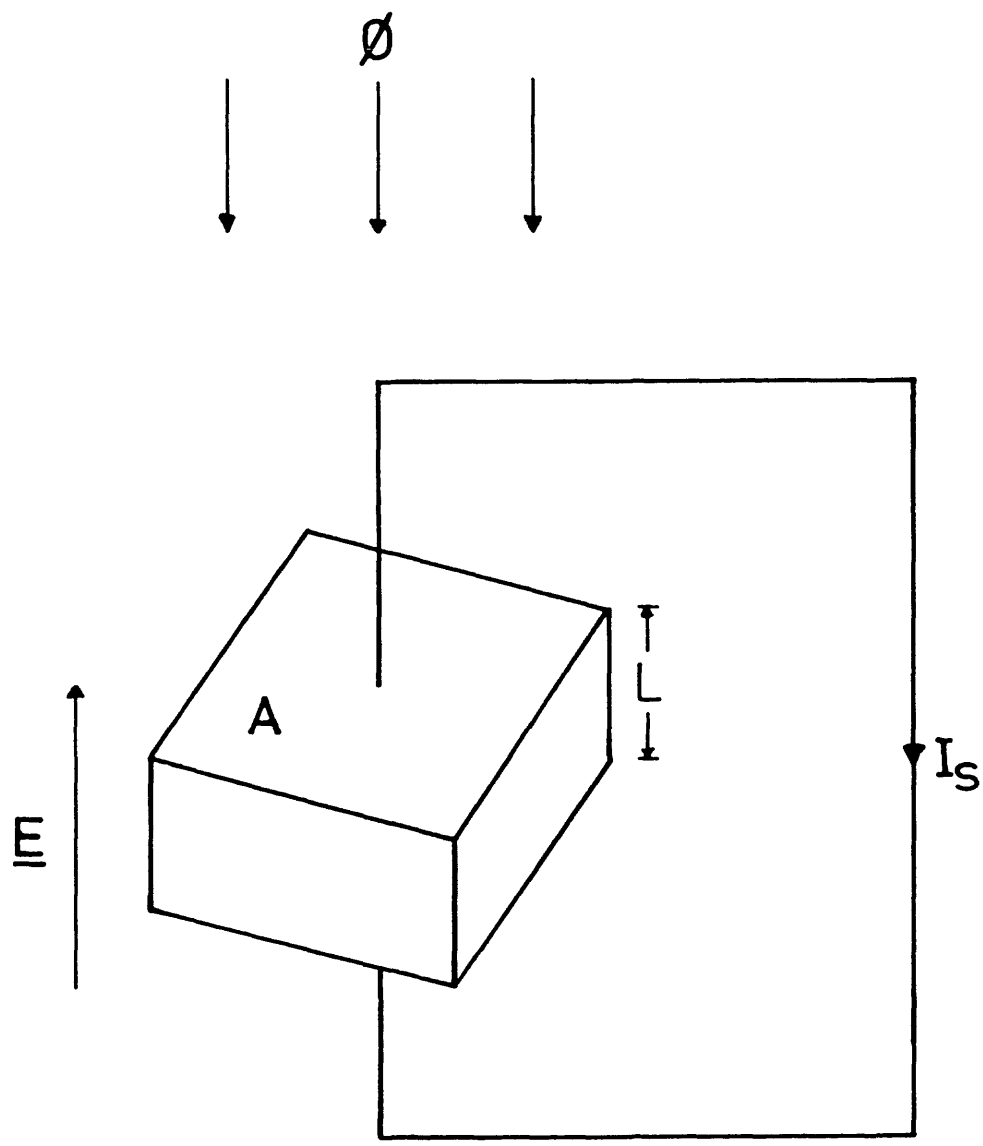


Figure 5. Idealized photoconductor. The detector has surface area A, and thickness L. A photon flux of ϕ $m^{-2}s^{-1}$, together with an electric field E, produces a short circuit photocurrent I_S.

where μ is the electron mobility. During their lifetime in the conduction band τ, the electrons travel a distance L_d, given by

$$L_d = v_d \tau \ , \tag{6}$$

and a charge eL_d/L per carrier is moved through the external circuit. If η is the efficiency of the conversion process, the quantum efficiency, then $\eta\phi A$ carriers are produced per second by the incident photon flux, and the short circuit photocurrent is thus

$$I_s = (\eta\phi A)\left[e \frac{L_d}{L}\right]$$

$$= \frac{\eta\phi A e \mu E \tau}{L} \ . \tag{7}$$

The photoconductive gain G, is defined (Kingston 1978) by the relation

$$G = \frac{L_d}{L} \ . \tag{8}$$

This quantity is the ratio of the number of carriers passing through the external circuit to the number of photons absorbed. In summary, the detector is acting as a photon-to-current converter, the efficiency of the process being characterized by the product ηG.

When used for astronomical observations, we are usually interested in small signals superimposed on a large background. Under these conditions, the response of a photoconductor is governed by the equation

$$\frac{d\Delta N}{dt} = \eta A \Delta\phi - \frac{\Delta N}{\tau} \ , \tag{9}$$

where $\Delta N = N(t) - \bar{N}$ is the instantaneous excess of conduction band electrons over the mean value, in the presence of a signal $\Delta\phi$. Usually, $\Delta\phi$ will be modulated in some fashion. Taking a sinusoidally-varying signal,

$$\Delta\phi = \Delta\Phi[1 + e^{i\omega t}] \ , \tag{10}$$

equation (9) can be solved to give

$$\Delta N = \eta A \tau \Delta\Phi \left[1 + \frac{e^{i(\omega t + \theta)}}{(1 + \omega^2\tau^2)^{1/2}}\right] \ , \tag{11}$$

where

$$\theta = \tan^{-1}(-\omega\tau) \ . \tag{12}$$

It can therefore be seen that the carrier recombination time determines the upper limit of operation. For an Si:As photoconductor $\tau \sim 10^{-8}$ s (MacGregor 1977) giving an upper limit to ω of $\sim 10^8$ Hz. Since the means of modulation in an infrared astronomical system is usually mechanical, this limit is never reached in practice, and so discussion can be confined to the region $\omega\tau << 1$. In this limit, equation (11) reduces to

$$\Delta N = \eta A \tau \Delta\phi \ . \tag{13}$$

Rewriting this in terms of the conductivity σ, gives

$$\Delta\sigma = \eta A \tau \left[\frac{\mu e}{AL}\right] \Delta\phi \ . \tag{14}$$

To calculate the signal voltage produced, we note that the resistance of the detector R_D, is given by

$$R_D = \frac{L}{\sigma A} \ . \tag{15}$$

Then

$$\frac{dR_D}{d\phi} = \frac{dR_D}{d\sigma} \frac{d\sigma}{d\phi}$$

$$= - \frac{R_D}{\sigma} \frac{\mu e \eta \tau}{L} \ . \tag{16}$$

If the detector obeys Ohm's law, the voltage responsivity is thus

$$\frac{dV}{d\phi} = \frac{dV_D}{dR_D}\frac{dR_D}{d\phi}$$

$$= - V_D \frac{\mu e \eta \tau}{\sigma L} , \qquad (17)$$

where V_D is the voltage across the detector.

2.3.3) Critique of the Elementary Case

From equation (7), the D.C. resistance of a photoconductor under constant illumination is

$$R_D = \frac{L^2}{\eta A \phi e \mu \tau} , \qquad (18)$$

i.e. independent of the applied electric field.

The most obvious failing of this relationship has been noted by Picus (1962) for Ge:Cu photoconductors - the devices are non-ohmic. This result holds for other photoconductors as well (see Chapter 3), and the case of a non-ohmic photoconductor as a circuit element is discussed in section 2.6. Picus tends to the view that the field dependence is present in the lifetime term τ, as a variation in the capture cross section. Bratt (1977) supports this, but also includes the field dependence of the carrier mobility as a possible contributor to this behaviour.

The capture cross section of an ionized donor σ_c, is given by

$$\sigma_c = \sum_j \sigma_c(j) S(j) , \qquad (19)$$

where $\sigma_c(j)$ is the capture cross section for the j th excited state, and S(j) is the sticking probability. The latter quantity is the probability that a carrier, once captured, will cascade down to the ground state before being excited again. Bratt shows that the field dependence enters into the S(j) term, this being a function of energy,

and leads to a lifetime

$$\tau \propto E^{1/2}. \qquad (20)$$

Although this has been seen for some photoconductors, Bratt notes that his own experiments (Bratt 1969) and those of others (e.g. Norton 1973) have shown a dependence

$$\tau \propto E^{x}, \qquad (21)$$

with $1 \leq x \leq 2$. The field dependence of the mobility arises when the carriers gain energy faster that they can lose it to the lattice (carrier heating, the principle on which the hot-electron bolometer works). This leads to an increase in the electron temperature, and a consequent decrease in the mobility of the form (Bratt 1977)

$$\mu \propto E^{-1/2}. \qquad (22)$$

Combining equations (21) and (22) leads to the result

$$\mu\tau \propto E^{x-1/2} \qquad (23)$$

which, with x around 1 to 2 leads to the observed non-ohmic behaviour.

2.3.4) Impact Ionization

Equation (17) suggests that for maximum voltage responsivity, the detector should be biased at as high a voltage as possible. However, as the electric field strength rises, there comes a point at which the carriers are able to gain enough energy to ionize neutral donors through impact ionization (Sclar et al. 1953). This effect is characterized by a sharp increase in the current through the detector. Although this effect is reversible, detectors are never operated in this regime for two reasons. Firstly, the breakdown liberates a large

number of carriers, which quite overwhelm the signal current. Secondly, the fact that the breakdown is sporadic means that excessive noise is produced. The critical electric field field varies with the type of photoconductor, but in general is higher for germanium based than for silicon based detectors.

2.3.5) Quantum Efficiency

The quantum efficiency η, has been referred to earlier as "the efficiency of the conversion process". For the case where only one electron is excited per photon absorbed, then this term reduces to the ratio of flux absorbed to incident flux. Taking the case of the detector in Figure 5, then if r_f and r_b are the front and back surface reflectivities respectively, it is easily shown that

$$\eta = \frac{\Phi_{absorbed}}{\Phi_{incident}} = \frac{(1 - r_f)(1 + r_b e^{-\alpha L})(1 - e^{-\alpha L})}{1 - r_b r_f e^{-2\alpha L}} , \qquad (24)$$

α being the absorption coefficient of the material. Maximum absorption occurs at the point $r_f = 0$, $r_b = 1$, where

$$\eta = 1 - e^{-2\alpha L} . \qquad (25)$$

Thus an optimum detector will have an anti-reflection coated front surface, and the rear surface backing onto a perfect mirror. If these geometrical considerations are satisfied, then maximizing η is reduced to increasing the αL product.

Recent experiments (MacGregor 1977) have hinted at η being dependent upon incident flux. That this is feasible can be seen by noting that

$$\alpha = \sigma_i(n_D - n_i) , \qquad (26)$$

where σ_i is the ionizing capture cross-section (which will be

wavelength dependent), and $n_D - n_i$ is the number density of unionized donors. As the photon flux increases, the density of ionized donors n_i, will also increase. Saturation will occur when

$$\eta A\phi = \frac{N_D}{\tau} \qquad (27)$$

(where $N_D = n_D AL$ is the total number of donor atoms in the detector), i.e. when all available donors are ionized all of the time. Further increase of the photon flux will not release any more carriers. Photoconductors are usually operated outside this regime though, and so this effect should be negligible. This has been confirmed by Luinge (1981), whose measurements on Si:As showed only a very slight drop in responsivity at high backgrounds, and by measurements by the author, described in the next chapter.

The wavelength dependence of absorption cross-section shows up in the shape of the spectral response curve of the photoconductor, shown in Figure 6. The response rises to a peak at the cut-off wavelength λ_c, and falls sharply to zero longwards of this. For this reason, all other factors being equal, an attempt is usually made to choose a photoconductor with a cut-off wavelength matching the wavelength of interest.

2.4) Noise Processes in Photoconductors

In a photoconductor-based detection system, noise arises in three places: the detector circuitry, the detector itself, and in the incident radiation stream. The first source, the circuitry associated with the detector will not be considered here. It is assumed that with proper design, the noise from the associated electronics can be made negligible in comparison to the other sources. In any case, the dependence of the noise on fundamental parameters (e.g. measurement

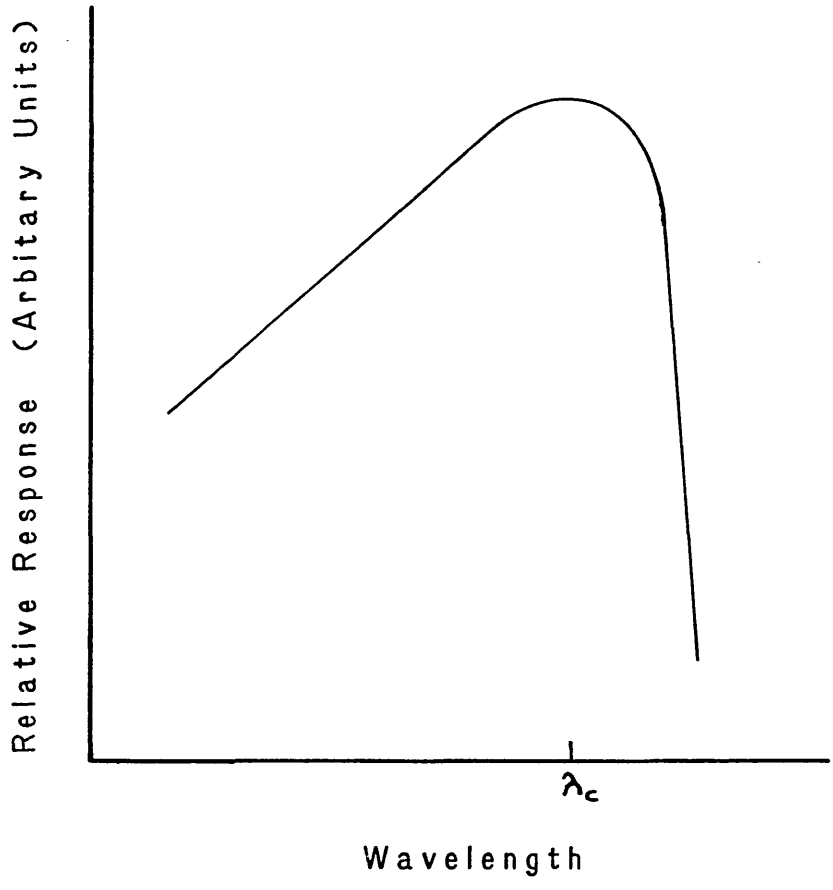


Figure 6. Variation of response with wavelength for a constant photon flux of an idealized photoconductor.

frequency) would be dependent on the type of circuit used.

In the detector itself, three types of noise are relevant for an extrinsic photoconductor: Johnson noise, generation-recombination noise, and 1/f noise. These are described in the following sections.

Finally, there is noise due to the fluctuations in the rate of arrival of photons in the stream of radiation incident upon the detector, namely photon-shot noise. It is this that places the absolute limit on the sensitivity of a detection system. This is discussed in section 2.4.5.

2.4.1) Johnson Noise

Even when there is no current through the detector, a noise voltage will still be present across the terminals, due to the random motion of the electrons within the device. This is the Johnson noise, present in all dissipative devices. The open-circuit noise voltage V_n, is given by (e.g. Smith et al. 1968)

$$V_n = (4k'TR_D\Delta f)^{1/2}, \qquad (28)$$

where Δf is the measurement bandwidth at a frequency f.

2.4.2) 1/f Noise

This is a form of noise, sometimes called current noise, in which the noise power is approximately inversely proportional to the measurement frequency f. The cause of it is unknown, but in general it is thought to be associated with surface phenomena. An approximate expression for this noise is (e.g. Kruse 1977)

$$V_n = \left[\frac{CR_D^2 I_D^{\alpha}\Delta f}{f^{\beta}}\right]^{1/2}, \qquad (29)$$

where $\alpha \simeq 2$, and $\beta \simeq 1$. C is the proportionality factor.

2.4.3) Generation-Recombination Noise

Generation-recombination (g-r) noise is caused by fluctuations in the generation and recombination rates of thermally excited carriers. Since this causes a fluctuation in the resistance of the device, it is observed as a noise voltage only when there is a current flowing. A generalized expression for g-r noise is complex, and depends upon the exact properties of the detector, but in the case of an n-type extrinsic photoconductor for use in infrared astronomy, there are two simplifying factors. These are 1) the detector is a one-carrier device, and 2) it will be operated at temperatures near absolute zero.

As a one-carrier device, simplification occurs in that fluctuations in the holes can be ignored, whereas operation at low temperatures (in addition to reinforcing condition (1) by minimizing the valence band - conduction band transitions) will mean that only a small fraction of impurity sites are ionized at any one time. The carrier generation is thus from an essentially infinite pool, and so Poisson statistics can be applied. The magnitude of recombination fluctuations is the same as that of generation, and so this leads to (Smith et al. 1968)

$$V_n = \left[\frac{4I_D^2 \tau R_D^2 \Delta f}{N_i(1 + 4\pi^2 f^2 \tau^2)}\right]^{1/2}, \qquad (30)$$

where N_i is the total number of ionized carriers.

2.4.4) Relative Contributions of Detector Noise Sources

The types of noise described in the preceding three sections add in quadrature to give the total noise due to the detector. Each type predominates in different measurement frequency regimes, an idealized noise spectrum being given in Figure 7 (Bratt 1977). The

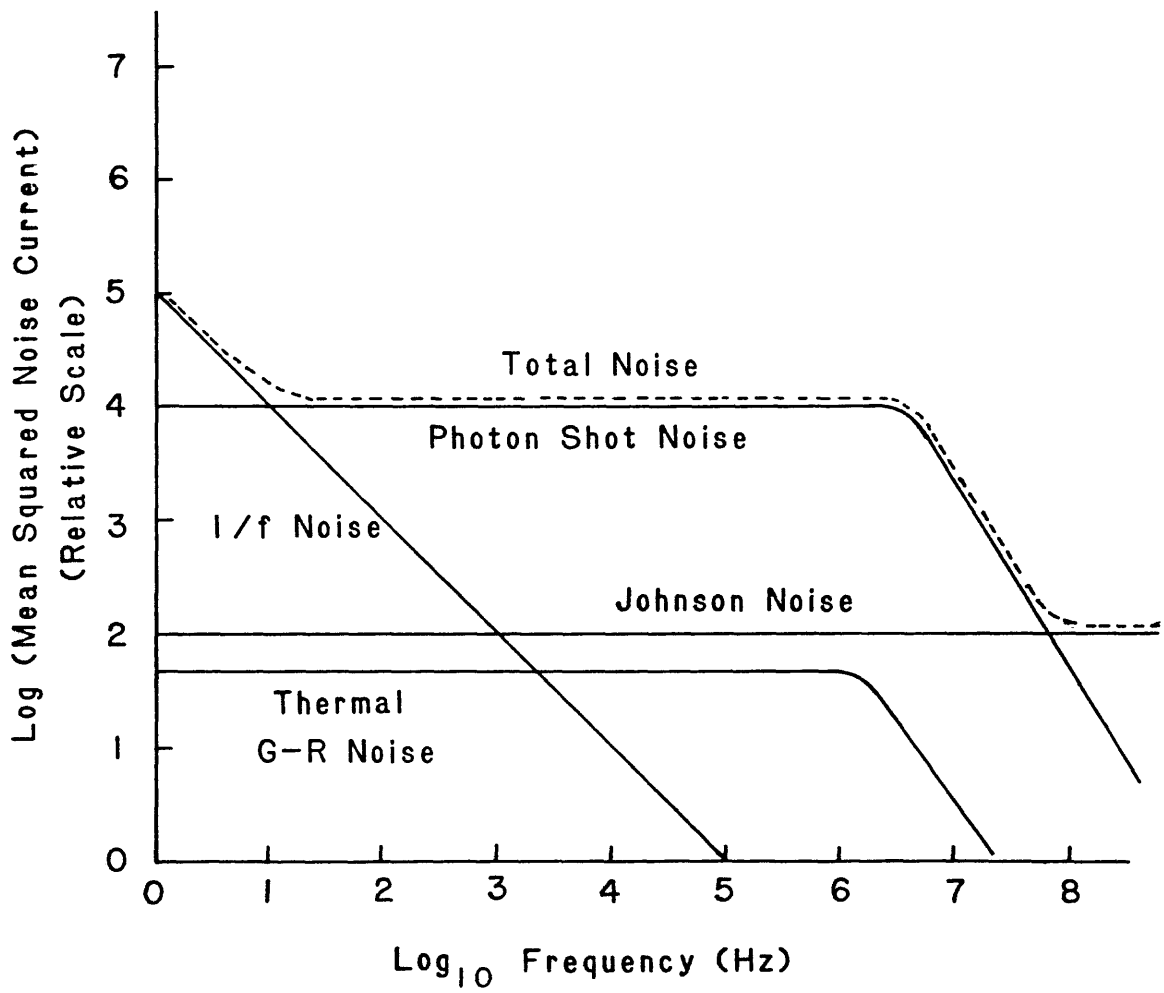


Figure 7. Noise spectrum for a photoconductor operating under a high background. The dotted line is the total noise. The actual crossover points from one type of noise to another will depend upon the detector involved. (Taken from Bratt 1977).

exact cut-off frequency for each type of noise varies with the composition of the detector.

2.4.5) Photon Shot Noise

When using a detector for astronomical work, the amount of radiation falling on it is set by the type of observations required, and by the telescope used. In this case the minimum noise is set by the fluctuations in the incident photon stream. The ultimate limit is, of course, where the fluctuations in the photons from the object being observed that determine the noise. In infrared astronomy however, the unavoidable background is usually many orders of magnitude greater than the signal, so it is background fluctuations that determine the limit (the so-called BLIP condition).

The shot noise equation has been given by many authors (e.g. Putley 1964, Smith et al. 1968). In essence, Bose - Einstein statistics give the photon density in an enclosure as

$$n_\nu = M_\nu \left[\exp\left[\frac{h\nu}{kT}\right] - 1 \right]^{-1}, \qquad (31)$$

where

$$M_\nu = \frac{8\pi\nu^2}{c^3} . \qquad (32)$$

$n_\nu d\nu$ is the number density of photons in the frequency interval $d\nu$. The mean square fluctuation will be

$$\overline{\Delta n_\nu{}^2} = n_\nu \left[1 + \frac{n_\nu}{M_\nu} \right] . \qquad (33)$$

Considering an area A on the wall of the enclosure to be the detector, then the number of photons striking this area per second will be

$$J_\nu = \frac{c}{4} A n_\nu . \qquad (34)$$

The fluctuations of photons absorbed by the detector will be

$$\overline{\Delta J_\nu^2} = \frac{c}{4}\eta A \overline{\Delta n_\nu^2}$$

$$= \eta J_\nu \left[1 + \frac{n_\nu}{M_\nu}\right] . \qquad (35)$$

The component of fluctuations at frequency f is

$$\overline{\Delta J_\nu^2(f)} = 2\eta J_\nu \left[1 + \frac{n_\nu}{M_\nu}\right]\Delta f \qquad (36)$$

(e.g. Smith et al. 1968, Kingston 1978). In the case of a detector working at 10 μm, where the background is at 300 K, the quantity n_ν/M_ν is negligible. Assuming a small detection bandwidth centred around the frequency ν, then equation (36) may be rewritten in terms of power received by the detector, viz.

$$\overline{\Delta P_\nu^2(f)} = \frac{2h\nu P_\nu}{\eta}\Delta f . \qquad (37)$$

If R_v is the voltage responsivity, then the noise voltage is

$$V_n = 2R_v\left[\frac{h\nu P_\nu \Delta f}{\eta}\right]^{1/2} . \qquad (38)$$

The extra factor of $\sqrt{2}$ comes from the recombination noise of the electrons excited by the background radiation.

2.5) Operating Considerations

2.5.1) Compensating Impurities

In addition to the impurities that give rise to the desired energy level structure, there are often others present that produce unwanted shallower levels (Figure 8a). At operating temperatures, these impurities may be ionized and, with a typical concentration of 10^{-3} to 10^{-2} that of the prime impurity (with current refining methods), this can cause a significant drop in responsivity.

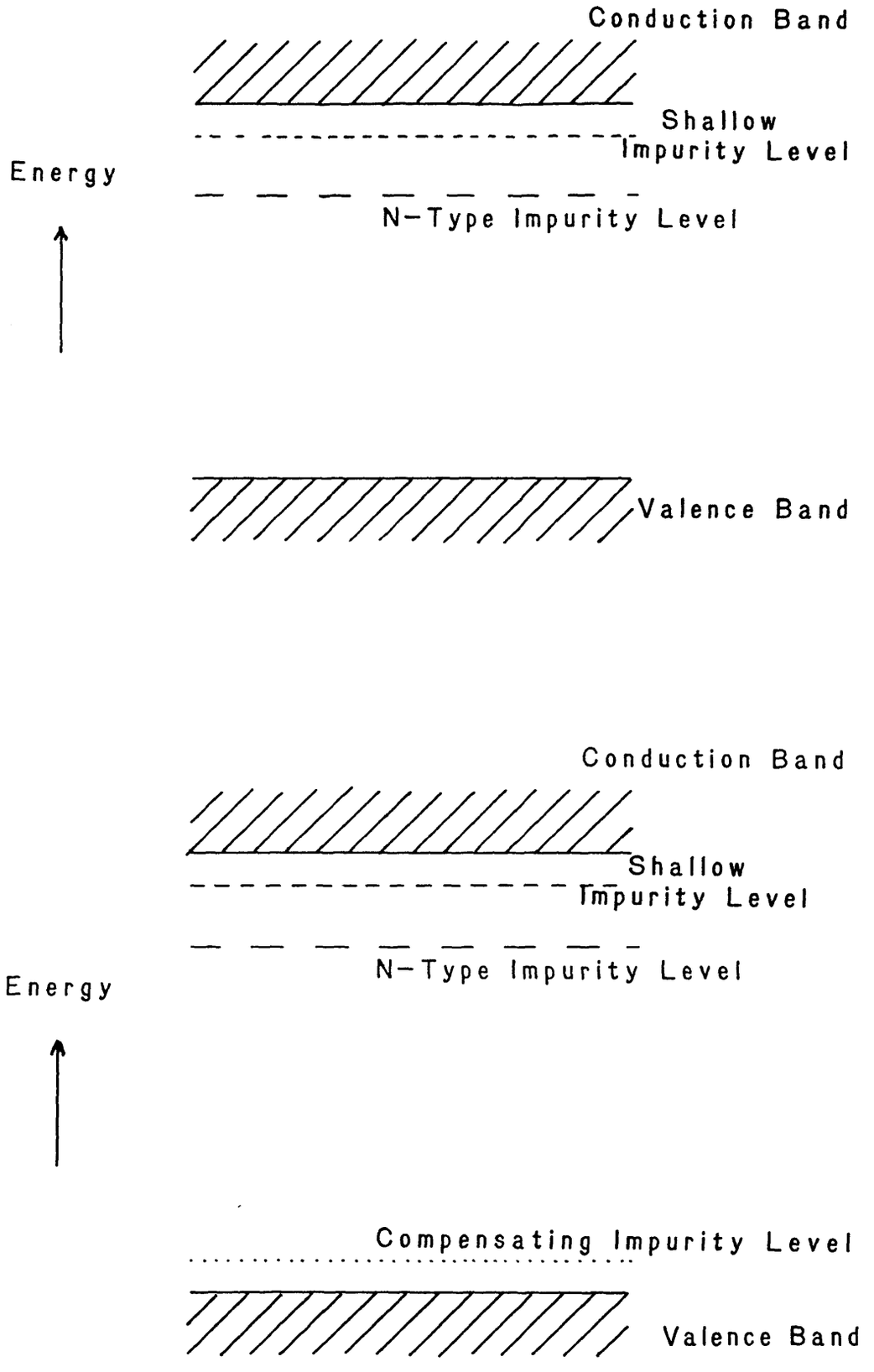


Figure 8. Energy level diagram showing the location of a) shallow impurity levels and b) compensating impurity levels in an n-type extrinsic photoconductor.

In order to offset this problem, compensating impurities may be added, which introduce a shallow acceptor level just above the valence band (Figure 8b). Electrons from the unwanted donor level fill this band (Levenstein (1965), Milton and Blouke (1971)), so reducing the problem. Exact matching of unwanted and compensating impurity concentrations is impossible, and so to prevent the existence of unwanted electrons, the detector is slightly over-compensated. Compensating impurities have a significant effect on the time constant of the detector, and this is discussed in the next section.

2.5.2) Time Constants

The quantity τ was introduced in equation (6) as the carrier lifetime. This is the average time between the excitation of an electron into the conduction band and its subsequent recombination with a recombination centre. This time constant is (Levenstein 1965)

$$\tau = \frac{1}{n_+ \sigma_i v_e} , \qquad (39)$$

n_+ being the number density of positively charged ion sites, v_e the electron velocity, and σ_i the capture cross-section. n_+ is best represented as the sum of two quantities: the number density of ionized sites due to the compensating impurities (n_c), and the number of ionized sites due to the incident radiation (n_r), i.e.

$$n_+ = n_c + n_r . \qquad (40)$$

If the intensity of radiation is very high, such that $n_r >> n_c$, then the time constant becomes dependent on the incident radiation, decreasing as the radiation level increases. In practice though, this condition is rarely met, $n_c >> n_r$ being more normal. Equation (39) then reduces to

$$\tau = \frac{1}{n_c \sigma_i v_e} \cdot \qquad (41)$$

Here, the recombination lifetime is controlled by the compensation impurity, and indeed, such tuning can vary τ by factors ~100 (Levenstein 1965).

In 1969, Williams noticed slow response effects not associated with the recombination time or the RC time constant of the detection circuitry. This he attributed to a sweepout of the majority carriers: if the lifetime of the electrons is long compared to the transit time across the detector, the electrons produced by the illumination may be swept out of the crystal. A space charge is left behind, which distorts the applied electric field. Carriers then diffuse into the crystal under the influence of the resultant field to restore charge neutrality. The relaxation time for this process τ_r, is given by

$$\tau_r = \frac{\xi_o \xi_r}{\mu \eta e} , \qquad (42)$$

where ξ_r is the relative permittivity of the detector material, and ξ_o is the permittivity of free space. For low fields, not all the charge will be swept out, but the relaxation time will remain the same. The importance of this effect is that it reduces the response of the detector with increasing modulation frequency, although this will only happen for a photoconductive gain of greater than unity (Milton and Blouke 1971).

2.5.3) Operating Temperature

The principal criterion for the choice of operating temperature is that thermal excitation of carriers must be negligible compared to the excitation due to incident radiation. The number of carriers in the excited state is governed by the Fermi-Dirac

distribution, viz.

$$\frac{n_2}{n_1} \simeq e^{-\Delta\epsilon/2kT} \quad . \tag{43}$$

If the detector is warm, then the limit on performance is due to g-r noise (section 2.4.3). As the detector is cooled, this is reduced and there comes a point at which noise due to the fluctuations in the background dominate. At this point, the detector is background limited, and further improvements are impossible. For doped-silicon detectors, this point is usually reached (for reasonable backgrounds) in the tens of Kelvin region.

2.6) Detection Circuitry

The "classical" circuitry for a photoconductor is shown in Figure 9a. The D.C. voltage at the measurement point is given by

$$V_L = \frac{R_L}{R_L + R_D} V_B \quad . \tag{44}$$

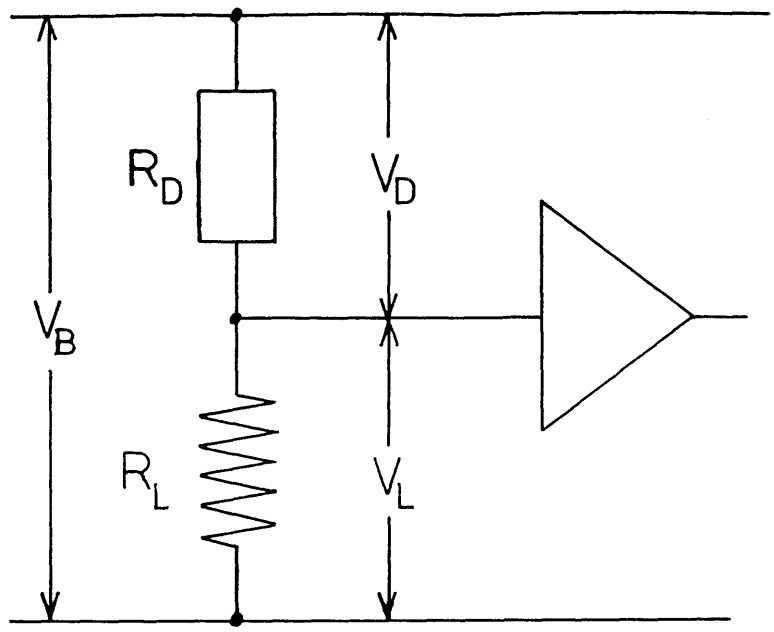
If the illumination level changes, causing a change ΔR_D in the resistance of the detector, then the consequent signal ΔV_L, is given by

$$\Delta V_L = -V_L \left[\frac{\Delta R_D}{R_L + R_D}\right] \quad . \tag{45}$$

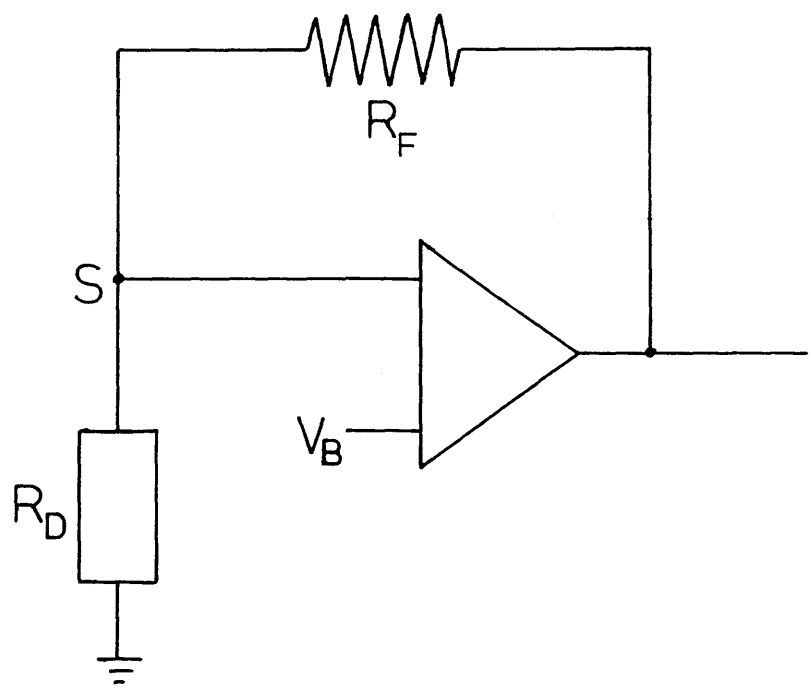
Since practically all infrared astronomical work is done with modulated radiation, and since photoconductors are in general non-ohmic devices, the A.C. resistance of the detector R_A, must be taken into account. R_A is defined by

$$R_A = \left[\frac{\partial V_D}{\partial I_D}\right]_\phi , \tag{46}$$

the subscript ϕ denoting a constant radiation level on the detector. A number of authors, amongst them Putley (1964) and Charlton (1964) have



a) "Classical" Circuit



b) Trans-Impedance Amplifier

Figure 9. Two types of detection circuitry for a photoconductor. The "classical" circuit (a voltage divider), with the following preamplification stage is shown in (a). The trans-impedance amplifier in (b) requires an operational amplifier as an intrinsic component of its operation.

applied this correction, to get

$$\delta V_L = \frac{R_L R_A V_B}{R_D(R_L + R_D)(R_L + R_A)}\left[\frac{\partial R_D}{\partial \phi}\right]_{V_D} \delta\phi \ , \qquad (47)$$

where the dependence of R_D on ϕ has been included.

With this circuit configuration, there remains the problem of how to choose the value of R_L: $R_L >> R_D$ (constant current mode), $R_L << R_D$ (constant voltage mode), or something in between. This problem has been considered by Penchina (1975) and Gopal (1981). Their main point is that the choice of load resistor depends on the noise predominating in the system. If the detector is limited by noise originating in the detector part of the circuitry, e.g. g-r noise, photon shot noise etc., then the value of R_L is irrelevant; both the signal and the noise are increased or decreased by the same proportion, and there is no change in the resultant signal-to-noise ratio. This is not so in the case where amplifier noise predominates, or if the limiting noise is due to Johnson noise in the detector - load resistor combination. Although a matched load resistor will give maximum signal, the signal-to-noise ratio can be increased by up to a factor of two if the load resistor is increased to the extent that the constant-current regime is entered (Penchina 1975).

If the photoconductor is operated under very low backgrounds e.g. a cooled telescope, then its resistance can be as high as $10^{12}\Omega$. In these circumstances, stray capacitance can impose a severe constraint on the response time, a few pF leading to time constants of the order of seconds. To overcome this, a trans-impedance amplifier has to be used (Figure 9b). The advantage that this has over the circuit of Figure 9a is that the summation point S is a virtual earth. Stray capacitance associated with the detector is unimportant, the time constant being determined by the capacitance associated with the

feedback resistor R_F. Since R_F may be a few dex smaller than R_D, this restriction is eased somewhat. A detailed description of the operation of this type of circuit has been given by Luinge et al. (1980).

The presence of a high impedance point at S also introduces two other problems. Firstly, unless the amplifier has an input impedance much greater than that of the detector - feedback resistor combination, severe loading by the amplifier will cause a sharp decrease in the signal. To overcome this, the amplifiers often have a J-FET or MOSFET input stage. The extremely high gate-source impedance of these devices is enough to prevent the loading. A practical amplifier using a MOSFET input stage has been given by MacGregor (1977). The other problem associated with the high impedance point is that of pickup noise. The only practical solution to this is to make all the leads as short as possible. In practice, this usually means mounting the MOSFET input stage actually in the liquid-helium cooled part of the dewar. Unfortunately, the temperature usually has a detrimental effect on such devices, often leading to sharp increases in noise, or malfunction (Young and Low 1979). At the moment, selection of a suitable MOSFET device is still very much a "try it and see" process, but work is going on into other ideas, such as heating the input stage during operation.

In conclusion then, a simple voltage amplifier is good enough if the detector is of relatively low impedance, but once the GΩ and above impedance is reached, a trans-impedance amplifier offers the only practical solution.

2.7) Photoconductors for the 10 μm Region

There are many photoconductors available for use in the infrared, but relatively few suitable for astronomical purposes, due

to the high sensitivity required. The impurity-doped photoconductors Ge:X and Si:X show both high gain and large internal resistance at normal background levels, and are particularly suitable for this type of work (Tolner 1977), more so than intrinsic photoconductors e.g. (Hg, Cd) Te.

Sclar (1976) has done a survey of Si:X detectors for use at 10 μm. He investigated the dopants Al, Ga, Bi and Mg and, with a 30^{o} field of view, was able to get background limited performance on all of them. Of the four dopants tested, Bismuth was the most successful, with a peak D^{*} of ~5 x 10^{10} cm $Hz^{1/2}W^{-1}$. This can be translated to noise-equivalent power (NEP), a much more useful measure of detector sensitivity in the astronomical infrared by the equation (Clark-Jones 1959)

$$NEP = \frac{(A\Delta f)^{1/2}}{D^{*}} . \qquad (48)$$

Taking a typical detector of area 1 mm^2, and a measurement bandwidth of 1 Hz, an NEP of 2 x 10^{-12} W $Hz^{-1/2}$ is obtained. However, this value does not represent the ultimate detector sensitivity, as it is a background limited value.

More recently, work has been concentrated on Si:As as a detector for this wavelength region. This material has a cut-off wavelength (and hence a peak of response) at 22 μm, but nevertheless, has extremely good characteristics at 10 μm. Luinge et al. (1980) did tests on some detectors for the IRAS satellite, and obtained a figure of 10^{-16} W $Hz^{-1/2}$ for the NEP at the backgrounds expected for the satellite observations. Tests near the peak of response (in the 18.5 - 22 μm region) show that the NEP can drop as low as 4 x 10^{-17} W $Hz^{-1/2}$ (Young and Low 1979).

There are also a number of doped germanium detectors

available for use in this waveband, some detectors matching the wavelength response of the silicon detectors very closely, e.g. Ge:Cu mimics the behaviour of Si:P (Bratt 1977). In particular, Ge:Cd has a response curve shape practically identical to that of Si:As, but the magnitude of the response is only about 20% that of the silicon detector (SBRC 1976). Interest has been particularly centred on impurity silicon detectors however, owing to the high solubility of the impurities in that material (Emmons et al. 1975). This leads to a high absorption coefficient and the consequent high quantum efficiency.

Of the remaining types of photoconductors for this region, probably the best-developed is the pseudo-binary alloy, $Hg_{1-x}Cd_xTe$. This is an intrinsic photoconductor, whose band gap varies with the proportion of cadmium. It is a very low impedance device, less than 1 KΩ (Palmer 1979), and is capable of being operated at 77 K (liquid nitrogen temperature). Unfortunately, it is these advantages that give it a low detectivity, of the order of 10^8 cm $Hz^{1/2}W^{-1}$ (Saur 1968).

In conclusion then, it can be seen that impurity-doped silicon detectors are the most sensitive photoconductors yet developed for this region of the infrared: of these, arsenic-doped silicon is currently the most sensitive.

CHAPTER 3

Design Construction, and Testing of a Photoconductor-Based Photometer for 10 μm Astronomy

3.1) Introduction

In Chapter 1, a programme of observing interacting galaxies at 10 μm was suggested, and in Chapter 2, an arsenic doped silicon photoconductor was put forward as being the most sensitive photoconductor yet available in the 10 μm band. This chapter describes the design and construction of a 10 μm Si:As based photometer for carrying out the described programme, and the tests carried out on it by the author at Imperial College between 1980 and 1981.

3.2) Photometer Design and Construction

Due to time constraints, the photometer design was neccessarily simple. It was based on far-infrared photometer designs of the Imperial College Far-Infrared Astronomy Group (e.g. Sugden 1978), but was designed for the Infrared Flux Collector (IRFC) at Tenerife. Two versions of the photometer were eventually built, the mark 1 to use the tertiary chopping arrangement described by Jorden et al. (1976). and the mark 2 to use a UKIRT-compatible television guiding system. Other than rotating the optical axis of the photometer through 90^{o} relative to the dewar, only minor changes in the basic design were made between the two versions, and these will be noted as they are encountered.

3.2.1) Mechanical Aspects

3.2.1.1) The Detector

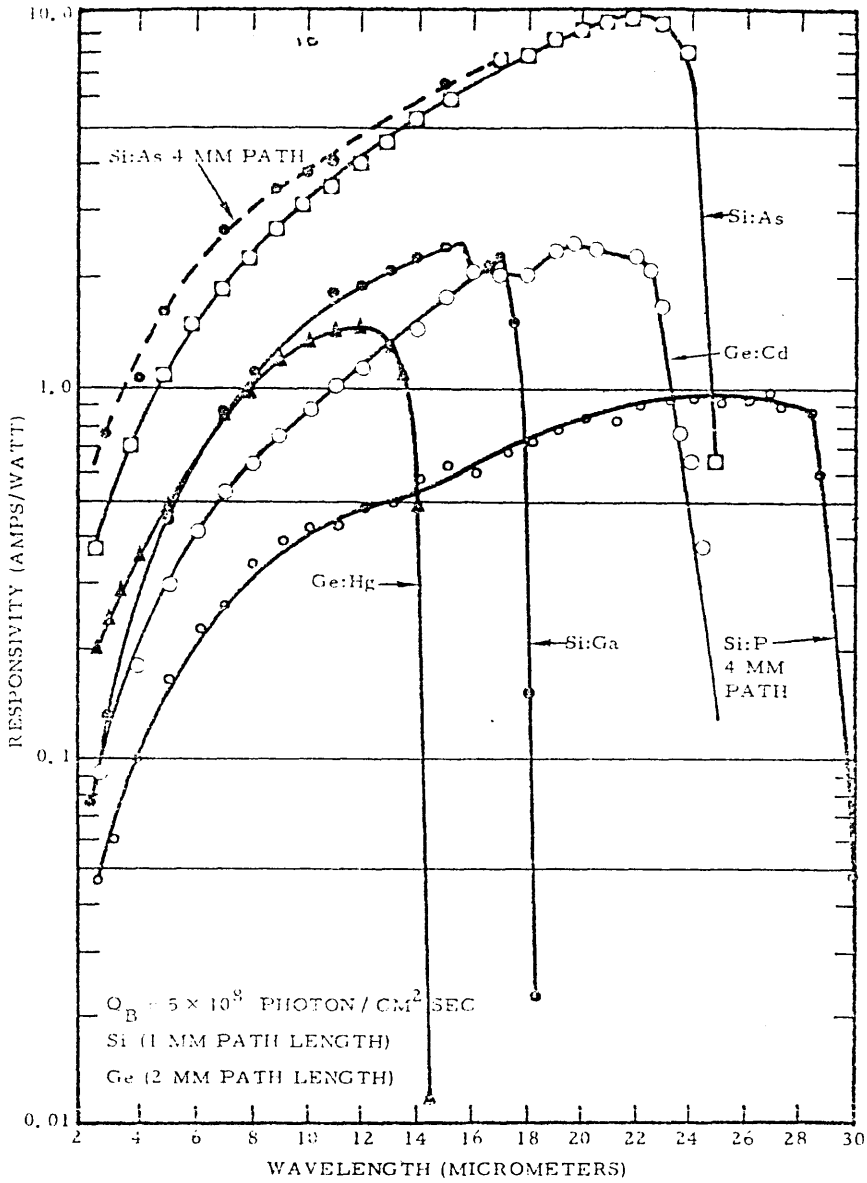
The detector used in the photometer was a Santa Barbara Research Center arsenic-doped silicon extrinsic photoconductor. The dopant concentration and other similar details are unknown since, despite requests, no specification sheet has ever arrived. However, an expected response curve for similar detectors under low backgrounds has been given in a technical bulletin (SBRC 1976) and is shown in Figure 10.

Figure 11 gives the dimensions and layout of the detector. Electrical contact is made through a transparent front electrode as one terminal, and through the mounting block as the other. The advantage of a reflecting rear surface was discussed in section 2.3.5.

3.2.1.2) Cooling System

As mentioned in section 2.5.3, during operation, the detector must be cooled below a few tens Kelvin in order to prevent the conduction band from being flooded with thermally excited carriers. The simplest way of achieving this is to attach the detector to a (heat-conducting) container of liquid helium. There are a number of commercially available dewars on the market for just such a purpose, and the one selected was an Infrared Laboratories model HD3.

The internal structure of the HD3 is shown in Figure 12. At the heart of the dewar is the liquid helium can, which has a copper work surface (the cold-face) on which the equipment is mounted. Surrounding this can, and the work space below it, is a shield to prevent excessive heat loading of the liquid helium. This shield is kept at 77 K by means of liquid nitrogen, held in the toroidal chamber



Spectral Responsivities for Detectors Fabricated by SBRC

The measurements shown above were made at a greatly reduced background of 5×10^8 photons/cm^2 and at an operating temperature of 8°K.

Figure 10. Expected response of extrinsic photoconductors under low backgrounds (SBRC 1976). As noted in section 2.3.5, the response rises to a peak at the cut-off wavelength, and then abruptly falls.

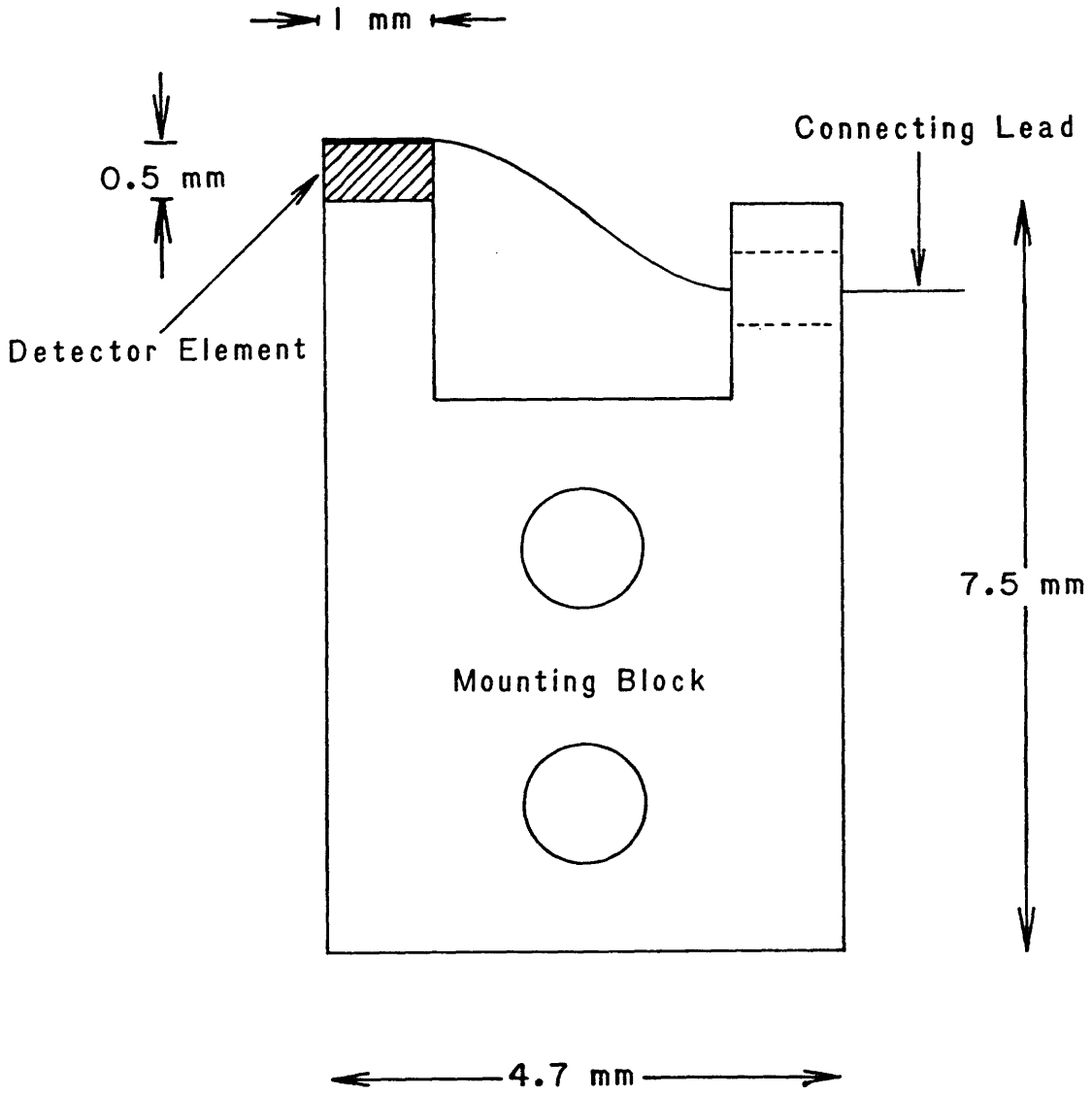


Figure 11. The Si:As photoconductor, with approximate dimensions. The detector element is surfaced by a transparent electrode. To give mechanical strength to the connecting lead, it passes through a block of insulating material in the mounting block. The second electrode is the mounting block itself.

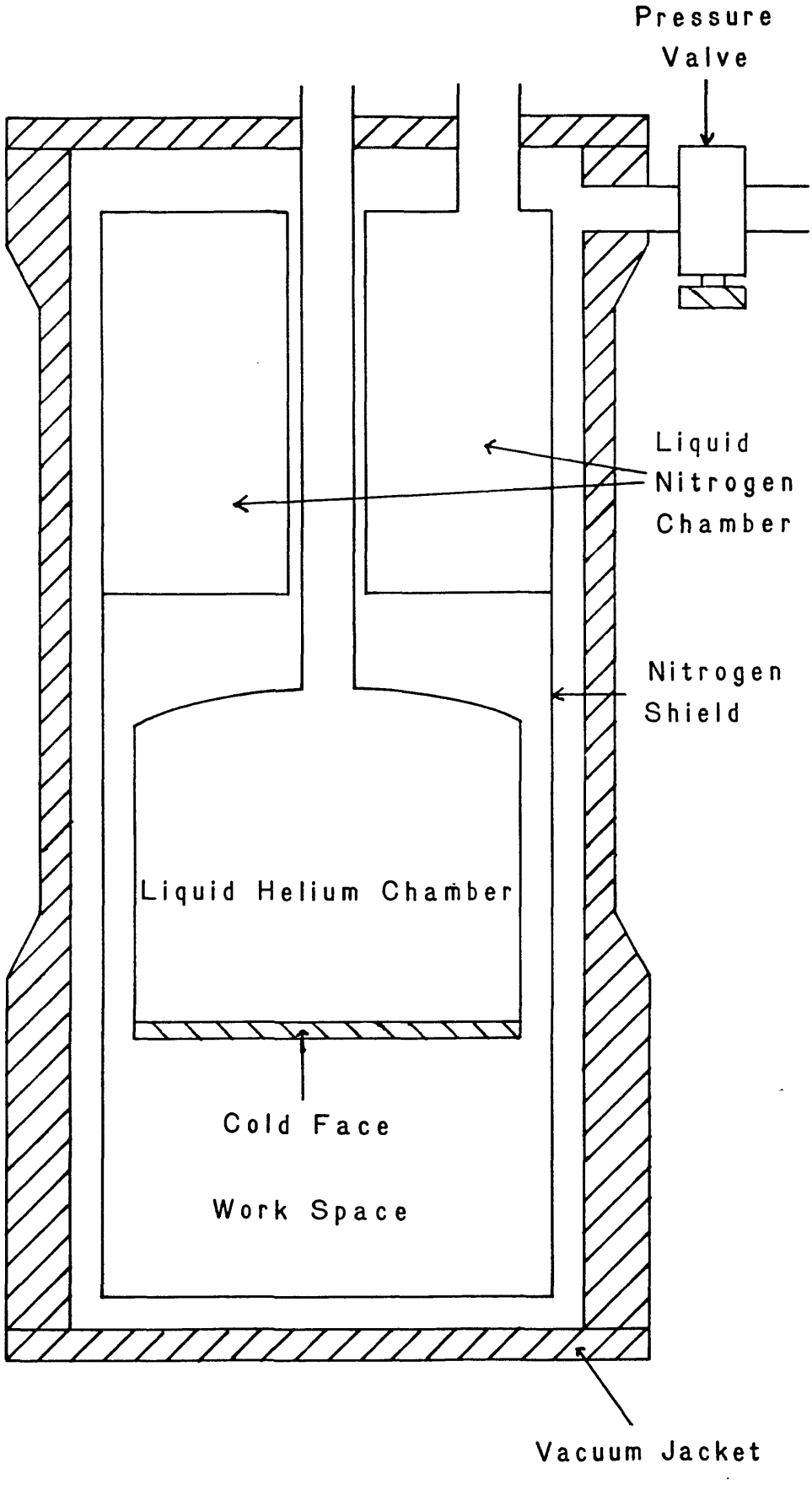


Figure 12. Simplified cross-section of an Infrared Laboratories HD3 dewar. The liquid nitrogen chamber is toroidal, and surrounds the liquid helium can filler tube.

above the liquid helium can. All surfaces, except the cold-face, are covered in aluminium foil to decrease emissivity and therefore to further reduce heat input. The whole assembly is enclosed in a vacuum jacket. During operation, the internal vacuum is less than 10^{-7} Torr.

To check the basic irreducible heat leak into the dewar, it was pumped down to operational pressure, and filled with the appropriate cryogens. The hold time for 1.05 litres of liquid helium was 90 hours, although the 0.8 litre nitrogen can required refilling every 28 hours. This indicates a heat leak to the liquid helium of ~8 mW, which is close to the expected heat loading by radiation from the nitrogen shield alone of 6.5 mW (assuming a wavelength-independent emissivity of 0.05 for the aluminium foil). In order to maintain a hold time for the helium of ~12 hours, long enough for a nights observing, the additional heat loading on the helium due to radiation input and conduction along the detector connections should not exceed ~50 mW.

The need for the detector to be cooled to liquid-helium temperatures also dictated the choice of the material out of which the photometer was made. The material had to have a relatively low specific heat capacity, or else an excessive amount of helium would be used in cooling the detector (this fact also constrained the photometer to use as little material as possible), and also have a high thermal conductivity, so that a minimal amount of time is spent waiting for the photometer to reach operating temperature. Experience has shown that oxygen-free copper is a suitable material, and so this was used for all parts of the photometer that were cooled to liquid-helium temperatures.

3.2.1.3) Mounting the Detector

The small size of the detector, combined with its delicate nature, made the mounting of it within the photometer difficult. An added problem was the need to keep the detector in thermal contact with, but electrically isolated from, the cold-face.

The size problem was overcome by attaching the detector directly to a bracket made of oxygen-free copper, and regarding the bracket as an integral part of the detector. Electrical isolation of the detector/bracket combination from the cold-face was achieved using the arrangement shown in Figure 13. The main body of the bracket is insulated from the photometer base (and hence the cold-face) by a sheet of mylar film. At the fixing points, the bracket is insulated from the fixing screw by a teflon collar. The spring and washer combination serve to keep the bracket firmly attached to the base despite thermal contraction of the parts when the photometer cools down.

The measured resistance of the insulation at 4.2 K was 490 MΩ. In the circuit used (section 3.2.4), this lay in parallel with the load resistor, of maximum resistance 3.8 MΩ, and so contributed an error of less than 0.8% to the measurements.

In the mark 2 photometer, the detector bracket was engineered so that it provided an integrating cavity around the detector element. No improvement in performance was noted because of this, but the arrangement was retained since it offered extra protection to the detector when the latter was being handled.

3.2.2) Optical Aspects

The optical system in the photometer has to do two things: it has to isolate the wavelength region of interest, and it has to focus the radiation on the detector. The various elements must also be able

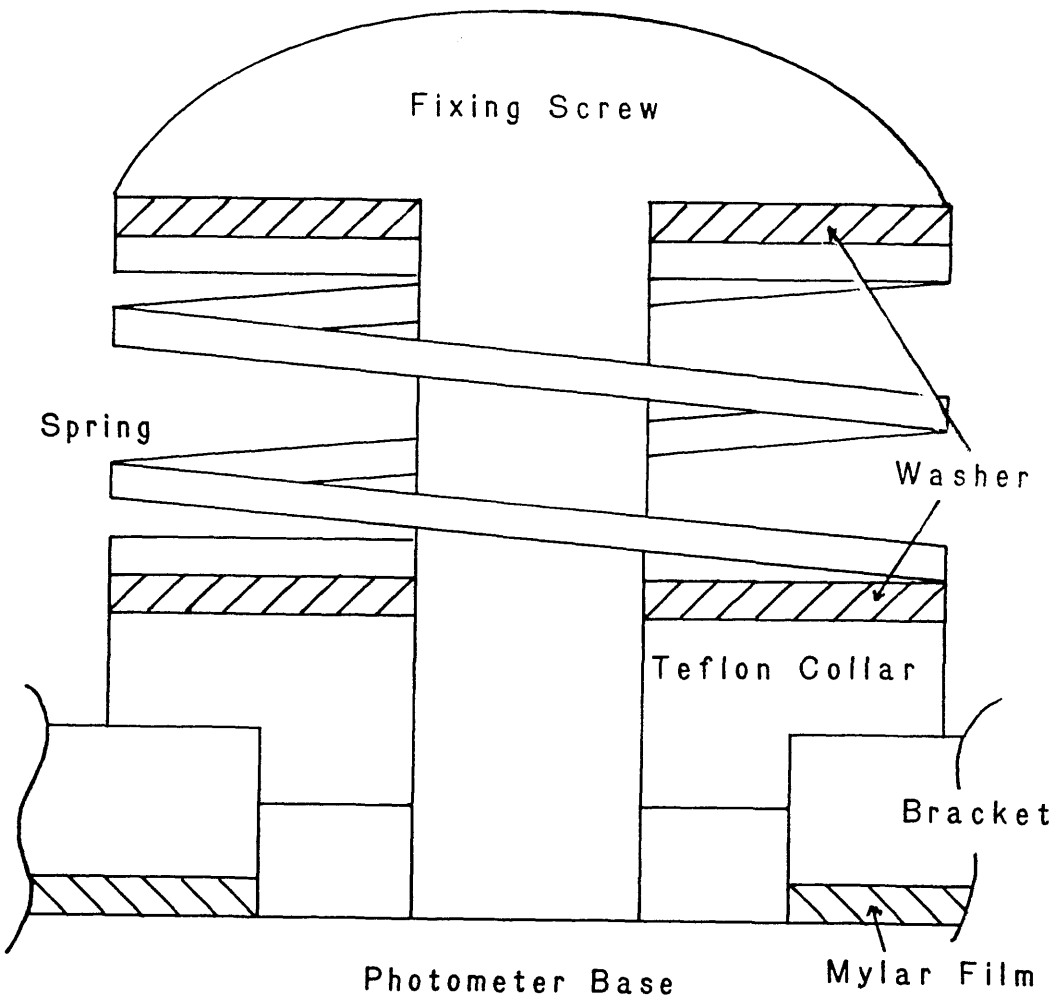


Figure 13. Detail of the detector bracket/photometer base connection. A sheet of mylar film provides the necessary electrical insulation, whilst ensuring suitable thermal contact.

to withstand the temperatures within the dewar.

3.2.2.1) Filters

To isolate the 10 μm wavelength region, a bandpass filter is required. Infrared bandpass filters (usually a multilayer dielectric) are normally designed and made in large batches for specific customers; any spares may then be sold on a retail basis. Thus the selection procedure boils down to scanning manufacturers catalogues and choosing the filter most closely satisfying the scientific requirements. The actual choice of bandpass filter may also be constrained by availability and cost. It was in this way that the bandpass filter was bought from OCLI Ltd. The transmission curve is shown in Figure 14.

A drawback with bandpass filters is that there are spectral leaks, and so long and short wavelength blocking is still needed. With the particular filter used, the dielectric of the filter provided the required short-wavelength blocking. Long-wavelength blocking was obtained with a barium fluoride filter.

3.2.2.2) The Fabry Lens

The optics of the telescope image an area of sky onto the focal plane. Since the plate scale is set by the f-number of the telescope, without some form of focal plane optics the maximum area of sky that could be observed in any one observation would be set by the size of the detector. A second reason why focal plane optics are needed is the possibility that there is a sensitivity variation across the detector. If this were the case, and the source were imaged directly onto the detector, then the signal measured would be highly dependent on the exact position of the image on the detector. For these two reasons, it is desirable to have some form of optical

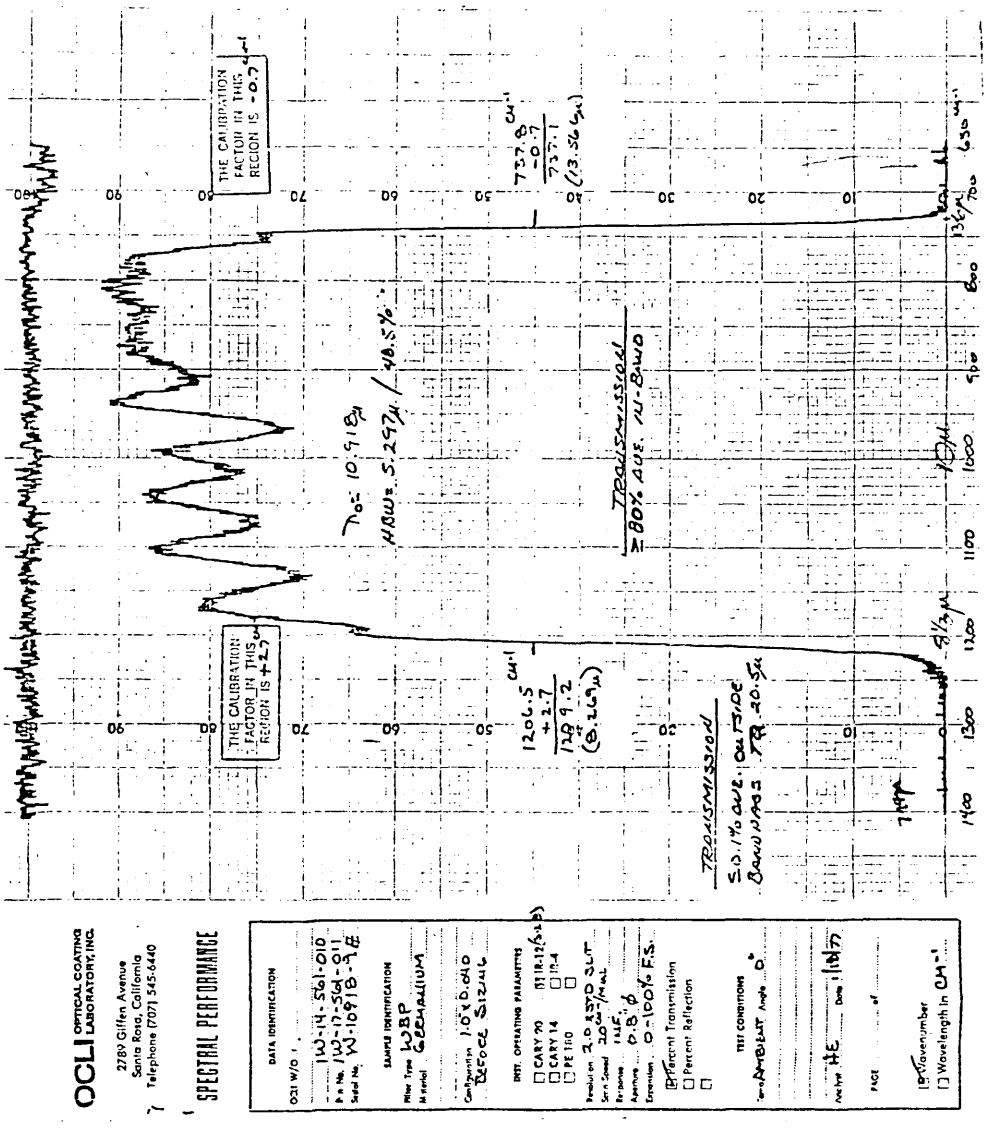


Figure 14. Transmission curve supplied with the OCLI filter.

arrangement that blurs the image over the entire detector.

The blurring may be achieved by an optical configuration that images the primary mirror of the telescope on to the detector. Essentially, there are two ways that this can be done: an arrangement of mirrors, and an arrangement of lenses. A mirror configuration, although having the advantage of no chromatic aberration, was rejected because of the complicated internal arrangement of the photometer that would be required, and the consequent increase in cost and time this would involve. A lens arrangement, on the other hand can be made quite quickly and cheaply, and can lie entirely on the optical axis of the photometer, thus eliminating the need for an awkward internal configuration. The simplest form of such a lens arrangement is a single plano-convex lens, because: 1) there is only one component, and 2) within that component, non-coincidence of the optical axes of the two surfaces is avoided. Accordingly, it was decided (at least initially), to proceed with the single lens design for the focal plane optics.

Potassium chloride was selected for the lens material. It has an almost constant transmission of 0.92 over the 10 μm band, and a nearly constant refractive index. It was this latter consideration that prompted against the use of barium fluoride as a combined long-wavelength cut-off filter and Fabry lens. Near the cut-off wavelength, not only does the transmission decrease, but the refractive index alters as well, which would make the image wavelength-dependent.

As a first estimate of the lens required, the simple thin lens formulae,

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \qquad (49)$$

and

$$\frac{h_v}{h_u} = \frac{v}{u} \qquad (50)$$

were used, where u and v are the object and image distances respectively, h_u and h_v the object and image heights, and f the focal length of the lens. The detector is 1 mm × 1 mm in area, and so allowing for diffraction, and for a margin of error, h_v was taken to be 0.9 mm. The telescope for which the photometer was being designed, the IRFC, has an f-number of 13.8 (IRFC users manual 1976) which, with the diameter of the telescope being 60 inches (h_u), gives u = 828". Substituting these figures into the above equations, gives f = 12.4 mm. Taking the average refractive index of potassium chloride to be 1.45 in the 10 μm band (Smakula et al. 1967), then for a plano-convex lens, the formula

$$\frac{1}{f} = \frac{n - 1}{r} , \qquad (51)$$

where n is the refractive index, gives a radius of curvature r of 5.6 mm. A more thorough examination, using a ray-tracing program for a thick lens, showed this quick calculation to be suprisingly good. A potassium chloride lens, of radius of curvature 5.6 mm and thickness 1.3 mm, will have, with the entrance pupil limited to 1.5 mm (15" on the IRFC), a circle of least confusion 0.9 mm in diameter at a distance of 11.35 mm behind the plane surface of the lens. There will be some diffraction, which will blur the circle by about 0.1 mm (~λf/aperture), but this falls within the confines of the detector.

3.2.2.3) The Vacuum Window

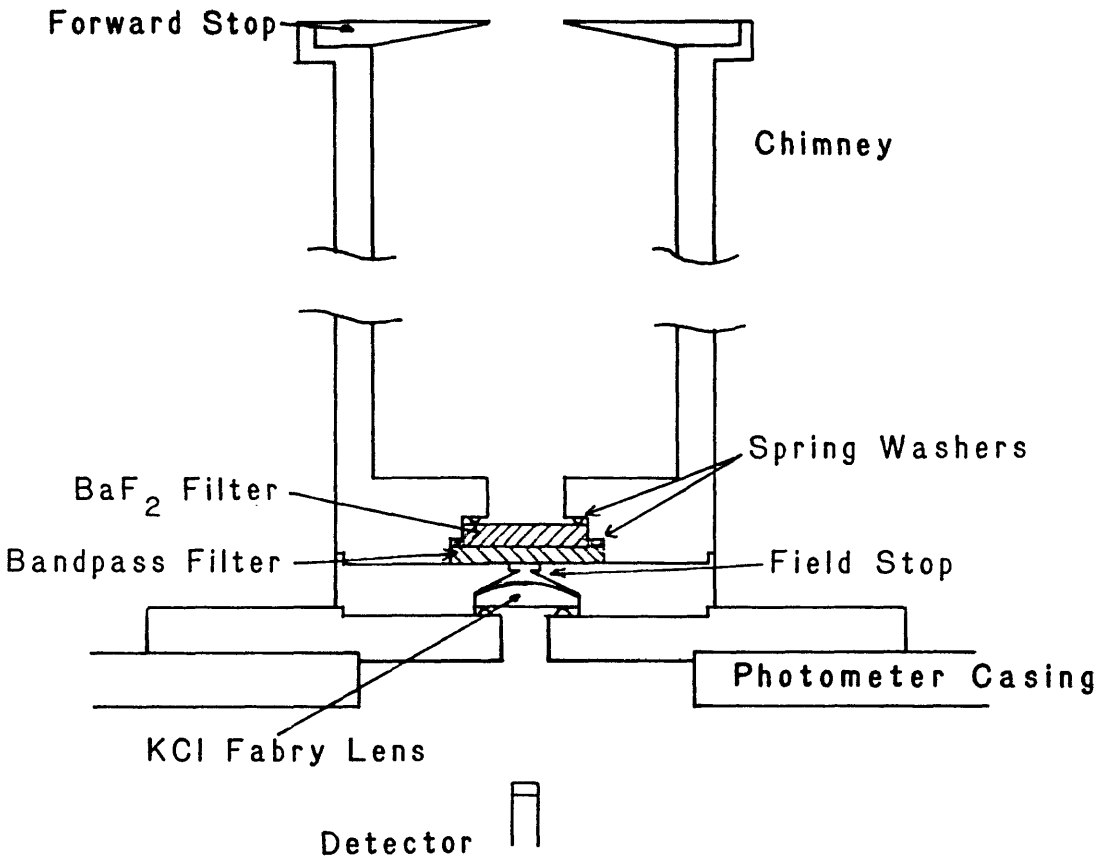
One element in the optical system that is unavoidable is the window in the vacuum jacket. Here, transmission characteristics are secondary to mechanical considerations. The window must be large

enough to allow entry of the cone of radiation from the telescope, but at the same time must be strong enough to resist a pressure differential of one atmosphere. It must also be flexible enough to resist any (reasonable) shocks; a sudden loss of the dewar vacuum while liquid helium was still present could have catastrophic consequences. Finally, it must be able to withstand exposure to the atmosphere without degradation. This last requirement effectively rules out the alkali halides, selected for the Fabry lens and long-wavelength blocking filter, since this class of materials tends to be highly hygroscopic. The substance chosen was KRS-5, a thallous bromide - iodide compound. This acts as a neutral-density filter, but although it has a transmission of only 0.72 in the 8 - 14 μm band, it satisfies the above requirements. It does, however, have the hazard of being relatively toxic.

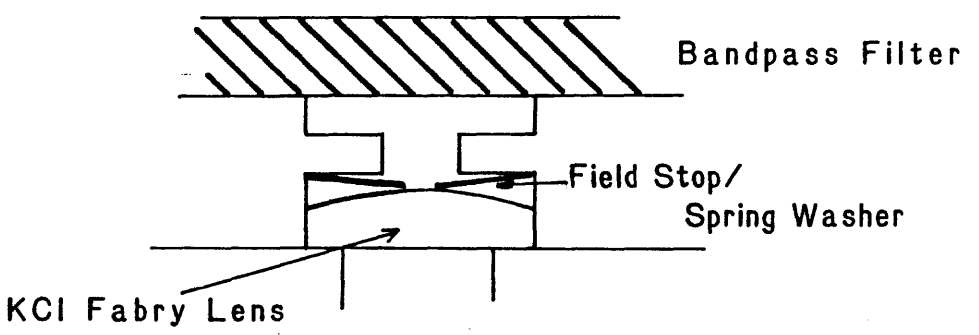
3.2.3) Constructional Details

A schematic layout of the photometer is shown in Figure 15a. The material used was oxygen-free copper (for reasons noted above), and the parts were screwed together with brass screws, to avoid problems associated with thermal contraction. All external surfaces of the photometer were gold plated, to reduce the heat leak to the nitrogen shield, and all internal surfaces (bar those in contact with other surfaces, or with the optics) sandblasted, and painted with 3M Nextel paint, to eliminate internal reflections.

The optics of the photometer were housed in a removable "chimney" assembly. With engineering tolerances of a couple of thousandths of an inch, this could be removed and replaced with the alignment intact: a desirable feature, as both the barium fluoride and potassium chloride elements are hygroscopic, and this allowed their



a) Photometer Design



b) Mark 2 Aperture Details

Figure 15. Photometer constructional details. The optical configuration of the (mark 1) photometer is shown in (a). Other than rotating the base through 90° for use with the television guiding system, the only significant change between mark 1 and mark 2 is shown in (b). The field stop is made out of a thin foil, and is bowed slightly in order to keep the Fabry lens in place.

insertion to be delayed until the last minute. To allow for thermal contraction, the optical components were of a somewhat sloppy fit in their holders, and so phosphor-bronze spring washers were inserted to keep them in place. An improvement made in the mark 2 photometer was to combine the function of the spring washer and aperture stop (Figure 15b). This has the advantage of ensuring that the back of the lens is normal to the optical axis, as well as allowing a large number of apertures to be made very cheaply.

Alignment for the mark 1 was achieved by moving the casing (to which the chimney assembly was fixed) relative to the detector, which was fixed to the photometer base. The photometer was assembled without the filters, and a beam of light passed down the photometer. The light was reflected from the front surface of the detector back up the photometer. In this manner, the position of the detector relative to the aperture could be ascertained, and the photometer casing adjusted. There was no photometer-lens distance adjustment. This was found to be unsatisfactory in practice, and so for the mark 2, the base and casing were made as an integral unit. The detector lay fixed on the optical axis of the photometer, but adjustment was still possible along that axis. The detector-lens distance could be accurately set using engineer's blocks.

3.2.4) Electronic Aspects

Inital measurements of the resistance of the detector soon showed that a trans-impedance amplifier was not needed. Accordingly, the "classical" circuit, described in section 2.6 was implemented, using a preamplifier designed for the Imperial College stabilized balloon telescope (Kessler 1981). These amplifiers are of very low noise (~10 nV $Hz^{-1/2}$ at 20 Hz), and have good low-frequency

characteristics, having an almost constant gain of 1000 over the range 5 - 100 Hz.

As mentioned in section 2.6, if the detector is background limited (as was the design aim), then the choice of load resistor is irrelevant when trying to maximize the signal-to-noise from the detector. Where the choice does matter is in the linearity of response. As can be seen from equation (47), the detector resistance (both A.C. and D.C.) enters into the responsivity. Thus if, during the course of observations, the background changes significantly, then errors will creep into the signal estimation.

If we take equation (47), substitute for $\partial R_D/\partial\phi$ from equation (18), and take R_L small enough so that $R_L << R_D$, and $R_L << R_A$, then we get

$$\delta V_L = - \frac{R_L}{R_D \phi} V_B \delta\phi \ . \qquad (52)$$

Noting that $R_D \propto \phi^{-1}$ (from equation (18)), gives a responsivity independent of background.

In practice, having R_L so small is not such a good idea, as although both the signal and the noise are reduced, the region where preamplifier and other electronic noise predominates can be entered. The photometer system thus becomes electronics - rather than background - limited. Other considerations are also relevant to the choice of load resistor. To reduce the Johnson noise, the load resistor is usually mounted on the cold face. Some resistors generate excess (current) noise when cooled to liquid-helium temperatures, and these must be avoided.

In the final system, the load resistor was a 3.8 MΩ wire-wound resistor, of the type used for the photometers on the Imperial College balloon telescope (Sugden 1978). Although this was

rather high, being ~5 - 15% of the D.C. resistance of the detector as measured at the telescope (this figure depending on the weather) it was a low noise resistor. Three other mitigating factors are firstly, the resistance of the detector is set by the background radiation. As section 3.3 shows, this is expected to come mostly from the telescope and dewar window, and so should be almost constant over the period of observation. Secondly, most observations will be done at, or near, the zenith. With good observing conditions, the sky background shouldn't change much over the path swept by the telescope beam, and so responsivity changes should be small. Finally, comparative photometry is being attempted, rather than absolute photometry, so nightly variations in the sky (or telescope) background are removed in the calibration.

To minimize Johnson noise, the load resistor was mounted on the cold face. Connection between the detector, load resistor and the outside was by means of small diameter constantin wires, to reduce heat input. Microphonics were reduced by making these wires as short as possible, and fastening them down with GE varnish wherever possible.

3.3) Limiting Sensitivity on the IRFC

Before embarking on a programme of measuring and improving the response of the detector, it is worthwhile calculating the attainable sensitivity limit. From section 2.4.5, this will occur when all the noise in the detector is due to photon shot noise from the background radiation. The actual amount of background will depend on the telescope and details of the photometer used, and so the figure derived here is only applicable to the IRFC.

From equations (5) to (8), the current when a signal flux of

Φ_S photons s^{-1} is incident on the detector is

$$i_S = eG\eta\Phi_S \ . \qquad (53)$$

The noise current for a background flux Φ_N is given by combining equations (5) through (8) and equation (36) (noting, as before, that the term n_ν/M_ν is negligible), viz.

$$i_N = 2eG(\eta\Phi_N\Delta f)^{1/2} \ . \qquad (54)$$

As before, the noise includes a factor of $\sqrt{2}$ due to recombination noise. For a signal to noise ratio of unity, $i_S = i_N$ and so, with $\Delta f = 1$ Hz,

$$\Phi_S = 2\left[\frac{\Phi_N}{\eta}\right]^{1/2} \ . \qquad (55)$$

The limiting sensitivity is thus

$$S_L = 2\frac{\langle h\nu\rangle}{A_p T_o \Delta\nu}\left[\frac{\Phi_N}{\eta}\right]^{1/2} \ , \qquad (56)$$

where T_o is the overall transmission of the telescope/photometer system, $\langle h\nu\rangle$ the mean photon energy in the passband, A_p the area of the primary mirror, and $\Delta\nu$ the frequency width of the passband. In infrared astronomy though, only half the time is spent looking at the source. Thus the above value must be doubled if a given signal-to-noise ratio is required in a given time. The limiting sensitivity in this case is S_{LIM}, given by

$$S_{LIM} = 2S_L \ . \qquad (57)$$

To calculate Φ_N we note that the background radiation has three sources. The vacuum window, the telescope, and the sky. The cold optics within the photometer contribute a negligible amount of background radiation. If the photometer has a field stop of area A_f,

and the optics limit the radiation to a cone of solid angle Ω_t, then the number of photons reaching the detector from the dewar window is

$$\Phi_w = \int N(\lambda,\theta_w) T_c \epsilon_w A_f \Omega_t d\lambda \ , \qquad (58)$$

where $N(\lambda,\theta_w)$ is the Planck function (in photons $m^{-2}s^{-1}sr^{-1}$) for a temperature θ_w, ϵ_w is the window emissivity, and T_c is the transmission of the cooled optics. Strictly speaking, both the emissivity and transmission are wavelength dependent. However, apart from the sky emissivity, over the wavelength region of interest these quantities should only be weakly dependent on wavelength. Where they do vary significantly (as in the case of the transmission of the cooled optics, where the transmission of the bandpass filter ranges between 0.70 and 0.95), a mean value will be used; a "ball park" figure is all that is required, and so fine variations are unimportant.

The radiation reaching the detector from the telescope is similarly

$$\Phi_t = \int N(\lambda,\theta_t) T_c T_w \epsilon_t A_f \Omega_t d\lambda \ , \qquad (59)$$

but the emissivity is now ϵ_t, the temperature θ_t, and the transmission of the window T_w, has also to be taken into account. The contribution from the sky is

$$\Phi_s = \int N(\lambda,\theta_s) T_t T_c T_w \epsilon_s(\lambda) A_p \Omega_s d\lambda \ , \qquad (60)$$

which takes the effective sky temperature to be θ_s, and the effective sky emissivity to be $\epsilon_s(\lambda)$. The overall transmission of the telescope is T_t, and the solid angle in which the object is observed is Ω_s. This last quantity is set by the size of the field stop. The total amount of radiation reaching the detector is therefore

$$\Phi_N = \Phi_w + \Phi_t + \Phi_s \ . \qquad (61)$$

Inserting figures: The IRFC has a 60 inch diameter primary mirror, giving $A_p \simeq 1.5\ m^2$. The f/13.8 configuration of the telescope gives a solid angle Ω_t of 4.1×10^{-3}sr, and also means that a 10" field of view requires a 1 mm diameter field stop, giving $A_f \simeq 7.9 \times 10^{-7} m^2$. The 10" also gives Ω_s, the solid angle on the sky, a value of 1.8×10^{-9} sr.

T_c is the product of the transmission of the bandpass filter, (averaged out at 0.82 in in the passband), the barium fluoride long-wavelength blocking filter (0.74), and the Fabry lens (0.92), giving a total transmission of 0.56. The $N(\lambda,\theta)$ term is the Planck function, and the window and telescope temperatures will be taken to be 293 K (room temperature). The sky temperature will be taken to be 260 K (Low and Rieke 1974). The transmission of the KRS-5 window T_w, is a constant 0.72 over the wavelength range. Allowing for some reflection, a value of 0.1 for the emissivity of this component seems reasonable. Substituting these figures into equation (58) gives the contribution from the window as

$$\Phi_w = 4.0 \times 10^{11} \text{ photons } s^{-1} \ . \qquad (62)$$

The emission from the telescope and chopper is somewhat more complex, as there are four mirrors to be taken into acount. If the reflectivity of each mirror is R, then the emissivity of each mirror is (1 - R). For a four mirror system, this will lead to an overall reflectivity of R^4 (corresponding to T_t), and an overall emissivity ϵ_t of $(1 - R^4)$. If R is taken to be around 0.95, then $T_t = 0.81$, and $\epsilon_t = 0.19$. Inserting these values into equation (59) gives

$$\Phi_t = 5.4 \times 10^{11} \text{ photons } s^{-1} \ . \qquad (63)$$

Finally, there is the contribution from the sky. Over the 8 - 14 μm region, the atmospheric transmission is of order unity, apart from the 9.6 μm ozone absorption feature, and a slow tail off longwards of 12.5 μm (Figure 16). A numerical integration of the $\epsilon_S(\lambda).N(\lambda,260)$ product over the passband gives, from this graph, a value of 1.36×10^{20} photons $m^{-2}s^{-1}sr^{-1}$. Including the attenuation of the various other components, the total sky contribution is

$$\Phi_S = 1.3 \times 10^{11} \text{ photons } s^{-1} . \qquad (64)$$

Inserting these values into equation (61) gives the background as

$$\Phi_N = 1.1 \times 10^{12} \text{ photons } s^{-1} . \qquad (65)$$

Taking the mean photon energy to be that of a photon of wavelength 10.7 μm (see section 3.4.3), and taking the quantum efficiency to be unity (which will give an upper limit to the sensitivity), this corresponds to a background limited NEP, given by the equation

$$NEP_{BL} = 2\sqrt{\Phi_N}\langle h\nu \rangle , \qquad (66)$$

of $\sim 3.9 \times 10^{-14}$ W $Hz^{-1/2}$. $\Delta\nu$ is the width of the 8 - 13 μm waveband defined by the filters ($\sim 1.3 \times 10^{13}$ Hz), and so by equation (57), the limiting sensitivity at the IRFC will be

$$S_{LIM} \simeq 1.2 \text{ Jy } 1\sigma 1s . \qquad (67)$$

In other words, with the detector background limited at the IRFC, and with a quantum efficiency of 100%, in one second of observing, a 1.2 Jy source would be expected to be seen with a signal-to-noise ratio of 1.

3.4) Measurements Made on the Photometer

The overriding goal of all work done on the photometer was to

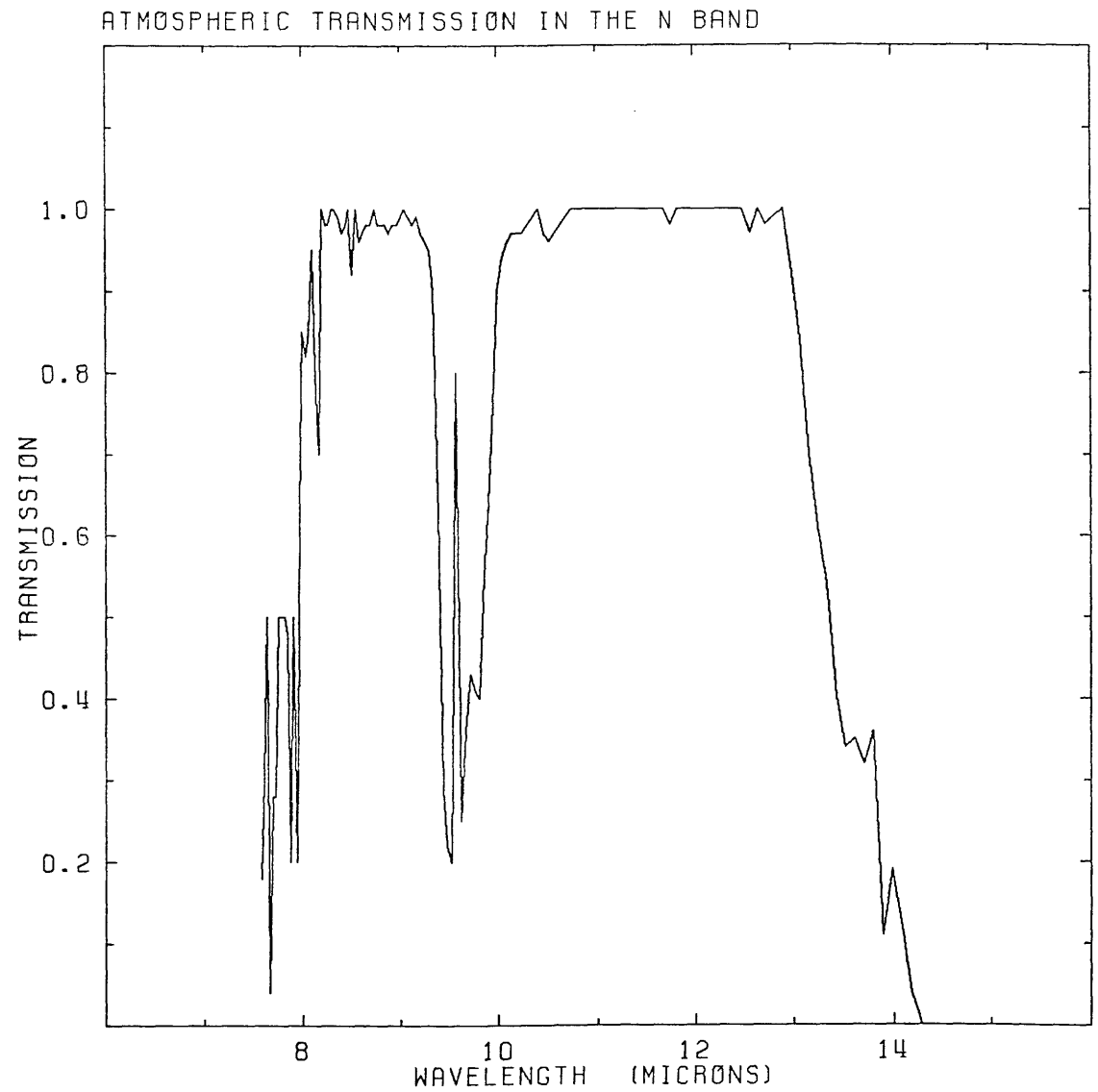


Figure 16. Transmission of the atmosphere in the 8 – 14 μm band (data taken from Kyle and Goldman (1975)). The ozone absorption feature at 9.6 μm is clearly visible, and it is this that supplies most of the sky emission.

make it as sensitive as possible. For a given waveband and aperture size (set by astronomical requirements), this means ensuring that the photometer is background limited. This objective is approached in two ways: Firstly, by minimizing the noise of all the other components in the system. Secondly, maximizing the radiation falling on the detector. As can be seen from equation (56), the limiting sensitivity is inversely proportional to some function of the transmission of the photometer-telescope system (as this transmission will also enter into the Φ_N term). Once the photometer optics have been decided upon, this really means ensuring that the alignment of the optics is correct, and that all radiation entering the photometer is falling on the detector.

Over the period 1980/1981, 24 sets of measurements were made on the photometer system with the above objectives in mind. Although the only real method of adjusting the photometer performance was by means of optical alignment, these runs also had the objective of seeing how the detector behaved under a variety of conditions.

3.4.1) D.C. Resistance

At room temperature, the D.C. resistance of the detector is negligible, and so this measurement also provided a quick check that the detector was cold. Two standard sources were used throughout the set of measurements; a gold mirror, and a sheet of blackened aluminium. These were placed against the vacuum window, thus eliminating any extraneous background from the laboratory. The D.C. resistance was measured using the circuit shown in Figure 17. Voltages were measured across the load resistor, in order to avoid the detector resistance being shunted by the input impedance of the voltmeter.

A typical $V_D - R_D$ curve is shown in Figure 18. The detector

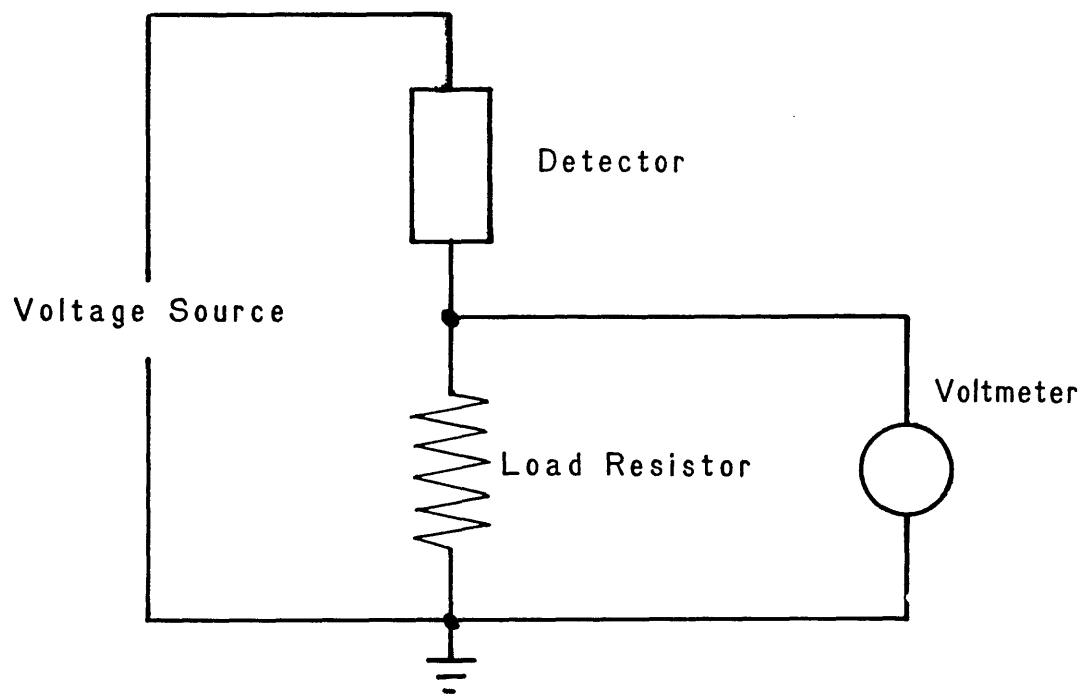


Figure 17. Circuit for measuring the D.C. resistance of the detector. Voltages are measured across the load resistor to avoid the shunting effect of the voltmeter.

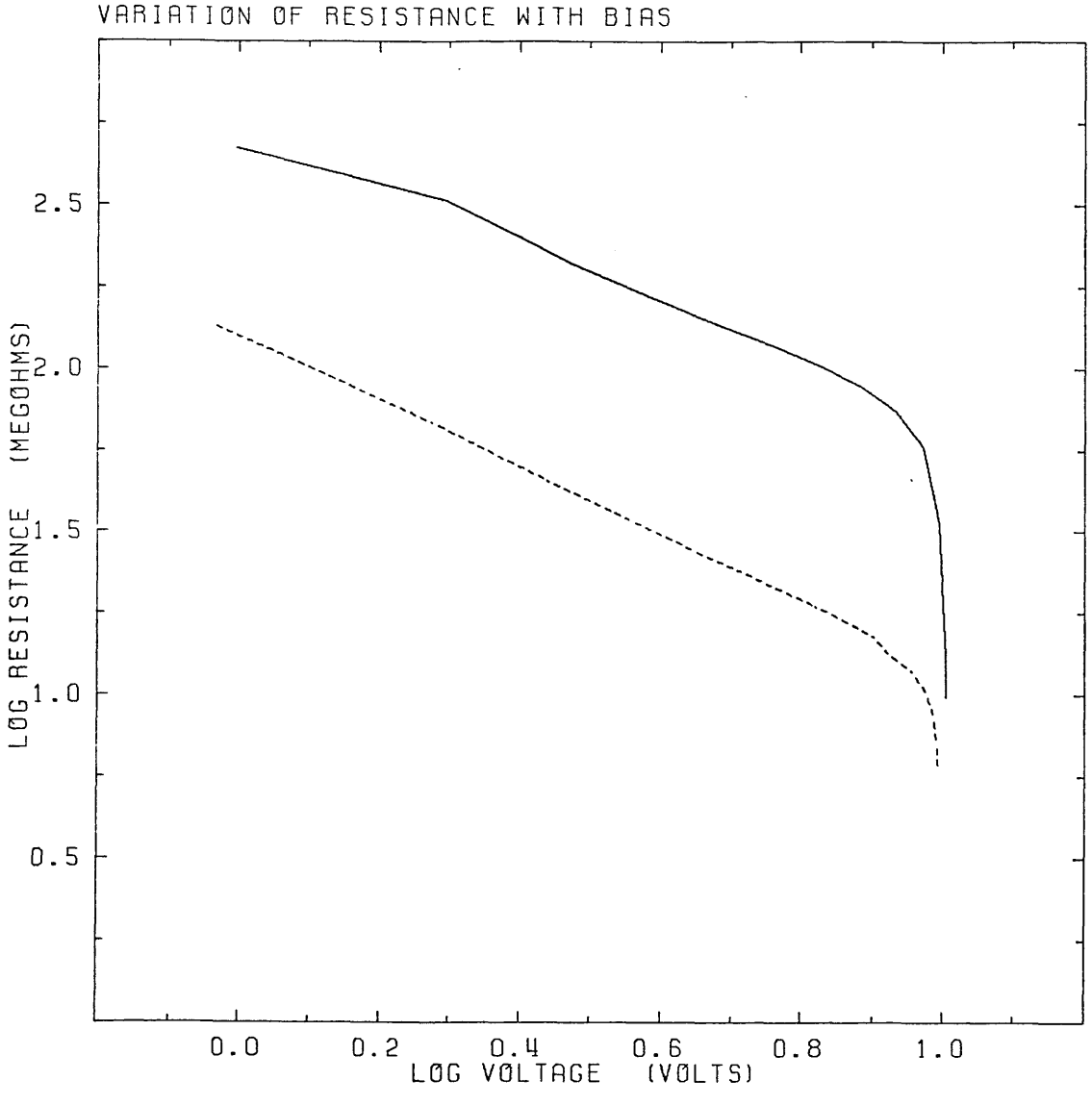


Figure 18. Variation of resistance with voltage across the detector. As the voltage increases, the resistance decreases, varying as V^{-1}. When the voltage rises above the breakdown voltage, the onset of impact ionization causes the resistance to drop sharply. The solid line is the resistance with the dewar window covered with a gold mirror, and the dashed line the resistance with the mirror replaced by a sheet of blackened aluminium.

resistance steadily decreases with increasing bias voltage, up to a limit of ~9 V. This is the behaviour described is section 2.3.3, and is due to the dependence of carrier lifetime and mobility on electric field. Beyond 9 V, impact ionization is starting to make itself felt, and so the resistance drops sharply. However, it was noted that the higher the photon flux on the detector, the less pronounced was the "knee" in the curve.

31 sets of resistance measurements were taken over the period 1980/81, using both versions of the photometer. From section 2.3.3, the resistance is expected to depend on the voltage across the detector by the relation

$$R_D \propto V_D^{-n} \ . \qquad (68)$$

Averaging over all of the runs between the detector voltages of 2 V to 9 V, n was found to be 1.03 ± 0.01, corresponding to a field dependence of the the carrier lifetime of the form $\tau \propto \phi^{1.5}$. The 2 V and 9 V limits were imposed on this calculation since, below 2 V, the resistance is getting so large that it cannot be accurately measured by the circuit of Figure 17, whilst beyond 9 V, the resistance is starting to be affected by impact ionization.

3.4.2) Variation of Quantum Efficiency with Background

In section 2.3.5, it was suggested that at very high radiation levels, the quantum efficiency may be a function of the amount of radiation falling on the detector, decreasing as the radiation level rises. In order to check this hypothesis, the variation of D.C. detector resistance with background was measured.

The variation between detector resistance and incident radiation has been given by equation (18), viz.

$$R_D \propto \frac{1}{\eta\phi} \quad . \tag{69}$$

If the dependence of quantum efficiency on radiation is assumed to be of the form

$$\eta \propto \phi^{-n} \quad , \tag{70}$$

then combining this equation with equation (69) gives

$$R_D \propto \phi^{n-1} \quad . \tag{71}$$

To measure this dependence, a Perkin-Elmer black-body was positioned in such a way that it completely filled the beam of the photometer, and the intensity of radiation incident on the detector varied by adjusting the temperature of the black-body. To avoid any ambiguities caused by the wavelength-variation of detector response, measurements were restricted to the temperature range 500 K to 670 K. Although there is a significant variation in flux from the black body over this temperature range, the spectral shape of the radiation is (nearly) constant at all temperatures. Measurements were made at 20^{o} intervals, the black-body being given time to settle before each reading. In addition, the voltage across the detector was adjusted to a constant value of 8.5 V before the measurement: a constant voltage avoids the contamination of the results by the field dependence of the mobility and the carrier lifetime terms.

The results are plotted in Figure 19. Since geometrical effects will be the same for all temperatures, the photon flux density has been plotted, rather than the actual number of photons reaching the detector. These two quantities will only differ by a constant factor. A linear regression analysis (e.g. Bevington 1969) gives a value for n of 0.06 ± 0.03. This small value, whilst not completely ruling out the possibility of a radiation dependence of η, does seem

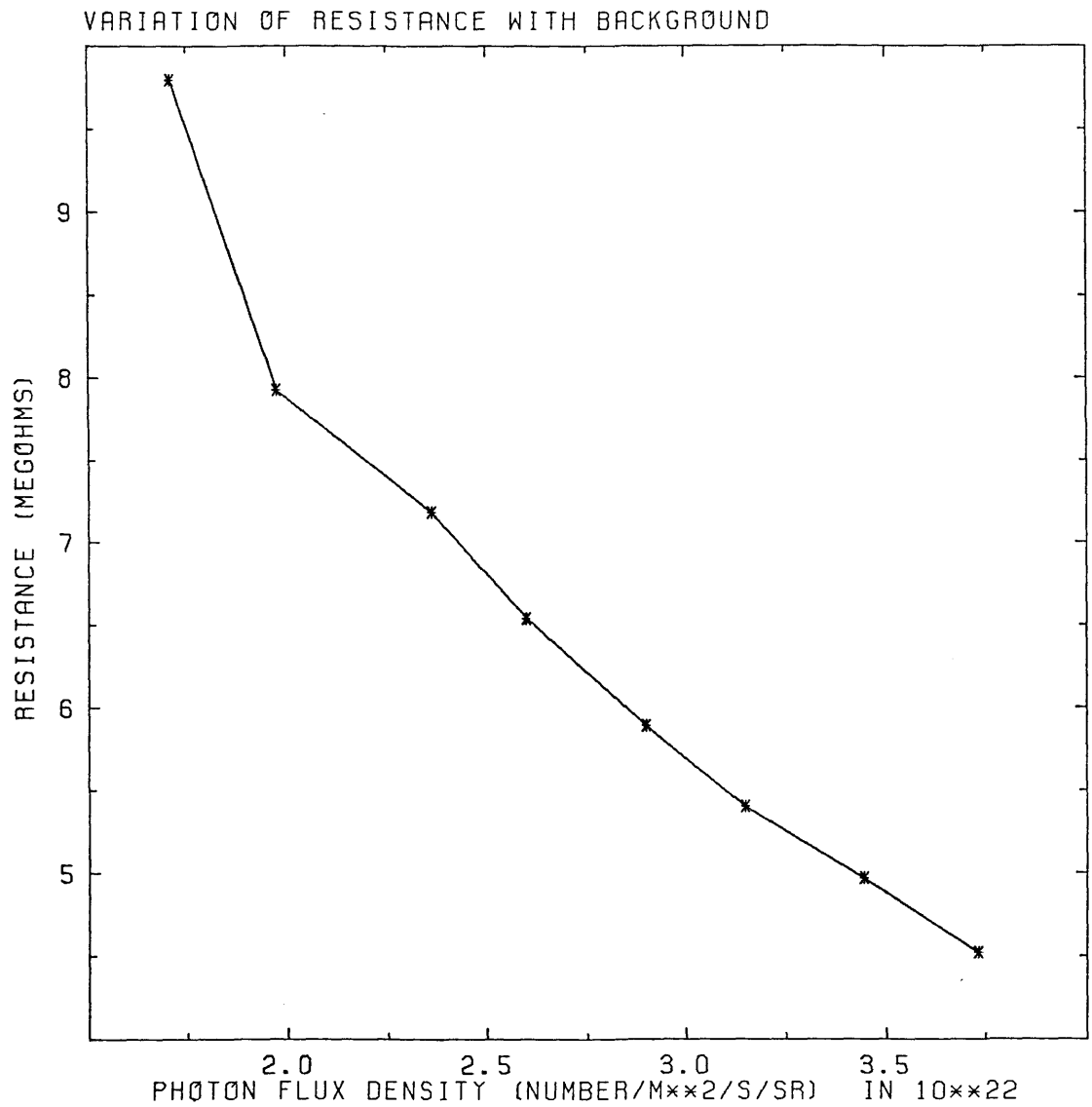


Figure 19. Variation of detector resistance with background. The abcissa is photon flux density from the blackbody (in photons m^{-2} s^{-1} sr^{-1}) rather than actual numbers of photons: the two quantities differ by a constant factor.

to argue against it.

3.4.3) Spectral Response

The spectral response of the photometer was measured in order to check whether the filters were producing the desired effect. This response should be the product of the filter transmissions, the detector response, and the transmission of the atmosphere. In laboratory measurements, it is not possible to get the effect of one air-mass of atmosphere, and so this latter component will be largely absent.

The signal was the 500 K black-body set a distance of ~1.2 m from the photometer, and chopped with a rotating blade chopper. An Oriel monochromator was used for wavelength selection, being set to a grating resolution of 0.24 μm. The output of the preamplifier was fed into an Ithaco Dynatrac phase sensitive detector (PSD), the reference frequency being taken from the chopper. The output from the PSD was fed to a chart recorder, from which measurements were taken. The bias of the detector was set to provide a maximum signal to noise ratio (see section 3.4.7).

The signal measured over the range 6.0 μm to 15.0 μm in 0.24 μm steps, although the step size was decreased for regions of steeply varying response. This spectral response, presented in Figure 20 has been corrected for the black-body spectrum in the measured range, by dividing the signal by the number of photons in the measurement band (the photoconductor is a photon - as opposed to power - device), and normalized to give a maximum response of unity. The shape of the monochromator pass band remains convolved in, but as we are interested only in overall shape rather than fine detail, this effect should not be too serious.

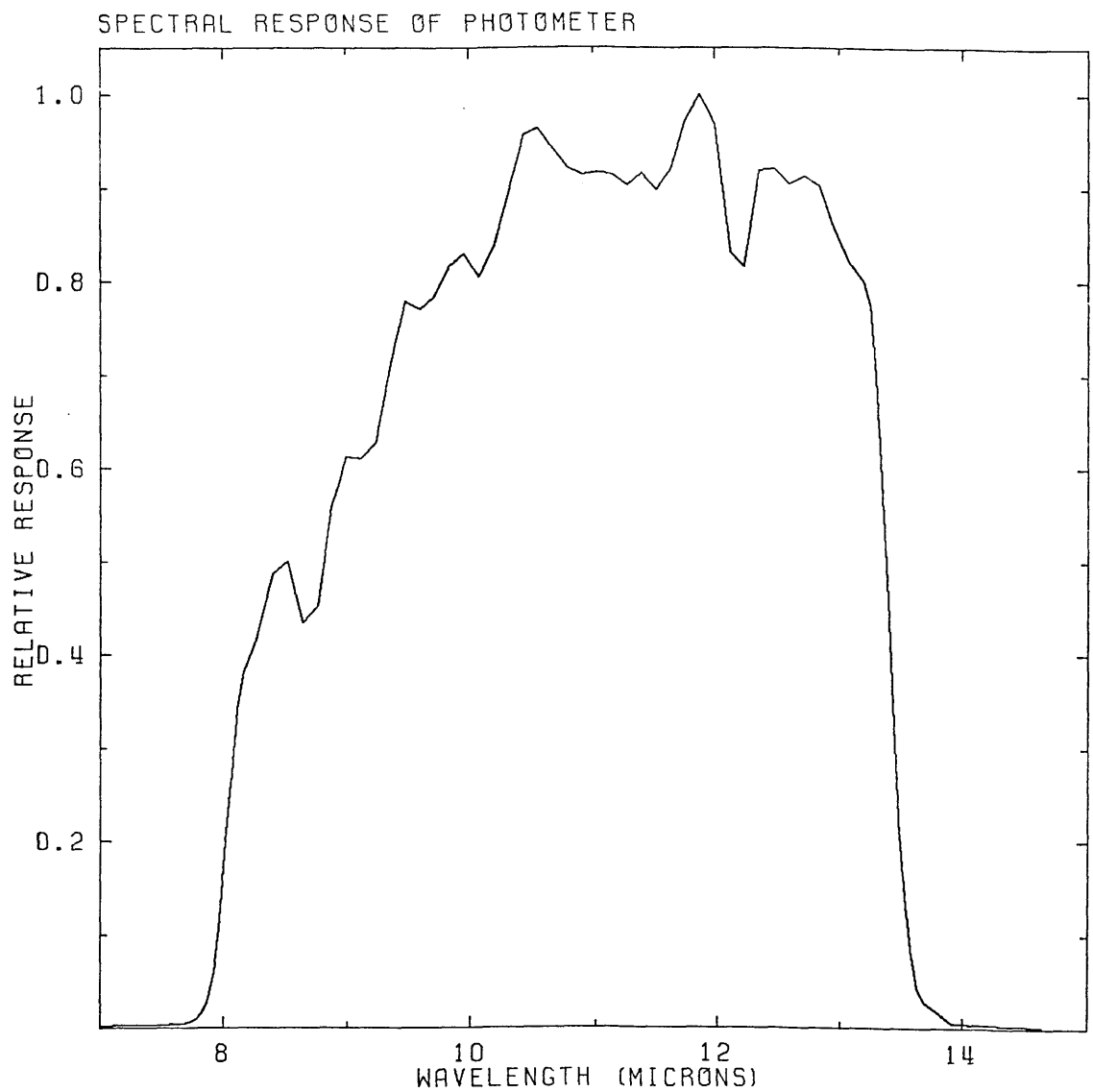


Figure 20. Spectral response of the photometer for a constant photon flux.

Comparison of this response curve with the transmission curve of the OCLI filter (Figure 14) shows the fall of detector response to short wavelengths. The large irregularities are due to the OCLI filter, and the shallow cut-off longwards of 12 μm is the barium fluoride filter making itself felt. There did not appear to be any transmission outside the limits shown in Figure 20. Overall, the response curve is what one would have predicted; a product of the transmissions of the filters, together with the (expected) response of the photoconductor. The effective wavelength for the system λ_{eff}, calculated from the relationships

$$\lambda_{eff} = \frac{c}{\nu_{eff}} \tag{72}$$

and

$$\nu_{eff} = \frac{\int \nu R(\nu) d\nu}{\int R(\nu) d\nu}, \tag{73}$$

(where $R(\nu)$ is relative response) works out to be 10.7 μm. The effective frequency is worked out first, as we require effective wavelength to estimate power, and photon energy is directly proportional to frequency.

3.4.4) Spatial Response

Measurements were made to determine the spatial response of the photometer (the beam pattern). For this, the bottom entry configuration was used, the equipment being set up as shown in Figure 21. The optics were designed to image a cone of rays over the whole of the detector, but if an object is placed in the field of view, they will produce a demagnified image of that object at a point between the Fabry lens and the detector. In this measurement, the object, (a

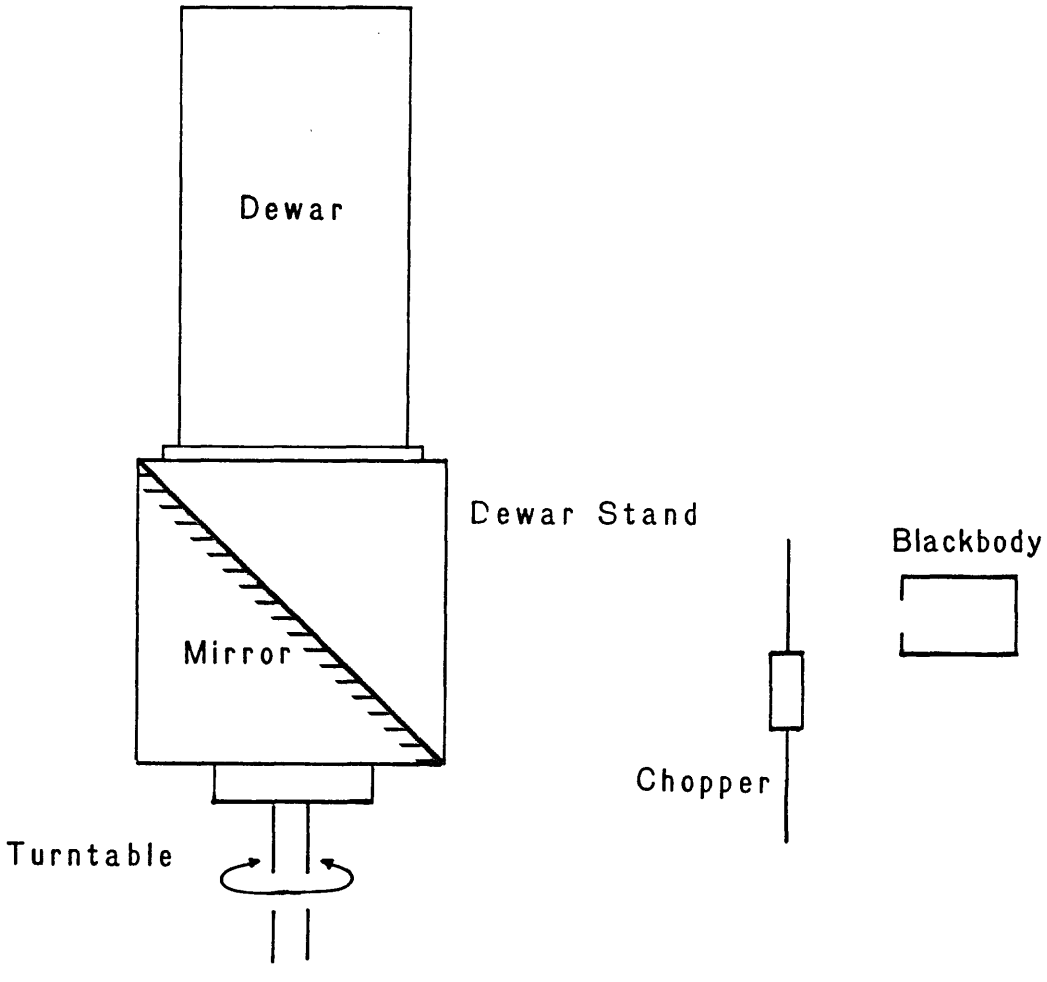


Figure 21. Configuration used for measuring the spatial response of the detector. As the photometer is turned about an axis parallel to the plane of the page, the image of the aperture of the blackbody moves across the detector in a straight line.

chopped black-body, stopped down to produce a strong signal which did not, however, overload the electronics) was placed at a distance of -1.2 m from the detector, or about 100 x the focal length of the lens. This produced an image less than 0.2 mm from the detector, which was accurate enough for the purposes of this measurement.

As the dewar is rotated, the beam from the source moves across the vacuum window, and the demagnified image moves across the detector. The signal measured is thus a convolution of the (slightly blurred) image of the source, and the spatial response of the detector. As the image of the source moves in a straight line relative to the detector, it is a fairly straight-forward matter to construct a sensitivity contour map for the photometer.

The measurement procedure was to set the dewar on the stand, and to finely adjust the position so that the output from the detector was a maximum. The dewar and mirror were rotated through half degree intervals at a time, and the signal measured in each position. The dewar was then rotated 90° relative to the mirror, and the procedure repeated. Since all the scans were exactly repeatable, the intensity of the signal was used to position the scans with respect to one another, and so to construct the contour map. Distinct asymmetries were present in the first measurements, and so the dewar was stripped down, the optical components realigned, and the scans repeated. The measurements were repeated two months later in a third run, this consisting of four scans at 45° relative to one another. The contour map drawn from this run is reproduced in Figure 22.

The important features of this map are the elongated ridge of high response, and the smaller maxima next to it. These were present in all maps, with the same orientation relative to the detector. What caused the irregular response is unknown. Misalignment of the optical

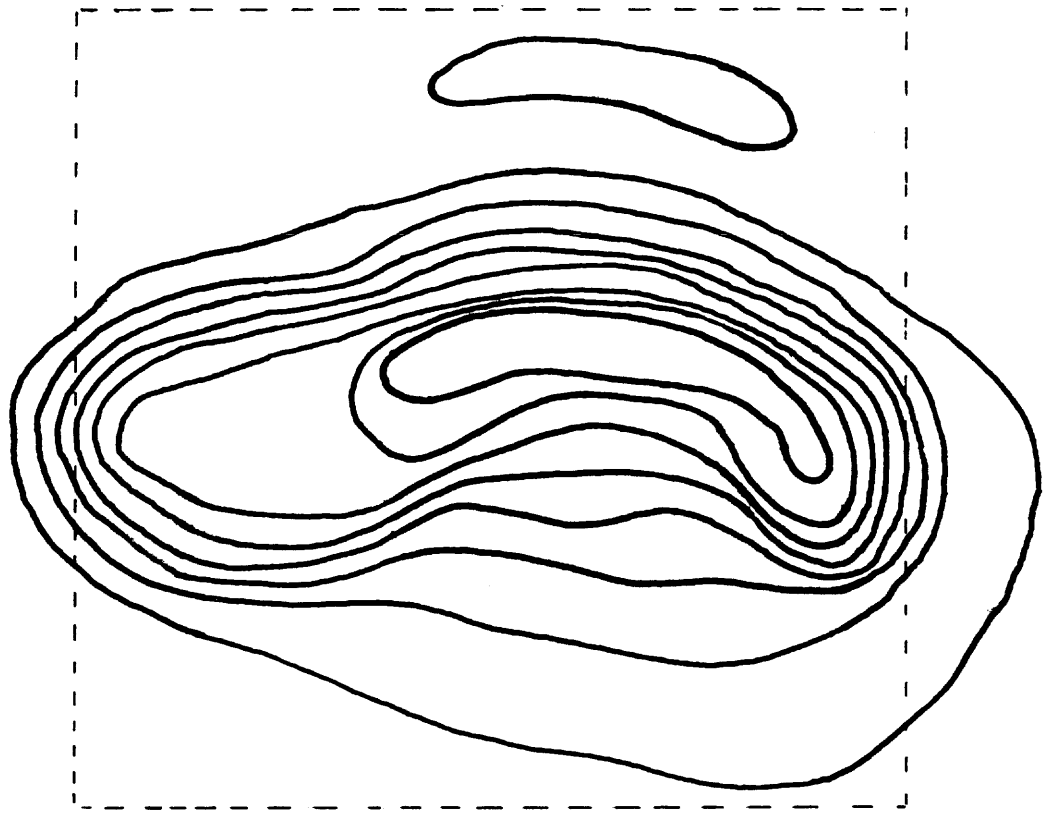


Figure 22. Spatial response of the detector. Contours are at 0.5V intervals, (measured on the output of the PSD), with the outside contour being 0.5V. The dashed line represents the detector.

components was at first suspected, but rejected for two reasons: firstly, these features were present in all runs, with the same orientation with respect to the detector. In between runs, the photometer had been realigned, and it seems unlikely that the same irregularities would be present, with the same orientation, in all cases. Secondly, misalignment cannot explain the secondary maximum. If what was being seen was the diffraction pattern from the circular aperture, and the primary maximum corresponds to the primary maximum expected in the diffraction pattern, then the secondary maximum should have an intensity only 1.7% that of the primary (Longhurst 1967). Although the secondary maximum is in roughly the correct position (ignoring the effect of the Fabry lens) for it to be caused by diffraction, the secondary maximum is about ten times stronger than the 1.7% expected.

It seems likely then, that the irregular response was a feature of the detector rather than of the optics. Dirt on the detector surface could have caused this, but a visual examination revealed nothing. An alternative, variation in the dopant concentration across the detector is possible, but there was no way of checking this. It seems unlikely though, that such severe gradients would exist over so small a distance.

A numerical integration over the contour map indicated that a factor of ~2 in responsivity was being lost. This can be regarded as an effective reduction in the quantum efficiency of 50%, and thus (by equation (56)) corresponds to a $\sqrt{2}$ loss in possible sensitivity. Some attempts were made to overcome this problem. A new Fabry lens was made in an attempt to image the primary mirror onto only the sensitive part of the detector, and a stop was made to mask the remaining part of the detector surface. Unfortunately, none of these attempts was

successful, the sensitivity undergoing degradation in every case.

3.4.5) Noise

These were made at the projected chopping frequency of observation, 20 Hz, using the circuit shown in Figure 23. The output from the preamplifier is fed into the PSD, the reference frequency being supplied by the PSD itself, and the output of this displayed on a chart recorder.

The noise obeys Gaussian statistics, and so we can use the fact that in a Gaussian distribution, 99% of the points will lie within 2.5σ of the mean, where σ is the standard deviation of the distribution: if 100 samples are taken, the peak-to-peak value should be around 5σ. Applying this to the measurement of noise voltages means that in a time 100 × the sample independence time, the peak-to-peak voltage will be 5σ, σ being the r.m.s. noise voltage. The measurement procedure thus reduces to letting the chart recorder run for 100 × the PSD integration time, and measuring the peak-to-peak noise voltage on the chart recorder trace.

Unfortunately, there is doubt as to the validity of some of the noise runs. It was discovered fairly late on in the series of measurements that a power supply used in some measurements may have been generating excessive noise. This was replaced by a battery for the last few measurements, (one of which is presented in Figure 24) and for use on the telescope. However, although the detailed measurements may have been contaminated, some generalizations can be drawn.

The fundamental question about the noise is "what is its origin?". The fact that the noise is different for different backgrounds seems to argue for photon shot noise being the major

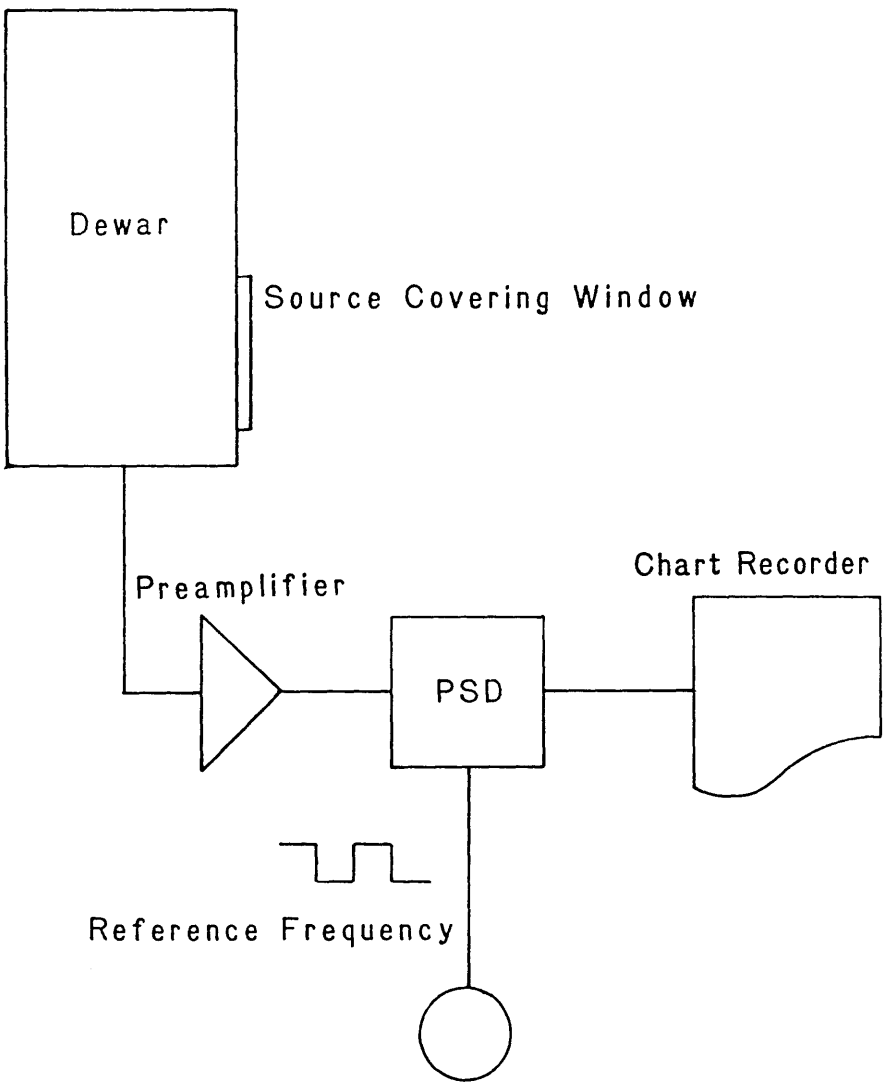


Figure 23. Configuration used for measuring the noise on the detector. The output from the preamplifier is fed into a PSD, together with a 20 Hz reference frequency. The output from the PSD is monitored on a chart recorder. In 100 × the integration time, the peak-to-peak output voltage is 5σ, where σ is the standard deviation of the noise.

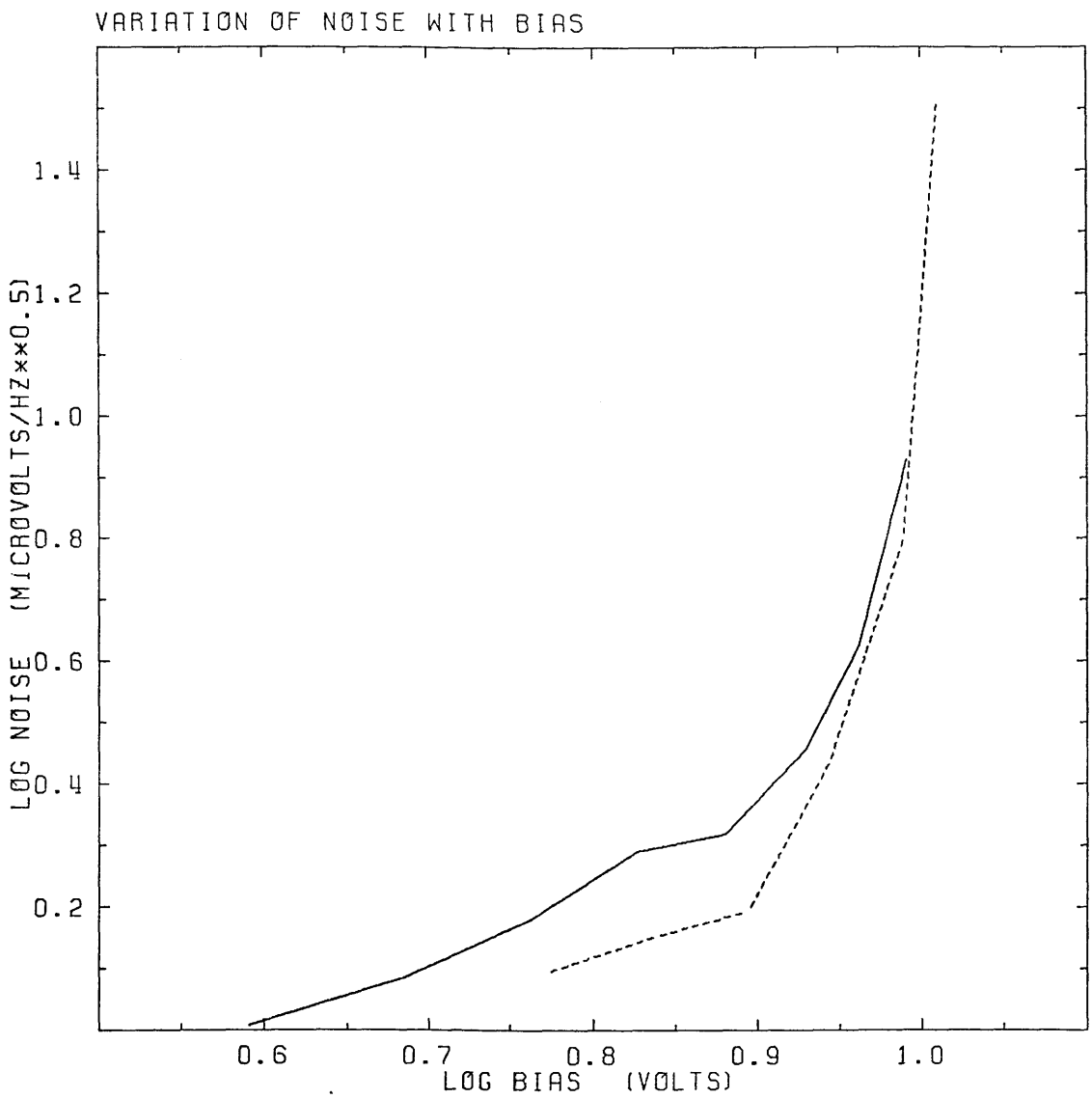


Figure 24. Variation of noise with bias. The noise units are $\mu V\ Hz^{-1/2}$, measured across a 3.8 MΩ load resistor. The solid line is the noise measured with the high-emissivity blackened aluminium sheet, and the dashed line is that measured with the low-emissivity gold mirror. The noise rises with increasing voltage until the breakdown voltage, beyond which it increases sharply.

contributor. Noise from the preamplifier can be eliminated as the cause, since this component had a noise voltages at the measurement frequency of ~10 nV $Hz^{-1/2}$, compared with the 1 - 3 μV $Hz^{-1/2}$ shown in Figure 24. It is not detector Johnson noise, as in the equivalent noise circuit, the detector is in parallel with the load resistor, so the measured Johnson noise voltage must be less than that from the load resistor alone. For a load resistor of 3.8 MΩ, at a temperature of 4.2 K, this works out to be ~30 nV $Hz^{-1/2}$. 1/f noise can be eliminated, as a quick check showed that the noise was independent of frequency up to 100 Hz. This leaves only g-r noise and photon shot noise, of which, according to Bratt (1977), the g-r noise should be negligible.

3.4.6) Responsivity

The responsivity is the voltage (or current) output per unit power input. Since the photometer was going to be used on the telescope with a modulated signal, all measurements in the laboratory were done with such a source (described above). The experimental arrangement was similar to that for the measurement of spatial response except, of course, there was no turntable. In this set of measurements, the variation of responsivity with bias voltage was measured.

Figure 25 shows the variation of responsivity with detector voltage for a selection of runs. This diagram includes runs where the photometer was out of alignment, these being the runs of low responsivity. The general trend to note is that the responsivity rises with increasing bias until the breakdown voltage is approached, and then falls. An expression for the responsivity has been given in equation (47) and, as can be seen, the responsivity depends not only

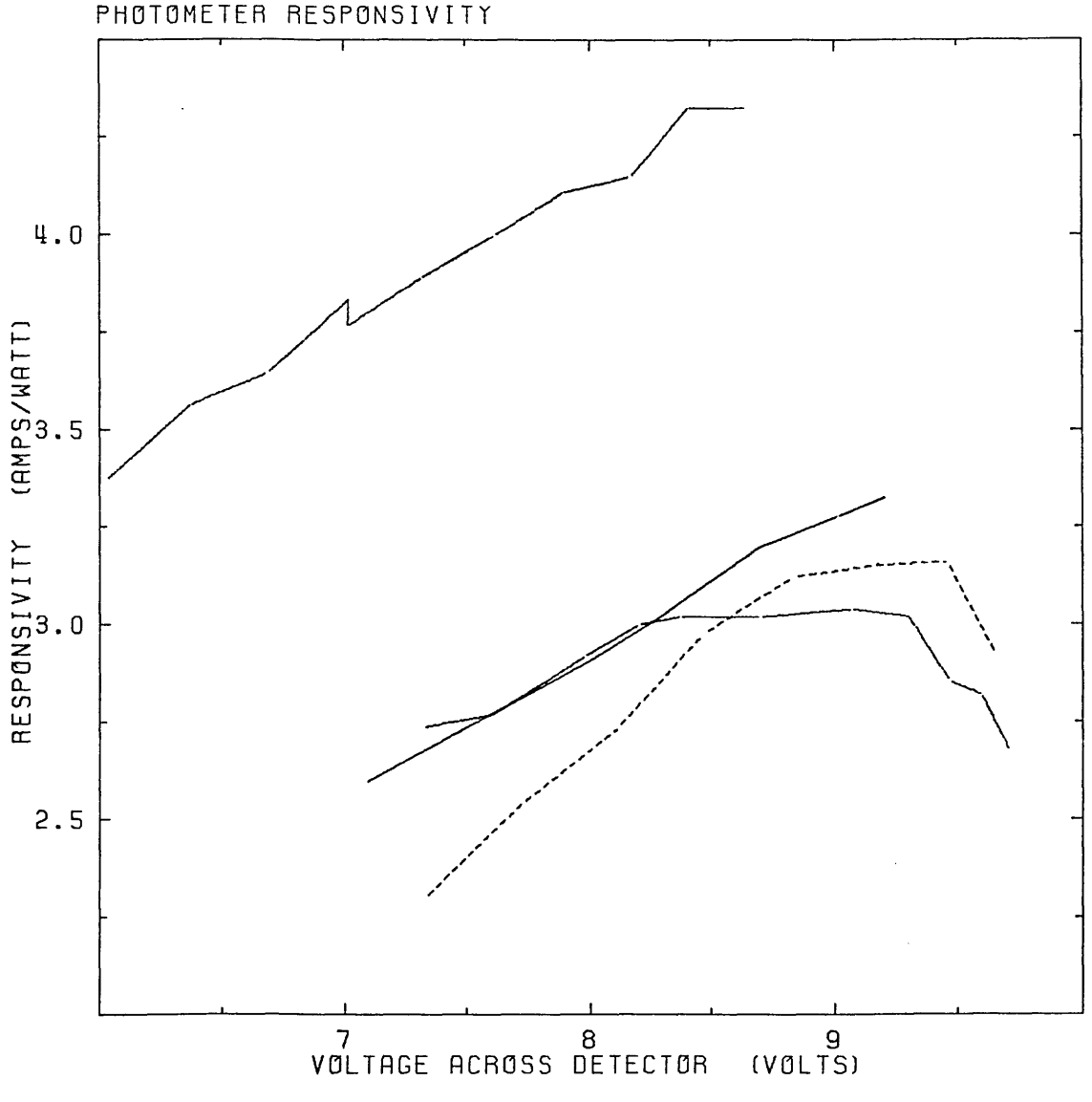


Figure 25. Variation of responsivity with bias. The responsivity rises to a broad maximum near the breakdown voltage, and then falls.

on the D.C. and A.C. resistance of the detector, but also on the magnitude of these quantities compared with the load resistor: there is quite different behaviour for resistances comparable with the load resistance, as to resistances much greater (or less) than the load. These particular runs were done with the A.C. and D.C. resistances of the same magnitude as the load resistor.

3.4.7) Photometer Sensitivity

The aim here was to find the optimum operating point of the photometer. As described above, both the responsivity and the noise, vary with the voltage across the detector but, as can be seen from the above figures, not in the same way. Thus the signal to noise ratio will vary with the detector voltage, and the optimum bias will be at the peak of the curve.

The signal to noise ratio measurements were a combination of the responsivity measurements and noise measurements. A "black body" background was used for the noise, as this was the most repeatable. A 1 mm diameter field stop was used for most of the runs. Although this leads to higher background fluxes than expected from the telescope, smaller stops will lead to significant diffraction, and uncertainties in the estimation of both the signal and the noise.

Figure 26 gives a selection of results from the same runs as Figure 25. Although not clear from this diagram, the NEP increased to lower biases. The sharp increase in the NEP towards the breakdown voltage is readily apparent. Over twenty such runs, under varying conditions, the mean operating voltage was 8.1 ± 0.9 V. This large spread in results (primarily due to uncertainties in the measurement of the noise component) indicated that it would be wise to determine to optimum bias for observations when at the telescope.

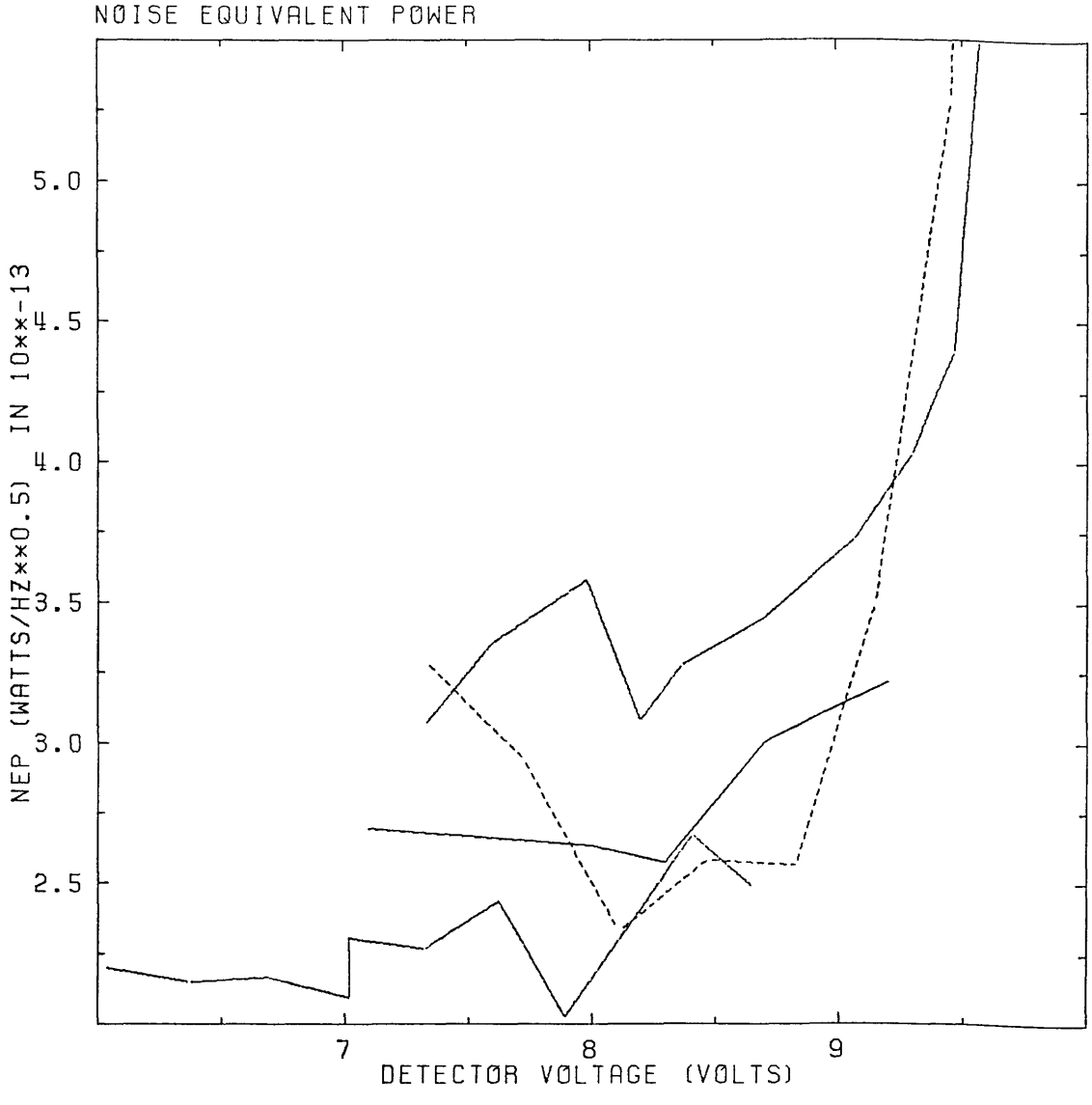


Figure 26. Variation of NEP with bias. As the breakdown voltage is approached, the NEP rises.

3.5) Discussion

The results presented above were taken from a number of the 24 experimental runs made with the photometer. These included a number of runs where the photometer was obviously out of alignment, but as the only effect of this was to reduce the background, these are equivalent to runs done with a small field stop. The best results in the entire set of laboratory runs produced the results in Table 1.

Comparing the sensitivity estimate in Table 1 with that obtained in section 3.3, we see that the photometer is a factor of 5 below the limit estimated for the IRFC. There are, however, several points to note. Firstly, the laboratory background (assumed a black body) is a factor of ~3 higher than that expected on the IRFC (this small factor is due to the large percentage of background radiation expected from the vacuum window). Thus the sensitivity of the photometer on the IRFC ought to be a factor of $\sqrt{3}$ higher than measured in the laboratory. Secondly, the spatial response of the detector in effect reduces the quantum efficiency by ~50%. As mentioned above, the reason for this response variation is unknown.

Finally, the limit obtained in section 3.3 was for a quantum efficiency of unity, which is virtually never the case. It is true (e.g. Low and Rieke 1974) that devices such as bolometers can have a quantum efficiency of nearly unity in the 10 μm band,and don't suffer the $\sqrt{2}$ increase in noise due to carrier recombination. These devices theoretically should be better than photoconductors. In practice though, they are more awkward to use, requiring operation at a temperature of around 1 K, and can be excessively prone to microphonics. This latter factor can greatly reduce their sensitivity.

Assuming that the laboratory is a black body, then the background under which the above tests were carried out was 3.4×10^{12}

Wavelength Passband ..... 8.4 μm to 13.2 μm

Effective Wavelength ..... 10.7 μm

The following figures apply to a background of ~3.4 × 10^{12} photons s^{-1} (~6.3 × 10^{-8} W) on the detector.

Maximum Responsivity ..... 1.6 × 10^{7} V W^{-1} (= 4.3 A W^{-1})

Minimum NEP ..... 2.0 × 10^{-13} W $Hz^{1/2}$

Expected Sensitivity on the IRFC ..... 6 Jy 1σ1s

Expected Limiting Sensitivity on the IRFC (assuming a factor of 3 reduction in background over the laboratory, and a $\sqrt{2}$ degradation in performance due to spatial response ..... 5 Jy 1σ1s

Table 1. Photometer parameters - best performance over the entire series of tests.

photons s^{-1} on the detector, or about 6.3×10^{-8} W. The theoretical NEP for this background is (from equation (66)) 6.8×10^{-14} W $Hz^{-1/2}$. With these figures, the quantum efficiency can be estimated from

$$\eta = \left[\frac{NEP_{calculated(\eta=1)}}{NEP_{measured}}\right]^2 \qquad (74)$$

(Young and Low 1979). Inserting the figures gives a quantum efficiency of ~12%, just lower than the 16 - 20% given by Young and Low for their Si:As detectors.

The performance indicated is however, disappointing when compared with other 10 μm photometers. Wolf and Lemke (1983) have reported a background NEP of 4×10^{-14} W $Hz^{-1/2}$ on a 10 μm broad-band Si:Ga photometer, with a 6" field of view; and the bolometer-based photometers at UKIRT are reported as having a 1σ1s sensitivity of ~0.3 Jy at 10 μm with a 5" aperture (UKIRT Observers Manual 1981, UKIRT Newsletter 1983). Allowing for the difference in optical configuration between UKIRT and the IRFC, this sensitivity is consistent with the background limited 1.2 Jy expected on the IRFC for this photometer, if the quantum efficiency were 100%.

Overall then, it seems that the photometer described in this chapter is background limited, but has a quantum efficiency of only 12%. Allowing for difference in background between the laboratory and the telescope, and for the effect of the spatial response pattern, a sensitivity on the IRFC of 5 Jy (1σ1s) is expected.

CHAPTER 4

Photometry on the Infrared Flux Collector

4.1) Introduction

In infrared astronomy at wavelengths longer than about 2 μm, a typical astronomical signal is orders of magnitude smaller than the background power radiated by the telescope. Special techniques are thus needed to measure this signal.

This chapter begins with a discussion of the signal extraction techniques, showing how they require specialized instrumentation and observing procedures. An overview of the data acquisition system developed for use at the Tenerife Infrared Flux Collector (IRFC) follows, with a detailed description of the software used to control and analyse the observations.

4.2) Chopping and Nodding

Photometry at 10 μm has been described as "...observing visually through a telescope lined with luminescent panels, and surrounded by flickering lights as though the telescope dome were on fire" (Low and Rieke 1974). In other words, the background emission is vastly greater than the signal from the object being observed. The technique of sky subtraction used in optical astronomy, where exposures are made first of the object, then of the sky, and the two subtracted does not work for two reasons: 1) The low frequency noise of the background will swamp the signal, and 2) the background is so large compared to the signal that possible shifts in the background during the observation will be much greater than the signal.

The method used in infrared astronomy is that of the matched filter. The input radiation from the object is modulated in some fashion, and measurement is made only of that component of the detector signal at the modulation frequency. If the modulation frequency is high enough, contribution from 1/f noise is minimized, and capacitative decoupling will remove the D.C. component. This technique is known as chopping, and although it has been briefly touched upon in Chapter 3, it will be discussed further in section 4.2.1.

Removal of background variations (both spatial and temporal) requires a similar approach, but at a much lower frequency. This technique is often called "nodding", and is described in section 4.2.2.

4.2.1) Chopping

Using a rotating blade chopper, as was used in the laboratory experiments on the photometer, is unsuitable for astronomical observations as both the background and the signal are modulated. The modulation in the background will thus be interpreted as signal. With the strong signal used in the laboratory, this didn't matter, but on the telescope the signal will be much less than the background. The best approximation to modulating the astronomical signal is to switch the telescope beam rapidly between the source and an adjacent patch of sky (i.e. to chop between them). When the signal measured in one position is subtracted from that measured in the other, the background cancels, and only the required signal passes through the decoupling circuitry. The higher the chopping frequency, the lower the temporal drift between the two beams, and the smaller the error.

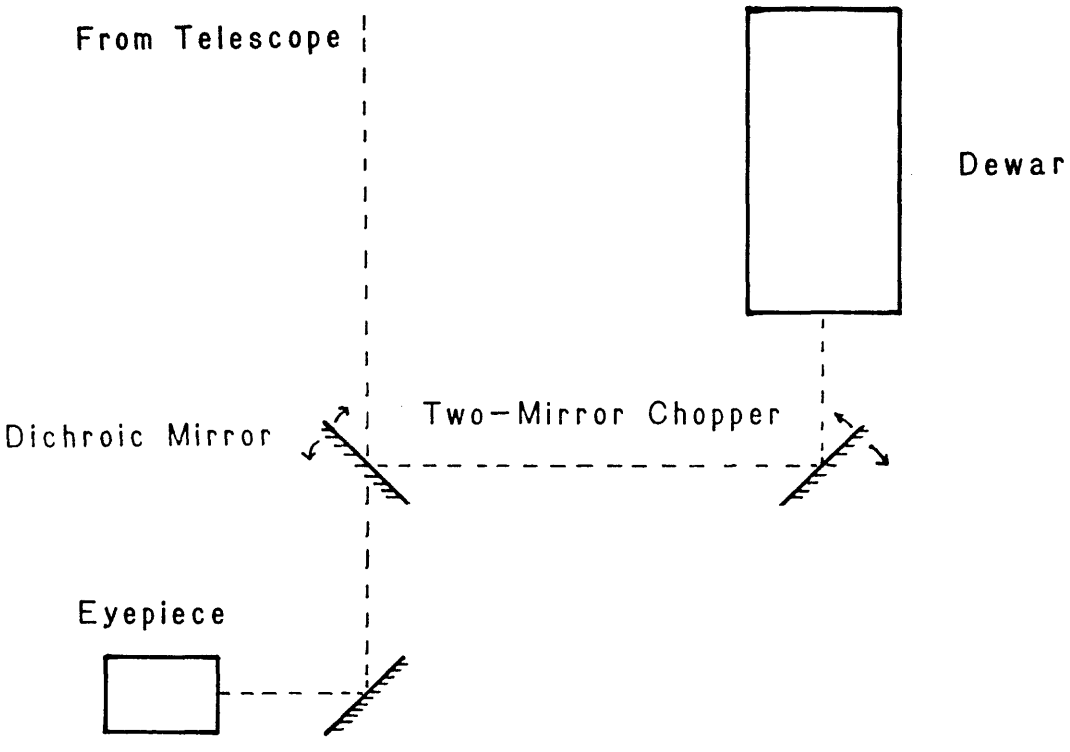
Inevitably though, the cancellation will not be exact.

Between the two beams, there will generally be some constant difference (e.g. the beams reflect off different parts of the telescope mirrors), and this will be indistinguishable from the signal. Removal of this offset is discussed in the next section.

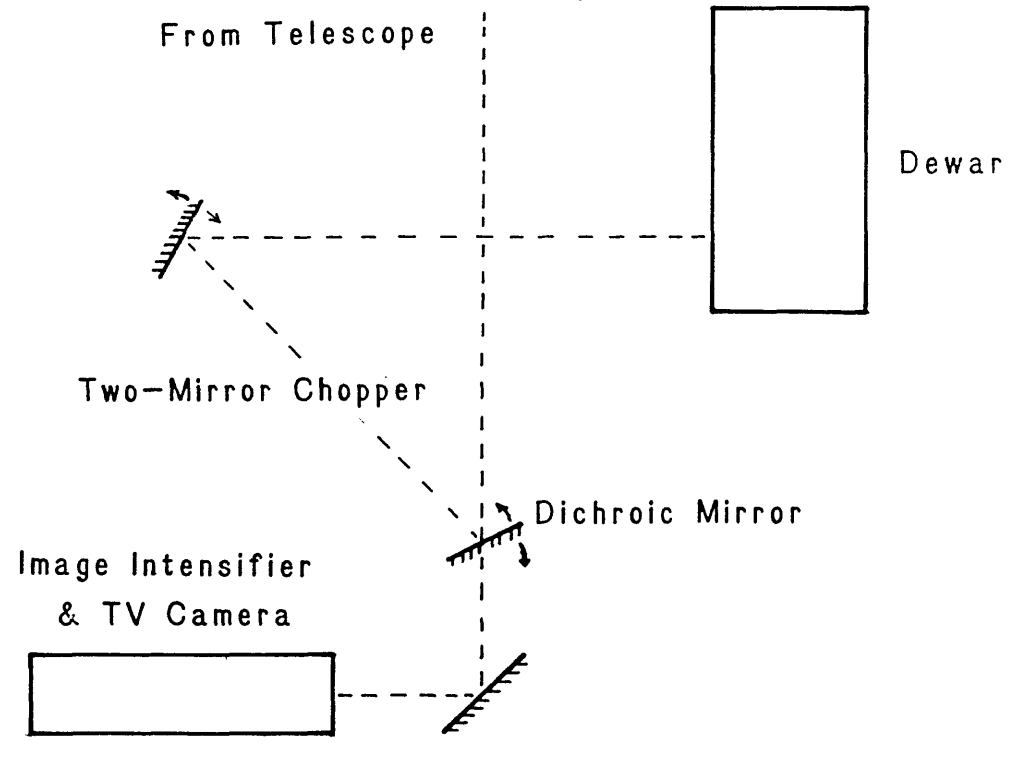
Chopping between the two adjacent patches of sky may either be done internally, within the photometer (e.g. the tuning fork chopper described by Rosen (1983)), or externally, by means of moving mirrors. The internal chopper has the advantage that the number of warm optical elements in the system is kept to a minimum, but also has the disadvantage in that large throws are not easy to obtain. An external arrangement reduces to three choices: chopping with the primary mirror, chopping with the secondary mirror, or using some form of tertiary mirror chopper. Bearing in mind that the chop frequency has to be around 10 Hz to avoid 1/f noise, one sees immediately that primary mirror chopping with a large telescope is not feasible. Secondary mirror chopping is by far the best, in that the number of mirrors required is kept to a minimum, and that it allows very large throws. The telescope however, must have been designed with secondary mirror chopping in mind.

Finally, there is tertiary mirror chopping. The arrangement used on the IRFC is has been described by Jorden et al. (1976), and is shown in Figure 27. As the dual mirrors vibrate, the image of the sky is moved back and forth across the focal plane optics of the photometer without defocussing. The detector is thus chopping between two patches of sky.

Whatever method of chopping is used, the output of the detector is fed (via a preamplifier) into a phase sensitive detector (PSD). This device performs the matched filtering by comparing the input signal with a reference waveform from the chopper drive.



a) Bottom-Entry System



b) Side-Entry System

Figure 27. Chopper configuration for (a) the bottom entry dewar, and (b) the side entry dewar. The chopper has been described by Jorden et al. (1976).

Included in this operation is a certain amount of smoothing, and the output is thus a measure of the difference between the two beams averaged over a short period of time. The actual value of "a measure" and "a short period" is dependent upon the degree of smoothing set.

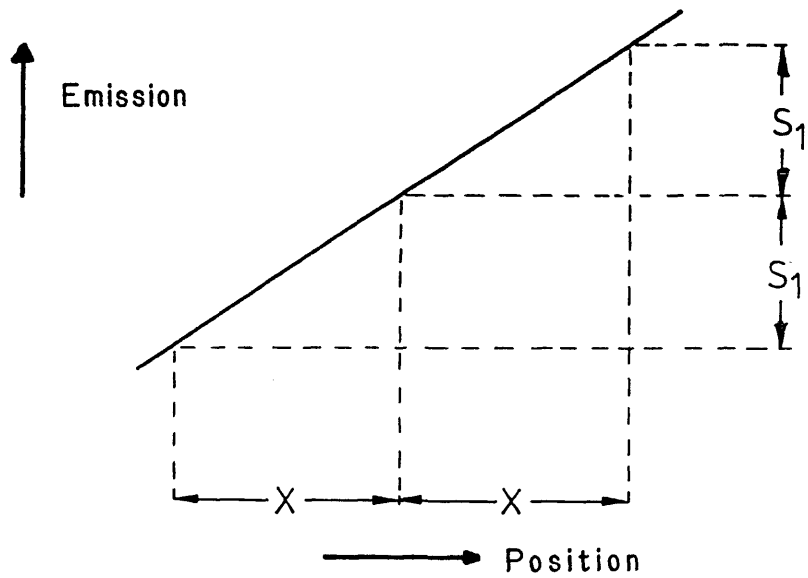
4.2.2) Nodding

If the background between the two beams is not identical, there will be a contribution to the signal from the difference between them. There are two sources for this offset, the telescope, and the sky itself. The former is a result of the beams reflecting off different parts of the telescope mirrors; a temperature or emissivity gradient across the optics will lead to the detected offset. Likewise, a similar gradient across the sky will also lead to such a signal.

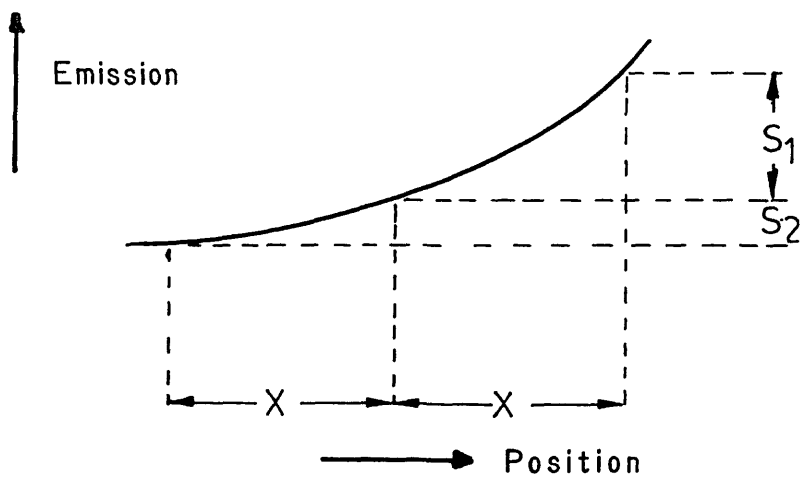
The key to removing this offset is to note that it is independent of the position of the two beams on the sky. To qualify this assertion, the instrumental and sky cases must be considered separately. Where the offset arises from the telescope, it is obviously independent of the position on the sky; the beam positions are fixed relative to the telescope, and the same contribution must arise regardless of where the telescope is pointing.

In the case of a sky contribution, independence of position must be taken to mean independence of position over a small region, as the emission/temperature gradient may vary significantly over large distances. A further point to note is that this independence is only a first order effect, sharp non-linear gradients giving rise to position-dependent offsets (Figure 28). Over the beam throws used in infrared astronomy though (of the order of minutes of arc), a linear model has been found to be a very good approximation.

Within this framework, extracting the signal proceeds very



a) Linear Gradient



b) Non-Linear Gradient

Figure 28. How a non-linear emission gradient on the sky can lead to a position-dependent offset. In (a), the gradient is linear, and so contributes an equal amount to the signal in both nod positions. A non-linear gradient (b) will lead to position dependent offsets.

much as chopping, albeit at a much lower frequency (~0.02 Hz). The whole telescope is moved (nodded), so that the source appears alternately in the two beams. The difference between them is then integrated over a period of time by a computer, and an estimate of the signal and noise derived. Any instrumental or sky offset is removed in the differencing.

The final point about nodding is, how often? If too long an interval is left between nods, serious drifts could arise. On the other hand, a typical nod will take about fifteen seconds (to allow for the telescope movement, and to let the observer fine guide it to the correct position), and so too short an interval will mean that most of the observing time is spent nodding. Experience has shown that a suitable nod period is about once every forty seconds (i.e. one nod every twenty seconds). This was the frequency selected for the observations described in the next chapter.

4.3) Signal Estimation Algorithms

A complication arises in estimating the signal when there is a temporal drift in the offset. In this case, the algorithm used in extracting the signal must try to avoid including the drift in either its estimation of the signal, or of the noise. Four algorithms were considered for use in the Tenerife observations, and these are described below.

In the subsequent sections, the following nomenclature will be used: A, B, C, and D refer to the output from the PSD measured when the telescope is in one of the two nod positions. The signal being measured is S which, in the absence of drift or noise, is the difference in PSD output between the two nod positions. σ is the standard deviation within a group of measurements, and E the

corresponding standard error. The subscript i differentiates between individual measurements of the PSD output made in a particular nod position, and the subscript j distinguishes between sets of nod position data.

4.3.1) Simple Differencing

This is shown schematically in Figure 29a. The nods are considered in complete cycles, and the mean and standard error for each nod position taken ($A_j \pm E_{A_j}$, $B_j \pm E_{B_j}$). The difference between them is taken to give an estimate of the signal viz.

$$S_j = B_j - A_j \tag{75}$$

$$E_j = (E_{A_j}^2 + E_{B_j}^2)^{1/2} . \tag{76}$$

Successive estimates of the summed to give the mean and standard error in the signal according to

$$\bar{S} = \frac{\Sigma S_j/E_j^2}{\Sigma 1/E_j^2} \tag{77}$$

$$E = \frac{1}{(\Sigma 1/E_j^2)^{1/2}} . \tag{78}$$

Intuitively, it can be seen that this algorithm is not very good in combatting drift. If there is drift present (as shown in Figure 29a) then it fails on two counts: firstly, the estimate of the signal is either an over-estimation or under-estimation of the true signal. Secondly, the noise associated with each data set (E_{A_j}, E_{B_j}) contains a component due to the drift within that set, and so the true noise is over-estimated.

4.3.2) Complex Differencing

Illustrated in Figure 29b, this algorithm tries to overcome

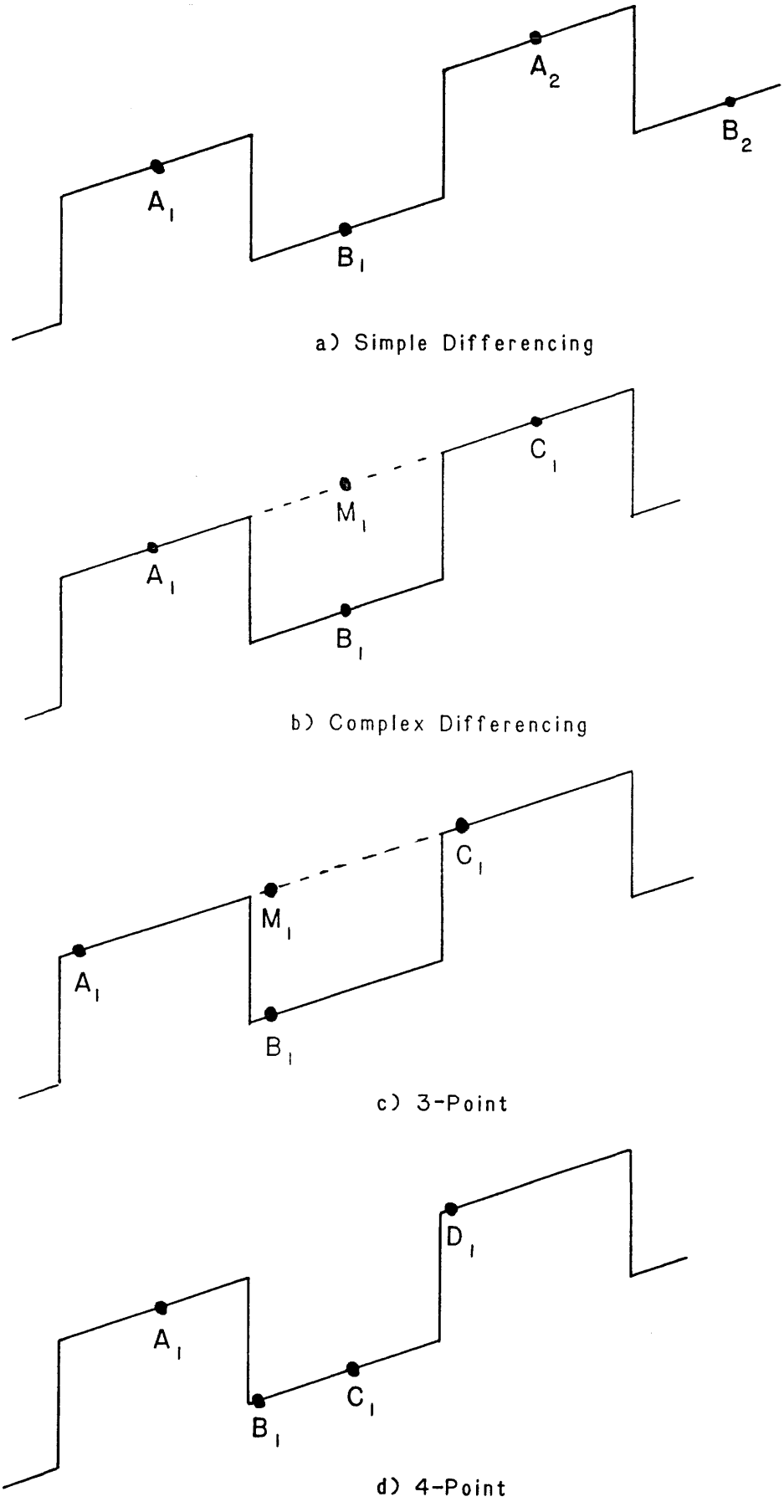


Figure 29. Signal estimation algorithms. The vertical axis is PSD output, and the horizontal axis time. The workings of the algorithms are described in the text.

the problem of over/under-estimating the signal by combining sets of data from the nod positions three at a time. Nod positions A and C are combined to give M, essentially an estimate of the signal that would be measured in this nod position when the telescope is in the other nod position measuring B. The value for M and its standard error are given by

$$M_j = \frac{1}{2}(A_j + C_j) \tag{79}$$

$$E_{M_j} = \frac{1}{2}(E_{A_j}^2 + E_{C_j}^2)^{1/2} . \tag{80}$$

This value is subtracted from the mean of position B to give an estimation of the signal viz.

$$S_j = (-1)^j(M_j - B_j) \tag{81}$$

$$E_j = (E_{M_j}^2 + E_{B_j}^2)^{1/2} . \tag{82}$$

The $(-1)^j$ is a parity term to ensure that the sign of S_j is correct. Successive estimations of the signal S_j are then combined as in equations (77) and (78). Although the signal is now estimated correctly, this algorithm still suffers from the drawback that the estimation of the noise in each nod position is still contaminated by the drift in each set.

4.3.3) Three-Point Algorithm

In this method, averaging the data points within a nod position has been discarded. Instead, corresponding points in three successive nod positions are combined to get an "instantaneous" estimation of the signal. These estimates are then summed for integration purposes.

Figure 29c shows the operation of this algorithm. Points A_i

and C_i are averaged to get M_i, analagous to the M_j in the Complex Differencing algorithm. M_i is then combined with B_i to get a best estimate of the signal

$$S_i = (-1)^i\left[\frac{(A_i + C_i)}{2} - B_i\right] \qquad (83)$$

(as with the previous algorithm, the $(-1)^i$ is a parity term). These estimates are then combined with equal weighting to get the signal $\bar{S}$,

$$\bar{S} = \frac{\Sigma S_i}{N} \qquad (84)$$

$$\sigma = \left[\frac{\Sigma(S_i - \bar{S})^2}{N - 1}\right]^{1/2} \qquad (85)$$

$$E = \frac{\sigma}{\sqrt{N}} \; . \qquad (86)$$

N here is the total number of estimations of S_i (one third the number of measurements of the PSD signal).

4.3.4) Four-Point Algorithm

This is similar to the three point algorithm, in that points are combined to get estimates of the signal, and successive estimates are averaged. The scheme is illustrated in Figure 29d. Points A_i and B_i, and points C_i and D_i are differenced to give two estimations of the signal, i.e.

$$S_{i_1} = A_i - B_i \qquad (87)$$

$$S_{i_2} = D_i - C_i \; . \qquad (88)$$

If there is a constant drift, then one combination is an over-estimation of the signal, and the other an under-estimation. Combination of the two yields

$$S_i = \frac{1}{2}(S_{i_1} + S_{i_2}) \; , \qquad (89)$$

a best estimate of the signal. Finally, successive best estimates are averaged as in equations (84) to (86), to get an estimate of the signal and the noise.

4.3.5) Comparison of Algorithms

To test these algorithms, each was implemented in Fortran, and run on sets of test data on the Cambridge Starlink VAX 11/780. The test data were generated by computer, NAG routines being used to add pseudo-random Gaussian noise where desired. The data varied in the amount of noise and drift, the latter quantity being either linear or quadratic.

Figures 30, 31, and 32 show a representative run, Figure 30 showing the actual raw data, which is contaminated by linear drift and Gaussian noise. Figure 31 shows the integrated signal as a function of the number of data points, and Figure 32 the standard error, also as a function of the number of points.

As expected, all algorithms managed to estimate the signal quite well, but where drift was present, the simple differencing method consistently got the wrong answer. Of the rest, the point by point methods tended to have lower standard errors than the complex differencing method, attributable to the inclusion of drift information in the estimation of the noise. Between the 3-point and 4-point algorithms, there was very little difference. The 3-point algorithm had marginally the better performance (in terms of accuracy) when the drift was a quadratic function with time, but this was highly sensitive to noise. When noise was added to the data, the 3 and 4-point algorithms were about equal. Thus the choice of the 4-point algorithm for data analysis is justifiable.

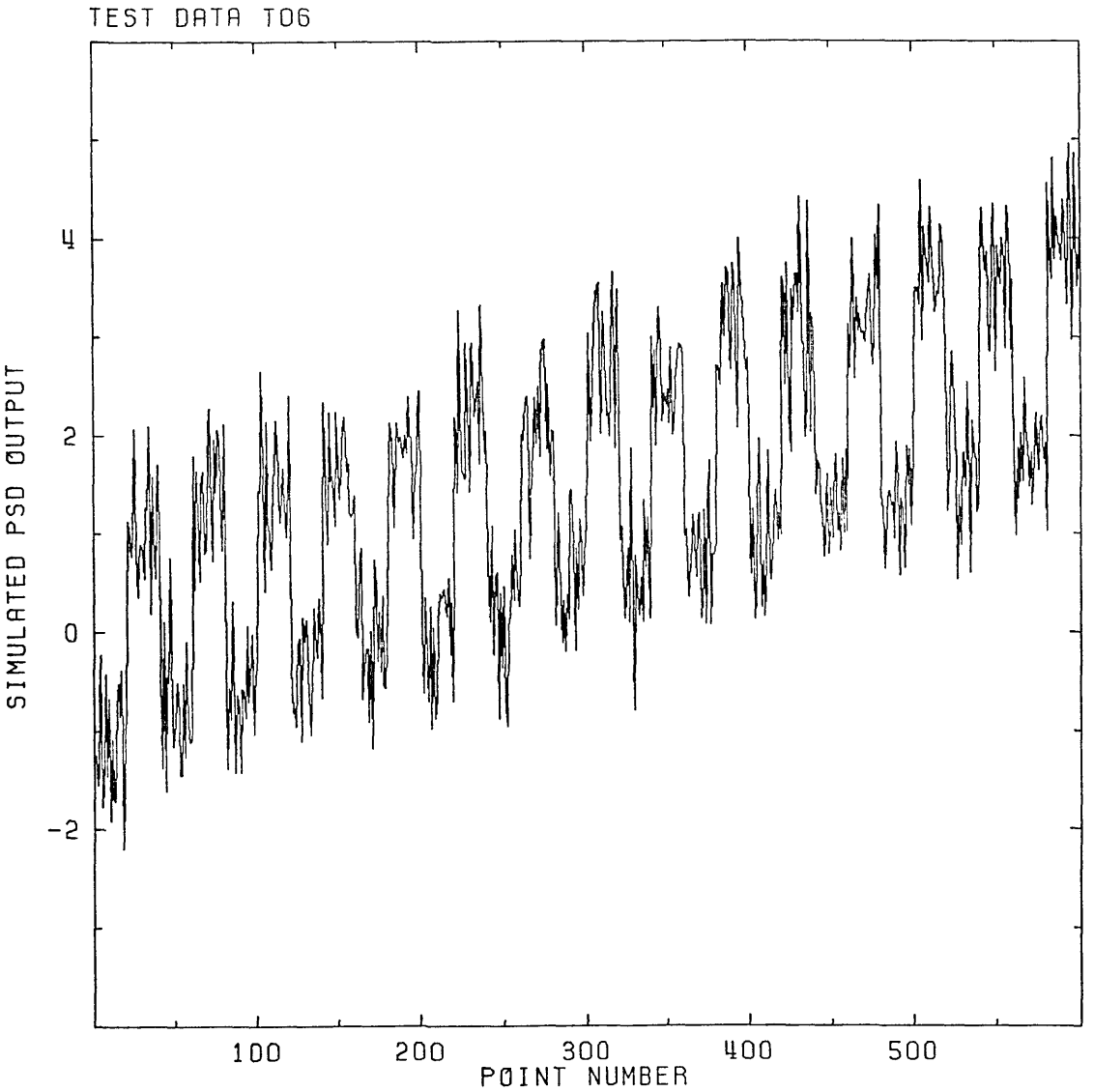


Figure 30. Typical set of test data used in evaluating the signal estimation algorithms. This set of data is a 2V signal, contaminated with linear drift of 5 mV per point, and pseudo-random Gaussian noise of standard deviation 0.5V.

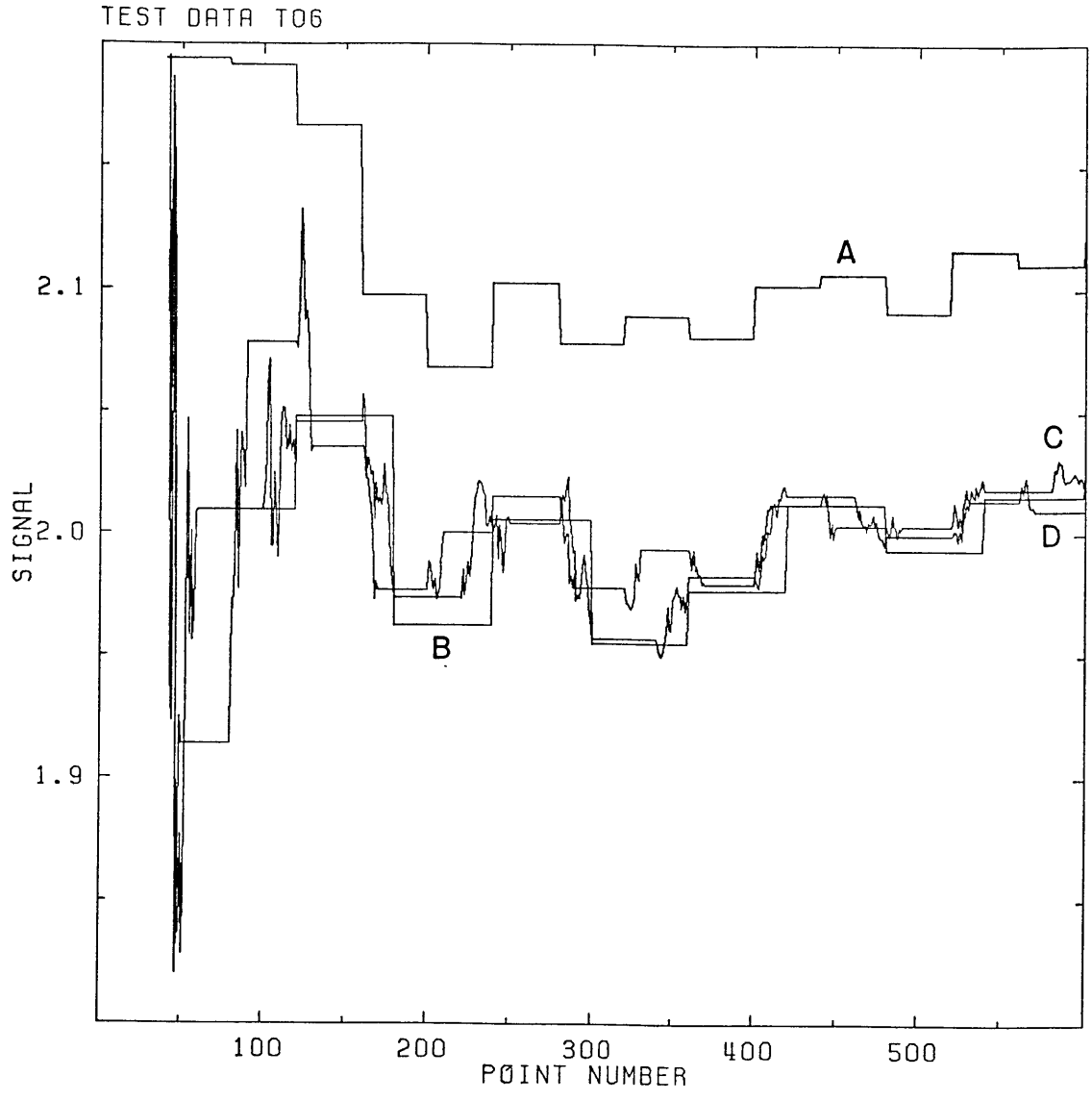


Figure 31. Estimation of the signal for each of the four algorithms as a function of the number of data points. The curves are labelled as follows: A - Simple Differencing, B - Complex Differencing, C - 3-Point, and D - 4-Point. The Simple Differencing algorithm has over-estimated the signal by over 0.1V, whereas the rest are close to the correct value.

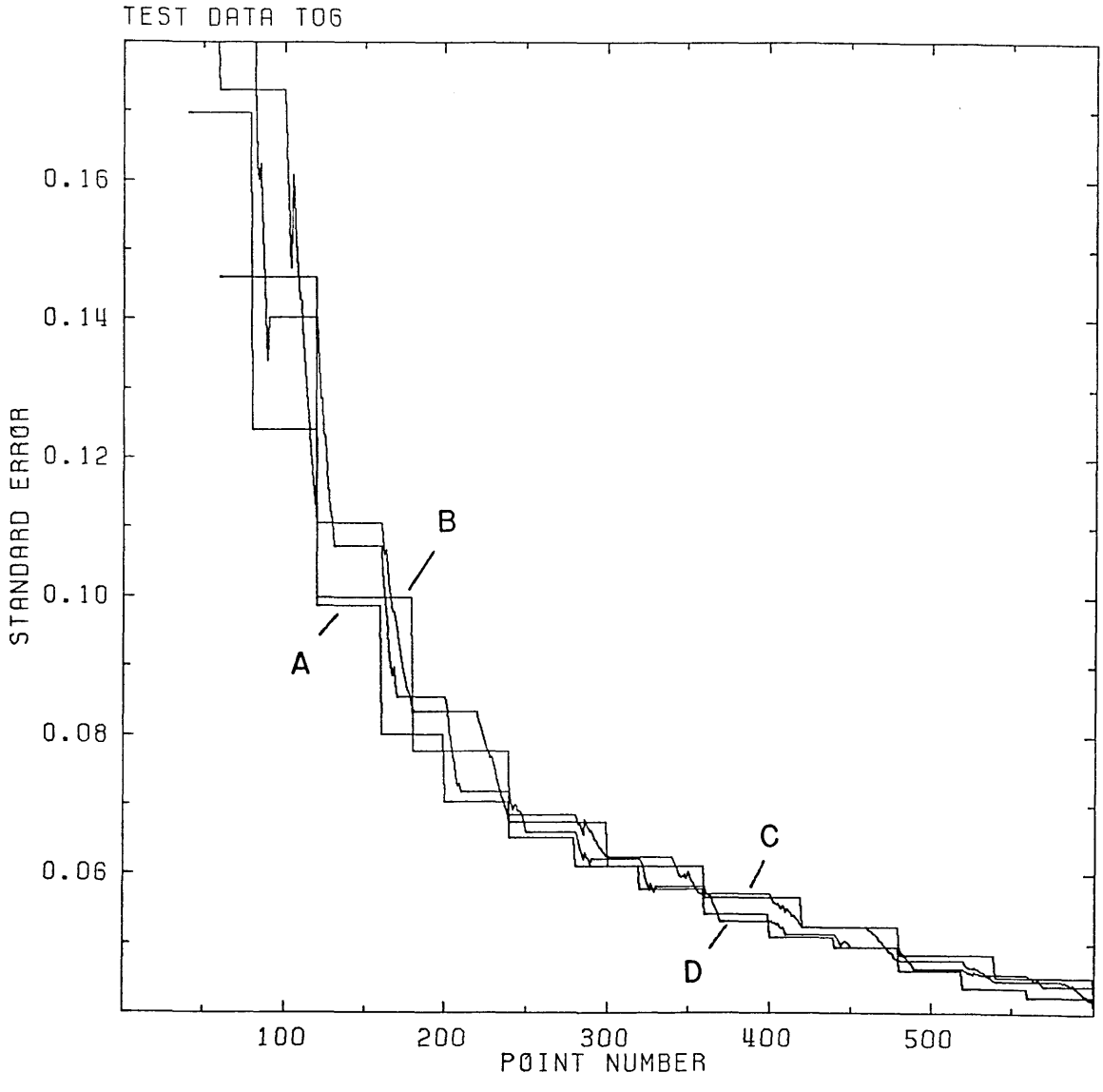


Figure 32. Standard error as a function of the number of data points. The curves are labelled as for Figure 31.

4.3.6) Non-Independence of Samples

The preceding sections have all assumed that each measurement of the PSD output (each sample) was independent of the rest. The PSD however, has a post-detection smoothing circuit, and so independence is not assured. In the PSD used, this smoothing circuitry was a double RC network, thus giving an equivalent noise bandwidth of

$$f_{eq} = \frac{1}{8RC} \cdot \qquad (90)$$

By Nyquist's theorem, this corresponds to an integration time of

$$t_{int} = \frac{1}{2f_{eq}}$$

$$= 4RC \cdot \qquad (91)$$

The integration time is the interval below which two successive samples are not independent. If the sample interval t_{samp} is smaller than this integration time, the noise is under-estimated, and so must be corrected. The key to doing this is to note that the noise is proportional to the square root of the number of samples, and that in a given time when n points are measured, there are only $n/(t_{int}/t_{samp})$ integration periods. It therefore follows that the noise is being under-estimated by a factor of $(t_{samp}/t_{int})^{1/2}$, and that dividing the noise by this factor restores it to the correct level. This correction is only required for cases when $t_{int} > t_{samp}$. For the reverse, $t_{int} < t_{samp}$, no correction is required.

4.4) The Tenerife Infrared Flux Collector

The observations described in the next chapter were performed with the aid of the Infrared Flux Collector (IRFC) in Tenerife. This telescope is sited at an altitude of 7800 ft on a volcanic ridge near the centre of the island. It has a 60 inch f/3 primary mirror and,

when working at the Cassegrain focus, the Dall-Kirkham mirror combination gives a focal ratio of 13.8, corresponding to a plate scale of 9.8 arcsec mm^{-1}. The telescope has an equatorial mount, and can cover the sky between -42^{o} and $+66^{o}$ in declination. Object acquisition is done manually, using the 4 inch and 10 inch finding telescopes. When tracking an object, fine guidance is by means of a handset.

As an aside, it is interesting to note that Tenerife was the site of the first astronomical observations in the infrared. In 1858, Charles Piazzi-Smythe managed to detect the moon in the near-infrared. The ruins of his observatory are close to the present day location of the IRFC.

4.5) Data Acquisition System

Figure 33 is a block diagram of the data acquisition system used with the IRFC. The signal from the detector, after amplification and phase sensitive detection with a reference waveform from the tertiary mirror chopper, is fed into a Data Precision 2540A1 digital multimeter (DMM). This is a $5\frac{1}{2}$ digit instrument (plus sign) and of these, the sign and the most significant $3\frac{1}{2}$ digits are read by the microcomputer, a Rockwell AIM-65. Since the PSD has an output swing of ±10 V, a 1999 bit resolution may be obtained (-9.99 V to +9.99 V; anything outside this range is considered to be an overload on the PSD).

The AIM-65 is the heart of the system, controlling the data analysis, data storage, and telescope control. The advantage of this system is that the architecture of the data logging set up is now almost entirely software controllable, with the consequence that any changes are relatively easy to make. The additional advantage of being

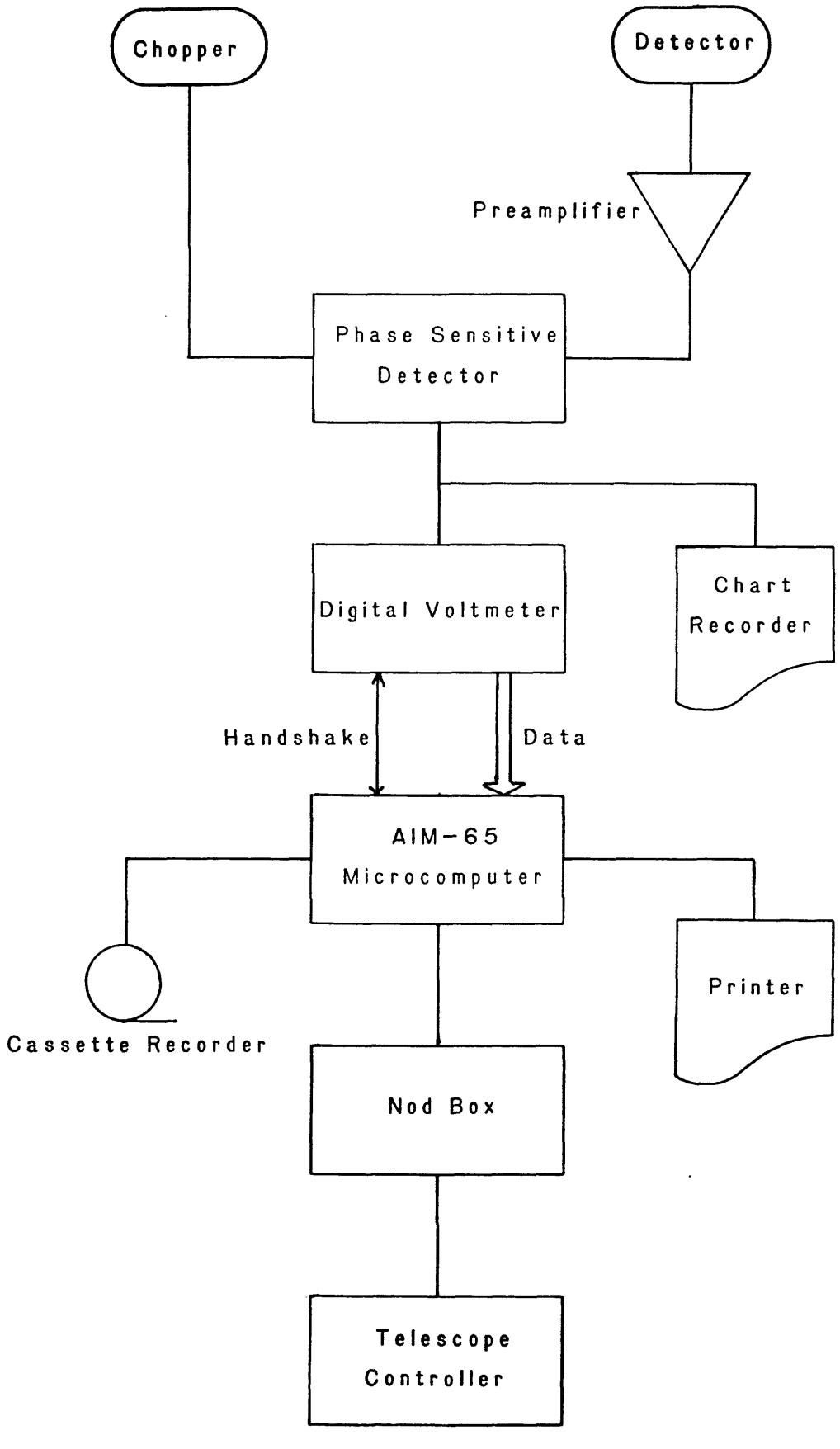


Figure 33. Block diagram of the data acquisition system.

able to test the entire system in the laboratory, and then take it in its entirety out to the IRFC is, of course, also very attractive.

4.5.1) The AIM-65 Microcomputer

The AIM-65 is an 8 bit microcomputer, based around the Rockwell 6502 microprocessor. It has an address space of 65 kB, and a 1 MHz clock gives it an average processing speed of one instruction per 3 - 4 μs. As configured in the data acquisition system, 20 kB is used for the I/O, the ROM-resident monitor, BASIC interpreter, and Assembler. There is 8 kB for user's programs, and a cassette recorder allows program and data storage. Experimental control and data input are handled by two 6522 Versatile Interface Adaptor (VIA) chips which give a total of 32 data lines, 8 handshake lines, and 4 timers. The microcomputer system was mounted in an attache case for ease of transport.

4.5.2) The Data Logging and Reduction Program

As the name suggests, this program was responsible for logging the data, and for doing real-time analysis. The core of the program is shown in Figure 34. After initialization, the program merely loops between two tasks. The first logs all the data, taking care of the handshaking with the DMM, reading the data, etc. The second part nods the telescope, writes the data to cassette tape, and then times out a delay while the telescope settles down, before the data is again sampled. The real-time analysis was done at the times when the microcomputer was not doing anything else.

To implement the program, five fundamental events were recognized:

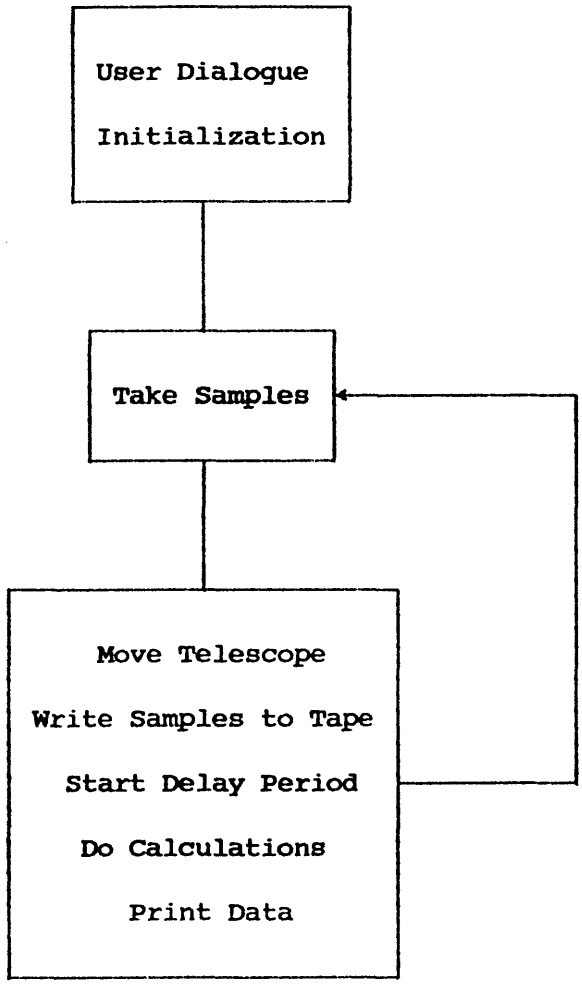


Figure 34. The core of the control and data analysis program. After initialization, the program loops between taking data and timing out the inter-position delay. Sampling and timing are done using interrupts, the data analysis occuring when the program is doing nothing else.

1) The time arriving for the microcomputer to request an analogue to digital conversion from the DMM.
2) A signal from the DMM announcing that data is available.
3) The time arriving to nod the telescope.
4) The time arriving to restart sampling.
5) A request from the user for some specified action.

The asynchronous nature of these events indicated that an approach using interrupts was needed, and so this part of the program was written in Assembler. The real-time analysis on the other hand, needed to do some quite high-level arithmetic, and so was written in BASIC.

4.5.3) Program Implementation

Sampling of the DMM is done on a regular basis, with a frequency set by the user (generally about once a second). The AIM-65 has a number of internal timers, each running at 1 MHz. In operation, a 16 bit counter is loaded with a number, and this counter decremented independently of the rest of the microcomputer until zero is reached. At this point, an interrupt is generated, the counter reloaded, and the countdown restarted.

Since the maximum interval that could be used was ~0.07 seconds, delays of seconds required extra programming. The approach taken was to load the counter with a suitable fraction of a second (0.05 seconds was used in practice) and on each interrupt decrement some preset memory locations. Thus the delay could be extended indefinitely. In the sequence of events described above, the microcomputer is always timing down to some event; either timing out the sample interval or timing out until the start of the next sampling period. Thus during operation the timers were left running all the

time; the actual event that was waiting for a timeout was indicated by a logical flag.

The signal from the DMM indicating that data is available is a truely asynchronous event. The Data Precision multimeter used has a delay of up to a quarter of a second between a conversion request and the appearance of data at the output. Rather than halt the program for this period, the appearance of data was also taken to require an interrupt. During this period, the timer is still counting down to the next sample, and is still interrupting. The AIM-65 has only one user interrupt vector, so a poll was needed on each interrupt to determine the source.

After all the samples for one nod position have been taken, the telescope nod is initiated, the count for delay between nod positions started, and the data dumped to tape. It is this last item that made the exact timing of this inter-position delay somewhat uncertain. There is no separate circuitry in the AIM-65 for formatting the data for the tape. Although the basic tones of 1.2 kHz and 2.4 kHz are produced in the system timers in the same way as the 20 Hz timeout signal, it is the AIM-65 that controls the pattern of tones on the tape. In order to avoid corrupting the output, the global interrupt disable had to be set for the couple of seconds that the write operation needed, and so shut out the 20 Hz timer.

The final asynchronous event noted above, is a request by the user for some action. This request was input to the AIM-65 by either a handset (requiring an interrupt), or by the keyboard (which was polled by the BASIC routine). The action taken here depended on the request made. Choices included pausing/restarting the program (e.g. if the guide observer was not ready at the end of the inter-position delay), a request for a printout of the latest signal estimate, etc.

The BASIC part of program conducts the initial dialogue with the user and starts up the Assembler routines. It then settles down in a loop, polling the keyboard (and the location in memory that a handset interrupt sets), and the data ready flag. This last item is a logical flag that informs the BASIC routine that a complete set of data has been taken. When the BASIC routine finds this flag set, it copies the data to its own internal storage. To allow for the possibility that this might be a "dud" set of data (e.g. the telescope did not settle down on the source in time), this data is not processed until immediately prior to reading the next data set. In the interval, the data is printed at a teletype, and the user may elect to reject this set from the calculations. Owing to memory limitations, only an accept/reject decision on the data is allowed. The data processing algorithm used to estimate the signal is the 4-point algorithm described in section 4.3.4.

The fact that nearly all operation is done under interrupts means that the BASIC and Assembler routines operate as virtually two independent programs; the BASIC routine merely looks at specific memory locations, and reads data that was placed there by the Assembler routines. This independence means that a minimum amount of effort need be spent when updating the program.

4.6) The Nod Box

Separation of the AIM-65 from the "noisy" electronics of the telescope drive system was the primary reason for the construction of this component. Historical reasons were responsible for the design. The control electronics of the telescope required a 0 V pulse of the order of milliseconds duration to trigger a change of state. The nod box receives a microsecond pulse from the AIM-65, and this triggers a

flip-flop/monostable/reed-relay circuit which brings the desired telescope control input down to 0 V for the required period of time.

The monostable was incorporated at a point when the AIM-65 had only 4 kB of memory and one VIA chip, which left only a handshake line to control the telescope. Although the AIM-65 has now been expanded, the telescope control part of system was left unchanged.

4.7) Summary

This chapter began with an explanation of why chopping and nodding were needed for infrared astronomy, and described some algorithms for reducing data obtained. The n-point algorithms proved themselves superior to the others, but there was little difference between the 3 and 4-point algorithms. The 4-point algorithm was selected as the basis of a data acquisition and real-time analysis system. An implementation of such a system, including a description of the data logging software was then presented.

This system was taken out to the IRFC in Autumn 1980, and Autumn 1981. The observations made there with this system are presented in the next chapter.

CHAPTER 5

Tenerife Observations

5.1) Introduction

The 10 μm photometric system, whose development was described in Chapters 3 and 4, was taken out to the 1.5 m Infrared Flux Collector (IRFC) in Tenerife for two observing runs, once during 1980 and once during 1981. This chapter describes the observations made at this telescope, and presents the sensitivity of the photometer derived from them.

5.2) Observing Trips

The photometer was taken out to the IRFC twice, during November/December 1980, and during October/November 1981. On both occasions, the weather was appalling. Of the 21 nights allocated in 1980, only 4 nights were completely clear, and one of these was spent in setting up the system. In 1981, 21 nights were again allocated, of which only 6 were completely clear. Of these, one night was lost to cryogen supply problems, and two were spent setting up.

Ground-based astronomy at 10 μm is particularly sensitive to the weather. Because the ambient (~300 K) temperature black body spectrum peaks at 10 μm, the presence of water and/or dust in the atmosphere can cause a sharp increase in the noise. Ideal observing weather is during the middle of a long spell of clear, dry weather, when such components have had a chance to precipitate out. Unfortuately, the clear nights in both trips did not meet such conditions, often coming sandwiched between two nights of 100% cloud

cover. With such weather, it was almost impossible to do any serious observing (although some was attempted), and so the trips were reduced to evaluating the equipment.

The 1980 trip was also a comissioning trip for the photometer and associated system. Use of the system during this period revealed a number of flaws in both the hardware (e.g. the load resistor was found to be producing excessive noise) and the software (e.g. a disparity of 0.1 seconds between the sample interval entered and that actually used). All these were corrected for the 1981 trip, and the experience gained was of inestimable value for that observing run. The nature of these flaws, combined with the very poor weather, meant that the 1980 results were useful chiefly for indicating improvements for the 1981 trip. Accordingly, only the 1981 results will be presented (with the knowledge that the sensistivity of the photometer in 1981 was certainly better than in 1980).

5.3) Data Reduction

As mentioned in the previous section, the weather was too bad to observe to observe any of the planned objects. In fact, only a total of 41 observations were attempted, and most of these were of standard stars, made for setting up the telescope, or were abandoned for one reason or another.

However, from the observations of these standard stars, it is possible to make an estimate of the sensitivity of the photometer. As mentioned in section 4.5.3., microcomputer memory limitations forced an accept/reject decision on each set of data as it was taken. The real-time analysis program could thus provide only a quick look at the data. In an attempt to clean up the data, it was transferred, via a PDP-11/CAMAC system, to the VAX 11/780 at the Cambridge node of the

Starlink network. The data was the edited to remove excessively noisy parts, using a purpose written editing program, and processed using the 4-point algorithm described in section 4.3.4. The judgement as to what constituted "excessively noisy" data was made by eye, but with a conscious bias towards retaining data.

5.3.1) Guiding Errors

One source of noise not mentioned in chapter 2 is noise due to guiding errors. Over the period of the observation, since the telescope will not track the source exactly, the source will, to varying degrees, move in and out of the telescope beam. On the IRPC, the fine guiding is done manually, and since the telescope is nodding once every 15 to 30 seconds, this problem will more noticable than on a telescope such as the Anglo-Australian Telescope, where the fine guiding is under computer control.

Also included in this category, though not directly associated with guiding, are the problems of seeing disk size and image flaring. If the seeing disk is larger than the aperture used, not all the energy from the source will fall on the detector. This is particularly acute if the object is a bright star, since this would be expected to be a point source and so the signal will be underestimated. Image flaring is when the size of the seeing disk fluctuates (due to atmospheric turbulence) during the course of an observation. If this flaring is larger than the aperture, then one has the problems mentioned above. Image flaring has a time scale of ~1 second and so shows up as noise on the output.

5.3.2) Extinction

Any stellar signal is attenuated by the atmosphere. From

Unsöld (1969), the intensity I, of a source at an elevation θ is

$$I = I_o \exp(-\tau \mathrm{Cosec}\theta) \ , \qquad (92)$$

where I_o is the flux density outside the earth's atmosphere, and τ is the optical depth of the atmosphere vertically above the point of observation. The quantity Cosecθ is the airmass, and is effectively the amount of atmosphere (in units of the quantity of atmosphere vertically upwards) through which the source is seen. The logarithmic dependence of extinction, combined with Pogson's formula for relating magnitudes to fluxes, means that equation (92) can be rewritten as

$$M = \alpha + \beta A \ , \qquad (93)$$

where M is the magnitude of a source, unattenuated magnitude α, at an airmass of A when the extinction is β magnitudes/airmass.

β is determined by measuring the signal at two different airmasses. With the signal from the electronics being proportional to the power incident on the detector (and hence the flux density of the source), then

$$\frac{S_1}{S_2} = \frac{I_1}{I_2} \ , \qquad (94)$$

where S_1, I_1 and S_2, I_2 are the signals and intensities at airmasses A_1 and A_2 respectively. Pogson's formula then gives the magnitude difference, from which the extinction per unit airmass can be derived.

5.4) Estimation of Photometer Sensitivity

The problems mentioned in section 5.3.1 make the estimation of the photometer sensitivity from a bright star very error prone. The noise in the output will be chiefly due to guiding errors, and will depress the estimate of sensitivity. However, if blank sky is

measured, the noise in the output will be "true" noise. Thus the sensitivity of the photometer can be worked out by using a calibration star to measure the responsivity of the photometer, and using blank sky noise to get the NEP.

5.4.1) Observations

The observations used to calibrate the photometer were taken towards the end of the observing trip, and the results (after analysis on the Starlink VAX) are presented in Table 2. On the 27th October, the weather was very windy, while on the evening of the 1st November, there was some visible cirrus cloud in the sky (and quite likely present all night). On both nights, the star β Peg was used as a calibration.

The problem of the poor weather is illustrated in the figures presented here. Note especially the second night, when the measured signal from β Peg increased by 30% between the two observations. Also, since the the average air mass during an observation marginally increased between the two observations (which were separated in time by only 10 minutes) an extinction correction can not be meaningful, and so none was made. The blank sky measurements are taken from fruitless attempts to observe some faint objects (Arp 190 and NGC 7469). The effect of guiding errors is readily apparent from these numbers; the highest standard deviation occurs in objects with a high signal.

5.4.2) Flux from β Peg

In order to evaluate the responsivity, it is essential to know the flux from the calibration star. Figure 35 gives the magnitude/wavelength relationship for this star, using data from the

Ref	Date	Object Name	Signal (μV)	Standard Dev (μV)	Number of Pts	Air Mass
G27	27/10/81	β Peg	221.5	4.552	136	1.000
G28	27/10/81	β Peg	239.4	2.375	76	1.004
G31	27/10/81	Arp 190	0.187	2.464	160	3.049
G32	01/11/81	β Peg	158.1	4.463	40	1.001
G33	01/11/81	β Peg	201.8	3.984	96	1.002
G36	01/11/81	NGC 7469	0.038	1.021	80	1.520
G37	01/11/81	NGC 7469	0.070	1.047	1000	1.746

Table 2. Observations from which the sensitivity of the photometer is derived. The columns are: (1) the reference number given to the observation, (2) the date of observation, (3) the name of the object being observed, (4) the signal, measured in μV across the load resistor, (5) the standard deviation in the signal (in μV), as calculated by the 4-point algorithm, (6) the number of points used in the calculation, and (7) the average airmass during the observation.

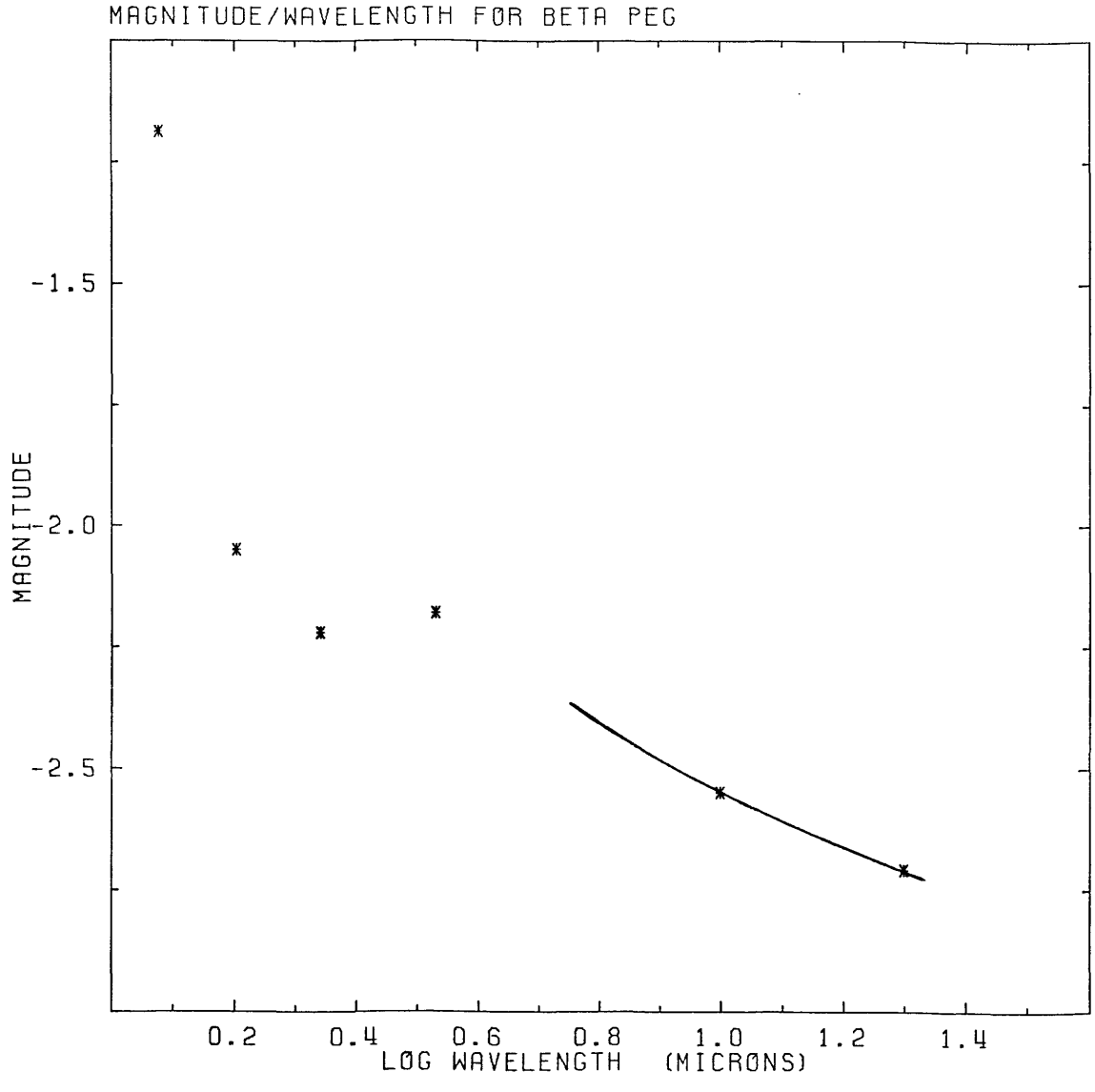


Figure 35. Relation between magnitude and wavelength for β Peg, using data from the unpublished Caltech list, and from Simon et al. (1972). The solid line is the best fit by eye through the data points in the 10 μm region.

unpublished Caltech list, and from Simon et al. (1972). The solid line is the best fit by eye through the data points, and gives a 10.7 μm magnitude of -2.56.

The relation between magnitude and flux density assumed for the calibration is shown in Figure 36, and is taken from Thomas et al. (1973), and Gehrz et al. (1974). Although the assumption on which some of the data is presented (i.e. that α Lyr is a zero magnitude star at all wavelengths) is still open to question (e.g. Wisniewski and Johnson 1979), the errors should be much less than the errors expected in the data reduction. The solid line in Figure 36 is a least squares fit to the data points. From this line, the flux density of a 0.0 magnitude star at 10.7 μm works out to be 30.5 Jy, indicating that the 10.7 μm flux density of β Peg is 322 Jy.

5.4.3) Results

To estimate the sensitivity, each night will be considered separately. This allows for the difference in extinction between nights. Taking the first night, the mean signal for β Peg is 236 μV (after weighting with the standard deviations). This implies a responsivity of 0.73 μV Jy^{-1}, and substituting this into the value for the noise gives a standard deviation of 3.4 Jy. Doing the same calculation for the second night gives a responsivity of 0.57 μV Jy^{-1} and a standard deviation of 1.8 Jy.

To convert the standard deviation (which is an estimate of the noise) into a sensitivity requires knowledge of the data reduction algorithm. In the four point method used, the standard error on the signal is given by the relation

$$E = \sigma\left[\frac{N}{4}\right]^{-1/2} \qquad (95)$$

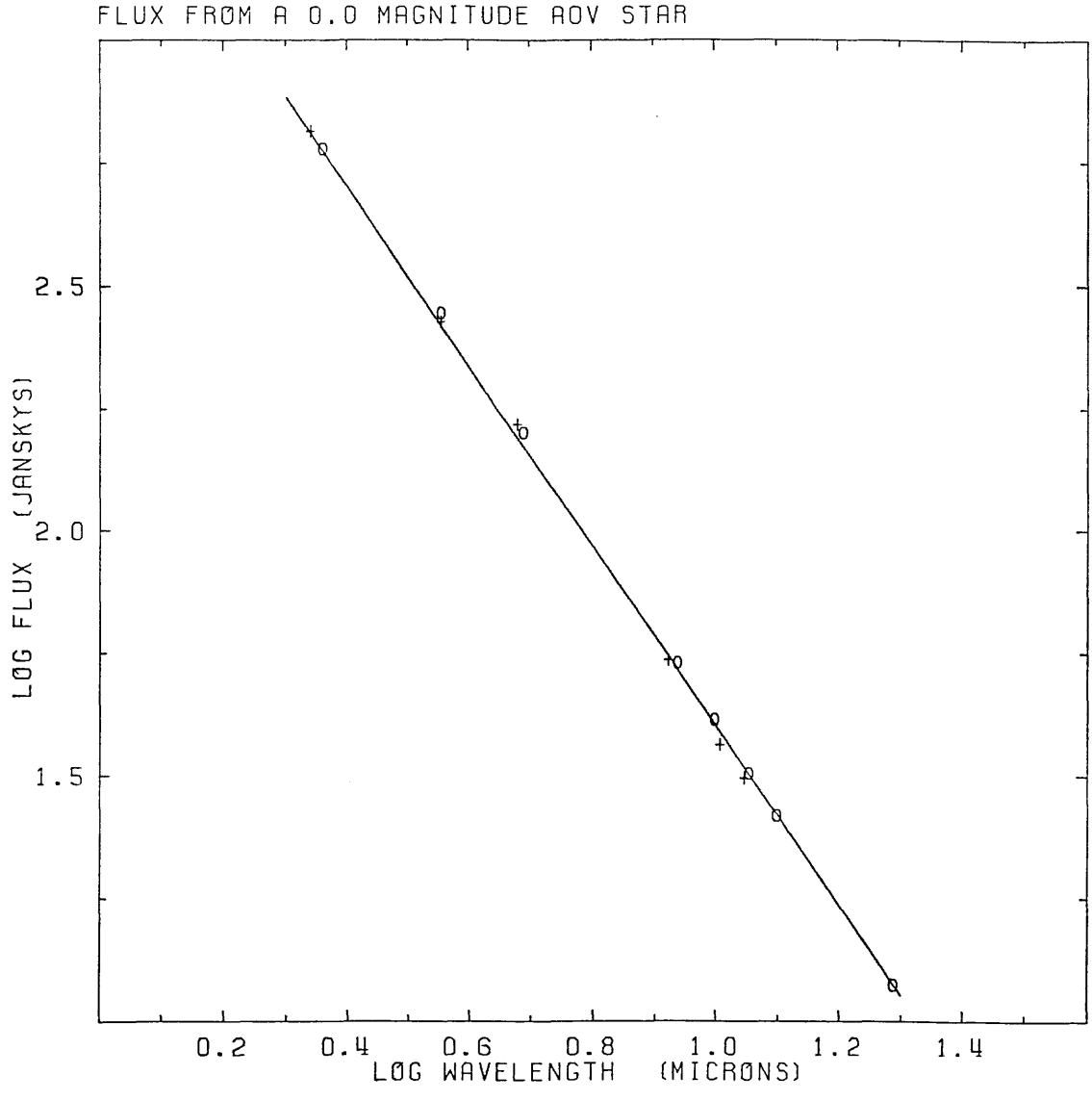


Figure 36. Relation between flux density and wavelength for a 0.0 magnitude AOV star. The crosses are points taken from Thomas et al. (1973), and the circles data from Gehrz et al. (1974). The solid line is a least-squares fit to the data points. (With acknowledgements to Martin Kessler).

(N being the total number of samples used), as each independent estimate of the signal uses four points. This value has been obtained over a time Nt_{samp} seconds (note that the standard deviation has already been corrected for sample independence). Since

$$E \propto \sqrt{time} \ , \tag{96}$$

It follows that in t seconds, the standard error would be

$$E(t \text{ seconds}) = E(N \text{ samples}) \times \left[\frac{Nt_{samp}}{t}\right]^{1/2} \ . \tag{97}$$

Substituting t = 1 s, and noting that t_{samp} was set to one second for the observations, leads to the results of

$$\text{Sensitivity (27 Oct)} = 6.7 \text{ Jy} \quad 1\sigma 1s \tag{98}$$

$$\text{Sensitivity (1 Nov)} = 3.7 \text{ Jy} \quad 1\sigma 1s \ . \tag{99}$$

5.5) Discussion

As can be seen from the factor of 1.8 between the figures, unless the photometer performance altered significantly between the two nights (and there was no reason why it should have), the figures can do little more than to put limits on photometer performance.

Converting the responsivities to Amps per Watt gives values of ~1.0 A W^{-1} and ~0.8 A W^{-1} respectively. These are below the laboratory figure of 4.3 A W^{-1}, but since the latter did not take into account the degradation of 50% expected because of the spatial response of the photometer (section 3.4.4), and the figures presented here do not take account of sky and telescope attenuation, the two sets of figures are not irreconcilable. It is clear from the closeness of the two figures however, that the cause of the sensitivity difference between the two nights is due to the difference in noise.

If the observations in which the noise is measured are analysed in the same way as the laboratory runs (i.e. finding the peak-to-peak voltage in 100 × the integration time), then the run G31 leads to a noise voltage of 3.9 $\mu V\ Hz^{-1/2}$. Adopting a mean value from the results of this analysis on G36 and G37 gives 1.7 $\mu V\ Hz^{-1/2}$. Since the field stop used on the telescope was the same as that used in the laboratory, it is possible to compare these measurements with the laboratory runs directly. Over the four laboratory runs immediately prior to taking the photometer to the IRFC, the mean noise measured at optimum bias was ~3.1 $\mu V\ Hz^{-1/2}$. The noise measured on the second night thus corresponds to that expected if the background in the laboratory were ~3 × that on the IRFC. The actual sensitivity measured is ~25% better than that anticipated from the laboratory measurements, but bearing in mind the inherent uncertainty in noise measurements, these figures are consistent.

As mentioned above, the discrepancy between the two nights lies mainly in the measurement of the noise. Since the measurement G31 was taken at a much larger airmass than the rest, it does not seem too biased to put more weight on the observations taken on the second night. With that in mind, we can cautiously say that the photometer, when on the IRFC, has a sensitivity around 3.7 Jy 1σ1s.

CHAPTER 6

10 μm and 20 μm Observations of NGC 3227

6.1) Introduction

In Chapter 1, it was argued that the gravitational interaction of galaxies passing close to one another could induce a burst of rapid star formation. It was also argued that this process would be detectable in the 10 μm band, and a programme of 10 μm observations of interacting galaxies was proposed. Unfortunately, circumstances prevented such a programme from being carried out. Nevertheless, this chapter examines the validity of such a hypothesis for one such galaxy, NGC 3227 (one of the Arp 94 pair), using 10 μm and 20 μm data generously supplied by R. D. Joseph and collaborators.

NGC 3227, together with NGC 3226, form an M51-type pair of galaxies, listed as object 94 in the Arp (1966) atlas. It has a bright semi-stellar nucleus, surrounded by a faint spiral structure containing patchy HII regions. Long exposure photographs have indicated that one arm of NGC 3227 connects with the elliptical companion NGC 3226 (Rubin and Ford 1968, hereafter RF). NGC 3227 is also a Seyfert galaxy, and has been variously described as type 2 (e.g. Weedman 1977) and type 1 (e.g. Rieke 1978), with recent observations (Ward 1983) favouring interpretation as a weak Seyfert 1.

6.2) Observations and Results

The 10 μm and 20 μm observations of NGC 3227 were made on the night of 22 February, 1983, with the 3.8 metre telescope of the UKIRT facility, Hawaii. The instrument used was the "Big Bertha" photometer

(described in the UKIRT Observers Manual (1981)), the measurements being made with an 8" aperture and sky subtraction achieved by chopping 11" in declination. Calibration was by means of comparison with the standard stars μ UMa and β Gem, the magnitudes of which are given in Table 3. Conversion to flux densities was made using the calibration curve of Figure 36.

The results are listed in Table 4. Although the observations were made with a signal to noise ratio of around 8, uncertainties in both the extinction measurements, and in the flux calibration have led to a cautious error of 20% being adopted.

Figure 37 puts these measurements in the context of other infrared observations of the same object. As NGC 3227 has a bright semi-stellar nucleus, the intensity distribution will almost certainly not follow the de Vaucoleurs relationship. This makes it difficult to compare observations of different apertures. Because of this, the points shown in Figure 37 (taken from Rieke (1978) and McAlary et al. (1983)) are for apertures close to the 8" used to take this data (being 7" and 9.1" respectively). The sole exception is the 10.6 μm point from Rieke (1978), which was taken with a 4.9" aperture.

A simple power law of index -1.6 fits the 10.5 μm and 20 μm points, and this was used in integrating the flux density over the 8 - 13 μm band, to give a flux of 4.2×10^{-14} W m^{-2}. Substituting this into equation (3) yields, to within a factor of 2 or 3, a bolometric flux of 6.1×10^{-13} W m^{-2}. With a distance to NGC 3227 of 12 Mpc (RF), this implies bolometric luminosity of $\sim 10^{36}$ W (2.7×10^{9} $L_{\odot}$). This figure applies to that part of the nucleus within the 8" aperture used for these observations. To get the mass within this aperture, we note that RF give a mass interior to a radius of 620 pc (21" aperture) of 3.0×10^{9} $M_{\odot}$. As this was derived from that part of the rotation curve

	N	Q
μ UMa	-0.95	-1.01
β Gem	-1.19	-1.24

Table 3. Magnitudes of calibration stars, taken from Gehrz et al. (1974).

λ (μm)	Flux Density (Jy)
10.5	0.31 ± 0.06
20.0	0.87 ± 0.17

Table 4. Flux densities obtained for NGC 3227 using an 8" aperture.

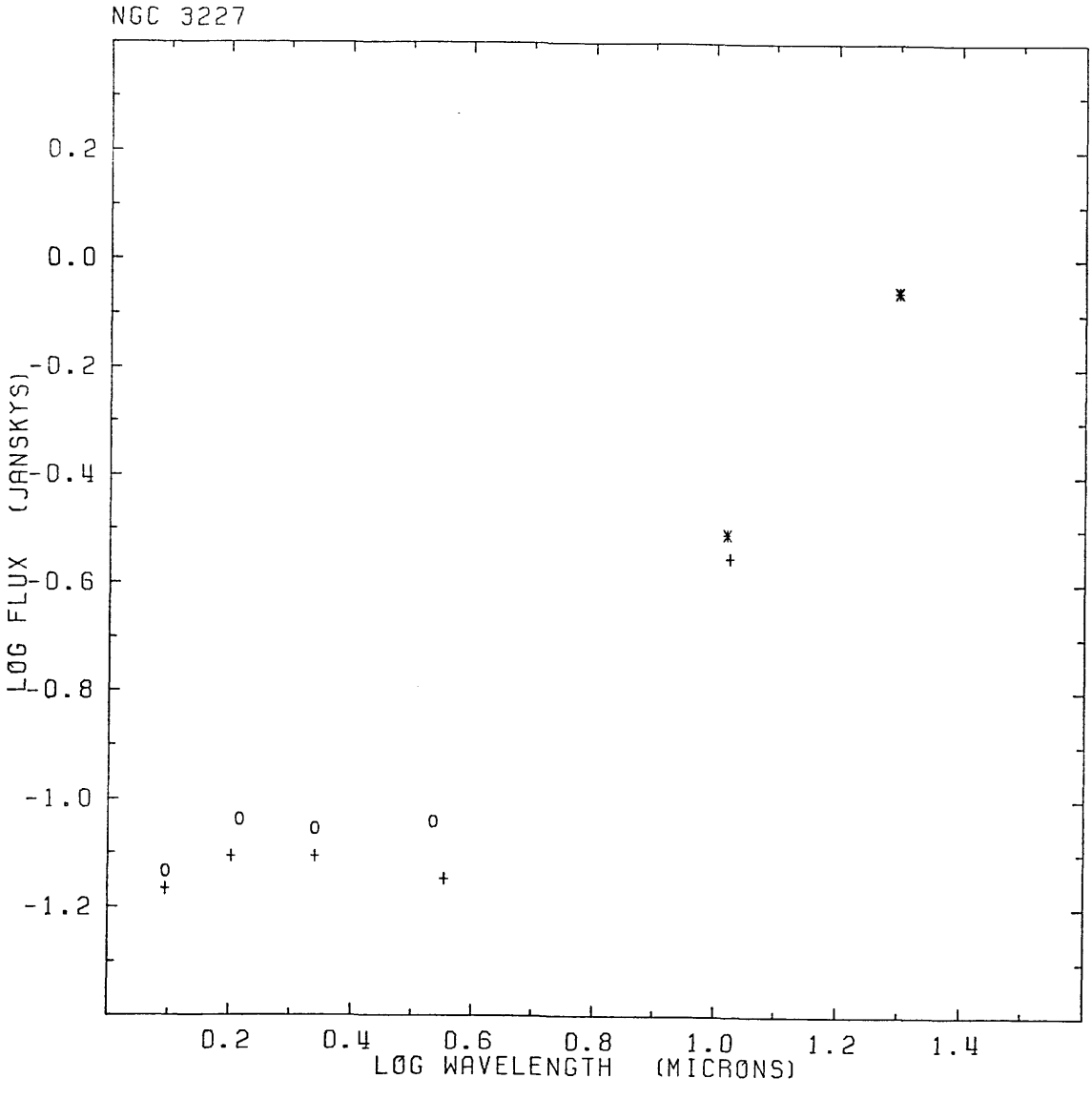


Figure 37. Infrared observations of NGC 3227. Crosses are data taken from Rieke (1978), circles from McAlary et al. (1983), and asterisks are data presented here.

that approximated to rigid body rotation, then we can use the relationship

$$M(r,\ r < 620\ \text{pc.}) \propto r^3 \qquad (100)$$

to get the mass observed with the 8" aperture as $1.6 \times 10^8\ M_\odot$. Combining this with the measurement of bolometric luminosity gives

$$\frac{M}{L} \simeq 0.06\ . \qquad (101)$$

6.3) Discussion

Measurements by RF give recessional velocities (corrected for the rotation of our galaxy) of 1069 km s^{-1} for NGC 3227 and 1243 km s^{-1} for NGC 3226, from which they adopt the mean, 1156 km s^{-1}, as the systematic velocity of the system. Adopting a Hubble constant of 100 km s^{-1} Mpc^{-1}, this places it at a distance of ~12 Mpc.

It is fairly certain that the two galaxies are interacting, and that the pair is not just a chance superposition on the sky. According to calculations made by RF, NGC 3227 has a mass out to 3.9 kpc of $\sim 2.6 \times 10^{10}\ M_\odot$, and NGC 3226 a mass about ten times greater. To be in a circular orbit around each other, NGC 3227 must have an orbital velocity of >450 km s^{-1} with respect to NGC 3226. If the transverse component of velocity just equals the radial velocity (of 173 km s^{-1}), then the galaxies will merge in $\sim 4 \times 10^7$ years. Thus, either the transverse velocity is much larger than the radial velocity, or the system is of fairly recent origin. The connecting arm seems to argue for the latter. The measured separation between the two galaxies is 7.3 kpc, and if the connecting arm is coplanar with NGC 3227, then the mean distance of the arm from NGC 3227 is ~5.6 kpc. Based on these figures, NGC 3226 should produce a significant-

perturbation of the arm. The regular shape belies this and so either the arm is much further away from NGC 3226 than it seems (e.g. if the arm is not coplanar with the galaxies), or that it is a relatively recent feature, and has not yet had a chance to disintegrate. Taken together, these facts seem to indicate that NGC 3227 has recently interacted with NGC 3226. Assuming then, that the galaxies have interacted, what is the cause of the Seyfert activity in NGC 3227?

McAlary et al. (1979), note that NGC 3227 is an unusual Seyfert 1, in that the JHKL spectrum, rather than being entirely non-thermal, lies midway between those galaxies whose spectrum is purely non-thermal, and those galaxies where the spectrum is purely stellar, and suggest that a mixture of non-thermal and stellar processes are responsible for the near-infrared emission.

Joseph et al. (1983) have analysed the JHKL colours of a sample of interacting galaxies from the Arp atlas. They find that most have H - K colours (roughly) consistent with those expected for normal galaxies, but about half (including NGC 3227) have a strong K - L excess. They consider four possible mechanisms for this:- reddening, optically thin free-free emission, non-thermal emission, and the addition of a warm (50 - 100 K) black-body component. The first is discounted because reddening sufficient to produce the strong K - L excess would also give a strong H - K excess, something which is not seen. The second is rejected because an extrapolation of a presumed free-free component falls short of the 2695 MHz flux density by a factor of 20. Addition of non-thermal component is also dismissed because of radio observations. Extrapolation of radio data into the infrared falls short of explaining the excess over the stellar 3500 K black-body curve by a factor of 100. They are thus led to the conclusion that their abnormal near-infrared colours are due to the

re-radiation by dust of stellar radiation. The correlation between interacting galaxies and star formation lead them to postulate that the stellar radiation is due to stars formed as a result of an interaction.

The data presented here lend support to the view of McAlary et al. (1979) that the near-infrared spectrum of NGC 3227 requires a two component model to explain it. If the 10 μm and 20 μm spectrum is extrapolated into the near-infrared using a a simple straight line between the two points (a power law of index -1.6), subtracted from the data of McAlary et al. (1983), and the result converted to magnitudes using the calibration curve in Figure 36, then the resulting colours are H - K = 0.37, K - L = 0.47, which is only marginally outside the area for "normal" galaxies on the (H - K, K - L) diagram (Joseph et al. 1983). It thus seems likely that there is a stellar component to the near-infrared radiation. The problem thus reduces to identifying the far-infrared component.

Certainly, an optically-thin free-free component cannot cause the spectrum to rise towards the far-infrared, and an optically-thick free-free component fails to explain the radio emission. Dismissing reddening for the reasons given above leaves either a black-body component, or a non-thermal component. In the case of NGC 3227, it is not so easy to reject the non-thermal option. The objection raised by Joseph et al. (1983) that extrapolation of the radio into the infrared yields too small a flux is overcome if there are two components, one producing the infrared radiation, and one the radio. This model however, has to explain why there is no trace of radio emission from the infrared component. A Balmer line full-width zero-intensity of many thousands of km s^{-1} (Ward 1983) argues for non-thermal interpretation. Elvis et al. (1978) have discovered X-ray emission

from NGC 3227 of ~10^{35} W, and McAlary et al. (1979) have noted that the ratio of 3.6 μm to 2 - 10 keV flux agrees well with the general 3.6 μm/X-ray correlation that they have noted for Seyfert 1 galaxies. They suggest that this may be due to inverse Compton scattering of the infrared photons.

If a black-body interpretation is taken, then fitting a single black-body curve to the 10 μm and 20 μm data points yields a temperature of ~230 K. This is not so high as to preclude thermal emission by dust. In any case, a distribution of (cooler) dust temperatures could also fit the spectrum. If the emission is due to dust, then the question remains as to what is the power source. In view of the association of interacting galaxies with star formation, a burst of star formation is a possible candidate. The recent discovery of a supernova on the edge of the nucleus (Aksenov and Pronik 1983) lends weight to this theory, and the M/L ratio of 0.06 is well within Struck-Marcell and Tinsley's limit of 0.5. If this is the case, then the star formation rate is, by equation (1), around 0.5 $M_\odot$ yr^{-1}.

Without some non-thermal component in the galaxy though, it is difficult to explain the X-ray luminosity. If the X-ray emission were produced by O stars, then taking the X-ray luminosity of an O star to be ~10^{26} W (Seward et al. 1979) requires NGC 3227 to be populated with ~10^9 such stars, too many for the luminosity measured for the galaxy. Supernovae are ruled out since a typical typical supernova remnant in our own galaxy outputs ~10^{41} J in the X-ray region over a period of ~10^4 years (Stewart et al. 1982, Fabian 1984), or an average power output of ~3×10^{29} W. To provide the observed X-ray luminosity thus requires the nucleus to contain ~3×10^5 supernova remnants, and these to be produced at the rate of ~30 per year. If a burst of star formation is responsible for the infrared

luminosity, then this rate is too high when compared with the star formation rate calculated above. This leaves close binary pairs. These have luminosities ~10^{31} W, and so only ~10^{4} are needed to explain the X-ray luminosity. Kriss et al. (1980) need a similar number to explain the X-ray luminosity of Seyfert 2 galaxies. In their words: "...This is perhaps no more unusual than a supermassive compact object at the center". Whether or not this is true, observations of extreme Seyfert 1 galaxies (e.g. NGC 4151, Rieke and Lebofsky 1981) tend to favour the supermassive object theory on account of the short-term variability.

It seems likely therefore that there is some non-thermal component in NGC 3227. Whether this is responsible for the infrared radiation directly is still open to question. If the infrared radiation is due to dust, then it may be that the dust is heated by such a component. Alternatively, the dust may be heated by a starburst, such a process coexisting with the non-thermal component.

To make any progress with these ideas, it is thus crucial to make further observations of NGC 3227. A search for the 10 μm silicate feature, and far-infrared observations out to 100 μm should help to resolve the question of the thermal (dust re-radiation) versus non-thermal (power law) infrared spectrum. Measurements of the intensity ratios of strong optical lines (Feldman et al. 1982) should help identify the physical processes active within the galaxy.

6.4) Conclusions

10 μm and 20 μm observations of NGC 3227 have been presented, and show that the spectrum is still rising into the far-infrared. The origin of the infrared luminosity is uncertain, and may be due to dust, or to a non-thermal component. If the former, then a non-thermal

component is needed to explain the X-ray luminosity and the Balmer line widths. If the latter, then a contrived model is needed to explain the apparent independence of the infrared and radio spectra. The possibility that both thermal and non-thermal processes are operating within the galaxy is not ruled out. Further observations are needed.

CHAPTER 7

Conclusions

The main conclusions of this thesis can be summarized as follows:-

1) It is quite probable that galaxy-galaxy interactions will induce a burst of star formation in one or both of the interacting galaxies.

2) This star formation should show up primarily in the far (~100 μm) infrared but, if far-infrared measurements are not available, it should be possible to estimate the strength of the star formation burst from an extrapolation of 8 - 13 μm measurements.

3) The Si:As photoconductor-based photometer built to make such measurements had a sensitivity on the IRFC of 3.7 Jy $1\sigma 1s$ and, although background limited, was a factor of ~3 lower in sensitivity than comparable bolometer-based 10 μm photometric systems. The quantum efficiency of the detector was calculated to be 12%.

4) Unless the quantum efficiencies of arsenic doped silicon photoconductors are significantly improved, a bolometer still seems best for broad-band 10 μm work, notwithstanding the cooling and microphonics problems.

5) 10 μm and 20 μm measurements of NGC 3227 made at UKIRT indicate that a two-component model is needed for the infrared spectrum, comprising a 3500 K stellar component, and a component that rises into the far-infrared. Further observations are needed to determine whether the latter is due to a black-body component (dust emission), or to some non-thermal process.

References

Aarseth, S.J., and Fall, S.M., 1980, Ap. J., 236, 43.

Adams, T.F., 1976, Bull. Amer. Astr. Soc., 8, 497.

Adams, T.F., 1977, Ap. J. Suppl. Ser., 33, 19.

Aksenov, E.P., and Pronik. V, 1983, IAU Circular No. 3887.

Alladin, S.M., 1965, Ap. J., 141, 769.

Allen, R.J., Ekers, R.D., Burke, B.F., and Miley, G.K., 1973, Nature, 241, 260.

Allen, R.J., and Sullivan, W.T., 1973, Astr. Ap., 25, 187.

Alloin, D., and Duflot, R., 1979, Astr. Ap., 78, L5.

Arp, H., 1966, Atlas of Peculiar Galaxies (Pasadena: California Institute of Technology).

Baldwin, J.A., Spinrad, H. and Terlevich, R., 1982, M.N.R.A.S., 198, 535.

Balick, B., and Heckman, T., 1979, A. J., 84, 302.

Barbieri, C., and di Tullio, G., 1979, Astr. Ap., 74, 110.

Becklin, E.E., Beckwith, S., Gatley, I., Matthews, K., Neugebauer, G., Sarazin, C., and Werner, M.W., 1976, Ap. J., 207, 770.

Beckman, J.E., and Moorwood, A.F.M., 1979, Rep. Prog. Phys., 42, 87.

Bergvall, N., 1981, Astr. Ap., 97, 302.

Bergvall, N., Ekman, A., and Lauberts, A., 1981, Astr. Ap., 95, 266.

Bevington, P.R., 1969, "Data Reduction and Error Analysis for the Physical Sciences", (McGraw-Hill), p. 92.

Bierman, P., 1976, Astr. Ap., 53, 295.

Bierman, P., Clarke, J.N., and Fricke, K.J., 1979, Astr. Ap., 75, 19.

Blaauw, A., 1964, Ann. Rev. Astr. Astrophys., 2, 213.

Blitz, L., 1980, in "Giant Molecular Clouds in the Galaxy", ed. A.M. Solomon, and M.G. Edmunds (Pergamon Press), p. 211.

Bothun, G.D., and Schommer, R.A., 1982, A. J., 87, 1368.

Bratt, P.R., 1969, Unpublished data.

Bratt, P.R., 1977, in "Infrared Detectors II", Semiconductors and Semimetals 12, ed. R.K. Willardson and A.C. Beer, (Academic Press), p. 39.

Burke, B.F., and Miley, G.K., 1973, Astr. Ap., 28, 379.

Carnevali, P., Cavaliere, A., and Santaneglo, P., 1981, Ap. J., 249, 449.

Charlton, D.E., 1964, IR. Phys., 4, 157.

Chaterjee, T.K., 1981, Ap. Space Sci., 76, 491.

Clark-Jones, R., 1959, Proc. IRE, 47, 1481.

Condon, J.J., 1980, Ap. J., 242, 894.

Condon, J.J, Condon, M.A., Gisler, G., and Puschell, J.J., 1982, Ap. J., 252, 102.

Dostal, V.A., and Metlov, G.V., 1977, IAU Symposium 79, 117.

Dostal, V.A., and Metlov, G.V., 1979, Sov. Astr., 23, 1.

Einasto, J., Kaasik, A., and Saar, E., 1974, Nature., 250, 309

Elmegreen, B.G., 1978, in "Infrared Astronomy", ed. G. Setti, and G.G. Fazio, (Dortrecht:Reidel), p. 327.

Elmegreen, B.G., and Lada, C.J., 1977, Ap. J., 214, 725.

Elvis, M., Maccacaro, T., Wilson, A., Ward, M.J., Penston, M., Fosbury, R., and Perola, G., 1978, M.N.R.A.S., 183, 129.

Emmons, R.B., Hawkins, S.R., Cuff, K.F., 1975, Opt. Eng., 14, 21.

Eneev, T.M., Kozlov, N.N., and Sunyaev, R.A., 1973, Astr. Ap., 22, 41.

Faber, S.M., Gallagher, J.S., 1979, Ann. Rev. Astr. Ap., 17, 135.

Fabian, A.C., 1984, Private Communication.

Fazio, G.G., 1978, in "Infrared Astronomy", ed. G. Setti, and G.G. Fazio, (Dortrecht:Reidel), p. 25.

Feldman, F.R., Weedman, D.W., Balanzo, V.A., and Ramsey, L.W., 1982, Ap. J., 256, 427.

Few, J.M.A., Madore, B.F., and Arp, H.C., 1982, M.N.R.A.S., 199, 633.

Fischer, J., Simon, M., Benson, J., and Solomon, P.M., 1983, Ap. J. (Letters), 273, L27.

Fitch, W.S., Pacholczyk, A.G., and Weymann, R.J., 1967, Ap. J. (Letters), 150, L67.

Gallagher, J.S., and Ostriker, J.P., 1972, A. J., 77, 288.

Gehrz, R.D., Hackwell, J.A., and Jones, T.W., 1974, Ap. J., 191, 675.

Gehrz, R.D., Sramek, R.A., and Weedman, D.W., 1983, Ap. J., 267, 551.

Gerhard, O., 1981, M.N.R.A.S., 197, 179.

Gerola, H., and Seiden, P.E., 1978, Ap. J., 223, 129.

Gopal, V., 1981, IR. Phys., 21, 31.

Gottesman, S.T., and Weliachew, L., 1977, Ap. J., 211, 47.

Harper, D.A., 1975, in "HII Regions and Related Topics" (Lecture Notes in Physics 42), ed. T.L. Wilson, and D. Downes, (Springer-Verlag), p. 343.

Harwit, M., and Pacini, F., 1975, Ap. J. (Letters), 200, L127.

Hawkins, M.R.S., 1983, Nature, 303, 406.

Herbst, W., and Assousa, G. A., 1978, in "Protostars and Planets", ed. T. Gehrels, (University of Arizona), p. 368.

Hickson, P., Richstone, D.O., and Turner, E.L., 1977, Ap. J., 213, 323.

Hodge, P.W., 1968, A. J., 73, 846.

Holmberg, E., 1941, Ap. J., 94, 385.

IRFC Users Manual, 1976, 2nd Edition, Science and Engineering Research Council.

Jones, B., and Tremaine, S., 1979, Nature, 277, 516.

Jorden, P.R., Long, J.F., MacGregor, A.D., and Selby, M.J., 1976, Astr. Ap., 49, 421.

Joseph, R.D., Meikle, W.P.S., Robertson, N.A., and Wright, G.S., 1983, Preprint, Submitted to M.N.R.A.S.

Kaufman, M., 1979, Ap. J., 232, 717.

Karachentsev, I.D., and Karachentsev, V.E., 1975, Sov. Astr., 18, 428.

Kessler, M.F., 1981, Ph.D. Thesis, University of London.

Khachikyan, E.Y., and Weedman, D.W., 1971, Astrophys., 7, 231.

Kingston, R.H., 1978, "Detection of Optical and Infrared Radiation", (Springer Series in Optical Sciences 10), (Springer Verlag).

Kittel, C., 1971, "Introduction to Solid State Physics", 5th edition, (Wiley and Sons).

Kolesnik, I.G., Silich, S.A., and Fomin, P.I., 1979, Sov. Astr. Lett., 5, 303.

Krienke, O.F., 1975, A. J., 80, 492.

Kriss, G.A., Canizares, C.R., and Ricker, G.R., 1980, Ap. J. 242, 492.

Kruse, P.W., 1977, in "Optical and Infrared Detectors", (Topics in Applied Physics 19), ed. R.J. Keyes, (Springer Verlag) p. 5.

Kyle, T.G. and Goldman, A., 1975, "Atlas of Computed Infrared Absorption Spectra", (N.C.A.R., Boulder, Colorado, TN/STR-112)

Lada, C.J., Blitz, L., and Elmegreen, B.G., 1978, in "Protostars and Planets", ed. T. Gehrels, (University of Arizona), p. 341.

Larson, R.B., 1978, in "Infrared Astronomy", ed. G. Setti, and G.G. Fazio, (Dortrecht:Reidel), p. 137.

Larson, R.B., and Tinsley, B.M., 1978, Ap. J., 219, 46.

Lauberts, A., 1974, Astr. Ap., 33, 231.

Lebofsky, M.J., Sargent, D.G., Kleinman, S.G. and Rieke, G.H., 1978, Ap. J. 219, 487.

Levenstein, H., 1965, Appl. Opt., 4, 639.

Lin, C.C., and Shu, F.H., 1964, Ap. J., 140, 646.

Longhurst, R.S., 1967, "Geometrical and Physical Optics", 2nd edition, (Longmans, Green and Co. Ltd.).

Loose, H.H., and Fricke, K.J., 1981, in "The Most Massive Stars", ed. S. D'odorico, D. Baade, and K. Kjar, (E.S.O. publications), p. 269.

Loren, R.B., 1976, Ap. J., 209, 466.

Low, F.J., and Rieke, G.H., 1974, in "Methods of Experimental Physics", (Academic Press), 12, 415.

Luinge, W., 1981, Preprint.

Luinge, W., Wildeman, K.J., Van Duinen, R.J., 1980, IR. Phys., 20, 39.

Lynds, B.T., 1980 Ap. J., 238, 17.

Lynds, R., and Toomre, A., 1976, Ap. J., 209, 382.

MacGregor, A.D., 1977, Imperial College of Science and Technology, Detector Testing Contract, Final Report.

Madore, B.F., 1980, in "The Structure and Evolution of Normal Galaxies", ed. S.M. Fall, and D. Lynden-Bell, (Cambridge University Press), p. 239.

Malin, D.F., and Carter, D., 1980, Nature, 285, 643.

Mathews, W.G., and Baker, J.C., 1971, Ap. J., 170, 241.

McAlary, C.W., McLaren, R.A., and Crabtree, C.R., 1979, Ap. J., 234, 471.

McAlary, C.W., McLaren, R.A., McGonegal, R.J., and Maza, J., 1983, Ap. J. Suppl. Ser., 52, 341.

Metik, L.P., and Pronik, I.I., 1979, Astrophys., 15, 27.

Miller, R.H., and Smith, B.F., 1980, Ap. J., 235, 421.

Milton, A.F., Blouke, M.M., 1971, Phys. Rev. B, 3, 4312.

Mueller, M.W., and Arnett, W.D., 1976, Ap. J., 210, 670.

Norton, P., 1973, Unpublished Data.

Opik, E.J., 1953, Irish Astron. J., 2, 219.

Ostriker, J.P., 1980, Comm. Astrophys., 8, 177.

Ostriker, J.P., Peebles, P.J.E., and Yahil, A., 1974, Ap. J. (Letters), 193, L1.

Palmer, J.M., 1976, IR. Phys., 19, 243.

Panagia, N., 1974, in "HII Regions and the Galactic Centre" (8th ESLAB symposium), p. 163.

Penchina, C.M., 1975, IR. Phys., 15, 9.

Penston, M.V., Penston, M.J., Selmes, R.A., Becklin, E.E., and Neugebauer, G., 1974, M.N.R.A.S., 169, 357.

Petritz, R.L., 1959, Proc. IRE, 47, 1458.

Picus, G., 1962, J. Phys. Chem. Solids, 23, 1753.

Putley, E.H., 1964, Phys. Stat. Sol., 6, 571.

Rieke, G.H., 1978, Ap. J., 226, 550.

Rieke, G.H., and Lebofsky, M.J., 1978, Ap. J. (Letters), 220, L37.

Rieke, G.H., and Lebofsky, M.J., 1981, Ap. J., 250, 87.

Rieke, G.H., Lebofsky, M.J., Thompson, R.I., Low, F.J. and Tokunaga, A.T., 1980, Ap. J., 238, 24.

Rieke, G.H., and Low, F.J., 1975, Ap. J.(Letters), 200, L67.

Roberts, W.W., 1969, Ap. J., 158, 123.

Roos, N., and Norman, C., 1979, Astr. Ap., 76, 75.

Rosen, D.L., 1983, Ph.D. Thesis, University of London.

Rubin, V.C., and Ford, W.K., 1968, Ap. J., 154, 431.

Sandage, A., 1961, "The Hubble Atlas of Galaxies", (Carnegie Institute, Washington).

Sastry, K.S., and Alladin, S.M., 1979, Astr. Ap., 63, 203.

Saur, W., 1968, IR. Phys., 8, 255.

SBRC, 1976, Technical Bulletin No. 208, Santa Barbara Research Center.

Schweizer, F., 1976, Bull. Amer. Astr. Soc., 8, 496.

Sclar, N., Burstein, E., Turner, W.J., and Davisson, J.W., 1953, Phys. Rev., 91, 215.

Sclar, N., 1976, IR. Phys., 16, 435.

Searle, L., Sargent, W.L.W., and Bagnuolo, W.G., 1973, Ap. J., 179, 427.

Seiden, P.E., and Gerola, H., 1979, Ap. J., 233, 56.

Seiden, P.E., Schulman, L.S., and Gerola, H., 1982, Ap. J., 253, 91.

Seward, F.D., Forman, W.R., Giacconi, R.E., Harnden, F.R., Jones, C.,

and Pye, J.P., 1979, Ap. J. Lett., 234, L55.

Seyfert, C., 1943, Ap. J., 97, 28.

Simon, T., Morrison, D., and Cruikshank, D.P., 1972, Ap. J. (Letters), 177, L17.

Smakula, A., Swinehart, C.F., Whitaker, G., O'Dell, E.W. and Johnson, R.W., 1967, "Harshaw Optical Crystals", (The Harshaw Chemical Company).

Smirnov, M.A., and Komberg, B.A., 1980, Astrophys., 16, 255.

Smirnov, M.A., and Tsvetkov, D.Y., 1981, Sov. Astr. Lett., 7, 86.

Smith, R.A., Jones, F.E., and Chasmar, R.P., 1968, "The Detection and Measurement of Infra-Red Radiation", (Clarendon Press).

Solinger, A., Morrison, P., and Markert, T., 1977, Ap. J., 211, 707.

Spight, L., and Grayzeck, E., 1977, Ap. J., 213, 374.

Spitzer, L., 1958, Ap. J., 127, 17.

Stewart, G.C., Fabian, A.C., Terlevich, R.J., and Hazard, C., 1982, M.N.R.A.S., 200, 61P.

Stocke, J.T., 1978, A. J., 83, 348.

Stockton, A., and Bertola, F., 1980, Ap. J., 235, 37.

Struck-Marcell, C., and Tinsley, B., 1978, Ap. J., 221, 562.

Sugden, K.C., 1978, Ph. D. Thesis, University of London.

Sunder, G.S., and Chaterjee, T.K., 1982, Ap. Space Sci., 81, 479.

Telesco, C.M., Becklin, E.E., and Wynn-Williams, C.G., 1980, Ap. J. (Letters), 241, L69.

Telesco, C.M., and Harper, D.A., 1975, Bull. Amer. Astr. Soc., 7, 436.

Telesco, C.M., Harper, D.A., and Lowenstein, R.F., 1976, Ap. J. (Letters), 203, L53.

Theys, J.C., and Spiegel, E.A., 1977, Ap. J., 212, 616.

Thomas, J.A., Hyland, A.R., and Robinson, G., 1973, M.N.R.A.S., 165, 201.

Tinsley, B.M., 1968, Ap. J., 151, 547.

Tinsley, B.M., and Larson, R.B., 1979, M.N.R.A.S., 186, 503.

Tolner, H., 1977, "Photoconductors" (Report prepared for LIRTS definition team).

Toomre, A., 1974, IAU Symposium 58, 347.

Toomre, A., and Toomre, J., 1972, Ap. J., 178, 623.

Tremaine, S., 1980 in "The Structure and Evolution of Normal Galaxies", ed. S.M. Fall, and D. Lynden-Bell, (Cambridge University Press), p. 67.

UKIRT Newsletter, 1983, No. 9, Science and Engineering Research Council.

UKIRT Observers Manual, 1981, Science and Engineering Research Council.

Unsöld, A., 1969, "The New Cosmos", (Springer-Verlag).

Van Woeden, H., Mebold, U., Goss, M., and Peterson, B.A., 1979, Astr. Ap., 76, 230.

Vardanyan, R.A., and Melik-Alverdyan, Y.K., 1975, Astrophys., 11, 12.

Viallefond, F., Allen, R.J., and De Boer, J.A., 1980, Astr. Ap., 82, 207.

Villumsen, J.V., 1982, M.N.R.A.S., 199, 493.

Von Kapp-Herr, A., Haslam, C.G.T., and Wielebinski, R., 1977, Astr. Ap., 57, 337.

Ward, M.J., 1983, Private Communication.

Weedman, D.W., 1977, Ann. Rev. Astr. Astrophys., 15, 69.

Weliachew, L., Sancisi, R., and Guelin, M., 1978, Astr. Ap., 65, 37.

White, S.D.M., 1978, M.N.R.A.S., 184, 185.

White, S.D.M., 1979a, M.N.R.A.S., 189, 831.

White, S.D.M., 1979b, Ap. J. (Letters), 229, L9.

White, S.D.M., 1980, Phil. Trans. Roy. Soc. Lon. A296, 347.

Williams, R.L., 1969, J. App. Phys., 40, 184.

Wisniewski, W.Z., and Johnson, H.L., 1979, Sky and Telescope, 57, 4.

Wolf, J., and Lemke, D., 1983, Astr. Ap., 119, 294.

Wright, A.E., 1972, M.N.R.A.S., 157, 309.

Woodward, P.R., 1976, Ap. J., 207, 484.

Wynn-Williams, C.G., and Becklin, E.E., 1974, Pub. A.S.P., 86, 5.

Young, E.T., and Low, F.J., 1979, Proc. SPIE., 172, 184.

APPENDIX

ON HELIUM ABUNDANCES IN GALACTIC HII REGIONS:

IS THERE A "GEOMETRIC EFFECT" IN THE

RADIO RECOMBINATION LINE OBSERVATIONS

R. D. Joseph and S. A. Morris

Abstract

He^+/H^+ abundances deduced from radio recombination line observations of galactic HII regions are found to range from about zero to 12% by number. This variation has been explained by a "geometric effect": that He^+ Strömgren spheres are smaller than those for H^+ for those HII regions with apparently low He^+ abundance.

We have re-examined the radio data, comparing apparent He^+ abundances obtained with telescope beams both smaller and larger than the HII regions observed, and we find the data to be incompatible with such a "geometric effect". We conclude that the suggestions of strong selective dust absorption or a deficiency of He-ionizing photons in the spectra of early-type stars to produce such a "geometric effect" are unfounded, and that radio recombination line techniques are apparently not yet reliable for measuring He^+ abundances in the Galaxy.

A.1) Introduction

Churchwell, Mezger, and Hutchmeier (1974), reporting on their extensive observations of hydrogen and helium radio recombination lines in 39 galactic HII regions, found that the measured He^+/H^+ line intensity varied from <0.01 to ~0.12. Since, for a wide range of conditions expected to hold in these HII regions, this line intensity ratio should equal the number density abundance ratio $y^+ = n(He^+)/n(H^+)$, these results imply there is a variation in He abundance ranging from zero to 12%

To explain these large apparent departures from an assumed universal He abundance of ~10%, Churchwell et al. suggested an observational effect could be responsible. If the He^+ Strömgren sphere were smaller than the H^+ Strömgren sphere, then a He^+ abundance less than the true abundance would be found because the He recombination line was emitted from a smaller volume than was the H line. The He^+ and H^+ Strömgren radii should be equal for a simple dust-free HII region ionized by a star of spectral type O8 or earlier. However, if dust were present inside the nebula, and its absorption cross-section for He-ionizing photons ($\lambda < 504\mathring{A}$) were substantially larger than that for H-ionizing photons ($\lambda < 912\mathring{A}$), then the He^+ Strömgren sphere could be significantly smaller than that for H^+. Churchwell et al. (1974) found some evidence to support this explanation in a (weak) inverse correlation between y^+ and infrared excess for some HII regions which had been observed in the far-infrared. Alternatively, a deficiency in He-ionizing photons in the spectra of the early-type stars ionizing the nebula, or ionization by more stars of later spectral types, could also make the He^+ Strömgren sphere smaller than the H^+ Strömgren sphere (Balick and Sneden 1976, Panagia and Smith 1978).

This "geometric effect" hypothesis, viz, that He^+ Strömgren spheres smaller than those for H^+ are responsible for the large variations in He abundance inferred from radio recombination line measurements has been reaffirmed and elaborated upon by several authors: Mezger, Smith and Churchwell (1974), Krügel (1975), Smith and Mezger (1976), Mezger and Smith (1976), Panagia and Smith (1978), Churchwell, Smith, Mathis, Mezger and Hutchmeier (1978), Emerson and Jennings (1978), Chaisson, Lichten and Rodriguez (1978), Thum, Mezger, Pankonin and Schraml (1978), and Thum, Mezger and Pankonin (1980). Lockman and Brown (1978, 1982) on the other hand, have argued that there is no "geometric effect", and the variations in apparent He^+ abundance are likely to be due to instrumental effects or data reduction techniques (cf. also Chaisson et al. (1978)).

If the "geometric effect" hypothesis is correct, whether it is due to strong selective absorption by dust, or due to a fall-off in the far-ultraviolet spectra of stars ionizing the gas, then it has consequences of major astrophysical importance for the ionization structure of HII regions, emission line intensities, and inferred elemental abundances. Because of the central importance of this idea and its increasing, although still-debated currency, it seems worthwhile to re-examine the radio recombination line data to investigate quantitatively the extent to which the observations are, themselves, consistent with the "geometric effect" hypothesis.

A.2) Beamsize Effects in the He^+/H^+ Observations

Churchwell et al. (1974) reported not only their own extensive observations on the He^+/H^+ line ratios in HII regions, but also collected together other radio recombination line ratios in the literature. Most of this data was obtained with half-power beam widths

of 4 and 6 arcmin, using chiefly the 109α lines. Additional measurements, and corrections to some data have been given by Churchwell et al. (1978), and the results they present are no doubt the most reliable body of data on He^+/H^+ line ratios presently available. Both these papers also list the half-power angular sizes of each HII region at 5 GHz, and one-third of the 91 separate He^+/H^+ ratios they tabulate were obtained with from measurements with telescope half-power beamwidth less than the half-power angular size of the HII region.

These small-beam measurements may be used to test whether the observations themselves are consistent with the "geometric effect" hypothesis. If the hypothesis is indeed valid, and the He^+ Strömgren spheres are smaller than those for H^+, then the He^+/H^+ line ratios for the small-beam measurements should be systematically larger than the line ratios obtained when the telescope beamwidth was larger than the HII region. For the 31 small-beam measurements tabulated in Churchwell et al. (1978), the average $\langle\theta_b/\theta_s\rangle = 0.75$, where θ_b and θ_s are the half-power angular sizes of the telescope beam and source (HII region) respectively. Therefore, if there is a "geometric effect", the average apparent He^+ abundance $\langle y^+\rangle_{small}$ for the small-beam measurements should be significantly larger than $\langle y^+\rangle_{large}$ for the large-beam measurements.

We have computed the weighted means of the apparent He^+/H^+ ratios for the small-beam and large-beam observations separately using the formula

$$\langle y^+\rangle = \frac{\Sigma y_i^+/\sigma_i}{\Sigma 1/\sigma_i} \qquad \text{(A-1)}$$

where y_i^+ and σ_i are the apparent abundance ratios and their errors tabulated in Churchwell et al. (1978). Following these authors, we

have taken weighting proportional to $1/\sigma_i$ as most appropriate, since the errors are expected to arise as much from systematic effects as from random noise. Using this data we find:

For telescope beam-size < source size : $\langle y^+ \rangle_{small} = 7.8\% \pm 1.9\%$

For telescope beam-size ≥ source size : $\langle y^+ \rangle_{large} = 7.8\% \pm 2.1\%$

Thus, in direct contradiction to what the "geometric effect" hypothesis predicts, there is no significant increase in the average He^+/H^+ line intensity ratio for those measurements in which the telescope beam is smaller than the HII region. In fact, if this calculation is repeated for the original data of Churchwell et al. (1974), on which the "geometric effect" hypothesis was based, the result is the same; the difference between the average He^+ abundance for the small-beam and large-beam observations is insignificant:

$$\langle y^+ \rangle_{small} = 8.2\% \pm 1.9\%$$

$$\langle y^+ \rangle_{large} = 8.3\% \pm 2.3\%$$

We conclude that there is no statistically significant evidence for a "geometrical effect" in the recombination line data itself. If there is any tendency for the He^+ Strömgren spheres to be smaller than those for H^+, then it must produce variations in apparent He^+ abundance less than the standard errors in the means calculated above, i.e. $\Delta y^+ < 2\%$.

A.3) Discussion

The "geometric effect" hypothesis was invoked to account for apparent variations in He^+/H^+ abundances in galactic HII regions from zero to ~10%. The analysis presented above shows that there is no

evidence for this effect in the data. In fact, a "geometric effect" sufficiently large to account for the reported variations is inconsistent with the radio data itself.

We must therefore conclude that the radio recombination line data presented to date are of insufficient reliability for measuring He abundances and their gradients in the Galaxy (cf. Churchwell et al. 1978, Thum et al. 1980). Moreover, use of this recombination line data to infer either strong selective absorption by interstellar dust in the far-ultraviolet (cf. Mezger et al. 1974) or a substantial deficiency of He-ionizing photons in the spectra of early-type stars (cf. Balick and Sneden 1976, Panagia and Smith 1978) is certainly unwarranted. Whether there is a weaker geometrical effect, similar to some published models of dusty HII regions (cf. Sarazin 1977) which might produce apparent variations $\Delta y^+ < 2\%$ is not so clear. However it does seem certain that the best recombination line data available at present are far from being sufficiently precise to test for it.

The reasons why the He^+/H^+ line ratios do vary so widely are not so clear. Lockman and Brown (1978) show that some of the commonly-used approximations in reducing radio-recombination line data, and especially the assumption that the line is formed in local thermal equilibrium, may be dominated by systematic effects. These can produce trends which appear, erroneously, to be physically or astronomically significant. Chaisson et al. (1978) call attention to the systematic effects which may be introduced by the technique chosen for removing baseline curvature from a recombination line spectrum. Lockman and Brown (1982) have recently made observations of both the 76α and 85α transitions in eight of these HII regions, and they find that the apparent He^+ abundance seems to vary with the inverse of the principal quantum number of the transition.

Whether or not these or other effects are responsible for the large variations in apparent He^+ abundance which characterize the large body of radio recombination data presently available, it is clear that any "geometric effect" present is too small to account for these variations. If the initial promise of the radio recombination line technique for measuring He^+ abundances in the Galaxy is to be realized, it is evident that formidable and poorly understood observational complexities must be resolved before reliable He^+ abundances can be obtained from radio recombination line observations.

Acknowledgements

We should like to thank Dick Jennings, Jim Emerson, and Michael Rowan-Robinson for helpful discussions on this problem. S.A.M. was supported by a U.K. SERC studentship.

References

Balick, B., and Sneden, C., 1976, Ap. J., 208, 336.

Chaisson, E.J., Lichten, S.M., Rodriguez, L.F., 1978, Ap. J., 221, 810.

Churchwell, E., Mezger, P.G., and Hutchmeier, W., 1974, Astr. Ap., 32, 283.

Churchwell, E., Smith, L.F., Mathis, J., Mezger, P.G., and Hutchmeier, W., 1978, Astr. Ap., 70, 719.

Emerson J.P., and Jennings, R.E., 1978, Astr. Ap., 69, 129.

Krügel, E., 1975, Astr. Ap., 38, 129.

Lockman, F.J., and Brown, R.L., 1978, Ap. J., 222, 153.

Lockman, F.J., and Brown, R.L., 1982, Ap. J., 259, 595.

Mezger, P.G., Smith, L.F., and Churchwell, E., 1974, Astr. Ap., 32, 269.

Mezger, P.G., and Smith, L.F., 1976, Astr. Ap., 47, 143.

Panagia, N., and Smith, L.F., 1978, Astr. Ap., 62, 277.

Sarazin, C.L., 1977, Ap. J., 211, 772.

Smith, L.F., and Mezger, P.G., 1976, Astr., Ap., 53, 165.

Thum, C., Mezger, P.G., Pankonin, V., and Schraml, J., 1978, Astr. Ap., 64, L17.

Thum, C., Mezger, P.G., and Pankonin, V., 1980, Astr. Ap., 87, 269.