15/9/78

HIGH TIME RESOLUTION STUDIES OF THE HYDROXYL AIRGLOW

A Thesis for the Degree of Doctor of Philosophy

in the University of Adelaide

submitted by

Robert William Basedow B.Sc. (Hons)

The Mawson Institute for Antarctic Research

September 1977

Awarded March 1978

CONTENTS

SUMMARY

A 50 mm scanning Fabry-Perot spectrometer constructed at the Mawson Institute for Antarctic Research, University of Adelaide, has been used to obtain rapid 1.0 nm resolution spectra of the (8 - 3) vibration-rotation band of the hydroxyl night airglow at λ733 nm. The temperature of the neutral atmosphere near 85 km is inferred from the distribution of the relative intensities of the lines within the band.

The theory of the Fabry-Perot interferometer is reviewed, consideration being given to the optimum choice of design parameters for the instrument and to the most suitable hydroxyl band for observation.

The construction and operation of the spectrometer are described in detail. Servo-control of the plate separation and parallelism is not used. The results of tests of the piezoelectric elements which form the basis for the scanning system are presented, with emphasis on finding a configuration which conferred suitable linearity and stability. Methods used to maximize the signal-to-noise ratio, including cooling the detector, quantum efficiency enhancement devices and pulse counting are described also.

Rotational temperatures are extracted from the raw data by fitting a nonlinear model of the recorded spectrum by least squares methods. Included in this model is an empirically derived representation of the effect of atmospheric water vapour absorption on the measured background spectrum. The analysis also yields estimates of the relative intensities and doublet ratio temperatures and the statistical uncertainties of these quantities.

Observations have been made at Mt. Torrens (34° S, 139° E) on

fourteen clear nights spanning eighteen months, from 14 April, 1975 to 11 November, 1976. Rotational temperatures were obtained every 10 minutes with an accuracy of ±2 to 5%. The fit to these spectra is not ideal and the results are initially examined to determine the errors incurred. However generally good agreement is found with the results of other observers, and the parameter estimates are considered to have little if any bias. The nature of the deficiency in the model remains obscure.

The mean nocturnal hydroxyl band intensity, doublet ratio and rotational temperatures are found to decrease from dusk till the middle of the night, and then to increase until just prior to dawn, when the temperatures again begin to decrease. The variation of the first two quantities lags that of the rotational temperature by about 2 hours and 0.5 hours respectively. A small intensity peak near local midnight was also present on several nights. The behaviour is considered briefly in terms of changes in the eddy diffusion coefficient and solar atmospheric tides. Power spectral density functions are calculated for each of the above three quantities and show narrow features of the type expected of gravity waves. Evidence in these spectra that the relationship between the relative change in temperature and intensity is not constant is explained as being due to changes in the degree of quenching of excited hydroxyl molecules by oxygen atoms.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to the best of the author's knowledge and belief, it contains no material previously published or written by another person, except when due reference is made in the text.

(R.W. Basedow)

ACKNOWLEDGEMENTS

I wish to express my appreciation to the personnel of the Mawson Institute for Antarctic Research, University of Adelaide, for their support and cooperation during the course of this work. I am particularly indebted to the Institute's Director, Dr. F. Jacka, for initially suggesting the research project and making available the facilities of the Institute, and for helpful criticism, advice and guidance throughout.

I wish also to thank Mr. D. Creighton, with whom were had many hours of instruction and discussion on the technical and engineering aspects of the project, and who was responsible for the design and construction of much of the associated electronics. His expertise and interest have been of immense value. Thanks are due too to Dr. P.W. Seymour, who undertook the duties of supervisor during the absence of Dr. Jacka.

A number of others have contributed to various aspects of the work. Mr. F. Koltai and Mr. F. Fone cheerfully assisted in the workshop, Mr. G. Brimble devoted considerable time to programming the minicomputer, Dr. P. Wilksch advised in matters of the theory of Fabry-Perots and Fourier transforms, and measurements of the piezoelectric elements were conducted in cooperation with Mr. T. Cocks. The thesis was typed by Mrs. D. Darwent. Their ready assistance is acknowledged with thanks.

Finally the support, encouragement and confidence of my wife Elizabeth contributed significantly to the completion of this undertaking, and are deeply appreciated.

The work was supported in part by a Commonwealth Postgraduate Scholarship.

LIST OF TABLES

CHAPTER 1

THE HYDROXYL AIRGLOW AND ATMOSPHERIC TEMPERATURES

1.1 Introduction

Minor constituents including the hydroxyl molecule play an important role in the photochemistry of the mesosphere and lower thermosphere. In the past, studies of the hydroxyl airglow have provided useful checks of the aeronomic models of the region. However, improvements in instrumentation have allowed increasingly high time resolution measurements to be carried out, and have shown that the emission can be used to study dynamic processes of the neutral atmosphere such as gravity waves and the diurnal temperature variation. In regard to the latter the hydroxyl airglow provides a particularly convenient tool, since other methods rely on incoherent radar scatter, rocket flights or the measurement of Doppler line widths using very high resolution spectrometers. In this chapter the relationship between the spectroscopic temperature and that of the neutral atmosphere at the emission height will be discussed. The spectroscopy and characteristics of the hydroxyl emission will also be briefly presented and the initial objectives of the research project outlined.

1.2 The Hydroxyl Airglow Emission - a Résumé

Emissions from the hydroxyl (OH) molecule constitute one of the dominant features of the night airglow, being observed to extend from 382 nm to 4470 nm in 45 bands, which are often called Meinel bands after the man who first resolved their structure in the airglow so as to enable their identification. These bands increase rapidly in brightness towards the infra-red, and have a total intensity of about

1 to 1.5 MR[+]. Rocket flights have shown that the emission in the atmosphere is confined to a layer between 6 and 20 km thick, centred on an altitude of about 85 km - that is, in the region of the mesopause. Observations of the behaviour of the emission intensity and spectral temperature, as well as discussion of the aeronomic processes involved, have been reviewed by Chamberlain (1961), Wallace (1962), Shefov (1972) and Vallance Jones (1973).

The radiation arises from transitions between the vibrational-rotational energy levels of the hydroxyl $X^2\pi$ electronic ground state. Each band is identified by the notation (v' - v"), where the vs are the vibrational quantum numbers associated with the transition, and the prime and double prime denote upper and lower state respectively. Each band consists of the characteristic three branches corresponding to the selection rule $J' - J'' = \Delta J = +1, 0$ or -1, where J is the rotational quantum number, and the branches are denoted R, Q and P respectively. Each rotational line is in fact a spin doublet, the ground $^2\pi_{3/2}$ state lying about 138 cm^{-1} below the $^2\pi_{1/2}$ state. The subscripts 1 and 2 are used to distinguish the members of the doublet. A further splitting of the spin doublets into lamda doublets, which have a separation of about 0.4 cm^{-1}, has not yet been observed in the airglow. Since the interaction of rotational and electronic motion is intermediate between Hund's case (a) and case (b) coupling (Herzberg, 1950), the rotational energy levels are characterized by the quantum number K, where $K = J - \frac{1}{2}$ for the $^2\pi_{3/2}$ state and $K = J + \frac{1}{2}$ for the $^2\pi_{1/2}$ state. Thus a transition from a $^2\pi_{3/2}$ state with quantum numbers v', K' to one with quantum

---

+Throughout this work the unit of radiance used will be the rayleigh, symbol R, which is equivalent to $(10^{10}/4\pi)$ photon $m^{-2}s^{-1}sr^{-1}$.

numbers v", K" = K' + 1 is the $P_1(K")$ line of the (v'- v") band. This is illustrated in the energy level diagram of Figure 1.1.

The intensity in photon $s^{-1}m^{-3}$ of a spectral emission line from a radiative transition between two states m and n is given by

$$I_{n \to m} = N_n A_{n \to m} \qquad \ldots(1.1)$$

where N_n is the population density of the initial state and $A_{n \to m}$ is the Einstein transition probability for spontaneous emission. If it is assumed that the initial state is in thermal equilibrium, N_n can be described by a Boltzmann distribution, and allowing for the (2J + 1)-fold degeneracy of the state, the intensity per unit volume of a line can be written as

$$I_{v'J' \to v"J"} = \frac{N_{v'}}{Q_{v'}} (2J' + 1) A_{v'J' \to v"J"} \exp \frac{-F_{v'J'}hc}{kT} \qquad \ldots(1.2)$$

where $F_{v'J'}hc$ is the energy of the upper level relative to the v = 0, J = ½ level, T is the rotational temperature and k is the Boltzmann constant. $N_{v'}$, the total population density of the upper vibrational level and $Q_{v'}$, the rotational state sum, which is a weak function of temperature, are usually considered to be constant for the lines of a branch.

In the absence of a knowledge of its value, the Einstein coefficient has been approximated by

$$A_{v'J' \to v"J"} = \text{Const.}\ \nu^3\ S_{j"}\ F_{v'J' \to v"J"} \qquad \ldots(1.3)$$

where ν is the wavenumber of the emission line and S and F, called the line strength and the Herman Wallis factor, together represent that part of the transition matrix element which is dependent on J. In this work the Einstein coefficients of Mies (1974) have been used, whereas

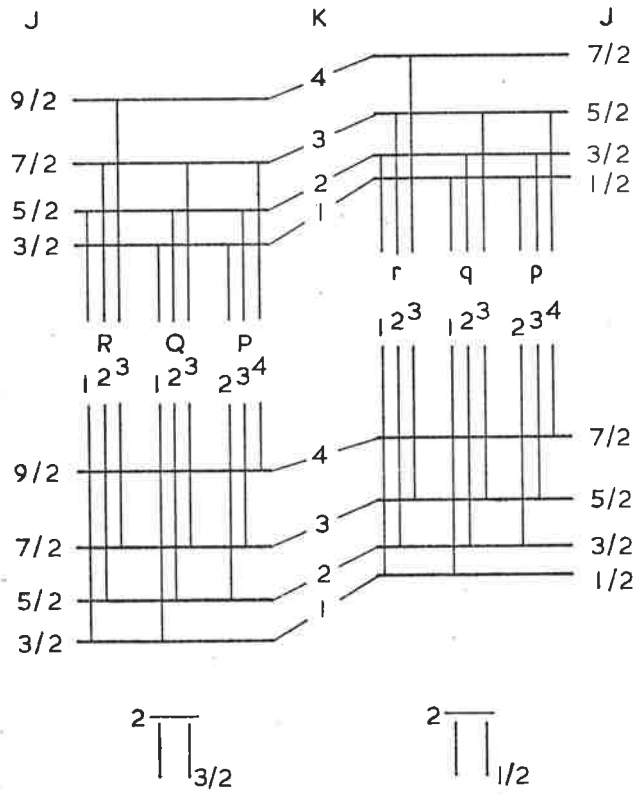


Figure 1.1 Rotational energy level diagram for a band of the hydroxyl spectrum.

most other authors have used line strengths calculated by Benedict et al. (1952) and an $\bar{F}$ factor of unity. The term values $F_{v'J'}$ used have been calculated from the formulae of Hill and van Vleck (1928) together with the spectroscopic constants tabulated by Wallace (1960). The values of the Einstein coefficients and term values, henceforth denoted simply as $A(J)$ and $F(J)$, relevant to this work appear in the Appendix.

Figure 1.2 illustrates the typical band structure with a spectrum of the (8 - 3) band at high resolution calculated from equation (1.2) for a rotational temperature of 200 K. The progression of rotational lines has been truncated at K = 5, and for clarity the weak Q_2 and R_2 branches and lamda-type doubling have been omitted.

It happens that the relative intensities of the airglow OH bands can also be described approximately using a Boltzmann distribution. However, the distribution is not characterized by a meaningful kinetic temperature, the value obtained being about 8000 K on average. This is so because of the way in which the vibrational levels are initially populated, because redistribution of the population across the v levels by collisons takes place relatively slowly, and because the rate constants for quenching reactions are v-dependent.

It is generally considered that the excited OH molecules are primarily formed by the reaction

$$H + O_3 \rightarrow OH^*(v \lesssim 9) + O_2 \ (+3.32 \text{ eV}) \qquad \ldots(1.4)$$

first proposed by Bates and Nicolet (1950) and Herzberg (1951). Recycling of the hydrogen atoms is then effected principally through the reaction

$$OH + O \rightarrow H + O_2 \ . \qquad \ldots(1.5)$$

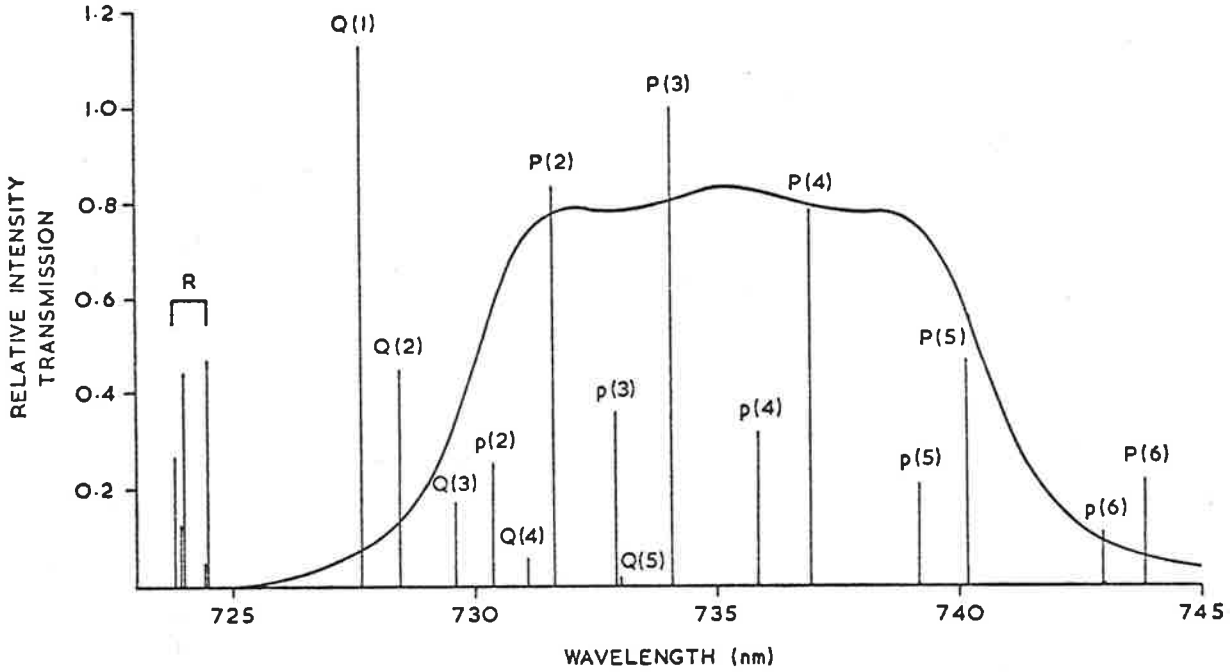


Figure 1.2 High resolution spectrum of the OH (8-3) band calculated for a temperature of 200K and normalized to the $P_1(3)$ line intensity. Superimposed is the transmission profile of the FPI prefilter measured with a grating monochromator at 1 nm resolution and in a 5^{o} beam.

A significant feature of this mechanism is that it provides an explanation for the absence of airglow bands with $v \gtrsim 10$, since the energy of the v = 9 level is 3.23 eV, whereas that of the v = 10 level is 3.49 eV. Other reaction schemes have also been suggested. Krassovsky (1956) has proposed

$$O_2^*(v \lesssim 27) + H \rightarrow OH^* + O \qquad \ldots(1.6)$$

in which the OH would be excited up to the ninth level at the emission altitude. However, because of the criticisms of this mechanism by Bates and Moiseiwitsch (1956) it is no longer thought to significantly contribute to the OH* population. Breig (1970) and Nicolet (1970) have found that another process which could be important is

$$H + O_2 + M \rightarrow HO_2 + M \qquad \ldots(1.7)$$

$$O + HO_2 \rightarrow OH^* + O_2 \qquad \ldots(1.8)$$

in which the OH* is expected to be formed in levels $v \lesssim 6$.

Models of the complex photochemistry of the mesosphere and lower thermosphere have been constructed by a number of workers including Gattinger (1971), Hesstvedt (1970), Hunt (1971) and Shimazaki and Laird (1970), and a review of the topic has been published by Strobel (1972). Considerable effort has been directed towards understanding and predicting the altitude profiles of the concentrations and emission rates of various reacting species, since a comparison between these calculations and observations provides a sensitive test of the accuracy of the model. In the case of the OH emission reasonable agreement with rocket measurements has been obtained by assuming that the ozone reaction is the only source of excited hydroxyl radicals, and by modelling a hydrogen-oxygen atmosphere in which the effects of the diurnal variation in solar illumination and molecular and eddy diffusion have been taken

into account. However, precise comparisons cannot be made because there is considerable variability evident in the measured profiles. For example the layer half width and peak emission height have been measured to be respectively 4.5 and 78 km (Tarasova, 1963), 20 and 83 km (Packer, 1961) and 21 and 97 km (Baker and Waddoups, 1967). These and other altitude profile results are usefully summarized by Vallance Jones (1973).

1.3 Interpreting the Hydroxyl Band Spectrum

1.3.1 Methods of Determining the Rotational Temperature

Once the assumption has been made that the population of the vibrational levels of the excited OH molecule can be described by a Boltzmann distribution (Section (1.3.2)), and hence that the line intensities within an emission band are described by equation (1.2), several methods are available for determining the corresponding temperature.

(a) The approach most commonly adopted is the graphical or "log plot" method. From equation (1.2) it can be seen that the slope of $\ln\left[\frac{I}{(2J+1)A}\right]$ versus $F(J)$ is equal to $-\frac{hc}{kT}$. Thus the rotational temperature may be readily determined using the relative intensities of two or more lines, all of which arise from either the $^2\pi_{3/2}$ or $^2\pi_{1/2}$ levels of a given band. Because they are fairly easily resolved compared with the lines from the Q and R branches, and are considerably brighter than the P_2 lines, the first four or five P_1 lines are best suited to this method. The temperature derived is quite sensitive to errors in the measurement of the relative line intensities, making an accurate determination of the background necessary.

(b) Since the selection rule $\Delta S = 0$ prohibits radiative transitions between the $^2\pi_{3/2}$ and $^2\pi_{1/2}$ spin doublet levels, the ratio $\frac{N}{Q}$ which is the

constant in equation (1.2) will be different for P_1 and P_2 lines, and log plots for these branches will give two parallel lines separated by $\ln\left(\frac{N_1 Q_2}{N_2 Q_1}\right)$. However, if the populations of the two spin states are part of the same Boltzmann distribution, either because of the nature of the excitation mechanism or because of collisional redistribution, $\frac{N_1}{Q_1} = \frac{N_2}{Q_2}$. In this case

$$\frac{I_{P_2(K")}}{I_{P_1(K")}} = \frac{(2K'+1)A_{2,v'J'\rightarrow v"J"}}{2(K'+1)A_{1,v'J'\rightarrow v"J"}} \exp\left[-\frac{hc}{kT_D}\left(F_{2,v'J'} - F_{1,v'J'}\right)\right] \qquad \ldots(1.9)$$

where T_D is the doublet ratio temperature (see for example Harrison and Kendall, 1973a).

(c) A method which does not rely on an accurate determination of the background has been employed by Shagaev (1974). He used the expression

$$f(T) = \frac{I_i - I_j}{I_k - I_j} \qquad \ldots(1.10)$$

which from equation (1.2) can be seen to be, for any three lines i, j and k of a given band, independent of background, providing that this is the same for all three lines.

(d) Because the Einstein coefficients of equation (1.2) vary from line to line in a band, the total intensities of any two branches of a band are different functions of temperature. This allows a determination of temperature from the ratio of the total intensities of two branches in which the rotational structure is unresolved (Meinel, 1950), providing that the factor $\frac{N}{Q}$ is the same for both. Takahashi et al. (1974) have shown that, in the case of the (8 - 3) band, the ratio $\frac{I_Q}{I_R}$ is the most sensitive function of temperature.

(e) When the rotational structure of a band is incompletely resolved, the measured spectrum may be compared with a set of synthetic

spectra, calculated at different temperatures and suitably scaled (Shemansky and Vallance Jones, 1961; Noxon, 1964; Harrison, 1969). The synthetic spectrum is calculated by convolving the instrument response to a delta function with a source spectrum given by equation (1.2). Such a straightforward approach is only possible if the relative intensity and spectral distribution of any background contribution is known. As well it usually involves the assumption that the spin doublet states are in thermal equilibrium.

1.3.2 The Rotational Temperature as a Measure of Atmospheric Temperature

Deductions made about the chemical and dynamic processes of the atmosphere on the basis of hydroxyl rotational temperatures frequently rely heavily on the assumption that such temperatures reflect more or less accurately the kinetic temperature of the atmosphere at the emission height. Evidence to support this assumption can be divided into three broad categories, namely

(a) the excited state lifetimes,

(b) the shape of the observed spectra, and

(c) the temperature values obtained.

Each of these will be considered briefly in turn.

(a) The number of collisions undergone by an excited OH molecule before it is deactivated is given by the product of the collision frequency and the lifetime of the excited state. When deactivation occurs as a result of both radiative transitions and quenching by other molecular species, the effective lifetime τ of the vibrational level v' is given by

$$\tau^{-1} = \sum_{v''=0}^{v'-1} A_{v' \to v''} + \sum_{m} k_{Mv'} [M] \qquad \ldots(1.11)$$

where $A_{v' \to v''}$ is as defined in Section 1.2, $k_{Mv'}$ is the rate constant

The corrected radiative lifetime published by Benedict and Plyler (1954) is 0.03 s, not 0.104 s. The number of collisions per lifetime from the figures of Potter et al is ~ 830, not 700.

One of the examiners of this thesis has pointed out the existence of more recent measurements of the rate constants for quenching of OH* by O_2 and N_2 [Streit, G.E. and Johnston, H.S. (1976) Reactions and quenching of vibrationally excited hydroxyl radicals. *J.Chem. Phys.* *64*, *95-103*]. For the v' = 9 band at 200K these are 1.3×10^{-13} and 1.6×10^{-13} cm^3 mol^{-1} s^{-1} respectively. However, the revised lower and upper estimates for the number of collisions per OH* (v'=9) lifetime, when calculated at 85 km where $[O_2]$ and $[N_2]$ are greater, are ~24 and ~260 - not significantly different from the figures quoted in the text.

for quenching by the species M, and the square brackets denote concentration. The radiative lifetime has been measured or calculated by a number of workers. When normalized to the $1 \rightarrow 0$ transition by means of Murphy's (1971) relative transition probabilities the estimates are 1.0 s (Potter et al., 1971), 0.118 s (d'Incan et al., 1971), 0.104 s (Benedict and Plyler, 1954) and 0.05 s (Mies, 1974). Worley et al. (1972) and Kaufman (1964) have measured the quenching rate constants of OH by air (nitrogen and oxygen molecules) and oxygen atoms, which when taken with data from the CIRA (1972) for the concentrations of these species, give estimated quenching lifetimes at 90 km of 2.8 s and 0.1 s respectively. If a collision frequency at the same height of 8×10^3 s^{-1} (Nicholls, 1971) were adopted, these figures would give between 26 (based on $\sum_v A_{9 \rightarrow v}$ from Mies) and 700 (based on $A_{1 \rightarrow 0}$ from Potter et al.) collisions per lifetime for the hydroxyl molecule. The number of collisions required to establish thermal equilibrium is uncertain, depending as it does on the initial population distribution. However, the calculations by Nicholls et al. (1972) of the effective emitting population of the eighth vibrational level, assuming an initial v' = 9 population confined to the seventh rotational level, suggest that even 10^2 collisions per lifetime may be insufficient (see following).

(b) Evidence of non-thermal behaviour obtained from the spectral distribution of the hydroxyl emission is also somewhat ambiguous. If the $^2\pi_{3/2}$ state population is not in thermal equilibrium the logarithmic plot described in section (1.3.1) will not be a straight line. Shefov (1961) has found no measurable deviation from a straight line for $K \leq 5$ in spectra which presumably included bands with upper vibrational level v' = 9. Sivjee et al. (1972), on all but a few occasions, likewise found no deviation up to K = 5 in the (8 - 3) band. The exceptions occurred at high K values, and were attributed to a particularly broad non-isothermal

hydroxyl emission layer and steep atmospheric temperature gradient. This effect has been considered by Nicholls (1971) among others, and has been found to be generally negligible for $K \lesssim 6$. On the other hand Harrison et al. (1971), by comparing low resolution recordings of the (4 - 1) and (5 - 2) bands with synthetic spectra, have found a persistent "high temperature tail" in the $^2\pi_{3/2}$ level population. They measured the rotational temperature using pairs of P_1 lines [$P_1(2)$, $P_1(K)$] for K = 3 to K = 6 and obtained average (winter) temperatures of 211, 224, 259 and 287 K respectively. Nicholls (1971) has found a similar effect in the higher resolution spectra of Broadfoot and Kendall (1968). From the data published by Shefov and Sivjee et al. it is difficult to see how an effect of this magnitude could remain undiscovered. Nicholls has considered the effects of simultaneous radiative and collisional relaxation on the spectra obtained from the hydroxyl molecule. Using Potter's absolute transition probabilities and initial non-Boltzmann populations in the eighth and ninth vibrational level as suggested by laboratory experiments (Velculescu, 1970), he was able to obtain apparent emitting populations for the v' = 8 level in very good agreement with Broadfoot and Kendall's (8 - 3) and (8 - 4) band spectra.

Wallace (1961) showed that, neglecting collisions during the radiative lifetime, the process of radiative cascading would result in a marked increase in rotational temperature with decreasing vibrational level. He took the absence of such behaviour in the airglow to be evidence that the radiative lifetime was long enough to allow effective collisional redistribution of the populations before they re-radiated. Nicholls in fact found that the rate of relaxation, the total number of collisions undergone by a population, and the mean radiative lifetime, all increase with decreasing vibrational level, so that spectra emitted from low levels should appear more nearly isothermal. Although Shefov

In connection with the first sentence of the second paragraph, it may be noted that although Polanyi and Woodall (1972) [Polanyi, J.C and Woodall, K.B. (1972) Mechanism of rotational relaxation. *J. Chem. Phys.* 56, *1563 - 1572.*] have shown that the optical selection rules governing transitions between rotational states are not applicable to collision induced transfers, at least for hydrogen halides, no similar investigation has been carried out with respect to spin doublet levels.

(1961, 1972) has apparently found no indication that this occurs, he has observed that on the average bands with v' = 9 have rotational temperatures 30 to 40 K higher than bands with v' = 3. (This difference could possibly be reduced by about 10 K - see following). Nicholls has suggested that failure to account for a high temperature tail in upper level bands could produce just such an effect.

It is usually assumed (Nicholls, 1971) that collisions will only cause those rotational transitions allowed by the optical selection rules, and in particular that the selection rule forbidding transitions between the spin doublet levels will hold. The $^2\pi_{1/2}$ and $^2\pi_{3/2}$ levels should therefore achieve equilibrium independently, so that in general the doublet ratio temperature should be unrelated to the rotational temperature of the P_1 and P_2 branches. Again airglow observations give conflicting results. Meinel (1950), using the (6 - 2) and (5 - 1) bands, and Blackwell et al. (1960), using the (9 - 3) and (8 - 2) bands, found a doublet ratio temperature more than 90 K lower than T_{Rot}, whereas a careful study by Harrison and Kendall (1973a) of the (4 - 1) band revealed no significant difference. This last result suggests that either the states are initially populated according to a Boltzmann distribution, or that some violation of the spin selection rule occurs, and the radiative lifetime is long enough to allow a redistribution between the two spin states.

Evidence has also been found for the occasional existence of non-thermal populations apparently differing in character from that discussed above. Gattinger and Vallance Jones (1973) cite the appearance in some spectra of a feature at the position of the $P_1(8)$ line of the (2 - 0) band, and which has an intensity well in excess of that predicted by the rotational temperature of an adjacent band. They also point out

anomalies in spectra obtained by Shamansky and Vallance Jones (1961). In addition, apparently inflated K' = 2 level populations have been found by Armstrong (1975 and private communication) in (6 - 1) and (8 - 3) band spectra on occasions when the emission layer was influenced by a large gravity wave.

(c) Groundbased measurements of hydroxyl rotational temperatures do not always appear to coincide well with atmospheric temperatures expected at the height of the hydroxyl emission peak (Takahashi et al., 1974; Vallance Jones, 1973). Meriwether (1975) has calculated the expected temperature difference assuming a broad non-isothermal OH emission layer centred on the mesopause. He showed that the rotational temperature can be between 0.5 and 10 percent high (winter to summer variation), depending on the layer width and the ambient temperature gradient. Taken with data from the CIRA (1972), and considering latitude and seasonal effects, this implies that OH rotational temperatures should fall in the range 140 - 240 K. Most workers have found this to be the case, but some measurements (Krassovsky et al., 1961; Fiocco et al., 1970) exceed this range quite appreciably. The agreement can be improved somewhat by using the Einstein coefficients recently published by Mies (1974) in place of the line strengths of Benedict et al. (1952) which are usually used to derive hydroxyl temperatures. Lower temperatures are obtained thereby, the percentage reduction increasing with increasing temperature, being about 8% at 250 K and 10% at 300 K.

Often quite rapid temperature fluctuations, of between 20 and 50 K, and with a time scale of less than 15 minutes, are observed (Armstrong, 1975; Shagaev, 1974). It is possible to account for their magnitude, at least in part, by a consideration of adiabatic processes (Krassovsky, 1972). The change in emission layer height dz required to give a change in gas temperature dT by adiabatic compression is

$$dz = \frac{\gamma}{1-\gamma} H(z_o) \frac{dT}{T} \qquad \ldots(1.12)$$

where $H(z_o)$ is the pressure scale height and γ the ratio of specific heats at the emission height z_o. A height variation of ±2 Km is therefore sufficient to produce a 25 K temperature change which, superimposed on the maximum expected winter value, could result in measured rotational temperatures close to 270 K. There is also considerable inhomogeneity in the OH airglow layer brightness. Kieffaber and Peterson (1973) and Kieffaber (1974) have measured r.m.s. spatial and temporal fluctuations of 2 - 11 percent and 2 - 8 percent respectively in the (4 - 2) and (5 - 3) band intensities. As the discussion in part (a) of this section suggests, the associated variations in the concentration of the quenching species could contribute to temperature fluctuations, particularly for high v' bands, though in this case a non-Boltzmann emitting population would be expected.

In summary, it would appear that the rotational populations of the vibrationally excited hydroxyl molecule are usually in quite good thermal equilibrium with the surrounding neutral atmosphere. Exceptions to this equilibrium state appear to occur, however. They can take the form of a shift in the population towards higher rotational levels, particularly, it would seem, for higher vibrational levels. There is also evidence that individual rotational lines may become enhanced. The values of the rotational temperatures observed, after correcting the line strength values used, are generally a fair indication of the atmospheric temperature at the height of the OH emission peak, though errors of several percent can be introduced by temperature nonuniformity over the width of the layer. Adiabatic changes in the emission layer height and spatial structure traversing the observer's field of view have been suggested as contributing significantly to the rotational temperature fluctuations.

The second sentence of the second paragraph should read: "Such measurements, together with intensity behaviour for the same period, could lead to a better understanding of the dynamic processes, such as eddy diffusion variations and changes in the concentration of quenching species, which affect the formation of excited hydroxyl molecules."

1.4 Objectives

This investigation of the hydroxyl airglow behaviour was begun with several objectives in mind. In the first place there existed no hydroxyl airglow observations made in southern latitudes, so that measurements at Adelaide (34° S) could represent a useful addition to the existing body of data with respect to global distribution.

In the second place there had been, at the time the research project was initiated, very little study of the behaviour of the OH rotational temperature with time (see however, Taranova, 1967 and Fiocco et al., 1970). Such measurements, together with intensity behaviour for the same period, could lead to a better understanding of the question, as yet unresolved, of the exact mechanism by which excited hydroxyl molecules are formed. It was therefore decided to try to obtain OH rotational temperature measurements at a sampling rate of four per hour or better.

A resolution in time of this order also suggested the possibility of detecting oscillations in the atmosphere, particularly internal gravity waves, which are expected to have a period of between about five minutes and several hours (Hines, 1960). The use of rotational temperatures to observe such phenomena can avoid several difficulties which arise in the interpretation of intensity variations. The latter are very susceptible to changes in atmospheric transmission or instrument sensitivity, whereas such changes are much less likely to distort a temperature value determined from the intensity ratio of two lines only one or two nanometers apart, as is the case for the log plot method. In addition, low resolution branch or band intensity determinations can be easily affected by contaminating emissions, while variations in the intensity of just one line can be simply a reflection of changes

in the distribution of the emitting population of a vibrational level.

As the discussion in Section 1.3.2 shows, it is not certain that the OH rotational level populations are always in thermal equilibrium with the surrounding neutral atmosphere. Since this is of some concern in interpreting both the actual value and any variations of rotational temperature, it was decided to adopt a method of temperature determination which would indicate whether or not equilibrium had been established. This required that the individual lines of a P branch be resolved - a procedure which would also allow an investigation of the doublet ratio temperature. It was considered that these two advantages were more important than the improved sampling rate which would be achieved if the branch intensity ratio method of determining temperature were used.

The above requirements could not be met by any of the spectrometers available to the Mawson Institute, so that a suitable instrument had first to be designed and built. Diffraction grating instruments, in conjunction with both photographic and photoelectric recording methods, have been most frequently used by other workers in the field. A combination of grating spectrograph and image intensifier has been used by Russian workers (Taranova, 1967), permitting temperature determinations as often as every 30 seconds. Tilting filters (Fiocco and Visconti, 1974; Meriwether, 1975) and the field widened Michelson interferometer (Baker et al., 1973) have also been employed.

For this study a 50 mm diameter Fabry-Perot interferometer was chosen. Its construction and use are in principle straightforward in comparison with other types of interferometer or diffraction spectrometer of equal light gathering power. Compared with fixed wavelength interference filters it offers the versatility to observe other bands of the Meinel system, and seemed better suited to the observation of

the several lines required for a determination of rotational population distributions. Since it appeared that, at the time this project was initiated, a Fabry-Perot had not been used for this type of airglow study, its suitability in this regard could be evaluated and compared with other instruments.

Finally, it appeared likely that a 150 mm Fabry-Perot spectrometer under construction at the Mawson Institute would eventually become available to permit simultaneous measurement of [OI] λ557.7 nm Doppler and OH rotational temperatures. Since the oxygen emission arises in the main from a layer centred on about 95 km this offered the possibility of checking to some extent the absolute OH temperature values, and in the longer term, of examining the [OI] and OH temperature behaviour for correlations.

CHAPTER 2

PRINCIPLES AND DESIGN OF A FABRY-PEROT INTERFEROMETER

2.1 Introduction

From the time of Babcock's 1928 measurements of the atomic oxygen line in the aurora, the Fabry-Perot interferometer (henceforth abbreviated to FPI) has frequently been used to study atmospheric emissions. The development by Jacquinot and Dufour (1948) of the instrument in scanning photoelectric form, the appreciation of its superior light gathering power compared to diffraction grating and prism spectrometers (Jacquinot, 1954), and improvements in the technology of producing very flat reflecting surfaces of low absorptance have contributed to its success. Sophisticated variable spacing FPIs have been constructed to obtain neutral atmosphere temperatures and winds from high resolution measurements of the [OI] λ557.7 nm and λ630.00 nm emission lines (e.g. Bower, 1974; Hernandez and Mills, 1973). At low resolution the FPI is familiar as the fixed spacing multilayer dielectric interference filter commonly used for photometry (Eather and Reasoner, 1969). However, tunable low resolution interferometers capable of being scanned over quite large wavelength intervals, for example over an emission band, seem to have been used but little (see however Drummond and Gallagher, 1973; Shepherd, 1960). The basic principles of the FPI are to be found in most textbooks on optics (e.g. Born and Wolf, 1975). Chabbal (1953, 1958), Ramsay (1969), Wilksch (1975) and others have examined the way in which imperfections and physical restraints broaden the instrument response. Chabbal has also discussed the conditions required to maximize the transmitted flux for a given resolving power. A useful review of these topics is presented by Jacquinot (1960).

In this chapter the important theoretical aspects of the FPI will

be briefly presented, and used as a basis for a discussion of the selection of certain parameters of the spectrometer constructed in the course of this research project.

2.2 Summary of Basic Theory

2.2.1 The Ideal Spectrometer

Figure 2.1 illustrates schematically the essentials of an FPI in the configuration used in this study, and commonly employed by others. The etalon itself consists of two optically flat discs (plates) of high quality glass, the adjacent faces of which are the substrates for very uniform evaporated reflective coatings, either metallic or of the multi-layer dielectric type. These surfaces have a high reflectance R, an absorptance A which is as low as possible, and a transmittance T. Their parallelism and separation are precisely maintained. The etalon is shown illuminated by an extended, uniform source of spectral radiance $B(\lambda)$. A monochromatic ray of wavelength λ from this source, incident on the etalon at some angle θ', undergoes multiple reflections between the surfaces, such that a set of parallel rays, each of which is retarded in phase with respect to its predecessor, is transmitted and reflected. The phase difference is

$$\delta = 4\pi \mu \ell \xi / \lambda + 2\gamma \qquad \ldots(2.1)$$

where μ is the refractive index of the medium between the plates,

ℓ is the plate separation,

$\xi = \cos\theta$, where θ is the angle of the ray between the plates, and is related to θ' by Snell's Law, and

γ is the phase change experienced by the wavefront upon reflection. The reflected radiation is seldom of interest (see however Section 3.1.2.2), but the transmitted beam is focussed by a lens, forming a set of

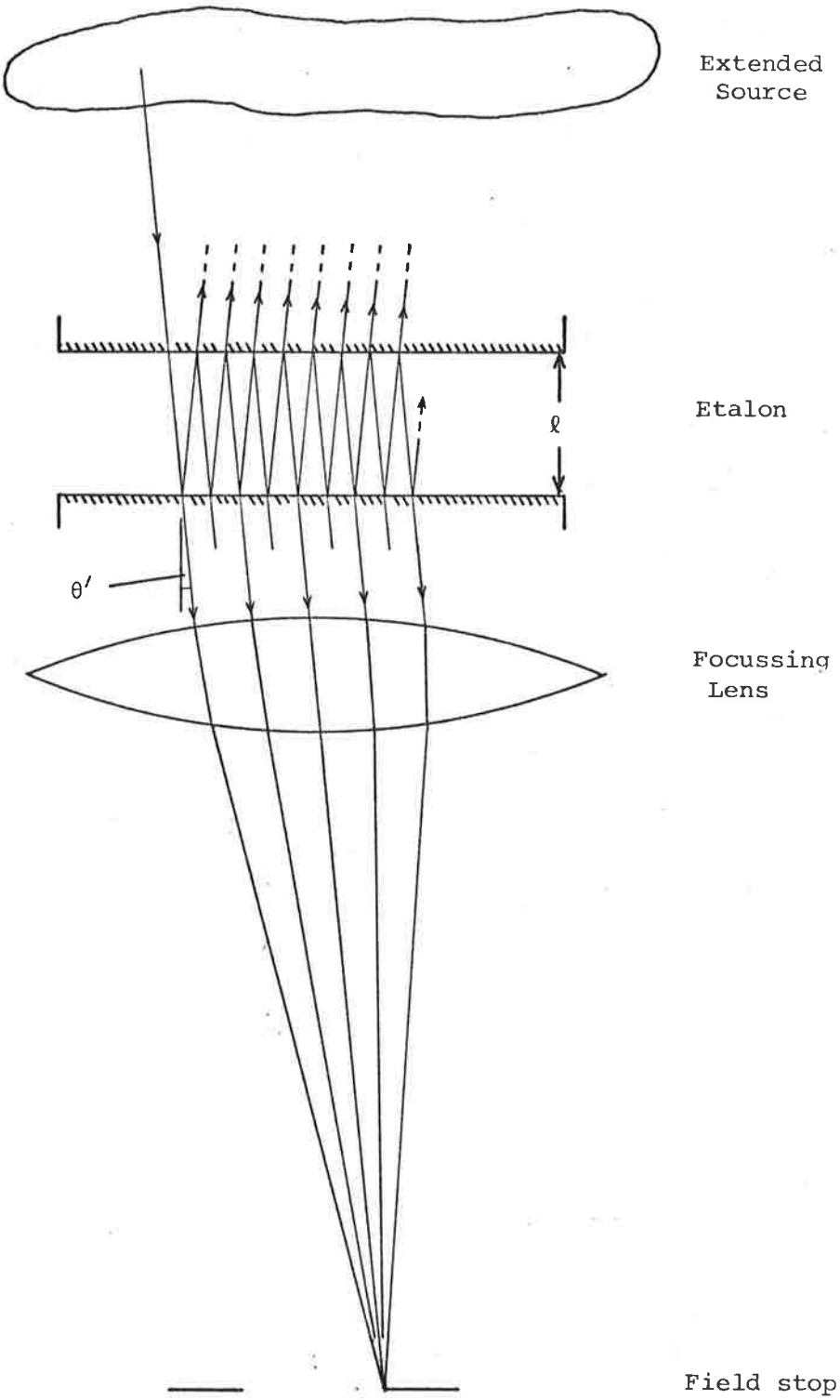


Figure 2.1 Illustration of the principle of a Fabry-Perot spectrometer.

interference rings in the image plane. Since the etalon has axial symmetry these fringes take the form of circles concentric with the optic axis. One of the fringes is selected, and a field of view defined, by a circular aperture placed on axis in front of a detector.

The function which describes the transmission of the perfect etalon is a form of the Airy function

$$A(\delta) = \tau_A \left(1 + \frac{4\underline{R}}{(1-\underline{R})^2} \sin^2 \frac{\delta}{2}\right)^{-1} \qquad \ldots(2.2)$$

where

$$\tau_A = \left(1 - \frac{\underline{A}}{1-\underline{R}}\right)^2 . \qquad \ldots(2.3)$$

It is periodic in δ and its maxima, equal to τ_A, occur when

$$\frac{\delta}{2\pi} = 2\mu\ell\xi/\lambda + \frac{\gamma}{\pi} = m \qquad \ldots(2.4)$$

where m is an integer called the order of interference. The function describing the reflected radiation (when $\underline{A} = 0$) is

$$A_R(\delta) = \frac{4\underline{R}}{(1-\underline{R})^2} \sin^2 \frac{\delta}{2} \left(1 + \frac{4\underline{R}}{(1-\underline{R})^2} \sin^2 \frac{\delta}{2}\right)^{-1} . \qquad \ldots(2.5)$$

Figure 2.2 displays slightly more than one cycle of the Airy function for two sets of reflective coatings, one with $\underline{R} = 0.939$, the other with $\underline{R} = 0.882$, and both having the same absorptance $\underline{A} = 0.01$. The third curve corresponds to an uncoated pair of plates, for which $\underline{R} = 0.04$ and $\underline{A} = 0$. The desirability of using coatings of high $\underline{R}$ and low $\underline{A}$ is evident.

The monochromatic interference fringes described above are transmission maxima, for which, from equation (2.4), the order of interference decreases in steps of one as the angle of incidence increases. When viewing a polychromatic source using a fixed optical spacing $n\ell$ and angle of incidence, the etalon acts as a multiple

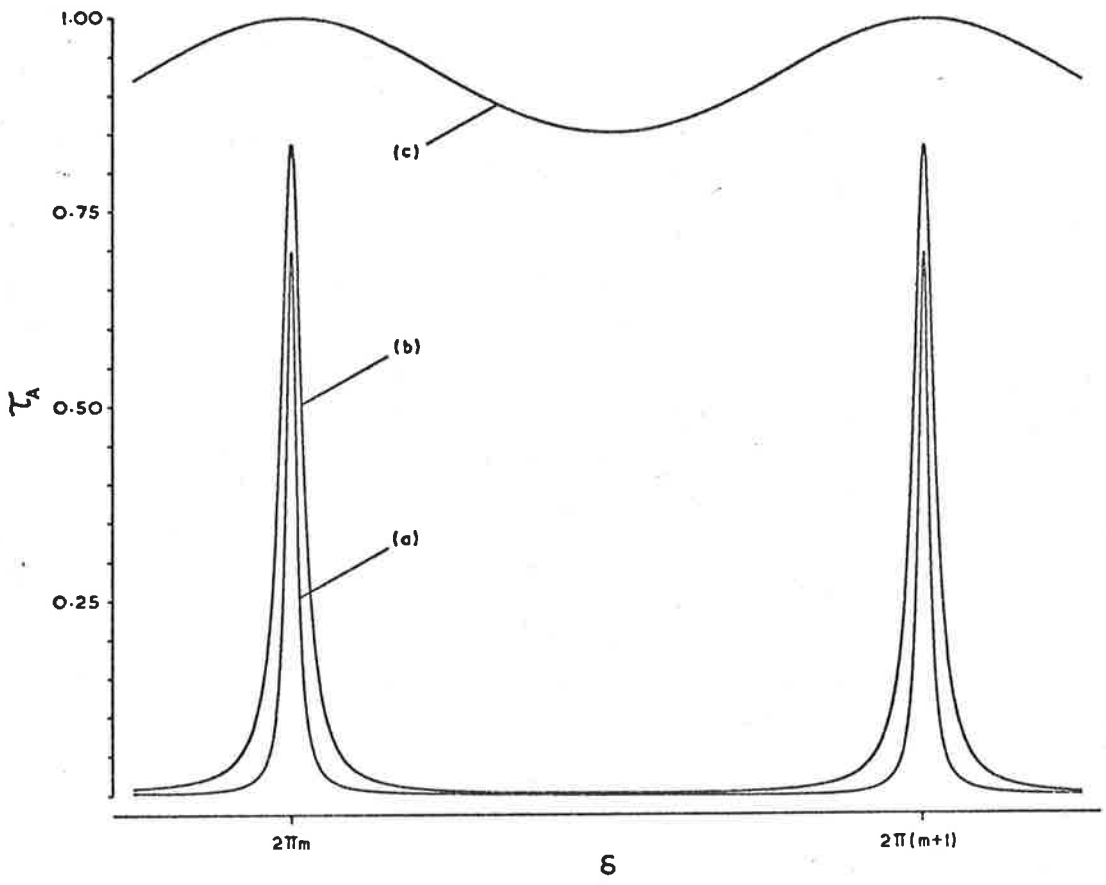


Figure 2.2 The Airy function graphed over a little more than one order for

(a) $\underline{R}$ = 0.939 and $\underline{A}$ = 0.01, giving N_R = 50

(b) $\underline{R}$ = 0.882 and $\underline{A}$ = 0.01, giving N_R = 25

(c) $\underline{R}$ = 0.04 and $\underline{A}$ = 0.0, corresponding to uncoated plates.

passband filter. Adjacent passbands, having wavelengths λ_1 and λ_2 and interference orders m_1 and $m_2 = m_1 + 1$ respectively, are separated by an interval called the free spectral range, which is given by

$$\Delta\lambda = \frac{\lambda_1\lambda_2}{2\mu\ell_1\xi_1}\left(1 + \frac{\Delta\gamma}{\pi}\right) . \qquad \ldots(2.6)$$

Since it is usual to operate an FPI at high values of m, the variable γ and its increment $\Delta\gamma$ in the interval $\Delta\lambda$ can be ignored. Also under these conditions $\lambda_1 \simeq \lambda_2$, so that

$$\Delta\lambda \simeq \frac{\lambda}{m} . \qquad \ldots(2.7)$$

When using the FPI as a spectrometer it is therefore necessary to provide a premonochromator, the bandwidth of which is narrow enough to exclude all significant radiation further than one free spectral range from the wavelength of interest. The interference pattern will then consist of fringes which increase in wavelength as ξ increases, each fringe corresponding to a spectral element in the source, and this pattern will be repeated at intervals of $\Delta\lambda$.

A spectrum could be recorded by photographing the fringe system in the image plane. Alternatively, for the configuration illustrated in Figure 2.1, appropriate to the use of a photoelectric detector, the transmission peak of the Airy function can be swept through the source spectrum by fixing the angle of incidence and varying the optical separation $\mu\ell$ of the plates (Section 2.3.4). The interval $\Delta\lambda$ is traversed by a change in $\mu\ell$ of

$$\Delta(\mu\ell) = \lambda/2\xi . \qquad \ldots(2.8)$$

At normal incidence ($\xi = 1$), $\Delta\lambda$ therefore corresponds to a separation change of half a wavelength. For this condition also the operating order m_o is equal to the optical spacing in units of $\lambda/2$.

The relationship between a, the width at half maximum amplitude of the Airy function, and the separation between adjacent orders is expressed as

$$N_R = \frac{\Delta\lambda}{a} \quad . \qquad \ldots(2.9)$$

The quantity N_R is called the reflective finesse or reflective definition. From equation (2.2) it may be readily shown that

$$N_R = \frac{\pi \sqrt{R}}{(1 - R)} \quad . \qquad \ldots(2.10)$$

Thus N_R depends only on the plate coatings, whereas a depends on the spacing also.

2.2.2 The Instrument Function

The distortion of the spectrum of a source by a spectrometer may in general be written as

$$Y(\lambda) = B * I \qquad \ldots(2.11)$$

where $Y(\lambda)$ is the recorded spectrum, $B(\lambda)$ is as previously defined and $I(\lambda)$ is called the instrument function or profile. The asterisk denotes the convolution operation - that is

$$B * I = \int_{-\infty}^{\infty} B(\lambda' - \lambda) I(\lambda') d\lambda' \quad . \qquad \ldots(2.12)$$

Unless I is narrow compared to the structure in the source spectrum, it is necessary to deconvolve the function Y in some fashion in order to obtain a faithful representation of B. A knowledge of the shape of the instrument function is therefore usually necessary. It may often be determined simply by measuring an isolated narrow spectral line, since if B approximates a delta function the properties of the convolution integral give $Y \simeq I$. In the ideal case discussed thus far, if the scanning aperture has negligible width, Y is also given by the Airy function. In practice, however, two factors contribute to the

broadening of the instrument profile. These are the effect of non-uniformity in plate spacing, arising either because of departures from perfect flatness, variations in phase change γ over the area, or lack of parallelism of the plates, and the effects of using a finite field of view.

The Defect Function.

The contribution of the first broadening effect is determined, following Chabbal (1953), by considering the etalon as consisting of many individual elementary etalons. Expressed as a fraction of the total area S, the area of one such element which has a spacing in the interval $\ell_o + x$ to $\ell_o + x + dx$, is

$$\frac{dS}{S} = D'(x)\, dx \quad . \qquad \ldots(2.13)$$

ℓ_o is the mean spacing of the plates. Equation (2.4) may be used to transform the independent variable to λ, resulting in a function $D(\lambda)$ known as the defect function. It is not in general analytic, but two simple cases have been investigated by Chabbal. He found that small scale randomly distributed departures from flatness result in a Gaussian defect function, while for a spherical curvature error D is rectangular. The transmission function of the etalon, called the etalon function $E(\lambda)$, is now described by

$$E(\lambda) = A(\lambda) * D(\lambda) \qquad \ldots(2.14)$$

and has a peak transmission

$$\tau_E = \tau_A \ \tau_D \quad . \qquad \ldots(2.15)$$

τ_D depends on the widths and forms of A and D, and is always less than one because the effect of surface defects is to broaden the Airy function

peak without altering the mean transmission.[+]

It is convenient to define a defect finesse, by analogy with the reflective finesse, as

$$N_D = \frac{\Delta\lambda}{d} \qquad \ldots(2.16)$$

where d is the width of D. Although $D(\lambda)$ is itself not periodic, N_D may be considered as that finesse which would be obtained for the etalon if the Airy function contribution (i.e. the width a) were negligible. Just as N_R depends only on the coatings, N_D depends only on the plate flatness, whereas the width d is a function of the plate spacing also. It is often referred to as the limiting finesse, because of the difficulty of polishing plates to a high degree of flatness.

As is shown in Section (2.2.4), a knowledge of D is required before an FPI's optimum parameter values (especially the coating reflectance R) can be determined. The function can be obtained exactly by measuring E and knowing A, which requires that the plates be coated. Often, however, it is sufficient to assume that the form of D is rectangular, and that when the maximum departure from flatness is λ/k, the defect finesse is $k/2$, as for a spherical curvature error.

The Aperture Function.

The etalon function is further broadened by the image plane aperture, which allows light from a finite range of incidence angles, and hence wavelengths, to reach the detector. By treatment similar to that used to evaluate the contribution of the surface defects, one obtains

---

[+] Chabbal has defined $A(\lambda)$ such that his τ^E corresponds to τ_D defined in equation (2.15).

(Wilksch, 1975)

$$\frac{d\Omega}{\Omega} = F(\lambda)\, d\lambda \qquad \ldots(2.17)$$

for the fraction of the total solid angle Ω which corresponds to a peak transmission in the interval $d\lambda$. $F(\lambda)$ is rectangular, and is known as the aperture function. The transmission function of the etalon/aperture combination is then

$$I(\lambda) = E(\lambda) * F(\lambda) \qquad \ldots(2.18)$$

and the peak transmission is

$$\tau_I = \tau_E \quad \tau_F \qquad \ldots(2.19)$$

where τ_F is analogous to τ_D.

For small angles of incidence θ, the width f of F is, from (2.4),

$$f = \frac{\lambda_o \Omega}{2\pi\mu^2} \qquad \ldots(2.20)$$

where Ω is the solid angle subtended by the source, and λ_o is the wavelength transmitted when $\theta = 0^\circ$. f is independent of ℓ, so that an aperture finesse defined by

$$N_F = \frac{\Delta\lambda}{f} \qquad \ldots(2.21)$$

does not correspond to N_A and N_D, in that it is not characteristic of the aperture only. For this reason f rather than N_F is often used to specify the properties of the aperture function. It may be noted that $F(\lambda)$ is always asymmetric with respect to λ_o, the wavelength transmitted at the order m_o, since from equation (2.4) λ decreases as θ increases from zero. It is therefore to be expected that $I(\lambda)$ will display some asymmetry also.

The instrument profile is thus the convolution of three functions,

A, D and F, and is in general not readily expressed analytically (see however Hernandez, 1966). Its width i depends on the component widths a, d and f and the forms of A, D and F, a relationship which will be denoted by the operator $\circledast$, thus

$$i = a \circledast d \circledast f \quad . \qquad \ldots(2.22)$$

A convenient approximation may be obtained for i by assuming that all the component functions are Gaussian in shape, in which case

$$i^2 = a^2 + d^2 + f^2 \quad . \qquad \ldots(2.23)$$

However, in the general case graphs and tables published by several authors, including Chabbal (1953) and Hernandez (1970a), must be used to estimate i more accurately. A finesse N_I, called the instrument or overall finesse, can be defined in the same manner as before.

2.2.3 Resolving Power and Light Gathering Power

The resolving power of a spectrometer is usually defined as

$$R = \frac{\lambda_o}{i} \qquad \ldots(2.24)$$

where i and λ_o are as previously defined. For an FPI working at an order of m_o, from equation (2.7)

$$R = m_o N_I \qquad \ldots(2.25)$$

- a result almost identical with that obtained when the Rayleigh criterion for resolution is used.

The light gathering power L (also called luminosity by Jacquinot, 1960) of a spectrometer is defined as the maximum light flux transmitted per unit radiance of the source. For an FPI

$$L = S\Omega \, \tau_C \tau_I \qquad \ldots(2.26)$$

where τ_C is the contribution to the overall peak instrument transmission due to absorption, scattering and reflection by optical components other than the etalon. From equations (2.15) and (2.19)

$$\tau_I = \tau_A \tau_D \tau_F \quad . \qquad \ldots(2.27)$$

It is best determined by direct measurement, but can be estimated by means of the graphs and tabulations referred to earlier. If the quantities N_R and N_D of an interferometer are fixed, and if the conditions for obtaining maximum transmitted flux for a given bandwidth are adhered to (see the following section), Jacquinot (1960) has shown that the product

$$L\mathcal{R} = \text{Constant} \quad . \qquad \ldots(2.28)$$

The necessary compromise between light gathering power and resolving power is discussed further in Section (2.3.3).

2.2.4 Optimizing the Instrument

Equation (2.22) expresses the fact that an instrumental bandwidth i can be obtained with an infinity of combinations of a, d and f. However, Chabbal's (1953) work shows that there exists a unique combination for any given i which results in maximum light gathering power. The choice of instrument parameters which gives such a maximum for the initial conditions pertaining to the present instrument (hereafter referred to as the OH FPI) is briefly discussed below.

The predetermined quantities are the operating wavelength λ_o, the free spectral range $\Delta\lambda$, the resolving power $\mathcal{R}$, the plate area S and the defect finesse N_D. The first two represent a choice of OH emission band and number of P lines to be scanned, while the third is dictated largely by the shape of the source spectrum (Section 2.3). S and N_D are not strictly beyond control in the design stages, but their present treatment

as constants will be justified shortly. According to equation (2.6), the plate separation ℓ_o is also determined once $\Delta\lambda$ and λ_o are known. Remaining to be chosen, then, of the set of possible variables, S, $\mathcal{R}$, N_R, N_D, ℓ_o and f, are only the reflective finesse N_R (that is the width a since ℓ_o is fixed), and the aperture function width f.

Because the hydroxyl P lines are very narrow compared to their spacing, it is appropriate to summarize those conclusions of Chabbal which apply to a monochromatic source. He found that the optimum values of a and f must satisfy the twin requirements of equation (2.22) and

$$a/f = a_o/f_o \quad . \qquad \ldots(2.29)$$

The quantities a_o and f_o are defined by

$$a_o * d = i \qquad \ldots(2.30)$$

and

$$d * f_o = i \qquad \ldots(2.31)$$

- in other words a_o is the Airy function width required to give an instrument width of i when f = 0, and similarly for f_o. Unfortunately there exists no general mathematical rule which allows a and f to be calculated once a/f is known. It may be necessary, therefore, to find a and f by trial and error, using equations (2.22) and (2.29) and tables or graphs such as Chabbal's (1953) Figure 5.5. Alternatively a fair approximation is

$$a/a_o = f/f_o \simeq 0.68 \quad . \qquad \ldots(2.32)$$

f is then set by choosing a field stop diameter appropriate to the focal length of the lens used, and a, since ℓ_o is fixed, is set by specifying the coating reflectance R.

Any improvements in the L$\mathcal{R}$ product can only be obtained by relaxing the initial conditions - that is by increasing N_D and S or decreasing $\Delta\lambda$.

As N_D decreased from infinity (i.e. d increases from zero), τ_D must also decrease from its maximum value of one. It is always best, therefore, to use as large a defect finesse as can be practically obtained. Once the plates have been coated N_D can only be increased by reducing the surface area - a method most suitable when the defect takes the form of a spherical curvature. However the surfaces obtained were always less regular than this, the variation of N_D with S being rather complex, so that improvements in L obtained by masking could not be readily predicted. Of course, L may also be increased by using larger diameter plates, but polishing technology is such that this is usually done only at the expense of surface flatness. For low light level applications it is best to choose S as large as possible, the limitations being expense and the defect finesse required. N_D should be significantly greater than N_R, and $\Delta\lambda$, and hence indirectly N_R, should be made no larger than necessary, since to do so not only keeps τ_A large, but also makes a suitably high value of N_D harder to achieve and maintain.

It is of interest in applications such as this which require a large free spectral range, to know the optimization conditions when ℓ_o is variable. The difference between this and the previous case is that the width of the Airy function may now be altered through ℓ_o as well as $\underline{R}$, so that the quantities τ_E and $\Omega\tau_F$ which determine L may be separately maximized. Chabbal finds that $\Omega\tau_F$ has a broad maximum when f = 1.15 e, and that the properties of the convolution integral are such that if $I = E * F$ and $f = e$, then $f \simeq 0.7\ i$. He also shows that again the lowest possible value of N_R should be used, because both τ_A and τ_D increase as N_R decreases. However a is now fixed by the requirement that $e = a \circledast d \simeq f \simeq 0.7\ i$, d being still an initial condition, so that, from equation (2.9), $\Delta\lambda$ must be reduced (by varying ℓ_o) together with N_R. Chabbal suggests, therefore, that if free spectral range is as

important as light gathering power, the product of the two should be optimized. Although more exact relations are given, he recommends that one should use $N_R \simeq N_D$, and $e = f \simeq 0.7\ i$.

2.3 Selection of Design Parameters

2.3.1 Operating Wavelength

The selection of a suitable operating wavelength was influenced by the characteristics of the emission from different bands, the spectral sensitivity and availability of the detector and the presence of contaminating emission or absorption bands. In regard to the first of these, two circumstances are of importance - the rapid increase in OH band intensities towards the infra-red, and the variation in line spacing from band to band. The detector chosen (Section 3.3.1) had a photocathode of the extended S-20 type. Its quantum efficiency q decreases rapidly with wavelength above λ500 nm, so that the product of line intensity and q has a maximum which corresponds to the (6 - 2) band at λ834 nm. Table 2.1 presents this result. The $P_1(3)$ line intensities I_{OH} were obtained using band intensities from Krassovsky and Shefov (1962) and a Boltzmann population distribution at 225 K (Chamberlain, 1961). Quantum efficiencies were obtained from EMI specifications. The characteristic increase with band head wavelength of the separation between adjacent P lines of a band is sufficiently rapid that the resolving power required of the spectrometer is decreased. Equation (2.28) suggests therefore that it is advantageous in this respect also to choose a band towards the infra-red. However the possible gain in light gathering power is somewhat limited by the increase in free spectral range required, which is such that the overall finesse would remain almost constant.

Of those bands which have high intensity-quantum efficiency

TABLE 2.1

Intensity-Quantum Efficiency Product ($I_{OH}q$) and Expected Signal-to-Noise Ratio (SNR) for the $P_1(3)$ line of several OH bands, assuming an instrument width equal to the $P_2(3)$ - $P_1(3)$ line spacing.

Band	$\lambda P_1(3)$ (nm)	q (%)		I_{OH}(R)	$I_{continuum}$(R/nm)	$I_{OH}q$ (relative)		SNR (relative)	
		S-20	Ext S-20			S-20	Ext S-20	S-20	Ext S-20
9-3	631	5.8	7.6	13	.03	.87	.53	.99	.78
6-1	655	4.8	6.7	12	.05	.67	.43	.82	.66
7-2	692	3.2	5.4	27	.08	1.00	.79	1.02	.91
8-3	734	2.0	4.3	43	.14	1.00	1.00	1.00	1.00
4-0	760	1.4	3.6	31	.22	.50	.60	.60	.66
9-4	782	.8	2.9	81	.29	.76	1.27	.85	1.10
5-1	799	.4	2.5	94	.39	.44	1.27	.62	1.05
6-2	845	<.1	1.6	169	.70	<.20	1.46	.41	1.11
7-3	892	-	.5	244	1.50	-	.66	-	.67

products, two were not selected because of the presence of contaminating radiation. The [OI] λ630 nm and λ636 nm lines overlap the (9 - 3) band, while the P(4) lines of the (9 - 4) band lie very close to the (5 - 1) band R branch, so that little information about the population distribution of the former could be obtained. Other emissions, particularly from N_2 and N_2^+, occur in this spectral region, but are of importance only during auroras, and can usually be disregarded at this latitude (35° S). The spectral radiance of the airglow continuum has been obtained by Gadsen and Marovich (1973), who find that over the region of interest here, the intensity increases logarithmically with wavelength, from about 0.1 R nm^{-1} at λ700 nm to 0.4 R nm^{-1} at λ800 nm. (Figures for Broadfoot and Kendall's 1968 spectra). Because of this, if an instrument bandwidth equal to the $P_2(3)$ to $P_1(3)$ line spacing of the band is used (Section 2.3.3), it is found that the signal-to-noise ratio (SNR) to be expected from the $P_1(3)$ line, the strongest in the P branch, increases rather less rapidly with wavelength than does its $I_{OH}q$ product (see Table 2.1).

Absorption bands of atmospheric constituents are of particular importance because the observing site has an elevation above sea level of only 0.6 km. The atmospheric absorption spectra of Taylor and Yates (1957) and Yates and Taylor (1960), measured with an instrument width comparable to that adopted for the OH FPI, show little or no significant structure in the region of the (9 - 3), (6 - 1) or (7 - 2) OH bands. Between λ700 nm and λ900 nm there occurs the strong (0 - 0) O_2 absorption band at λ762 nm which masks the (4 - 0) OH band, as can be seen in, for example, the airglow spectra of Broadfoot and Kendall (1968). In addition the spectra of Taylor and Yates, and the less empirical results of McClatchey et al. (1972) show several weak water vapour absorption bands in this interval. There is, however, considerable disagreement between these spectra in regard to both the position and strength of the bands

(Selby and McClatchey, 1975). Unfortunately there appear to have been very few measurements of these atmospheric bands below λ 1μ made at suitable resolution. Without extensive and laborious computation, the high resolution spectra of the solar atlases or published by Curcio et al. (1964) can serve only as a rough indication of the OH bands which should be avoided.

Finally the choice of band was influenced by the availability of detectors, a consideration entirely justified by the difficulty and delay actually experienced in obtaining a suitable photomultiplier tube. Several EMI 9558B tubes were already on hand - the spectral sensitivity of their lower efficiency S-20 photocathodes is such that the peak signal-to-noise ratio is shifted to about the (7 - 2) band, as shown in Table 2.1.

The (8 - 3) band at λ727 nm was eventually chosen. It appeared to be free of contaminating emissions, would permit the use of an available detector, and was expected to result in a signal-to-noise ratio from the preferred detector only slightly lower than the maximum attainable. The atmospheric absorption measurements cited above, though they show the presence of gaps between the water vapour bands, do not appear to locate these gaps with sufficient accuracy to clearly point to a more suitable band. However this aspect of the choice will be discussed further in Section (4.3.3).

2.3.2 Prefilter and Free Spectral Range

As discussed in Section (2.2.1), a monochromator of some description must be used in conjunction with a Fabry-Perot interferometer to eliminate radiation from outside the free spectral range. On this occasion, as is common practice, another broader bandwidth Fabry-Perot, in the form of a multilayer dielectric interference filter, was used.

Such a filter can be made much narrower than an ordinary absorption filter, and offers the well-known light gathering power advantage over, for example, a grating instrument. Clearly the ideal filter shape is rectangular, its width being slightly less than $\Delta\lambda$. The extent of the filter profile wings is of particular importance in the present application because of the proximity of the (8 - 3) and (4 - 0) Q and R branches. Therefore a three period filter, which had a steeper cutoff at both long and short wavelengths than the more common single period filters, was used. It had a width at half maximum transmission, when measured in a beam of 5° divergence, of 10.7 nm, a centre wavelength of λ735.3 nm and a transmission in excess of 50% for the first four P lines, with a peak value of 84%. In addition the transmission at the wavelengths of the R branches of the (8 - 3) and (4 - 0) bands was < 0.5% (see Figure 1.2).

According to the conclusions of Section (2.2.4) the free spectral range should be as small as possible. It was therefore chosen to be approximately equal to the larger of the wavelength intervals between the $P_1(5)$ line and the (8 - 3) band R branch ($\sim$ 16.3 nm), and between the $P_2(2)$ line and the (4 - 0) band R branch ($\sim$ 16.9 nm). With this choice the band could be scanned from the $P_2(2)$ to the $P_1(5)$ line ($\sim$ 9.8 nm), and the worst case of contamination from adjacent orders - the $P_1(5)$ line overlapped by the (8 - 3) band R branch - would give an error in measured intensity of less than 2%.

The final selection of the exact operating order was also influenced by the requirement that the He-Ne laser line at λ632.8 nm, which served as a wavelength reference, be suitably positioned in the scan interval. An order of $m_o = 44$ was therefore used, giving a free spectral range (calculated for the median wavelength of λ735.3 nm)

equal to 16.7 nm. The corresponding plate spacing in air was 16.2 μm.

2.3.3 Resolving Power

Ideally the resolving power of the spectrometer should be chosen sufficiently high that both the line and the background intensities could be determined without having to undertake a lengthy numerical analysis of the data. This is not feasible for an FPI, however, if the twin aims of rapid temperature measurements and an extended free spectral range are to be achieved.

The expression relating temperature and the relative intensities of two lines (which may be readily obtained from equation (1.2)) is such that a 1% error, in for example the $P_1(2)$ and $P_1(3)$ line intensities, results in a rotational temperature error of 2.4% at 200 K. It is desirable, therefore, that contamination from adjacent lines be reduced to less than 1%. Because of the extensive wings of the Airy function, this means that for the $P_1(2)$ line the bandwidth must be about one quarter of the line spacing, depending on the widths of the defect and aperture functions. To maintain a free spectral range of 16.7 nm, the overall finesse N_I must be about 60, a figure well in excess of the limiting finesse of 28 actually achieved. If the background signal is to be measured at more than one wavelength, as is desirable, this finesse must be increased still more. In the case of the OH FPI, even if the required finesse could be obtained, the associated low light gathering power could not be offset by stepping the instrument bandpass from line to line, owing to drift in the scanning system used (Section 3.1.1).

In these circumstances, the resolving power is dictated by the strength of the recorded signal as a function of instrument bandwidth, and the ability of the numerical techniques employed to accurately

recover relative intensities from the smeared spectrum. The former is not easily evaluated. The transmitted flux is not directly proportional to the bandwidth i as expressed in equation (2.28), because in the design stage the reflective finesse is still subject to choice. However Chabbal's (1953) curves can be used to obtain an estimate under the conditions that the defect finesse and free spectral range are held constant, and the optimization criteria of Section (2.2.4) are adhered to. As the resolving power is decreased by increasing both the aperture and Airy function widths, τ_A, τ_D and Ω (proportional to f) all increase, while τ_F decreases slightly. The product $\Omega\tau_A\tau_D\tau_F$ was thus found to increase more rapidly with i, for a range of i centred on the P line spacing, than would be expected if equation (2.28) applied. The accuracy of this calculation was limited chiefly because the form of E varies continuously between rectangular and Airy as the ratio a/d increases.

The data analysis involved modelling the recorded spectrum, and fitting this model using a least squares technique. The parameter estimates obtained contain errors due both to inaccuracies in the model and to statistical uncertainties, for which estimates are also calculated. In relation to the former, the predominant consideration when increasing the bandwidth is likely to be that OH lines, which contribute in the wings of the instrument profile but do not lie within the spectral interval scanned, must be more accurately accounted for. Of importance in the presence of noise is the degree of orthogonality of the functions comprising the model. As the bandwidth increases, the functions describing adjacent lines in the spectrum become more similar, and the corresponding parameters can therefore be determined with less certainty. On the other hand, the transmitted flux (proportional to $\tau_I\Omega$) increases steadily with bandwidth, improving the accuracy of the computation.

Using signal levels similar to those obtained in practice, a spectral interval equal to that eventually scanned (Section 3.5.1) and a series of bandwidths, model spectra to which normally distributed noise had been added were constructed. The one for which the analysis returned the smallest rotational temperature error was considered to have the optimum bandwidth. A width of almost 1.0 nm was in fact adopted, although the value obtained by this method was about 1.7 nm. For only a moderate increase (< 30%) in the standard deviation of the temperature, the higher resolution gives a spectrum more readily interpreted by visual inspection, and reduces errors which may be incurred by having to model parts of the spectrum outside the region scanned.

2.3.4 Scanning the Spectrum

Reference to equation (2.4) shows that there are three parameters of the interferometer which can be varied to effect a wavelength scan. They are the refractive index μ of the material between the plates, the angle θ between the incident light and the etalon normal and the spacing ℓ of the plates. Many workers (for example, Mack et al., 1963; Nilson and Shepherd, 1961) have used the first of these by cycling the pressure of the gas (usually air) within the etalon. If the dependence of the refractive index of the gas with pressure p is of the form $d\mu/dp = c$, where c is a constant, then, from equation (2.4)

$$\frac{d\lambda}{dp} = \frac{\lambda}{\mu}\frac{d\mu}{dp} = \frac{\lambda c}{\mu} \quad . \qquad \ldots(2.33)$$

The value of c for air is given by Jacquinot (1960) as 3×10^{-4} per atmosphere, but a somewhat higher figure can be obtained using a denser gas such as Freon. Nevertheless the wavelength range which can be scanned by a realistic pressure change is no more than a few tenths of a nanometer - only a small fraction of that required for the present application. Because, as equation (2.33) shows, the interval $d\lambda$ is

independent of plate spacing, the technique is best suited to high resolution studies.

The second method is usually referred to as spatial scanning. The required variation of θ can be obtained by tilting the etalon with respect to the incident beam, but to do so seriously distorts the instrument profile (Eather and Reasoner, 1969). More satisfactory arrangements are described by Shepherd et al. (1965). Most of these essentially involve inserting in the image plane, either sequentially or simultaneously, a set of annular apertures which correspond to regularly incremented wavelengths. Their drawbacks lie mainly in constructional difficulties - obtaining the correct radii and areas for the annuli so that the instrument profile remains undistorted, positioning the apertures to the required accuracy, and the limited number of samples which can be taken per free spectral range. On the other hand, at the expense of a variable bandwidth a more rigid and stable etalon can be built, with a solid dielectric between the reflecting surfaces which gives an increase in effective area proportional to the square of the refractive index. It is also possible to increase the instrument sensitivity by using a separate optical coupling and detector for each aperture. Hirschberg et al. (1971) have described a method of multiplexing the signal by chopping the light from each annulus at a different frequency.

Methods by which the plate spacing has been varied include purely mechanical devices such as an air-driven piston deforming a metal membrane (Shepherd, 1960), thermal expansion of a bronze cylinder (Roig, 1958), electrostriction (Leś and Leś, 1968) and magnetostriction (Slater et al., 1965). The piezoelectric effect has also been used by many workers; for example Hernandez and Mills (1973), Bates et al.

(1971) and Shepherd (1967). From equation (2.4), for separation scanning

$$\frac{d\lambda}{d\ell} = \frac{\lambda}{\ell} \qquad \ldots(2.34)$$

so that a change in ℓ of $\lambda/2$ results in a scan through one free spectral range irrespective of the mean spacing. A continuous scan over the large wavelength interval required for the present application is therefore quite readily made.

Separation scanning using piezoelectric elements was adopted for the OH FPI. In the three point support system used (Section 3.1.3) this choice made the instrument quite versatile in that the operating order could be easily changed, and provided a simple fine adjustment for separation and parallelism. It was also compact, relatively easy to construct, allowed a high scan rate and had a very low power dissipation, unlike magnetostrictive systems. Its principal disadvantages lie in the lower stability of a non-rigid etalon, particularly in regard to the maintenance of parallelism during a scan, and the hysteresis and creep of the piezoelectric material (Section 3.1.1.2). Other workers (e.g. Bower, 1974; Ramsay, 1966) have overcome these problems by servo-controlling the plate separation and parallelism.

CHAPTER 3

CONSTRUCTION AND OPERATION OF THE SPECTROMETER

3.1 The Etalon

3.1.1 The Piezoelectric Elements

3.1.1.1 Testing

It was decided at an early stage of this project that the interferometer should be separation scanned using piezoelectric elements, that the spacing and parallelism should if possible not be servo-controlled, and that the scan rate should be sufficiently slow that the parallelism could be checked during a scan by visual inspection of the fringe pattern. The advantages of this approach lie principally in the constructional simplicity and low cost involved, and the avoidance of resonance and rigidity problems. On the other hand a high degree of stability and linearity is required of the piezoelectric elements. Several workers, including Clarke et al. (1975), Hernandez (1970b), Jackson and Pike (1968) and Reay et al. (1974) have built etalons along these lines. However there exists little documentation of the low frequency electrical and mechanical properties of the piezoelectric materials commercially available, although some data have been obtained by Ramsay and Mugridge (1962) and Sacconi (1970).

An effort was therefore made to determine by measurement the best type of piezoelectric element and set of operating conditions. Since the variation in plate separation had to be sufficient both to scan over almost one order and to allow for parallelism and mean separation adjustments, a large piezoelectric strain constant was required of the material. This constant d is defined (in simplified

fashion) by the expression

$$d = \frac{\Delta \ell}{\ell} \frac{t}{\Delta V} \qquad \ldots(3.1)$$

where $\Delta \ell$ is the increase in the dimension ℓ and the drive voltage increment ΔV is applied across a thickness t. Two types of lead zirconate titanate ceramic material supplied by Vernitron Corp., designated PZT-4 and PZT-5H, the first in tube form and the second in both tubes and discs, were chosen for testing. Their strain constants d_{31} (tubes) and d_{33} (discs) appear in Table 3.2. The tubes were supplied prepolarized radially, the discs parallel to their axes, so that an electric field in the same sense as the polarization produced a length contraction when applied across the wall of a tube, and an increase in thickness when applied between the faces of a disc.

The tubes had an overall length and outer diameter of 12.7 mm and a wall thickness of 0.8 mm. Their ends were polished flat and parallel to within one minute of arc. The inner and outer walls were coated with silver electrodes to provide good electrical contact and a uniform electric field distribution. The arrangement used to mount a tube was as shown in Figure 3.5. Constructed of brass, it formed a pillar separating the upper and lower sections of the etalon, with the ceramic prestressed by means of a nut and spring washer at the top of the pillar. Electrical connection for the high voltage dc supply was obtained by a coil spring of gold-plated beryllium copper wire wrapped round the outside of the tube. A leaf spring of the same material bearing on the inner surface provided the earth connection. Thin fused silica washers, the faces of which had been polished flat and parallel, insulated the ends of the tube.

The discs had an ID of 4 mm, and OD of 12 mm and were 1 mm thick. Their faces were coated with a silver electrode. For testing purposes

a stack was made, consisting of a sandwich of two discs and three beryllium copper shim washers for electrical connection, cemented together with silver-loaded epoxy resin (Araldite).

The properties of the ceramics were investigated in two ways. One method employed a capacitative displacement transducer described by Wilksch (1975), which was sensitive to movements as small as 0.5 nm. These measurements were made with the transducer clamped in a jig above the ceramic mount, and the whole enclosed in a temperature stabilized container. The other method simply involved recording the response of the interferometer to a spectral line source as a function of the voltage applied to the ceramics. Neither arrangement was entirely satisfactory - the first suffered from flexure in the mounting jig, while results from the second were difficult to quantify because the individual responses of the etalon's three piezoelectric elements could not readily be distinguished.

The relevant properties of the ceramics, creep, hysteresis and strain constant variation, are discussed below. Further results are presented by Cocks (1977).

(a) Creep. When a voltage step is applied to a ceramic there is an almost instantaneous change in dimensions, followed by a much slower asymptotic change referred to here as creep. Typical behaviour is illustrated in Figure 3.1, a record of the transducer output versus time for a 0 - 80 V step applied to a PZT-4 tube. The magnitude of the creep can be measured (rather arbitrarily) as the slow length increase or decrease during the 10 minute interval following the voltage step, expressed as a percentage of the initial length change ($\delta x/x$ in Figure 3.1). It takes a value of about 20%, and varies little between PZT-4 and PZT-5H, or between tubes and discs. It does appear, however,

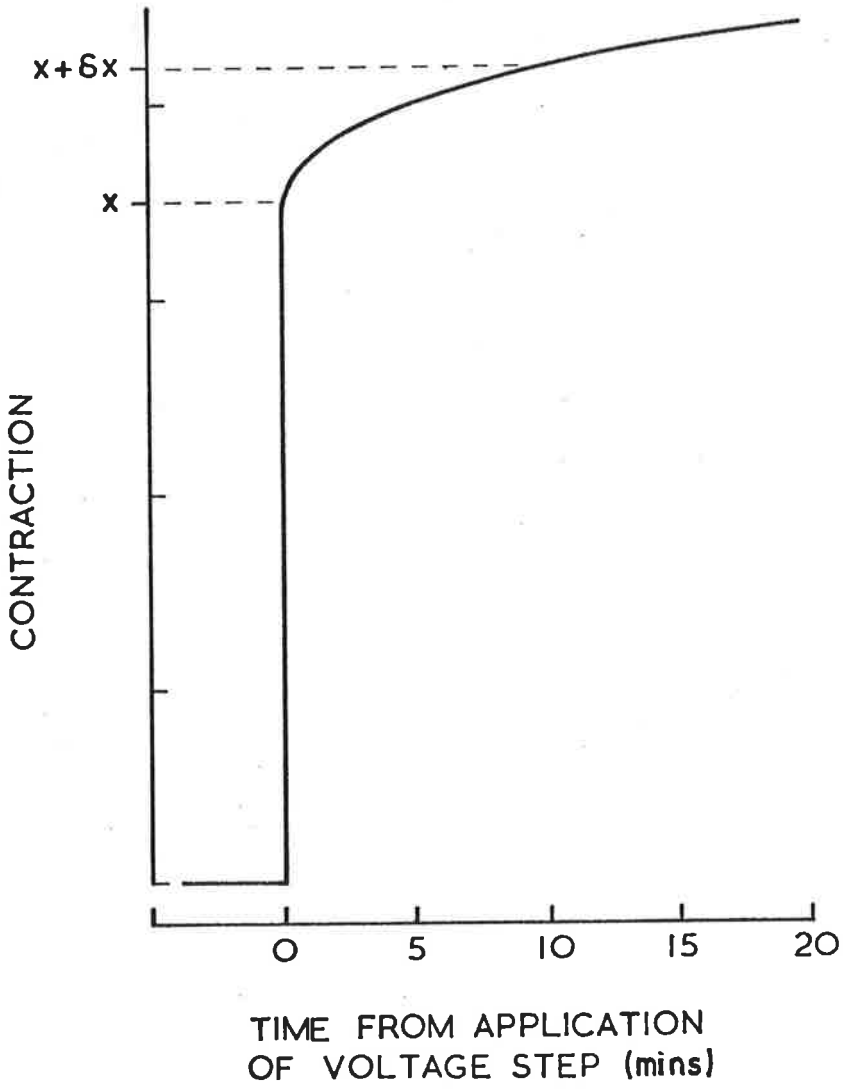
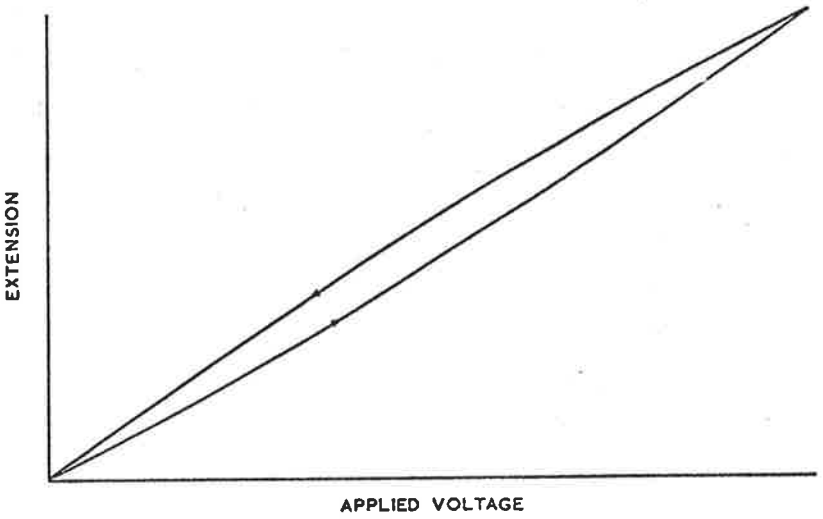


Figure 3.1 Creep behaviour of a PZT-4 tube on application of a 0-80v step.

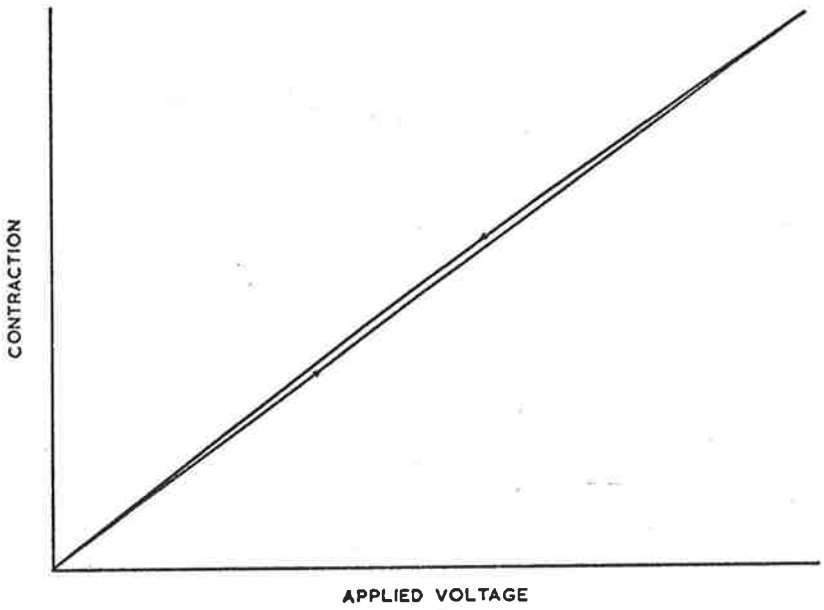
to lessen with increasing step size in the same proportion as the piezoelectric coefficient is reduced (see following). It is also found to significantly diminish if the desired voltage step is momentarily exceeded by about 20%.

Despite this, the creep phenomenon makes it difficult to quickly and reliably set the etalon to a given spacing. As a result a scan cannot be frozen at any given point and the parallelism checked, unless this is done very rapidly. Neither is it feasible to step the instrument profile from one OH line to the next as would be desirable if a narrow bandwidth were used. The best performance in the tests was obtained by cycling the etalon continuously - even a brief pause before beginning a scan disturbed the ceramic response over a sizeable fraction of that scan.

(b) Hysteresis. When the drive voltage to a ceramic is cycled between two values a closed hysteresis loop of displacement against voltage is obtained. In general the upward part of the curve (increasing voltage) is more linear than the downward part. Figures 3.2(a) and (b) show the hysteresis obtained with PZT-5H discs and a PZT-4 tube respectively using the transducer. They were measured with a 10^{-2} Hz symmetrical triangular scan voltage which gave a total displacement approximately equal to half a wavelength. No significant variation of linearity or hysteresis loop width with bias voltage or stress could be discerned in the transducer measurements, contrary to the findings of Hernandez (private communication). The hysteresis curves showed that the best linearity and minimum hysteresis could be obtained with PZT-4 tubes, but flexing of the jig restricted further interpretation of these results.



(a)



(b)

Figure 3.2 Hysteresis curves for a triangular drive voltage waveform of frequency 10^{-2} Hz and amplitude such as to give a length change of approximately one order (3×10^{-7} m).

(a) Two PZT-5H discs. $\Delta V = 245V$

(b) PZT-4 tube. $\Delta V = 242$ V

The hysteresis can also be observed in the interferometer response to a monochromatic line. It causes a displacement between the profiles recorded on the upward and downward legs of a symmetric scan. Figure 3.3 shows a measurement of the PZT-4 tubes using a 10^{-2} Hz scan encompassing slightly more than two orders of the Cd λ643.7 nm line. In addition to this peak shift, small differences in the convexity of the hysteresis loops of the three elements in the etalon made it impossible to obtain adequate parallelism in both directions of a scan.

The difference in finesse obtained from the first and second peaks of the upward leg of the curve in Figure 3.3 provides an indication of the nonlinearity of the scan. For PZT-4 tubes this difference is about 12%. A more useful measurement, covering a range of about one third of an order, was made using as a source a monochromator, the bandwidth of which was substantially less than that of the OH FPI. The position in the scan of the recorded profile was accurately determined using a simple modification of the data analysis computer routine. Unfortunately the result was affected by a periodic error in the wavelength calibration - an error which could have contributed greatly to the measured maximum departure from scan linearity of 2% of full scale. Nevertheless it appeared that the asymmetric scan employed both for this measurement and airglow observations distorted the ceramic response most within about the first fifth of its excursion, and it was estimated that the linearity was about 1% over the remainder.

(c) Variation in piezoelectric strain constant. Measurements made both with the capactitor transducer and the etalon show that the piezoelectric strain constants d_{33} and d_{31} vary significantly with bias voltage. Table 3.1 shows the fractional change $\delta d/d$ in the strain

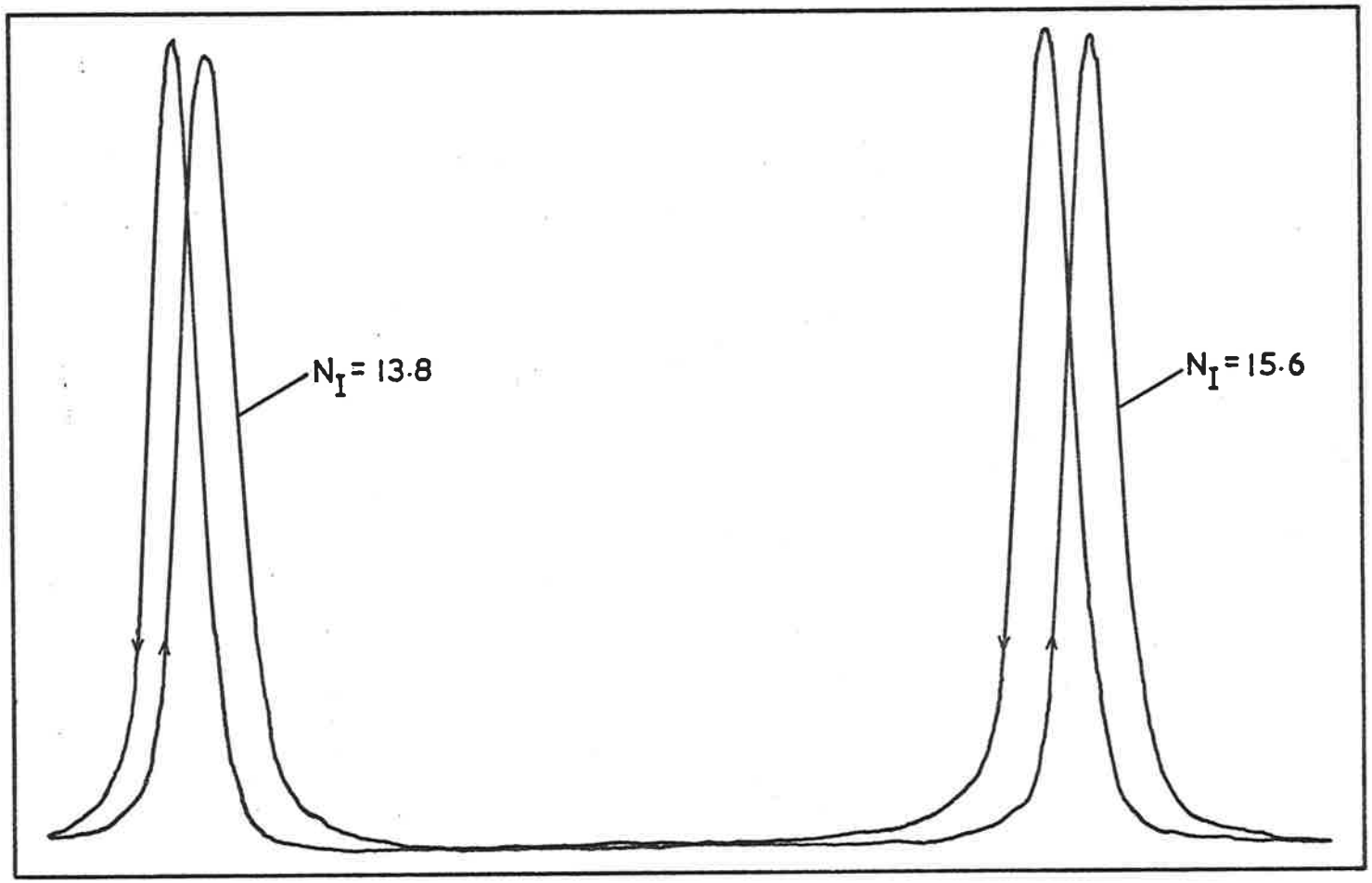


Figure 3.3 Finesse measurement and return scan using a Cadmium λ643.7 nm source, PZT-4 tubes and a symmetric 10^{-2} Hz waveform.

constant per volt of bias for the elements tested. The values are referred to the zero bias condition. They were obtained from hysteresis measurements made with a 10^{-2} Hz triangular drive waveform of fixed amplitude and variable mean. The entry designated 'stabilized' corresponds to a PZT-4 tube which had been held at 200 C for one hour, in accordance with the manufacturer's recommendations for improving the thermal stability of the material.

TABLE 3.1

Variation of Piezoelectric Strain Constant with Bias Voltage

Element	$\frac{\delta d}{d} \times 10^{-4}$ (volt^{-1})
PZT-5H disc	-1.5
PZT-5H tube	-4.5
PZT-4 tube	-1.7
Stabilized PZT-4 tube	+1.2

Although this was not measured, the behaviour of the etalon suggested that there was some disparity between the values of δ for different elements.

Absolute values of the strain constants were obtained by determining the voltage excursion which had to be applied to each element to scan the etalon in parallel fashion from one order of a spectral feature to the next. Considerable variation between elements of the same material, and between the experimental values and the manufacturer's data were encountered. The latter was to some extent expected, because stresses along other axes of the ceramic could not be avoided, although the tube mounts were designed to minimise these effects. The results for a set of three tubes of PZT-4 and PZT-5H material are presented in Table 3.2.

TABLE 3.2

Comparison of Piezoelectric Strain Constant Values for Several Elements (10^{-12} m.volt)

Element	Measured Value			Nominal Value
PZT-4 Tubes	-80	-80	-63	-123
PZT-5H Tubes	-178	-197	-260	-274
PZT-5H Discs				593 (d_{33})

3.1.1.2 Summary and Application

The following selection of the type and mode of operation of the piezoelectric elements used in the OH FPI was made on the basis of the foregoing investigation, with a view to obtaining maximum stability and linearity in particular.

(a) PZT-4 material was adopted because its behaviour is considerably more linear than that of PZT-5H. The penalty for this choice is a considerable reduction in strain constant.

(b) Tubes rather than discs were used, again because of their more linear response. The possibility of dimensional instability and additional stresses created by using a glued stack of discs is thereby avoided also. The mounting arrangement for the tubes was as described in the previous section.

(c) The etalon was scanned continuously to improve its stability and to avoid distorting individual scans. However a stabilization period had to be allowed for the creep to disappear after each electrical adjustment of the parallelism or separation.

(d) The scan rate was made sufficiently slow that the parallelism could be checked visually without interrupting the scan. A period of $\sim$30 seconds was in fact used - well above the value of about 1 second at which Hernandez (private communication) found significant self-heating effects.

(e) Voltage was always applied to the ceramics in the same sense as their polarization (Hernandez, 1970b).

(f) Only the upward (increasing voltage) leg of the scan was used because of differences between tubes in the shape of the hysteresis curves. To reduce the loss of observing time entailed in a symmetric scan, and to avoid jarring the plate mounts, the fall time of the scan waveform was made equal to one sixteenth of the rise time.

(g) A variable gain control was provided to electrically match the strain coefficients of the tubes.

(h) Large changes in the mean voltage applied to any element were avoided because of the voltage dependence of d_{31}. When this was done, the electrical gain had to be readjusted, and the wavelength versus voltage scale recalibrated.

Recent results by McKeith et al. (1976) suggest that better hysteresis and linearity behaviour could be obtained by using PZT-4 discs operated in the reverse biased mode. In order to improve the performance of an interferometer employing this method of scanning, further investigations of the electro-mechanical properties of piezoelectric ceramics are needed. These should include studies of

(a) Methods of eliminating creep,

(b) The frequency dependence of the piezoelectric properties, especially in regard to the relationship between strain constant and bias voltage,

(c) Differences in behaviour from one element to another with a view to obtaining a matched set (which could permit parallelism to be maintained during both parts of a scan), and

(d) The effect of mechanically loading an element.

3.1.2 The Plates

3.1.2.1 Description

The interferometer plates were made of Homosil (a high quality fused quartz) by J. Cole of Cole Precision Optics, and had a clear aperture of 50 mm and were 12 mm thick. They were cut from a larger disc to leave an encircling ledge 5 mm thick, projecting 0.6 mm beyond the clear diameter, and situated midway between the plate faces. This ledge allowed the plates to be mounted in the etalon with minimum distortion and no reduction in aperture. Each plate was slightly wedged to avoid unwanted interference between its faces. The surfaces to be coated were hand polished to a specification of better than $\lambda/100$, and in fact this figure could not be greatly improved upon without mechanical polishing.

The coatings were applied by J. Ward, of the Australian Department of Defence. They consisted of a 30 nm thick layer of silver followed by two quarter-wave layers, the first of magnesium fluoride at λ356 nm, the second of cerium oxide at λ500 nm. At the operating wavelength they had a transmittance $\underline{T} = 0.051$, a reflectance $\underline{R} = 0.940$ and an absorptance $\underline{A} = 0.009$, giving a reflective finesse $N_R \simeq 50$ and a peak transmission $\tau_A = 0.71$. Metallic rather than dielectric coatings were chosen primarily because they had a much larger useful operating range - about 300 nm instead of 50 nm - despite their somewhat higher absorptance. Unfortunately the uniformity of the coatings was not good, and the eventual defect finesse was about 28 rather than in excess of 50 as expected from measurements of the uncoated plates (see following section). In consequence the value of $\underline{R}$ was not in accordance with the criterion given at the end of Section 2.2.4, namely $N_R \simeq N_D$.

3.1.2.2 Testing

Measurements of the plate flatness were made at intervals during the polishing process to determine whether the specifications had been met, and if not to indicate the magnitude and form of the correction necessary. When the polishing was complete, the coating reflectance was specified according to the value of N_D, estimated on the basis of the maximum flatness error observed (Section 2.2.2). Since four 50 mm plates were being polished concurrently, similar procedures were also used to select the best combination and orientation of the finished (coated) set.

Most of the testing was done using fringes of constant optical thickness (Fizeau fringes). The optical assembly required to obtain these fringes in reflection is shown in Figure 3.4. Uniform monochromatic illumination is provided by means of a point source, a beam splitter and a collimating lens. The latter also forms a Fabry-Perot interference pattern in the plane of an on axis field stop, which is small enough to select only part of the central fringe. Thus if the (uncoated) plates are perfectly flat and parallel and appropriately spaced, the eye, when placed at the stop, will see a bright fringe uniform over the whole etalon, corresponding to an equal path length between the plates for all rays. Local departures from this condition caused by plate defects will then appear as darker areas on the plate surface. If the parallelism is now somewhat degraded, a series of parallel fringes will appear, each of which maps a contour of constant separation. Bends in the fringes show departures from flatness, the magnitude of which can be readily measured because the perpendicular distance between adjacent fringes corresponds to a change in plate spacing of half a wavelength. By varying the mean etalon

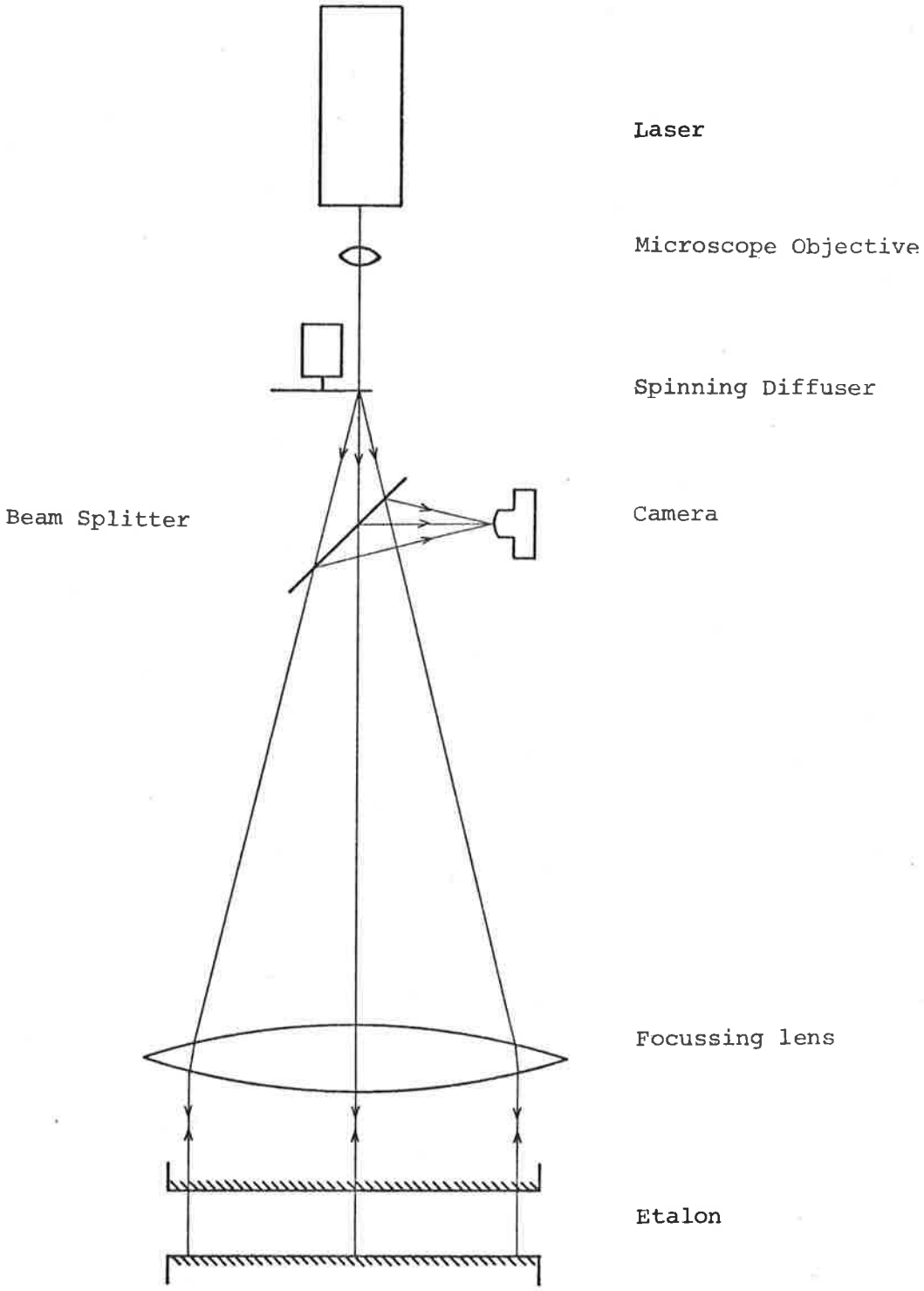


Figure 3.4 Optical assembly for viewing Fizeau fringes in reflection.

spacing the positions of the fringes may be altered, and the entire plate area examined.

The measurements may be conveniently separated into those made with reflected and with transmitted light.

(a) Measurements in reflection.

The Airy function for uncoated plates (Figure 2.2(c)) has a very low finesse because the reflectance of an air/glass interface is only about 0.04. The resulting broadness of the Fizeau fringes makes small plate defects difficult to detect, particularly if the fringes are viewed in transmission, when they appear as dark bands on a bright background. The situation is improved slightly by using reflected light. In this case the reflected intensity is described by equation (2.5), the fringe intensity is the complement of that obtained in transmission, and the pattern is one of bright bands on a dark background. Some estimate of the distinctness of the fringes can be made with the help of Michelson's expression (Born and Wolf, 1975), for visibility, namely $V = (I_{Max} - I_{Min})/(I_{Max} + I_{Min})$, where I is fringe intensity. Using equations (2.2) and (2.5) V is found to be approximately 0.08 and 1.0 for transmitted and reflected light respectively. Since much sharper fringes are obtained with mirrored surfaces, tests during the polishing process should ideally be done with the plates flash-coated with, for example, aluminium. However this proved to be very time-consuming, and involved the rather doubtful assumption that the temporary coatings did not distort the result.

The tests were made by placing the plates, unclamped, in the etalon, illuminating them with a He-Ne laser, the beam of which was spread somewhat with a short focal length lens (a microscope objective), and photographing the Fizeau fringes with a camera (Nikormat SLR and

Ilford FP-4 film) placed immediately behind the sampling aperture. The collimating lens had to be of long focal length and good quality - a Leitz 560 mm lens was used. Speckle in the laser beam lowered the image quality, particularly in later transmission measurements, and was therefore removed by spinning a diffusing screen in front of the diverging lens. A series of photographic enlargements of the fringes in various orientations was measured by hand to determine the degree of flatness error present. The practical limit to the sensitivity of this technique was about $\lambda/120$.

(b) Measurements in Transmission.

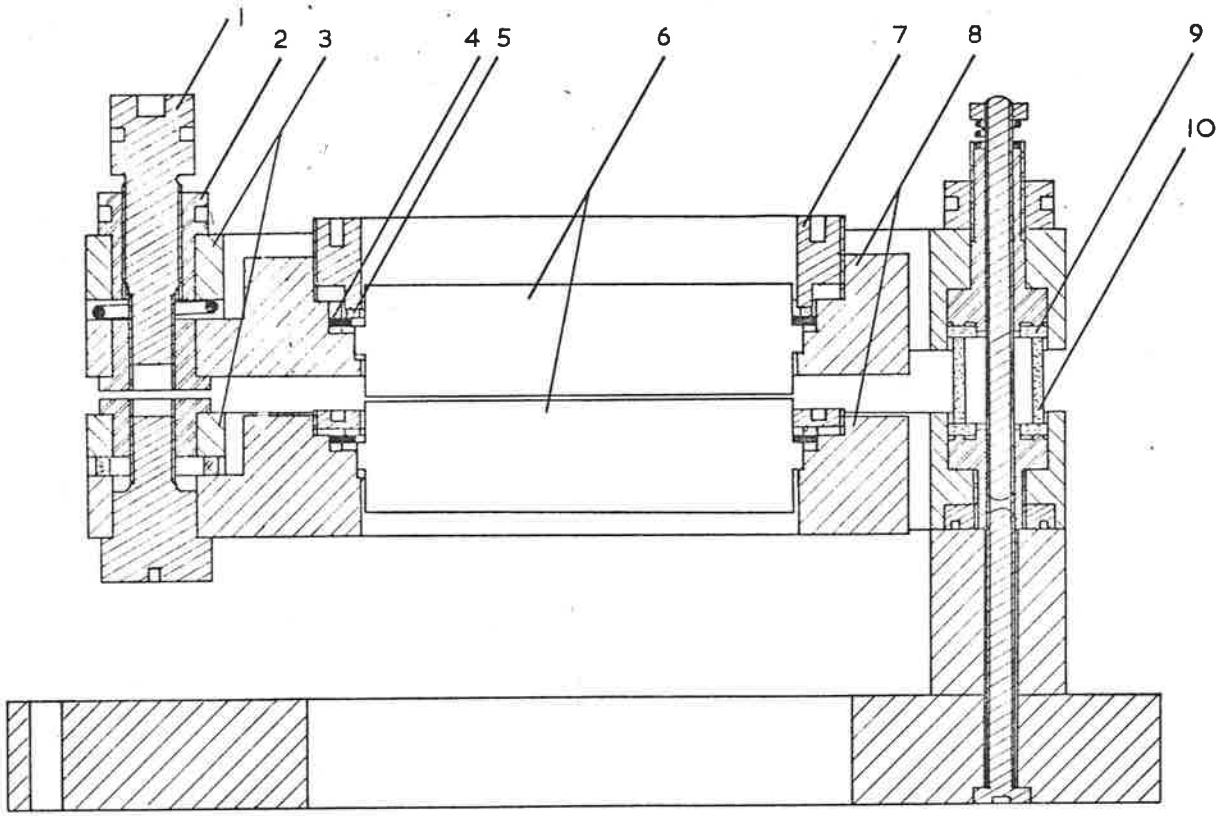
When the plates were coated, the associated absorption made it advantageous to view the Fizeau fringes in transmission. The technique was very similar to that just described, the chief difference being that the source and collimating lenses were placed on the other side of the etalon and the fringes were projected onto a translucent screen and photographed.

Two other methods may be used to obtain the defect function and hence the defect finesse. In the first the plates are set parallel and photographs of the resulting single fringe are taken at slightly different spacings. The contours of maximum brightness of each frame are then used to construct a histogram of plate area versus departure from flatness - an approximation to the defect function. This technique was not employed with the 50 mm plates, and further details are given by Wilksch (1975). In the second method the instrument profile $I(\lambda)$ is recorded using a narrow line source (for example the laser or an Hg198 lamp) and a pinhole field stop, so that the result is to a good approximation $E(\lambda)$. $D(\lambda)$ is then calculated by numerical deconvolution using the known values of $\underline{R}$ and $\underline{A}$ which define the Airy function.

3.1.3 Mechanical Structure

Since the interferometer was not to be servo-controlled, the etalon ideally had to be designed to be rigid, temperature insensitive and stress free in order to obtain maximum stability. A cross-section is shown in Figure 3.5. The etalon construction is based upon two pairs of concentric 14 mm square-section rings, labelled in the diagram as $I_{1,2}$ and $O_{1,2}$, where I denotes an inner ring, O an outer ring and the subscripts 1 or 2 refer to the upper or lower rings respectively. The rings I mount the plates, and are connected to the rings O by tongues which are set at 120° intervals round the etalon. O_1 and O_2 are held apart by the three piezoelectric ceramic tubes, each of which is positioned midway between two tongues, and is mounted as described in Section 3.1.1.1. The whole etalon sits on three pillars which are directly beneath the ceramics.

Mechanical adjustment of the plate parallelism is provided by the three double-threaded screws and spring washers which are used to clamp I_1 to O_1. A coarse adjustment is made by turning one of the nuts, which is free to rotate in its seating, while turning a screw only gives fine control. The original design called for a similar arrangement for the lower rings, so that the etalon could be aligned perpendicular to the optic axis, but it was found that packing shims were quite adequate for this purpose. The screw pitches are 72 and 76 threads per inch, so that one full turn of a screw results in a movement of 18.6 μm. This is sufficient to obtain fairly good parallelism, but requires a delicate touch. The last stage of adjustment is achieved electrically with the ceramic tubes, which have a 750 nm (two order) range for a 600 V drive voltage swing.



1. Double-threaded screw (fine adjust)
2. Nut (coarse adjust)
3. Outer support rings O_1 and O_2
4. Three-lug washer
5. Spring washer
6. Plates
7. Bezel ring
8. Inner support rings I_1 and I_2
9. Silica insulating washer
10. Piezoelectric ceramic tube

Figure 3.5 Cross-sectional diagram of the etalon

The double ring system was found to be perfectly adequate in shielding the plates from stresses associated with the ceramics or adjusting screws. Some difficulty was experienced, however, in devising a method of clamping the plates without distorting them, and even the arrangement eventually adopted was not completely satisfactory. The ledge of a plate rests on three pegs projecting from the inner wall of the inner ring and diametrically opposite the connecting tongues. A washer fits over the plate ledge such that three lugs on its internal diameter are positioned directly above the pegs. Above this washer is a spring washer and bezel ring which clamp the plate ledge between the pegs and lugs.

In order that the etalon respond as little as possible to temperature changes, the material used in its construction should ideally have had a very low thermal expansion coefficient. Invar is most suitable but is expensive and difficult to obtain, and instead bronze, which had been stress-relieved by heating after machining, was used. The desired temperature insensitivity was then achieved in the first place by recessing the ceramic tubes into the rings O_1 and O_2 so that the thermal expansion of the materials was compensated, and in the second place by controlling the temperature of the etalon chamber.

3.1.4 Electronic Circuitry

Most of the circuitry used in conjunction with the etalon was designed by others in the Mawson Institute (Bower, 1974; Creighton, unpublished; Wilksch, 1975) and will be described only briefly here. Heater and detection circuitry will be dealt with in Sections 3.2.3 and 3.3.5.

Two dual 12 V and one dual 15 V regulated supplies, each capable of delivering 800 mA, and based on a single integrated circuit, (μA723)

were required to power the heater controllers and the three ceramic drive units.

The high voltage dc ceramic supplies may be considered simply as programmable operational amplifiers which can deliver up to 600 V at 3 mA. Each is equipped with a front panel meter to display its output.

Because it was desired that the ceramics be perturbed as little as possible, a "keep alive" scan generator was built. In essence this consists of an up/down counter and digital-to-analogue convertor which are clocked by a fixed frequency oscillator to give a staircase asymmetric triangular waveform. A control unit disables the four least significant bits when counting down, making the fall time equal to one sixteenth of the rise time. Either 128 + 16 or 256 + 32 steps per period can be selected at a front panel switch. The unit accepts externally generated scan signals, and the amplitude of the scan, remote or local, can be continuously varied by means of a graduated potentiometer. To enable finesse measurements to be conveniently made, a front panel switch allows the amplitude to be increased by a fixed factor and at constant slope. A train of channel advance pulses (128 or 256 per period), which can be inhibited by a synchronization pulse until a scan begins, is provided so that the scan generator can be used in conjunction with a Nuclear Data Signal Analyser (Section 3.3.6). The output voltage is displayed on a front panel meter and fed (indirectly) to the 600 V ceramic supplies. A schematic diagram of the circuit is shown in Figure 3.6.

A network was also designed to sum the scan, parallelism adjustment and manual separation signals required by the etalon. The parallelism controls are arranged so that only two potentiometers,

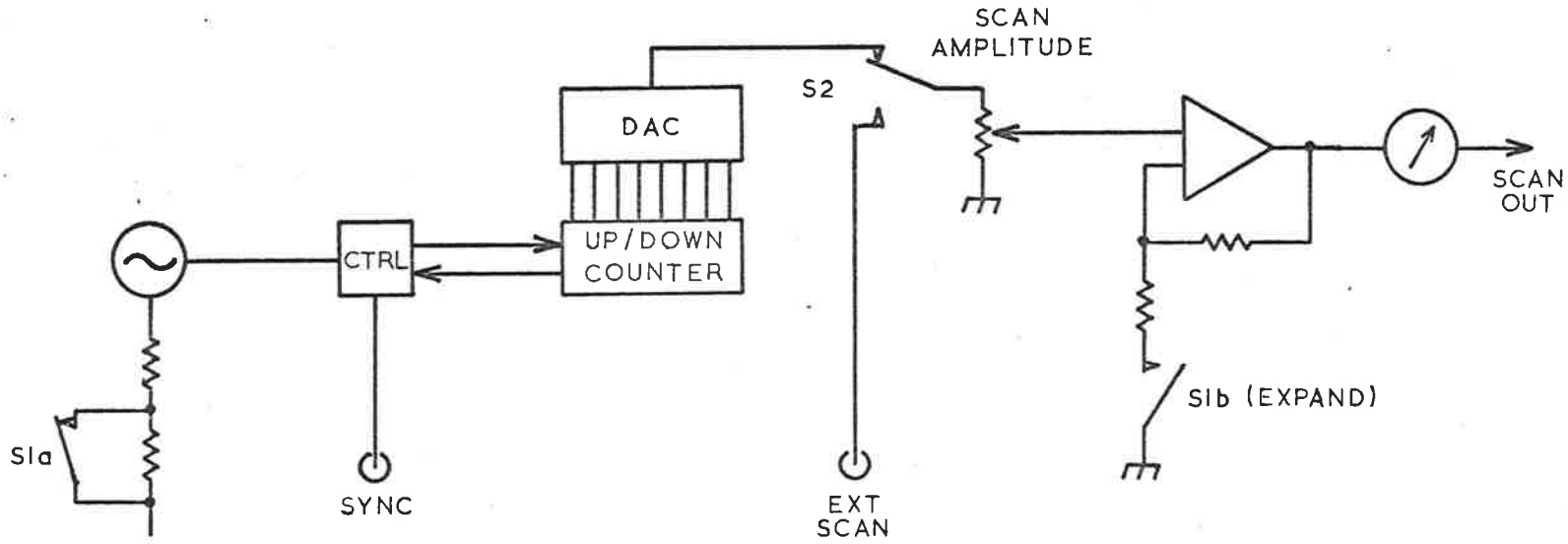


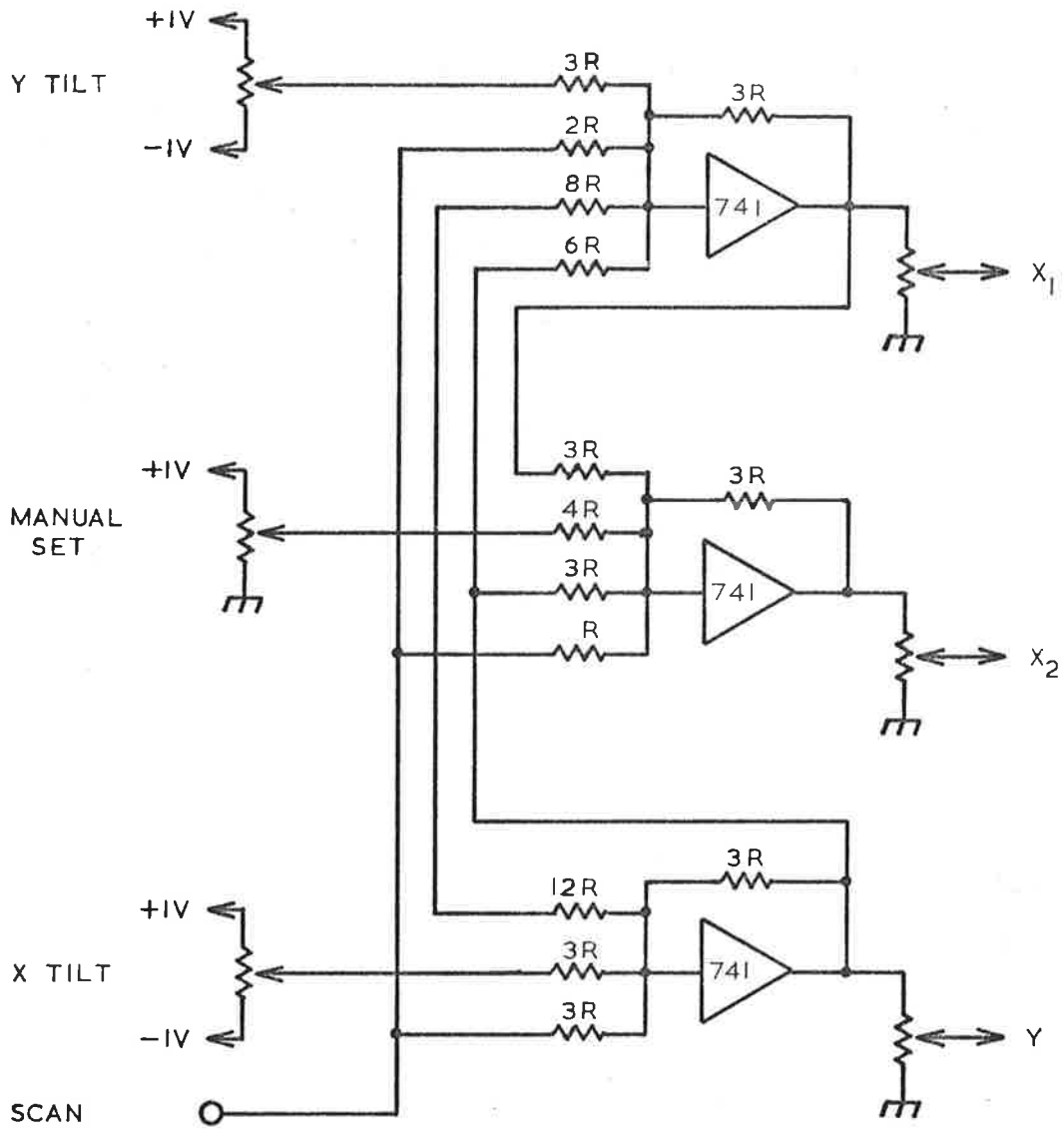
Figure 3.6 Schematic diagram of the 'keep alive' scan generator.

acting on orthogonal axes denoted X and Y, are required, and so that a parallelism correction does not alter the mean plate separation. This is done by driving the 600 V supplies with three operational amplifiers (μA741), which have weighted gains and cross-coupled feedback as shown in Figure 3.7. The manual separation control and scan signals - the former from another graduated potentiometer and the latter from either the keep alive generator or a remote source - are also summed at the operational amplifier inputs using appropriate resistor values. Variable gain for each of the ceramics is then obtained by including a variable resistor in series with the inputs to the 600 V supplies.

3.2 The Spectrometer Body

3.2.1 Optical System

Figure 3.8 is a sketch of the optical system of the spectrometer. The principal optic axis is oriented vertically. Light is reflected into the instrument by the periscope mirror, which is directed towards the south celestial pole to ensure that the stellar contribution to the background signal is both small and constant. Below it are two 200 mm f2.9 single component biconvex lenses separated by the sum of their focal lengths and positioned such that they image the etalon aperture at the mirror. With this arrangement the main body of the spectrometer is kept entirely within the observatory, and only the mirror projects above the roof. Beneath these again is the filter, which precedes the etalon so that broadening of its transmission profile by obliquely incident light is minimized and scattering inside the barrel is reduced. The collimating element, a 230 mm f4.3 single component plano-convex lens, is followed by an eyepiece lens, the function of which is to



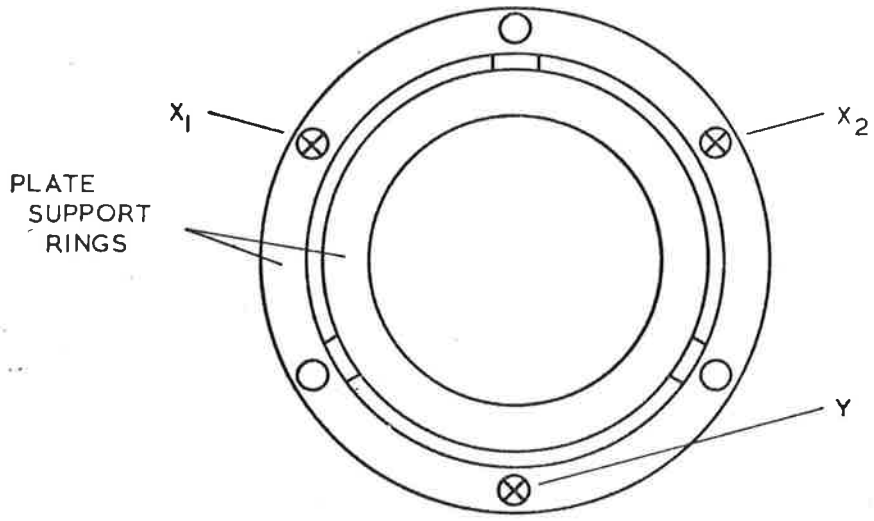


Figure 3.7 Schematic diagram of the parallelism and separation control electronics.

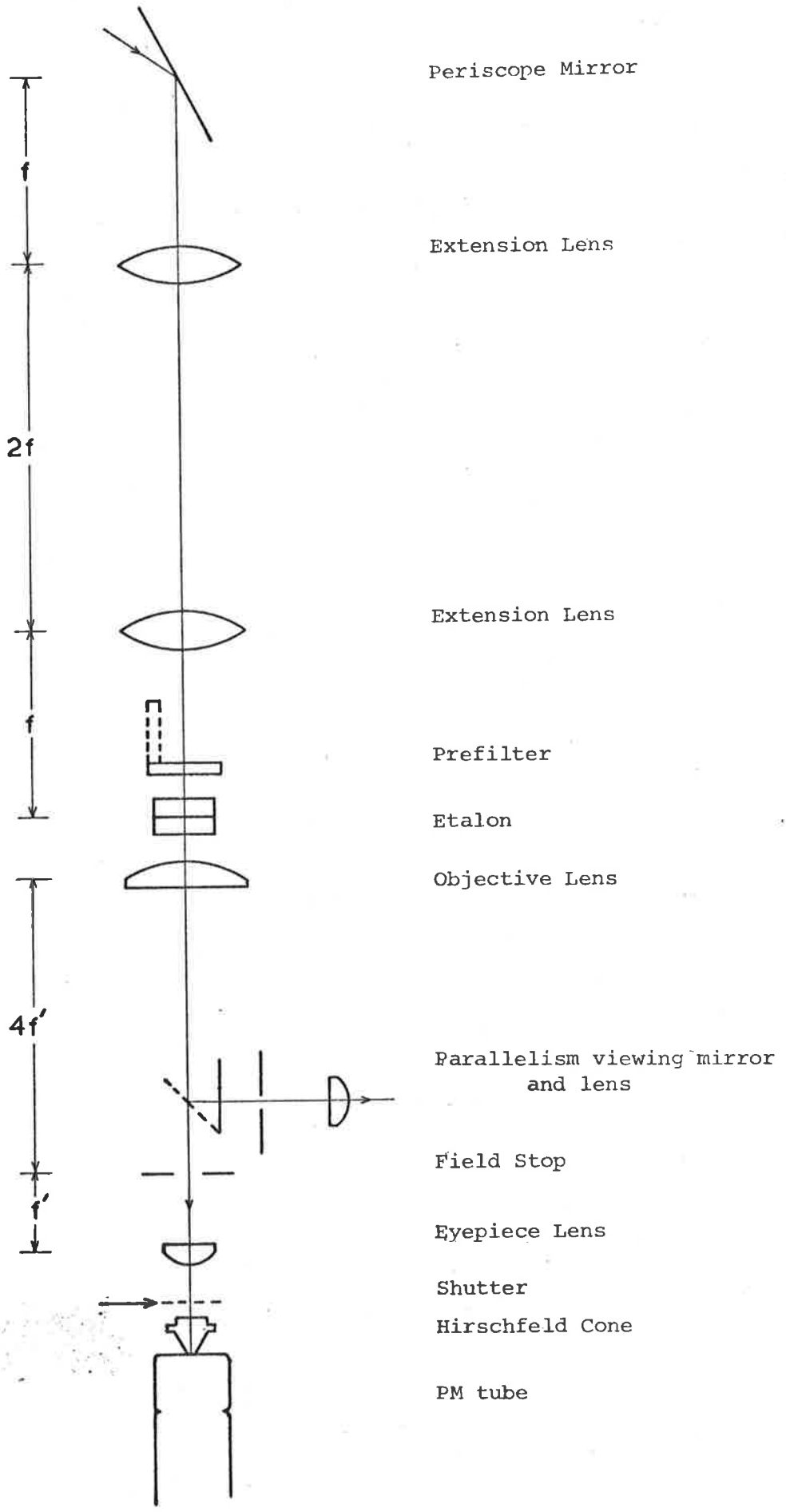


Figure 3.8 Schematic diagram of the spectrometer optical system.

deliver near-parallel light to a quantum efficiency enhancement cone (Section 3.3.4) placed over the photomultiplier. Should the cone not be used the lens ensures that inhomogeneities in the photocathode do not distort the signal during a scan. An 18.5 mm diameter field stop in the image plane of the collimating lens both selects the central fringe of the Fabry-Perot system, and defines a 4.6° (full angle) field of view. A 90° side-viewing system used to check the parallelism consists of an aluminium front-surfaced mirror, which can be lowered into the optical path above the field stop by means of a small dc motor, and an eyepiece lens. All the lenses are bloomed at λ735 nm to give a reflectance of about 1% at the glass/air interface.

3.2.2 Mechanical Structure

The spectrometer is constructed as follows. The etalon and the objective and eyepiece lenses of the small telescope which follows it are mounted inside a brass barrel, beneath which is slung a cylindrical photomultiplier chamber. This assembly is enclosed within a wooden box, some 500 mm square and 1200 mm high, which is bolted to an aluminium girder frame. The majority of the electronics is housed in a rack beneath the box, but the controls required for setting the parallelism, plate separation and ceramic gains are mounted near the top, conveniently close both to the side-viewing eyepiece, which projects from the box at eye level, and to the etalon itself.

The etalon is isolated to some degree from mechanical shocks by suspending the barrel and photomultiplier chamber from three coil springs working in compression. The movement is damped by means of oil-filled dashpots, and has a free period of about one second. This is sufficient to prevent misalignment being caused by all but severe shocks, though some elastic deformation still results from moderate

perturbations. However a final degree of isolation is obtained with the instrument standing on the substantially vibration-free concrete observatory floor.

Immediately above the barrel, and separated from it by a double-glazed window, is the filter enclosure. It can be opened to allow the filter to be pivoted upward out of the beam when parallelism or separation checks are made. The tube containing the two extension lenses and periscope mirror can also be separated just above the lower lens to allow the introduction of light sources. Two 10 W plate heaters are positioned next to the mirror to prevent condensation forming upon it. All the internal surfaces of the spectrometer are blackened to reduce the scattering of stray light.

3.2.3 Temperature Stabilization

To obtain the degree of stability required of the etalon, it was necessary to control its temperature to within narrow limits. To do this the inside of the containing box was insulated with polystyrene foam 75 mm thick, a 45W heater element of nichrome wire was wound on the outside of the barrel, and the etalon temperature was controlled at 30C. To improve the stability sections of the interferometer had been made deliberately massive, thereby increasing the thermal inertia. In addition, the temperature of the insulated filter enclosure was similarly, though less rigorously, controlled. This also had the effect of reducing the thermal drift of the filter profile to negligible proportions.

The electronic circuitry used to drive the heaters will be only briefly described (see Section 3.1.4). Current through a thermistor positioned close to the etalon is compared with a reference current,

which is set by a suitably chosen high stability resistor. The amplified difference is used in a negative feedback loop to provide trigger pulses to a pair of SCR's, which pulse modulate the 32V ac heater supply within the small proportioning range (±0.01C) of the controller, and act as simple on/off switches outside this range. The original design of the controlling circuitry included provision for rate feedback - that is, the heater current was made a function of the rate of change of temperature also. This facility is utilized in the OH FPI, though it appeared to have little effect on the ultimate stability attained. Two heaters were built, one for the barrel and the other for the filter enclosure. The error amplifier of a third monitors the etalon temperature, using a thermistor clamped to the upper outer (O_1) ring, and its output is displayed on a front panel meter. With the spectrometer installed in the air-conditioned (24 C) observatory a stability of better than 0.02 C was achieved.

3.3 The Detection System

3.3.1 Choice of Detector

The spectrometer was originally designed with a view to using a photomultiplier for a detector, since existing solid state devices (PIN diodes) had an area × solid angle product orders of magnitude smaller than that required. With the advent of large area diodes such as the UDT-500, 5DP and 10DP, the performance of these devices was reconsidered. However a brief calculation was sufficient to demonstrate their continued unsuitability. Taking into account the variation of airglow intensity with hydroxyl band and the current responsivity of the diode, a maximum signal is expected when observing the (4 - 1) band. For the $P_1(3)$ line the radiant power incident on the detector, assuming

an instrumental bandwidth of 2.0 nm, an overall transmission of 0.2 and OH intensities quoted by Krassovsky and Shefov (1962), is $\sim 4 \times 10^{-13}$ Watts. At the same wavelength the noise equivalent power of the least noisy diode, the PIN-5DP, is 7×10^{-13} W $Hz^{-\frac{1}{2}}$, although it seems likely that this figure could be reduced, perhaps by as much as a factor of 16, by cooling the diode to dry ice temperatures (Neiswander and Plews, 1975). Should this be possible, and without regard for amplifier noise or the probable necessity to chop the signal, the SNR of this configuration is only about 10, considerably less than the figure of 50 or more actually obtained with the photomultiplier.

The tube was an EMI type 9659B, and was selected primarily because of its high quantum efficiency. The only other tube known to be superior in this respect, the RCA C31034 with a nominal quantum efficiency of 11%, was unsuitable because of its small field of view. It appeared that the EMI 9658, which has a corrugated end window to enhance its sensitivity by multiple reflection, was not better than the 9659 since it did not permit the use of a Hirschfeld cone.

The photocathode material of the 9659 is a trialkali, designated extended S-20, with a nominal quantum efficiency of 4% at 735 nm and a diameter of 44 mm. The tube's eleven dynodes have CdS secondary emitting surfaces, and are arranged in a venetian blind configuration. The voltage divider network used is similar to that recommended for pulse counting, the cathode to first dynode voltage being defined by a zener diode, and the last three gain stages being decoupled. However the anode was earthed to permit current measurements to be made, for example when, for calibrating and evaluation purposes, signal count rates were used which exceeded the response time of the pulse amplifier. This did not seem to have a deleterious effect on the tube noise as is

suggested by the manufacturers. Over a period of time several 9659 tubes became available, and the one with the highest apparent quantum efficiency was selected. Table 3.3 is a list of these tubes together with the data supplied with each. The suffix A designates tubes selected for high sensitivity at λ735 nm. To give some idea of the variation possible, the table also lists the absolute quantum efficiency of each tube, obtained by comparing a signal from standard illumination at λ735 nm with that from a tube (serial number 20082) which had been precisely spectrally calibrated before purchase. These results are corrected to a common dynode gain using the data supplied.

TABLE 3.3

Test Ticket Data and Measured Quantum Efficiencies for a Selection of Photomultipliers

Type	Serial Number	EHT for 200 A/lm (volts)	Dark Current at 20 C (nA)	Cathode Sensitivity (μA/lm)	Quantum Efficiency at λ735 nm (%)
9558B	9964	1050	0.7	124	1.5
9659QA	19680	990	4.0	292	0.9
9659QA	19655	1010	4.0	273	2.9
9659QA	20082	840	4.0	275	4.1
9659B	24318	980	6.0	238	7.2

3.3.2 Photomultiplier Chamber Structure

The photomultiplier (PM) is housed in a fibreglass cylinder bolted to the bottom of the interferometer barrel. The outside of the cylinder is covered with aluminium tape both to make it light-tight and to retard heat transfer. Within is a cooling jacket of copper coils wound on a brass cylinder which closely surrounds the PM tube itself, and filling the space between this jacket and the outer wall is an insulating layer of plastic foam. The eyepiece lens at the end of the barrel and its fibreglass mounting disc help to prevent heat transfer between the

etalon and photomultiplier enclosures. Access to the tube is provided by a fibreglass lid sealing the bottom of the chamber. The resistor divider chain wired to the teflon socket of the tube is enclosed in a small compartment, the inside of which is lined, as is the cooling jacket, with a thin sheet of mu-metal to exclude stray magnetic fields. A quantum efficiency enhancement cone, positioned so that its upper surface coincides with the exit pupil of the spectrometer, is optically contacted to the face of the photomultiplier with a drop of oil, and held in position by a light perforated cap of insulating material.

At the top of the chamber is a shutter which is remotely operated by means of a worm drive and synchronous motor mounted outside the cylinder. Microswitches limit its travel and control a pair of front panel mounted lamps, which indicate whether it is open or shut. In conjunction with the limit switches on the side-view mirror movement, they also form a safety interlock, preventing the shutter from being inadvertently opened while the side-view system is being used or vice versa.

3.3.3 Cooling System

The weak signal expected from the airglow made it necessary to cool the photomultiplier tube in order to reduce the dark current. According to the manufacturer's data a dark count rate of about 30 sec^{-1} at -20 C might be expected for a trialkali photocathode of the type used. In practice this figure appears to vary considerably between tubes, and was found to be generally larger for 9659's than for the less red-sensitive 9558's. The tube used had a ticket voltage dark current of about 0.02 nA when cooled to -20 C, while under operating conditions - i.e. 1200 V and a lower level discriminator setting of 200 mV - the dark count rate was about 100 sec^{-1}, some 350 times less than the room

temperature value. In the present application, where the signal usually exceeds the dark current by a factor of 10, there was little to be gained by operating below -25 C.

It was also found that the response of the tubes increased with cooling. For example two 9659's, one of which was the tube used for the airglow observations, showed a 35% increase in going from 23 C to -20 C. Similar effects have been seen by others (Murray and Manning, 1960), but the behaviour of the quantum efficiency is not clear. According to the manufacturers the overall gain should increase and the photocathode sensitivity decrease with decreasing temperature.

The cooling system is shared with the Mawson Institute 150 mm FPI, and is a modification of the system originally used with that instrument (Wilksch, 1975). A ¼ HP refrigerator unit with a capacity of about 100 W at -40 C and located outside the observatory cools a small reservoir of brine - a mixture by weight of 60% methanol and 40% water. Lengths of insulated copper pipe from the reservoir terminate close to the OH FPI in two self-sealing dripless couplings. The final short connection to the photomultiplier chamber is made with silicone rubber tubing to maintain the mechanical isolation of the instrument. A small gear pump, driven by a dc printed circuit motor run from a current limited 12 V supply, circulates the coolant. The satisfactory operation of the system is monitored by a pressure gauge and a thermistor embedded in the cooling jacket close to the PM photocathode. Its performance is dependent on the ambient temperature, so that the PM temperature can vary between -10 C and -30 C from one night to another. This can also cause quite significant changes in the tube dark count rate during a night's observation.

After some months of operation it was found necessary to take

steps to remove water vapour from within the chamber. When the coolant was turned on any moisture present condensed on the tube and cooling jacket, leaving the warmer face of the enhancement cone unfogged. However this condensation was found to collect at the base of the tube, leading to poor electrical contact, an erratic dark current and eventual corrosion of the contact pins. The solution finally adopted was to hermetically seal the chamber and purge it with dry nitrogen for several minutes prior to cooling the tube.

3.3.4 Hirschfeld Cone

In the past several schemes (Gunter et al., 1970; Aarons and Manuel, 1969; Hirschfeld, 1968) have been devised for enhancing the quantum efficiency of photomultipliers without increasing their dark currents. The technique involves introducing the light beam into the photocathode window at an angle in excess of the critical angle, so that the beam is ducted across the tube face and multiple contacts with the sensitive layer occur. Following the success obtained by Bower (1974) based on the ideas put forward by Hirschfeld, such a device was constructed for the OH FPI, and attempts were made to improve it.

The Hirschfeld cone, as it will be referred to, is shown in Figure 3.9(a). Using the notation illustrated there, it can be shown by a straightforward geometric argument that the relationship between W and w is

$$W = \frac{2\, w \tan \frac{\theta_C}{2}}{\tan(\theta_C + \frac{\Delta\theta}{2}) - \tan \frac{\theta_C}{2}} \qquad \ldots(3.2)$$

θ_C is the apex angle of the cone, and $\Delta\theta$ is the divergence of the beam in the cone material, the refractive index of which is μ_C. The minimum cone angle is

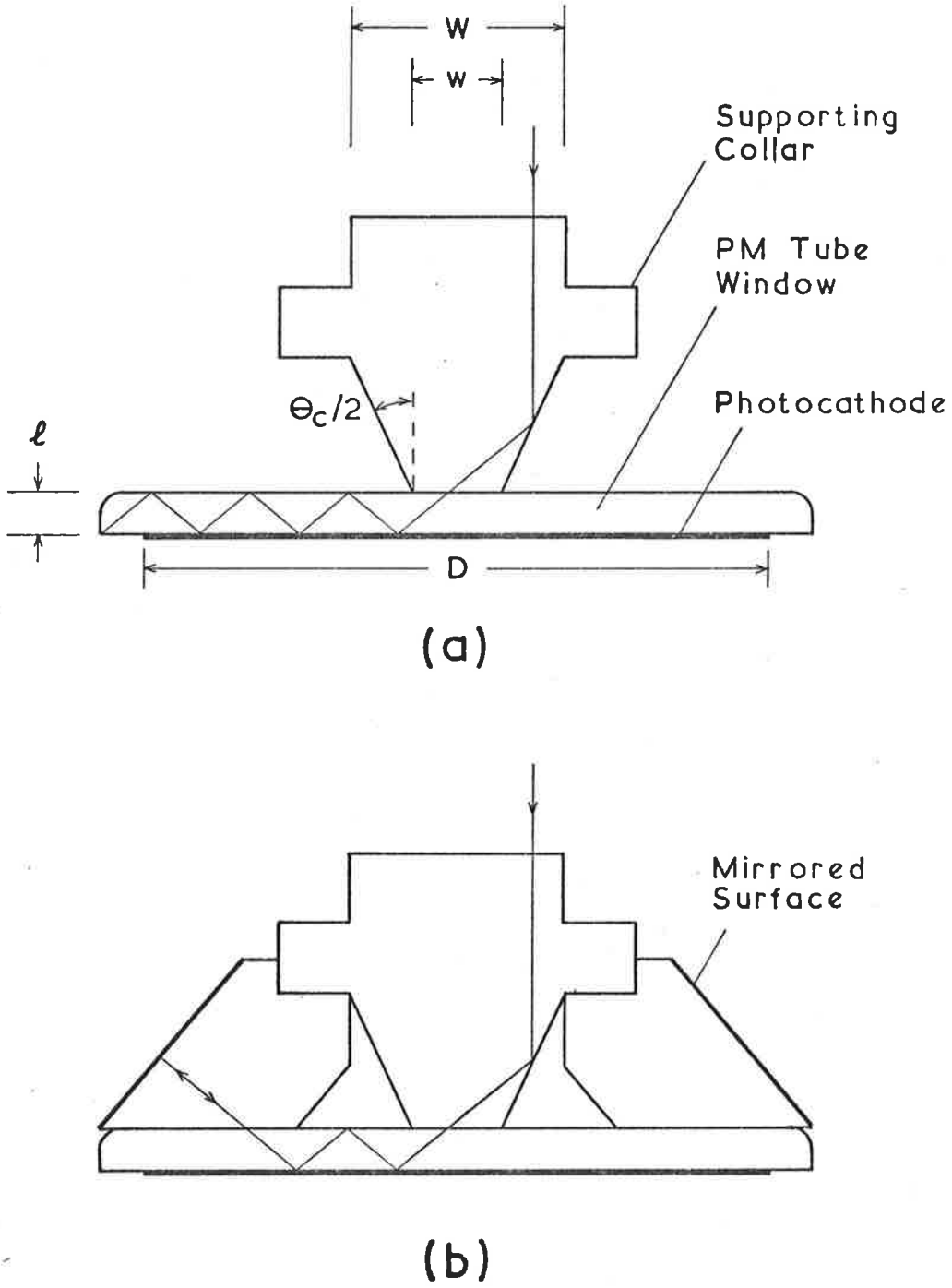


Figure 3.9 (a) The Hirschfeld cone
(b) The cone and annulus

$$\theta_{Min} = \arcsin\left(\frac{\mu_G}{\mu_C} \sin \theta_o\right) + \frac{\Delta\theta}{2} \qquad \ldots(3.3)$$

where μ_G is the refractive index of the window material and θ_o is the critical angle at the window/air interface - that is

$$\theta_o = \arcsin\left(\frac{1}{\mu_G}\right) \qquad \ldots(3.4)$$

The central portion of the beam will not be reflected from the walls of the cone, and will therefore undergo only one contact with the photocathode. Equation (3.2) shows that as θ_C is decreased, the fraction of the entrance area corresponding to this part of the beam decreases, and at the same time the number of bounces within the window increases, making it advantageous from this point of view to have a cone angle as close as possible to θ_{Min}. If ℓ is the thickness of the window, the maximum cone apex diameter is defined by

$$w_{Max} = 2\,\ell \tan\left\{\arcsin\left[\frac{\mu_C}{\mu_G} \sin\left(\theta_C - \frac{\Delta\theta}{2}\right)\right]\right\} \qquad \ldots(3.5)$$

which reduces to $2\,\ell \tan\theta_o$ if $\theta_C = \theta_{Min}$. A larger diameter would result in some rays being reflected from the photocathode back into the cone. Since the number of contacts z for a given beam is approximated by

$$z = \mathrm{Int}\left[\frac{\phi_{PM} - w_{Max}}{2\,w_{Max}}\right] \qquad \ldots(3.6)$$

where Int denotes the integral part of the bracketed expression and ϕ_{PM} the photocathode diameter, best results are obtained for tubes with a high ϕ_{PM}/ℓ ratio. On the other hand a large window thickness allows a larger diameter, and hence usually a smaller divergence beam to be used.

In the present case, however, the exit pupil of the spectrometer

(W + w = 12.7 mm) was already defined, as were the characteristics ϕ_{PM} (= 44 mm), ℓ (= 3 mm) and μ_G (= 1.467 at λ735 nm) of the photomultiplier. Accordingly a computer program was written to determine, by evaluating a more exact expression for z, the enhancement to be expected under these conditions and for different cone angles. It was found to be a monotonically decreasing function of θ_C, so an angle of 50° was adopted - a few degrees larger than θ_{Min} to allow a small margin of error. The agreement between the calculated (1.8x) and measured (Table 3.4) enhancements for three 9558B tubes was not particularly good, and in addition the result for the 9659B tube was considerably smaller than expected. It was found, however, that these measurements are quite susceptible to nonuniformities in the photocathode layer. For one tube not listed in the table an enhancement of 11x was obtained, which was caused by a small area of very low sensitivity at the centre of the photocathode. In regard to the low empirical enhancements found,it may be noted that although absorption at the photocathode was modelled in the calculations,its value was largely unknown, as was the refractive index of the thin layer of photocathode material itself.

TABLE 3.4

Quantum Efficiency Enhancement Obtained for Several Photomultipliers

Type	Serial Number	Enhancement
9558B	11609	2.8
9558B	9737	2.3
9558B	9964	1.8
9659B	24318	1.3

Hirschfeld has also suggested that the periphery of the photocathode window be surrounded by a mirrored cylinder, so that light reaching the edge could be redirected across the cathode. It

was not feasible to do this with the EMI tubes, the corners of which were rounded. Instead a truncated conical annulus, shown in Figure 3.9(b), was made from perspex, and its outer surface was coated with silver to obtain the desired reflection. However this device did not noticeably improve the enhancement obtained, although a modification of the above mentioned computer program (which took no account of skew rays) predicted an enhancement of about 3x. Another annulus, having a larger internal diameter than the first, likewise gave no improvement. These results suggest that there is scope for further investigation into enhancement techniques. The work of Jones (1976) indicates that better performance might be obtained at greater angles of incidence.

3.3.5 Detection Circuitry

A study made by Young (1969) of the comparative merits of the three basic signal detection methods - which he categorizes as direct current, pulse counting and the shot-noise-power method - has shown that at both high and low signal levels the second is the most efficient. When the signal and tube dark count rates are of the same order (as was the case for OH observations before the acquisition of a good 9659 tube), a significant advantage is obtained because pulse counting gives each photocathode event an equal weighting. In contrast sporadic large amplitude events, which are a characteristic of a photomultiplier's dark current behaviour, are integrated by the time constant associated with a current amplifier, leading to a serious degradation of the SNR. Pulse counting also provides the facility to discriminate against large and small pulses, thereby taking advantage of possible differences in the amplitude distribution of signal and dark current events to increase the SNR. A pulse

amplifier of fairly modest bandwidth was therefore built for hydroxyl observations, but was made readily interchangeable with an existing low current amplifier, which could be used for the high signal levels encountered when calibrating and testing out the spectrometer.

This analogue amplifier had been designed for several tilting filter photometers, and was suitable for measuring the 1 to 3 nA signals obtained with the OH FPI. It is based upon a Fairchild ADO-24 integrated circuit, connected such that the output voltage is proportional to the input current. A pair of front panel mounted rotary switches allow the gain and time constant to be selected over a range of 330 to 1 and 10 to 1 respectively, the former being done in such a way that the SNR remained constant.

The pulse amplifier/discriminator circuit is similar to that used in the 150 mm FPI (Bower, 1974). Its three stage preamplifier has a 1 kΩ input load resistor, a gain of 36 dB and a bandwidth of about 7 MHz. The lower level discriminator (LLD) consists of an integrated circuit comparator (μA710), the threshold of which is set by means of a trim pot. An upper level discriminator was not included because tests with a prototype had shown that its use did not significantly improve the SNR. The LLD output triggers a monostable (74121), which delivers a +5 V, 1.5 μs long pulse suitable for driving either of the two signal averaging instruments directly. The detection threshold of the circuit is 4 mV, referred to the input, when measured under test conditions similar to those encountered during OH observations - that is a 1 kHz repetition rate, 20 ns wide input pulses and a discriminator setting of 200 mV. Overload occurs above 2 V, while the noise threshold is 0.2 mV, again referred to the input. Under the same test conditions the pulse pair resolution is 1.5 μs, a limit set by the

monostable time constant, and which can be decreased to 70 ns by reducing the timing capacitor value. The slower response is, however, more than adequate to cope with the count rates to be expected from the airglow. A sufficient degree of noise immunity is obtained by careful circuit layout and by housing the amplifier in a shielded box immediately adjacent to the PM tube.

Attempts by Bower to measure the pulse height distributions of signal and dark counts from a 9558B tube in order to determine the optimum LLD setting met with little success. Instead he measured the SNR as a function of discriminator voltage, and a similar procedure was adopted in this case also. The measurements were made at -20 C and with a signal level - the difference between the count rate measured in the presence of light and the dark count rate - the same as that obtained from the OH band. At low LLD levels the dark count rate was comparable to or greater than the signal, and its non-Poisson nature made it necessary to determine the noise from the sample variance of a number of measurements. For large signal-to-dark current ratios the noise was with little error the square root of the total count rate. Curves of SNR versus LLD voltage for constant illumination and two PM tube voltages corresponding to 200 and 2000 A/ℓm, the latter being the rated tube maximum, appear in Figure 3.10. Similar results were obtained for two other tubes, a 9659B and a 9558B. They show that there is an advantage in operating the tube at as high a voltage as possible (see also Jonas and Alon, 1971). Because the observatory environment was electrically rather noisy, an LLD setting of 200 mV was adopted with little detriment to the SNR.

3.3.6 Signal Processing and Data Handling

Because of the relatively low intensity of the airglow and the

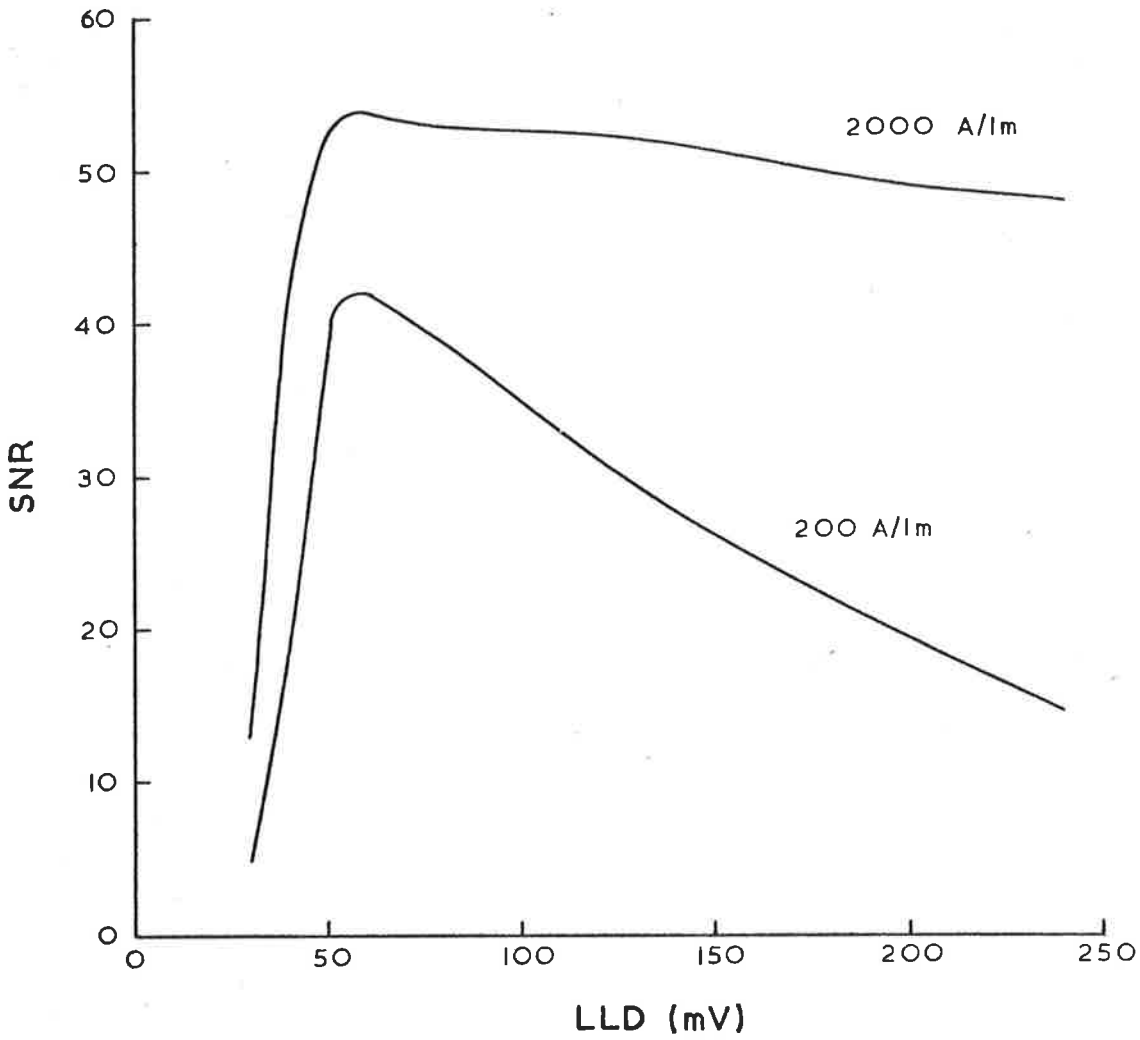


Figure 3.10 Signal-to-noise ratio (SNR) versus lower level discriminator setting (LLD) for a 9659B PM tube at an overall gain of 200 and 2000 A/lm.

speed with which it can vary, it was necessary to scan the interferometer quite rapidly, and to sum the results of each scan until a suitable SNR was obtained. Two instruments were available for this purpose, one of which, a Nuclear Data model 1100 Signal Analyser, was used only in the initial stages of the observation program. The other was a Data General Corp. Nova 2 minicomputer equipped with 8 K of memory and the necessary input/output circuit cards.

A diagram of the signal processing and data logging system is shown in Figure 3.11. A voltage controlled oscillator converted the analogue signal from the low current amplifier to a pulse train suitable for acceptance by the signal averager. When the pulse amplifier was used, its output was coupled to the Analyser through a pulse transformer, or to the Nova through an opto-isolator, to avoid ground loop problems. A ratemeter allowed the signal to be easily monitored as well as providing an analogue output if required.

The Analyser was used in multiscaling mode - that is, counts were accumulated in channel one at the beginning of a scan for a dwell time defined by the scan frequency and number of steps (128 or 256), after which a pulse from the scan generator advanced the channel address by one as the scan voltage was incremented to the next level. For most of the OH observations 128 channels were used per record,[+] although the Analyser had a capacity of 512 channels. At the end of a scan the process could be automatically recycled or manually interrupted. In the latter case the spectrum accumulated so far could then be displayed on an oscilloscope and/or read out via a Digideck digital tape

---

[+]The body of data which constitutes a single spectrum obtained by summing a number of scans will be referred to as a record.

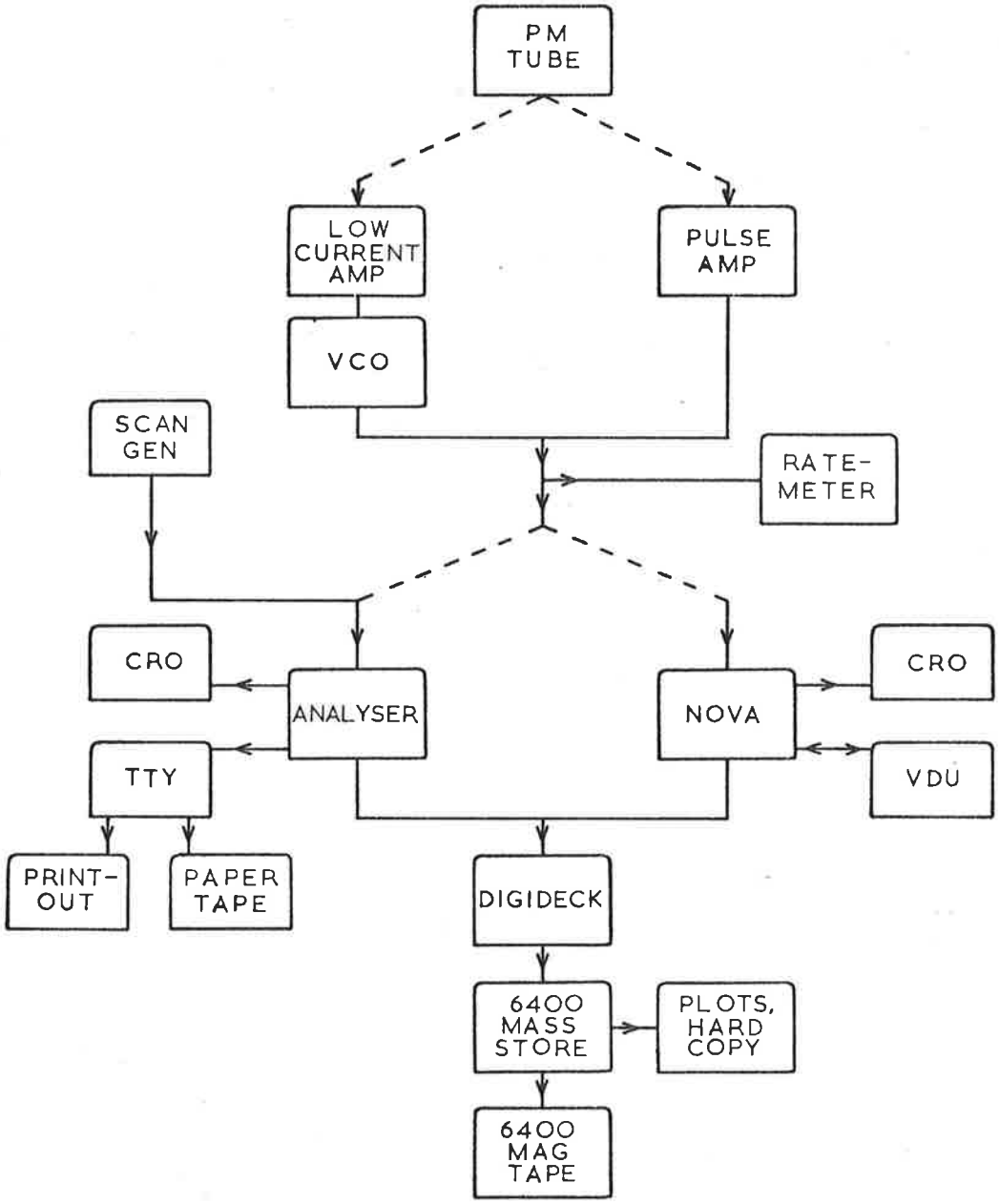


Figure 3.11 Block diagram of the signal processing and data logging system.

recorder onto a magnetic tape cassette. It was also possible to obtain a punched paper tape and/or hard copy from a Teletype model ASR33. However this facility was not used for routine observations, since the rapid readout and ease of handling and storage offered by the cassette system proved to be much more satisfactory.

The Nova was programmed to function in a manner basically the same as the Analyser - that is, it accumulated successive scans in a memory and output the record to tape. However its greater inherent flexibility allows several notable differences which will be briefly detailed. The execution of the data logging program is controlled from the keyboard of a visual display unit by means of a question-and-answer command structure. Since the efficient operation of the computer requires that the interferometer scan signal be internally generated, the user can select the dwell time, number of channels ($\leq$512) and number of scans per record. The Nova then remains the scan source for the full night's observations. Each record can be readily indexed with a formatted header which includes, for the OH observations, the date, start time and duration of the record, as well as details such as zenith and azimuth angles used. An autocycle mode is particularly useful because of the short (4.8 minutes) averaging times employed, allowing successive records to be regularly acquired under identical conditions without operator intervention. Each record is displayed on an oscilloscope as it is being accumulated. To reduce time lost between records two information buffers are used, one of which can be output to tape while counts are being accepted in another. It is eventually intended that an updated version of the program will be capable of servicing three similar experiments concurrently.

After a night's observations the data cassettes were taken back to the laboratory and copied onto mass storage on the University's CDC 6400 computer. This file was then edited to delete unwanted records, to correct errors generated in the transfer and to remove occasional severe noise spikes to which the detection circuitry was susceptible. The editing process produced a hard copy of the data as well as a line printer plot of each spectrum. All data was then stored on a 6400-compatible magnetic tape to await analysis. It was arranged such that each night's data was easily accessible, with the records indexed by the Nova-generated header but maintained in chronological order.

3.4 Adjustment and Performance

3.4.1 Parallelism

The degree to which lack of parallelism affects the width and peak transmission of the instrument profile of an FPI has been investigated theoretically by several authors, including Chabbal (1958) and Ramsay (1969). Their results are in approximate agreement although those of the latter were obtained for a square aperture only. Treating departures from parallelism as giving rise to a second defect function, Chabbal finds for the corresponding defect finesse

$$N_D' = \frac{n}{\sqrt{3}} \qquad \ldots(3.7)$$

where the maximum departure of the spacing at the edge of the plate from the mean is expressed as $\delta x' = \lambda/2n$. This compares with the case of a spherical defect, for which $N_D = n/2$ and the sagitta of the curvature is $\delta x = \lambda/n$. In order that lack of parallelism should not contribute significantly to the overall finesse, therefore, $\delta x'$ should

be rather less than δx.

Two methods were considered for visually checking the etalon alignment, one employing Fabry-Perot, the other Fizeau fringes. In the first case departures from parallelism could be deduced from the change in fringe diameters as the eye was scanned across the etalon along one of two perpendicular diameters. Changes could most easily be discerned if the spacing was chosen such that the exploratory fringe was just large enough to enclose a small dark spot. This method was satisfactory when working at an order of more than 100, being sufficiently sensitive to reveal the larger irregularities in the plate surfaces. However at the small spacing employed in the present application, the angular width and diameter of a fringe becomes too great for the accuracy required. The continuous motion of the Fabry-Perot pattern as the plates are scanned, and the difficulty of judging a trend across a plate diameter in the presence of surface defects emphasize this drawback.

The alignment technique adopted, involving the use of Fizeau fringes, has been referred to in Section 3.1.2.2, and the optical system required has been described in Section 3.2.1. A cadmium lamp was used as the source, since at the operating order it gave the best available distribution of lines through the scan. In addition the short wavelength of two of the lines (λ468 nm and λ480 nm) improved the sensitivity of the method. A check made with the red lines of the neon lamp did not reveal any other adverse dependence of defect function on wavelength. Since the Fizeau fringe of a given line mapped the plate defects as intensity contours, the alignment procedure simply consisted of adjusting the X and Y parallelism controls so that each fringe appeared and disappeared as uniformly as

possible over the area of the etalon during a scan. Misalignment was shown by a tendency for the fringes to move across the field of view. On occasions when the mean ceramic voltage had been altered (Section 3.1.1) this procedure was modified in that the parallelism controls were used only at the beginning of a scan, while adjustments at the end were made with the ceramic gain controls. This process converged to the optimum gain settings after only a few iterations. The precision of the Fizeau fringe method is assisted by the fact that continuous scanning is a necessity rather than a hindrance, and that information about the parallelism is obtained from the whole etalon area, rather than from two observations at opposite edges of the plate.

The accuracy with which the parallelism could be set is readily found by measuring the change in ceramic voltage required to produce a detectable difference in the Fizeau fringe pattern, and calibrating this voltage in units of length by scanning over one order. The results obtained for the OH FPI under field conditions - that is, with the instrument well temperature stabilized and a scan range of one third of an order - were a repeatability in setting the parallelism of about $\lambda/250$ ($= 2\delta x'$), a variation in parallelism from beginning to end of a scan slightly greater than this, and a maximum variation during the night of $\lambda/125$ at worst. From Chabbal's (1953) curves and measurements of the defect finesse (Section 3.1.2.2) it was therefore expected that the instrument peak transmission would vary by less than 1% over a scan.

The variation in the width i of the instrument profile over a period of 6 hours was found by recording at intervals the scan through the laser line, and determining the peak width by means of a least squares fit. The results appeared to be randomly distributed with a

variance of 3×10^{-4}% of the peak width. It was expected that i, as determined by the functions A, D and F, should be only slightly wavelength-dependent, but recordings of the laser line made at different mean etalon spacings showed that the measured width changed by ∿10% from beginning to end of the scan. This was considerably larger than could have been caused by parallelism variations of a magnitude not detectable by the Fizeau fringe method. In fact it was attributed to the nonlinearity of the scan - such a width change is obtained for a quadratic linearity error if the maximum departure from linearity is ∿1% of full scale - and had to be taken into account in the data analysis as described in Section 5.4.

3.4.2 Separation

The nature of the experiment did not require that the operating order be determined with any great accuracy, the chief concern being to obtain a free spectral range no smaller than that specified. Because the required order was so low (44), considerable care had to be exercised when first setting up the etalon to ensure that the coated surfaces did not touch. In fact the margin for error was somewhat less than the above figure would suggest owing to the presence of dust particles on the plates. The desired spacing was therefore approached cautiously from above. In doing so it was found convenient to use two methods to determine the order. At first the diameters of the Fabry-Perot fringes were measured, a procedure which proved to be rapid and accurate when the order was greater than ∿100. The photomultiplier housing was replaced by a sodium lamp, a wide-angle (92°) camera lens focussed on infinity was positioned above the etalon, and the circular fringe pattern was imaged on a translucent screen. The diameters Φ_i of successive fringes of the same component of the sodium doublet

could then be measured with a pair of vernier calipers, and the order m_o found from the small angle approximation

$$m_o \simeq \frac{8f^2\mu^2}{\Phi_{i+1}^2 - \Phi_i^2} \qquad \ldots(3.8)$$

Here f is the focal length of the camera lens.

At an order of less than about 100 the fringes became too large to be imaged, and the remainder of the adjustment was done by counting orders as the separation was decreased. A final check was made by illuminating the interferometer from above with both the He-Ne laser and an Hg198 source filtered to give a single narrow line at λ546.1 nm. The etalon was scanned over slightly more than two peaks of the laser line, and the position of the mercury line within this scan was determined in the usual fashion. A similar relative orientation is repeated every 6 orders, but is only obtained again almost exactly when the separation is changed by 63 orders. The order could therefore be precisely identified with little difficulty.

It was determined from the data analysis program that inaccuracies of 0.01 nm in the absolute wavelength scale of the spectrometer, which correspond to separation drifts of about 0.2 nm, could not be tolerated. Fortunately this degree of stability was only necessary for the duration of one record (usually 10 minutes), since the OH lines themselves could be used for wavelength calibration. Nevertheless, by recording a scan of the laser line at frequent intervals and determining the exact position by means of a least squares fit, the separation was found to drift less than 0.4 nm over a period of 6 hours, provided that the FPI temperature was well regulated. As long as the etalon did not suffer severe thermal or

kinetic perturbations, it appeared that after a "bedding down" period of about a week following major adjustment, the separation would remain within one fifth of an order of the set value indefinitely. It was thus usually necessary to verify the operating order before using the spectrometer merely by ensuring that the cadmium lines used in the parallelism adjustment occurred in correct sequence in the scan.

3.4.3 Finesse and Transmission

Because of the lack of a suitably narrow spectral source having an isolated line close to the operating wavelength of the instrument, finesse measurements were made at λ632.8 nm with the He-Ne laser. The figures obtained were therefore about 12% lower than would be expected at λ735 nm. The etalon finesse N_E, obtained with a pinhole field stop and by averaging the measurements from both peaks in the scan (Section 3.1.1), was ∿22. Since the reflectance of the plate coatings was specified to be 0.94, giving a reflective finesse N_R of 50, the defect finesse could be found from Chabbal's (1953) curves. Its value fell between 24 and 30, where these limits correspond to a defect function which is either rectangular or Gaussian. The exact shape and width of D(λ) could, if necessary, be found by one of the methods given in Section 3.1.2.2. The mismatch between N_R and N_D has been discussed in Section 3.1.2.1. To obtain the desired instrument profile width i = 0.98 nm, an aperture function width f = 0.60 nm was adopted. The measured overall finesse N_I was then ∿17.

An estimate of the instrument transmission τ_I was made on the assumption that the parameters of the coatings were in fact as quoted by their manufacturer. The mean instrument transmission over one complete order - also called the ordinal area A_Δ - was calculated from a scan through two successive peaks of the laser line. Since this

result is independent of broadening by the defect and aperture functions, and is given by the relation

$$A_{\Delta} = \frac{\underline{T}^2}{1 - \underline{R}^2} \quad , \qquad \ldots(3.9)$$

the laser record could be scaled in absolute transmission units. The peak transmission τ_I was found to be ∿29% at λ632.8 nm, in good agreement with an estimate of about 30% derived from Chabbal's curves on the basis of the above mentioned finesse measurements.

3.5 Airglow Observations

3.5.1 Observational Practice

All measurements of the hydroxyl airglow were made from the observatory at Mt. Torrens, the geographic coordinates and elevation of which are 34° 52.2' S, 138° 56.6' E and 590 m above sea level. They should ideally have been made on clear, dry, moonless nights, but since these conditions were extremely rare it was endeavoured to obtain useful results on all clear nights (which occurred only slightly more frequently). Concurrent observations with two λ630.0 nm and λ557.7 nm tilting filter photometers (Schaeffer, 1970) also served to detect auroral activity or the presence of clouds in the field of view. Late in the observational program the Mawson 150 mm FPI became available to measure temperatures and winds using the green oxygen [OI] emission, which arises from heights close to the OH layer. Most of the observations, both with the spectrometers and the photometers, were made at a zenith angle of 55° and an azimuth of 180° - that is, towards the south celestial pole - so that the stellar background was small and fairly constant. This choice also resulted in an increase in the signal strength over the zenithal value by a factor of about

1.7 due to the van Rhijn effect.

A 128 channel record and a dwell time of 200 ms were adopted for the observations. The resulting scan period was a compromise between being able to set the parallelism and avoiding the effects of rapid changes in the band intensity. The scan range eventually settled upon was 6.0 nm, and covered the P branch of the (8 - 3) band from the $P_2(2)$ line to just beyond the $P_2(4)$ line. The range was less than that for which the spectrometer had been designed because of the discovery, later confirmed by Armstrong and Meriwether (private communications), that at least one component of the lamda-doubled $P_1(4)$ line is quite strongly absorbed by water vapour. This effect could not be satisfactorily modelled in the nonlinear fitting procedure which was found necessary to analyse the data (Section 4.3.4), and the contribution of the $P_1(4)$ line to the record therefore had to be minimized. The averaging time used per block was 256 seconds (10 scans), which resulted in a temperature error of ±2 to 7%, depending on the band intensity. However, successive blocks could be summed on the 6400 computer to reduce this error if it were so desired.

Prior to the commencement of a night's observations the parallelism and separation had to be adjusted. Thereafter it was usually unnecessary to reset the parallelism, although it was convenient to check it when dark count measurements were being made. The separation adjustment consisted of suitably positioning the laser peak near the centre of the scan. This record was used as a start point for the absolute wavelength calibration procedure carried out by the analysis program, but because the OH line wavelengths were known no further such measurements were needed. It was important that these

adjustments be made some 15 to 30 minutes before the first OH observation to allow for the creep of the ceramics. To assist in maintaining good thermal stability the PM tube coolant was usually allowed to run continuously. However because of the sensitivity of this system to the outdoor temperature, dark count records had to be made at intervals of about two hours during the night.

3.5.2 Instrument Specifications

The following table summarizes the instrument parameters used for the majority of the OH observations.

TABLE 3.5

Prefilter:	type	3 period interference
	centre wavelength	735.3 nm
	bandwidth	10.7 nm
	peak transmission	0.84
Etalon:	plates	51.2 mm diameter Homosil
	working area (S)	1900 mm^2
	flatness uncoated	better than $\lambda/100$ at $\lambda 632.8$ nm
	coated	$\sim\lambda/56$ at $\lambda 735$ nm
	coatings	30 nm Ag + $\lambda/4$ MgF_2 at $\lambda 356$ nm + $\lambda/4$ CeO_2 at $\lambda 500$ nm
	reflectance (R)	0.940
	transmittance (T)	0.051
	absorptance (A)	0.009
	spacing (ℓ_o)	16.2 μm
	operating order (m_o)	44
	free spectral range ($\Delta\lambda$)	16.7 nm
Optics:	collimating lens	230 mm, f4.3
	field stop diameter	18.5 mm
	field of view (full angle)	4.6°
	solid angle	5.1×10^{-3} sr

Airy function:	finesse (N_R)	50
	width (a)	0.33 nm
	peak transmission (τ_A)	0.71
	ordinal area (A_Δ)	0.0222
Defect function:+	finesse (N_D)	∿28
	width (d)	0.60 nm
Etalon function:+	finesse (N_E)	22
	width (e)	0.76 nm
Aperture function:	width (f)	0.60 nm
Instrument function:+	finesse (N_I)	17
	width (i)	0.98 nm
	peak transmission (τ_I)	∿0.3
Detector:	type	EMI 9659B photomultiplier, S/N24318
	photocathode	trialkali extended S-20
	quantum efficiency at 23 C	7.2%
	. . . with Hirschfeld cone++	9.4%
	dark count rate at -20 C	∿100 s^{-1}
Data logging:	number of channels	128
	dwell time per channel	200 ms
	wavelength interval per channel	0.047 nm

+These measurements were made at the He-Ne laser wavelength of λ632.8 nm and corrected to λ735 nm.

++Probably less at -20 C.

CHAPTER 4

DATA ANALYSIS

4.1 Introduction

Before embarking on a description of the technique used to extract rotational temperatures from the hydroxyl observations, it is useful to summarize the character of the actual airglow emission and the recorded spectrum. Within the 6.0 nm wide wavelength interval covered by the spectrometer scan lie the first two P_1 lines of the OH (8 - 3) band, the first three P_2 lines and the $Q_1(4)$ line. Other Q branch lines are too weak to contribute significantly. Immediately outside this interval, and sufficiently close to be detected by the wings of the instrument profile, are the $Q_1(3)$ and $P_1(4)$ lines, the latter of uncertain intensity, being subject to atmospheric water vapour absorption. Also present is a continuous background spectrum which, because of the elevation of the observing site, local weather conditions and the resolving power of the spectrometer, was likewise found to be significantly influenced by water vapour.

A typical recorded spectrum, obtained as a low resolution computer plot of the raw data, is shown in Figure 4.1(a). The wavelength scale increases from right to left in steps of about 0.05 nm per channel. The $P_1(2)$ and $P_1(3)$ lines dominate the picture, but each P_2 line except the first, $P_2(2)$, which is masked by the edge of the prefilter, can be distinguished also. The structure of the water vapour spectrum cannot be discerned unless the background is very strong (for example in early twilight).

The extent of the wavelength scan is rather less than that for which the instrument had been designed (Section 1.4) owing to the effects of water vapour. However the increased time resolution achieved

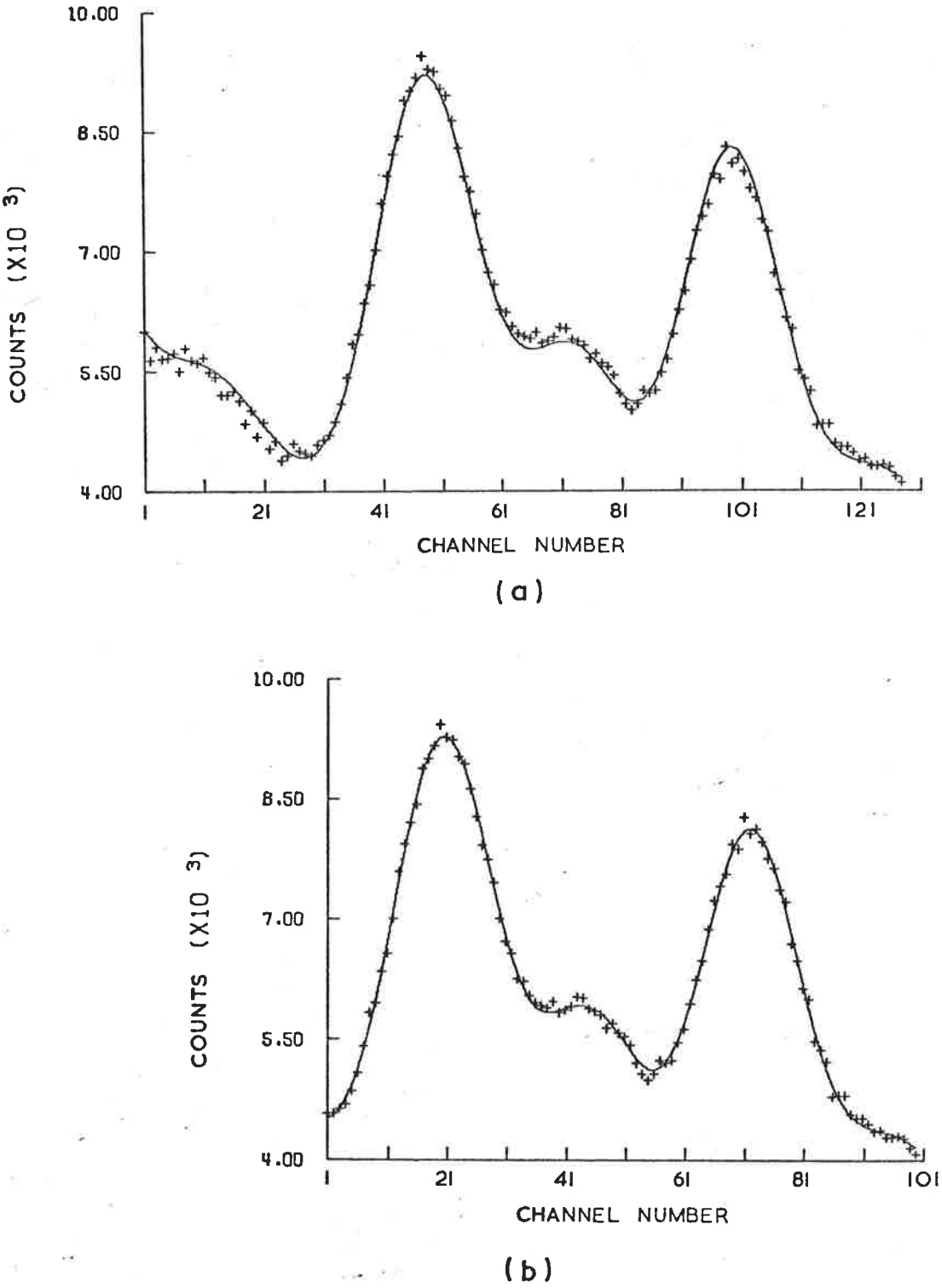


Figure 4.1 Plot of a recorded spectrum and its least squares fit

(a) Full 128 channels. $Q = 2.33$ $T_R = 177K$

(b) Truncated 100 channels. $Q = 1.02$ $T_R = 197K$

with the shorter scan to some extent compensates for the loss of information regarding the exact intensity distribution of the OH lines. In fact it was found that the data interval had to be still further reduced by numerically truncating the leftmost 28 channels in order to reduce to acceptable levels the influence of the $P_1(4)$ line and the initial nonlinearity of the scan (Figure 4.1(b) and Section 4.5).

In this chapter consideration will be given to the possible methods of analysing the recorded spectra with the aid of a digital computer. The model of the data and the least squares fitting procedure required in the method adopted will be presented, and the derivation and precision of the functions and constants required for use in the model will be discussed.

4.2 Selection of the Method of Data Reduction

Because of the large spectrometer bandwidth employed, it is not possible to obtain useful relative intensity information from the raw data - each OH peak contains counts both from adjacent spectral lines and the unknown airglow background. Several approaches to the problem will be considered briefly. The simplest, and one used by several other workers who have obtained low resolution OH spectra (Section 1.3.1), is to visually compare the recorded spectrum with synthetic spectra calculated for a number of different rotational temperatures. However this method requires that the shape and level of the background signal be known, is not very effective when only a small portion of the band is observed, and is less accurate than the least squares method to be described below.

For some problems (for example that encountered in the measurement of upper atmosphere winds and temperatures using the [OI] emission lines)

it has been possible to remove the effects of the instrument response $I(\lambda)$ from the recorded function $Y(\lambda)$ by direct deconvolution using the Fourier transform. If the source spectrum is $B(\lambda)$ then Y is described by equation (2.11). Since it is a property of the Fourier transform (Bracewell, 1965) that

$$Y' = B'I' \qquad \ldots(4.1)$$

B may be found, at least in principle, as

$$B = F^{-1}\{Y'/I'\} \qquad \ldots(4.2)$$

where Y', B' and I' represent the transforms of Y, B and I respectively, and $F^{-1}\{\ \}$ denotes the inverse Fourier transform operation. However when the width of I is large compared to the width of B, which in the case of the OH spectrum is effectively the P line width, complete recovery is fundamentally impossible (Horlick, 1972). The reason is that in the transform domain B is much broader than I, so that when performing the multiplication of equation (4.1) some of the information about B is lost. The numerical filter which must be applied to the quotient in (4.2) to reduce the effects of noise, and the division by small numbers can also result in considerable distortion of the deconvolved function. In practice a reduction in peak width at half height of about 25% can reasonably be expected for an isolated line - certainly much less than the factor of four or more required to reduce contamination from adjacent P lines to an acceptable limit.

The remaining alternative is to model the recorded spectrum as accurately as possible, and to determine all the unknowns of the model by means of a numerical fit to the data. For the present problem either a linear or a nonlinear model could be considered, and since they are similar in many respects, both will be described before discussing the

reasons for the choice of the second.

4.3 Modelling the Recorded Spectrum

4.3.1 Hydroxyl Lines

The contribution of the OH lines to the recorded spectrum is quite readily determined, since they are very narrow compared to the spectrometer bandwidth, and their wavelengths λ_k (or more importantly their relative spacings) are known. They therefore appear in the model as the sum of a number of instrument profiles suitably positioned on the data interval, and with relative amplitudes distorted only by atmospheric (aerosol) scattering, the prefilter and the spectral response of the PM photocathode. According to the curves of Harrison and Kendall (1973b), the effect of aerosol scattering is negligible, amounting to less than a 0.4% intensity change across the scan. In the linear model the measured relative intensities a_k are independent parameters of the fit, so that the OH line component of the recorded function can be written as

$$\hat{Y}_{OH}(\lambda) = \sum_k a_k \, I(\lambda - \lambda_k) \qquad \ldots(4.3)$$

the sum being over the number of lines considered to contribute significantly to the data. The rotational temperature can then be found by the log plot method described in Section 1.3.1.

In the nonlinear model on the other hand, the P_1 and P_2 line intensities are separately described by equation (1.2) using a single rotational temperature. That is, it is assumed that the $^2\pi_{3/2}$ and $^2\pi_{1/2}$ states are separately in thermal equilibrium, but, in consideration of the conflicting results obtained by others for the doublet ratio temperature (Section 1.3.2), no assumption is made about transitions between these states. The parameters of the fit are then the relative

intensities of the $P_1(3)$ and $P_2(3)$ lines, denoted by a_1 and a_2, and the rotational temperature a_3. Because they also arise from the $^2\pi_{3/2}$ state, the lines of the Q branch are described by the same temperature and have intensities fixed relative to the P_1 branch intensity. For this model the OH component of the recorded spectrum can therefore be written as

$$\hat{Y}_{OH}(\lambda) = a_1 \sum_{\ell} c_{\ell} \exp\left(\frac{c'_{\ell}}{a_3}\right) I(\lambda - \lambda_{\ell}) + a_2 \sum_{m} c_m \exp\left(\frac{c'_m}{a_3}\right) I(\lambda - \lambda_m) \quad \ldots(4.4)$$

where the first sum is of the P_1 and Q lines and the second is of the P_2 lines. The constants $c_{\ell,m}$ and $c'_{\ell,m}$ are obtained from equation (1.2) and the molecular constants given in the Appendix.

4.3.2 Background

The background emission, denoted by $C(\lambda)$, arises primarily from three sources, the airglow continuum, both direct and scattered starlight, and the zodiacal light. The measurements of Gadsen and Marovich (1973) and Broadfoot and Kendall (1968) show that at the wavelength of interest and for the 1 nm resolution employed, the first of these exhibits a smooth variation which increases with wavelength by about 5% over the scanned interval. The spectrum of the other two sources is also expected to vary slowly at this resolution, though the slope of the variation appears to be a little uncertain. As for the OH lines the effect of aerosol scattering is negligible, so that in the absence of atmospheric water vapour the background is expected to be a linear function of λ, that is

$$C(\lambda) = a_k + a_{k+1}\,\lambda \quad \ldots(4.5)$$

a_k and a_{k+1} being parameters of the fit. Its contribution to the

recorded spectrum can be written as

$$\hat{Y}_C(\lambda) = [C(\lambda)\ G(\lambda)]\ *\ I(\lambda) \qquad ...(4.6)$$

where $G(\lambda)$ represents the prefilter transmission function multiplied by the photocathode response.

Since the convolution integral in the above expression cannot readily be manipulated, and since the slope of the background was expected to be slight, the approximation

$$\hat{Y}_C(\lambda) \simeq C(\lambda)\ [G(\lambda)\ *\ I(\lambda)] \qquad ...(4.7)$$

was made. By calculating $\hat{Y}_C$ according to the above two expressions using the discrete Fourier transform it was found that the error incurred for reasonable values of C was less than 0.3% of $\hat{Y}_C$. This approximation also allows the use of $G * I$ as a fitting function, which can be measured directly and with much greater accuracy than can G alone (Section 4.5). Equation (4.7) is applicable to both the linear and nonlinear model.

Unfortunately it was found that the effects of atmospheric water vapour absorption were rarely if ever absent from the OH observations. This aspect of the background spectrum is discussed in the following section.

4.3.3 Water Vapour Transmittance

Poor values of Q (equation (4.20)) resulting from attempts to fit the functions developed thus far to the airglow data, and comparisons between spectra obtained using the sun and a tungsten filament lamp as sources, showed that the model was in some respect in error. A transmission function was therefore included to account for possible

atmospheric water vapour effects. If this function, measured at a resolution much less than the width of an individual absorption line, is denoted by $T(W, \lambda)$, where W is the total effective absorber amount present in the optical path, then the background contribution to the recorded spectrum becomes

$$\hat{Y}_C(\lambda) = [C(\lambda)\ G(\lambda)\ T(W, \lambda)] * I(\lambda) \quad . \qquad ...(4.8)$$

(T does not affect the OH contribution to $\hat{Y}$ except in respect to the $P_1(4)$ line). Again the convolution is intractable to the analysis, making an approximation of some kind unavoidable. Fortunately this can be done because the function G is flat over most of the data interval. Computer calculations were made using the high resolution water vapour transmission spectra of Curcio et al. (1964) as the function T, and the discrete Fourier transform to evaluate the convolution. They showed that for the largest expected background slope and water vapour absorption the approximation

$$\hat{Y}_C(\lambda) \simeq C(\lambda)\ [G(\lambda) * I(\lambda)]\ [T(W, \lambda) * I(\lambda)] \qquad ...(4.9)$$

resulted in an error which was generally much less than 1%. It was also found that a better approximation was obtained using the function (G * I) rather than G itself, thus permitting an additional source of error to be avoided.

The quantity which had to be determined was therefore $\overline{T}(W, \lambda) = T(W, \lambda) * I(\lambda)$ - that is the water vapour transmission spectrum measured with a resolution comparable to that employed for the OH spectrometer. The construction of a suitable function from published constants of the individual water vapour lines would have been extremely laborious, and was not considered. Three other approaches to the problem were attempted - to use existing results for the λ and

W dependence of the smeared transmittance function, to use this W dependence in conjunction with a measured λ dependence, and to obtain the entire function empirically.

There appear to have been very few measurements made of atmospheric transmittance spectra of the type required for this problem. Those of Taylor and Yates (1957) and Yates and Taylor (1960), while having an appropriate resolution, do not yield the required dependence on absorber amount. McClatchey et al. (1972), using laboratory data and line by line transmittance calculations degraded in resolution to 20 cm^{-1}, have obtained a nonanalytic functional relationship between $\bar{T}$ and W based on a model for the water vapour bands of the form

$$\bar{T} = f(\psi(\lambda)\Delta L p^{u}) \quad . \qquad \ldots(4.10)$$

Here ΔL is the effective path length (or absorber concentration), p is the pressure raised to a power $u \lesssim 1$ which was determined empirically by McClatchey et al., and $\psi(\lambda)$ is an averaged absorption coefficient. To deal with slant paths through the atmosphere an equivalent sea level absorber amount W (kg m^{-2}) may be defined such that

$$W = \sec \phi \int_{z}^{\infty} w(z) \left(\frac{p}{p_o}\right)^{u} dz \qquad \ldots(4.11)$$

where w(z) is the water vapour concentration in kg m^{-3} at a height z, p_o is the sea level pressure and ϕ is the zenith angle of the observation. The resulting transmittance function

$$\bar{T}(W, \lambda) = \bar{T}(\log W + \log \psi(\lambda)) \qquad \ldots(4.12)$$

is tabulated, together with values for log $\psi(\lambda)$, by Selby and McClatchey (1975). However, reliable values of atmospheric water vapour content corresponding to the periods of OH observation were not available, making it necessary that W be a parameter of the airglow model. In

addition it was not practicable to completely recalculate the function $\overline{T}$ for each new value of W according to equation (4.12) as frequently as was required by the fitting procedure. $\overline{T}$ was therefore approximated by

$$\overline{T}(x) = 1 - c_1 \exp(c_2 x) \qquad \ldots(4.13)$$

and least squares methods were used to determine the coefficients c_1 and c_2, so that a relationship of the form

$$\overline{T}(W, \lambda) = 1 - h(W) \quad H(\lambda) \qquad \ldots(4.14)$$

was obtained. On the basis of the atmospheric transmittance curves published by Taylor and Yates (1957) and estimates of W calculated from local meteorological records,[+] the approximation was confined to the region of the curve above $\overline{T} = 0.78$, giving an expected maximum error of less than 2% of $\overline{T}$ at nearly all points in the data interval. The function $H(\lambda)$ was then calculated from equations (4.12) and (4.13) and used in the model (equation (4.9)). However the fit, though improved, was still not considered to be satisfactory. In view of the fact that there is quite good overall agreement between Selby and McClatchey's (1975) results and those of Yates and Taylor (1960), and also because of considerable differences between the model and a spectrum of the daylight sky obtained with the OH spectrometer, it seemed likely that the λ rather than the W dependence of $\overline{T}$ was in error.

For an instrument resolution of 1 nm the sun (or moon) was a suitable source for obtaining an alternative absoprtion coefficient function $\psi(\lambda)$. Over the region of interest the sun's intensity

---

[+]Obtained by numerically integrating equation (4.11), using for w(z) radio-sonde measurements of absolute humidity made by the Bureau of Meteorology at Adelaide (40 km from the Mt. Torrens observatory) at 1000 U.T. and 2200 U.T. each day.

decreases by only 2%, and an examination of the Fraunhofer structure (Moore et al., 1966) showed that its spectrum in the absence of atmospheric absorption would be smooth to within about 1%. The recorded function Y_{Co} resulting from an observation of the sun is

$$Y_{C_o} \simeq C_o \cdot T_A \cdot [G * I] \cdot \overline{T}(W_o) \qquad \ldots(4.15)$$

where W_o is the amount of water vapour present in the path at the time, C_o is the unattenuated solar spectrum and T_A represents the spectrally flat aerosol transmittance. The function $H(\lambda)$ of equation (4.14) is then

$$H = \frac{1}{h(W_o)} \left[1 - \frac{Y_{Co}}{C_o \cdot T_A \cdot [G * I]} \right] . \qquad \ldots(4.16)$$

Since T_A was unknown it should strictly have been an additional parameter of the fit, but by generating a synthetic spectrum which had the maximum expected values of T_A and water vapour absorption, and fitting to this a model having $T_A = 1.0$, it was established that negligible error would be incurred by omitting T_A altogether. This both reduced computation and made the behaviour of the model easier to interpret. W_o was also unknown, but could be incorporated into the parameter describing W. The overall background component of the nonlinear model was then

$$\hat{Y}_C = [a_4 + a_5\lambda] [G * I] [1 - a_6 H] \qquad \ldots(4.17)$$

using the sloping background source spectrum of equation (4.5) for $C(\lambda)$.

The results obtained with this model suggested that it was still in some respect inaccurate. An attempt was therefore made to verify the dependence of $\overline{T}$ on W by measuring the solar spectrum as a function of zenith angle. The approximate relative absorber amount for each record was calculated from the van Rhijn function, assuming, on the

basis of the standard atmospheres presented by McClatchey et al. and local radio-sonde data, that the majority of the absorption would occur in a layer about 2 km thick at the earth's surface. Three of these solar spectra corresponding to a representative sample of absorber amounts (path lengths) were then selected and used to generate three functions $H(\lambda)$ according to equation (4.16). A series of background spectra of the form of (4.17) were constructed from each sample for a progression of values W (= a_6) and compared with the appropriately scaled solar spectra. It was found that quite good agreement could be obtained providing the calculated W values of the solar spectra were not considered to be exact. Indeed little else could be expected in view of the inaccuracy of the van Rhijn method, the lack of an exact datum for W and the assumption involved of the stability of the atmosphere with respect to W during the observing period. The approximations contained in equation (4.14) were therefore confirmed for a range of W of about 4, where W for the OH observations was expected to lie between 1 and 6. A representative H function was constructed for use in the analysis, and equation (4.17) was retained as the best available model of the effects of water vapour. On the basis of the above comparison of spectra the error in the transmittance $\overline{T}$ was expected to be generally much less than 5%. This figure would only be reached for W close to 1 or 6 and over the last few channels of a record. Figure 4.2 shows $\overline{T}$ for W $\doteq$ 1.5 and a 128 channel interval corresponding to 6.02 nm.

The full nonlinear model was thus the combination of equations (4.4) and (4.17), that is

$$\hat{Y} = \hat{Y}_{OH} + \hat{Y}_{C} \quad \ldots(4.18)$$

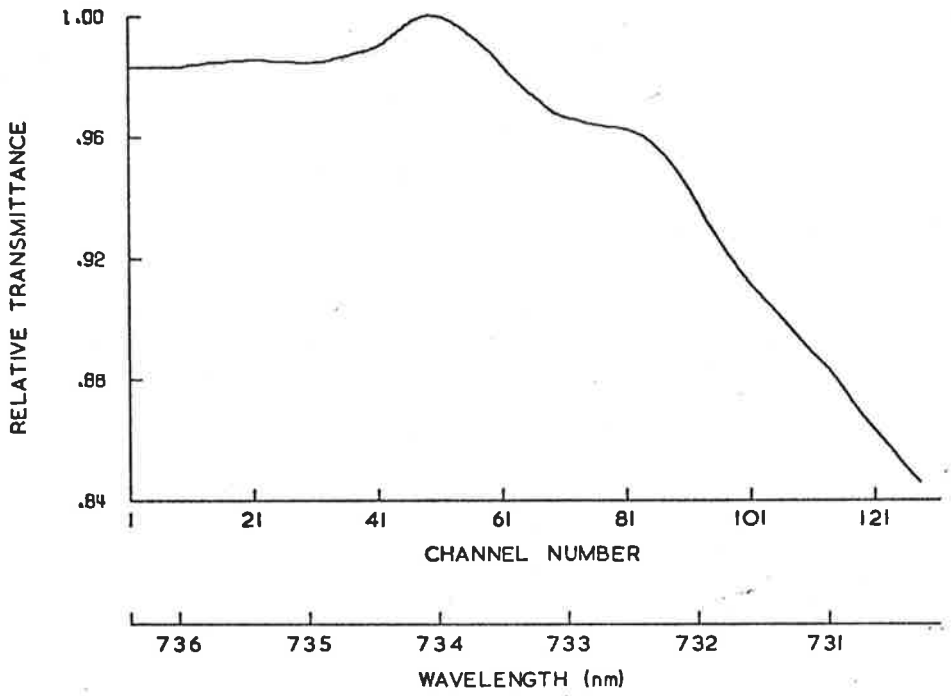


Figure 4.2 Water vapour transmittance function $\bar{T}$ (W,λ) over 128 channels (6.02 nm) for W = 1.5

4.3.4 Adoption of the Nonlinear Model

The linear model has two very desirable characteristics, namely that the computation required to fit it to the data is relatively straightforward and rapid, and that it yields the relative amplitude of each emission line independently, thereby both avoiding errors arising from, and providing additional information on, possible departures of the emitting population from thermal equilibrium. This allows the data interval to be extended beyond the $P_1(4)$ line despite its being masked by a water vapour line.

However it is apparent from equation (4.17) that as long as the total absorber amount is non-zero and unknown the model must be non-linear in the fitted parameters. Some experimentation with the OH data suggested that even if a reliable value of W were available, or if some other linearized representation could be used for $\hat{Y}_C(\lambda)$, satisfactory results would still be difficult to obtain. Because the $P_2(2)$ and $Q(3)$ lines and the principal features of the functions describing the filter and water vapour transmittance occur in the same region of the data interval, the corresponding amplitudes cannot be easily distinguished by the fitting procedure. (In other words the Q hypersurface (Section 4.4) is very elongate and shallow near the minimum with respect to two or more parameters.) As a result, for observations made when the background intensity was low, the linear fit would on occasions return negative amplitudes for the weakest lines in the data interval, or very large P_2 temperatures. In addition the estimate of the error in the P_1 temperature, which could be deduced only from the $P_1(2)$ and $P_1(3)$ lines, was several times larger than that obtained from the nonlinear fit.

4.4 Nonlinear Least Squares Parameter Estimation

4.4.1 Obtaining a Solution

The airglow temperature and intensity information contained in a recorded spectrum is extracted by fitting a model to the data according to the method of least squares, which assumes that the best estimate for the model parameters are those for which the sum of the squares of the deviations of the data from the fit is a minimum. The data is represented by the set of counts $\{Y_j | j = 1,...n\}$ in up to n = 128 channels corresponding to the wavelengths λ_j. The model can be generalized as

$$\hat{Y}_j = \hat{Y}(X_{j1},...,X_{j\ell},...,X_{jr};\ a_1,...,a_k,...,a_N) \qquad ...(4.19)$$

where the r independent variables $X_{j\ell}$ are themselves functions of the λ_j and describe the OH lines and the background and water vapour transmittances as modified by the prefilter. If the uncertainties of the data points are denoted by σ_j, the least squares condition requires that the quantity

$$Q = \sum_{j=1}^{n} \left(\frac{Y_j - \hat{Y}_j}{\sigma_j} \right)^2 \qquad ...(4.20)$$

be minimized with respect to the N parameters a_k. Since in this case $\hat{Y}$ is a nonlinear function of the a_k s, they cannot be directly determined as the solution of N simultaneous equations obtained by setting $\frac{\partial Q}{\partial a_k} = 0$. Instead it is best to use an iterative method of solution - that is, to conduct a search in parameter space according to some algorithm, evaluating Q at each of the search vectors $\underline{a}$ until the minimum of Q is approached sufficiently closely.

A number of techniques have been devised to cope with the convergence problems often encountered in nonlinear curve fitting

(Bevington, 1969; Jones, 1970; Powell, 1965). The method adopted here, proposed by Marquardt (1963), has been among the most successful (Bard, 1970). It performs an optimum interpolation between the gradient method, which chooses new search vectors in the direction of steepest descent of the Q hypersurface, and the Taylor series method, which approximates the fitting function in the neighbourhood of the minimum by a truncated Taylor series. A first order expansion of $\hat{Y}$ about $\hat{Y}_o$ corresponding to some origin vector $\underline{a}_o$ is

$$\hat{Y}(\lambda_j) = \hat{Y}_o + \sum_{k=1}^{N} \left[\frac{\partial \hat{Y}_o(\lambda_j)}{\partial a_k} \delta a_k \right] \qquad \ldots(4.21)$$

and is linear in the parameter increments δa_j, so that the usual multiple linear regression methods can be employed. The result is a set of normal equations which specify the vector $\underline{\delta a}$, and which can be written as

$$\underline{\delta a}\ \alpha = \underline{\beta} \qquad \ldots(4.22)$$

where α is the square symmetric curvature matrix of order N whose elements are

$$\alpha_{km} = \sum_{j=1}^{n} \left[\frac{1}{\sigma_j^2} \frac{\partial \hat{Y}_o(\lambda_j)}{\partial a_k} \frac{\partial \hat{Y}_o(\lambda_j)}{\partial a_m} \right] \qquad \ldots(4.23)$$

and the vector $\underline{\beta}$ has elements

$$\beta_m = \sum_{j=1}^{n} \left[\frac{1}{\sigma_j^2} \{Y_j - \hat{Y}_o(\lambda_j)\} \frac{\partial \hat{Y}_o(\lambda_j)}{\partial a_m} \right] . \qquad \ldots(4.24)$$

The solution to (4.22) is

$$\underline{\delta a} = \underline{\beta} \in \qquad \ldots(4.25)$$

where $\in$, the inverse of α, is known as the error matrix. Marquardt's method involves augmenting the diagonal terms of the curvature matrix by a factor ρ, to obtain

$$\underline{\delta a}(\alpha + \rho\zeta) = \underline{\beta} \qquad \ldots(4.26)$$

where ζ represents the identity matrix.

The search strategy employed is as follows. A start point $\underline{a}_o$ and an initial value of ρ are adopted. The latter should be small enough that it is essentially the properties of the analytical representation (equation 4.22)) which are first exploited. The quantities $\underline{\delta a}$, $Q(\underline{a}_o)$ and $Q(\underline{a}_o + \underline{\delta a})$ are then computed using $\underline{a}_o$ as the origin of the Taylor expansion. Whenever

$$Q(\underline{a}_o + \underline{\delta a}) < Q(\underline{a}_o) \qquad \ldots(4.27)$$

ρ is decreased by a factor of ν and a new solution found using this value of ρ and a new origin and start point equal to the current best estimate of $\underline{a}$. In so doing the linear expansion becomes a better approximation to the data over a smaller region surrounding the Q minimum. The process is repeated until the fractional change in successive values of either Q or all the $|a_k|$ fall below some pre-determined limit. However if the start point of an iteration is too far from the solution the linearized fitting function may not be a good approximation at the point $(\underline{a}_o + \underline{\delta a})$, and the result

$$Q(\underline{a}_o + \underline{\delta a}) \geq Q(\underline{a}_o) \qquad \ldots(4.28)$$

may be obtained. ρ is then increased in steps of a factor of ν and new solutions are found as before, but always using the same origin for the Taylor series, until the condition of equation (4.27) is again met, and ρ may once more be decreased. This procedure makes the diagonal terms of the matrix of equation (4.26) dominate, and the solution for $\underline{\delta a}$ becomes, for large ρ,

$$\delta a_k = \frac{\beta_k}{\rho\ \alpha_{kk}} \quad . \qquad \ldots(4.29)$$

That is, the new solution moves closer to the start point and towards

the direction of steepest descent of the Q hypersurface. Thus the algorithm chooses its own step size and direction for the search, and in so doing avoids the convergence problems often associated with the linearization method far from the minimum on the one hand, and with the gradient method close to the minimum on the other.

The computer program employing this algorithm was obtained from Bevington (1969), as were as many as possible of the other programs required, such as interpolation or linear regression. Similar though much less well documented routines were also available from the University computer program library.

4.4.2 Error Estimates

Errors associated with the data and analysis were treated in the usual fashion (Draper and Smith, 1966; Wolberg, 1967). Each term in the sum of squares for Q (equation (4.20)) was weighted according to the expected variance σ_j^2 of the data point. In doing so allowance had to be made for the non-Poisson behaviour of the PM tube dark current (Section 3.3.5). It was also assumed that there were no systematic errors present, and that all the error derived from the statistical variations in Y_i, and none was present in the $X_{j\ell}$. The variances of the parameters $\sigma^2_{a_k}$ were determined from the error matrix (Arndt and MacGregor, 1966), that is

$$\sigma^2_{a_k} = \epsilon_{kk} \qquad \ldots(4.30)$$

under the assumptions (not always satisfied in practice) that the values of Q were close to one.

4.4.3 Convergence and Conditioning

Several points regarding the convergence of the iterative method described above merit further attention. In the first place it is necessary that $\underline{a}_o$ be chosen so as to avoid subsidiary minima in the Q hypersurface. For this problem the surface was quite well behaved in the region of the solution. As a result it was possible to choose any reasonable value for $\underline{a}_o$ for use with the first record of a night's data, and thereafter to use instead the solution obtained for the previous record. In general the rapidity with which the search converged improved with the estimate of $\underline{a}_o$. Difficulties were only encountered when the start point and solution coincided too nearly, and an incremented vector $\underline{a}_o + \underline{\delta a}$ could not be found such that Q decreased. The above procedure was not always successful if an unmodified linear expansion method of fitting was used, however, and it was for this reason, and because a simple grid search or the gradient method converged very slowly, that Marquardt's algorithm was adopted.

The rate of convergence may also be improved by a judicious choice of ν, the initial value of ρ and to some extent the size of the increments Δa_k used to compute the derivatives in equations (4.23) and (4.24). It was found that for this problem the condition (4.28) occurred only rarely, making it advantageous to begin the search with ρ less than the value of 10^{-2} suggested by Marquardt. For the same reason there was little to be gained by using a different interpolation factor ν, as for example proposed by Tabata and Ito (1975), or to modify the algorithm to avoid reinverting the curvature matrix whenever ρ was increased, as suggested by Jones (1970). The values $\rho = 10^{-3}$ and $\nu = 10$ were therefore chosen. The increments Δa_k were set at $0.001\ a_k$ - that is, of the same order as the last correction δa_k made by the fit,

small enough to yield derivatives undistorted by rapid changes of a_k in the region of the Q minimum, and large enough to avoid the effects of roundoff error.

Some parameter estimation problems are such that while the iterations converge nicely for a given set of coefficients in the normal equations, slight changes to these coefficients can result in large changes in the solution (e.g. Wilkinson, 1967). Of the several tests (Conte, 1965) which can be applied to indicate the presence of such ill-conditioning, as it is termed, the most direct was carried out by examining the results of a fit to a number of synthetic spectra. These were constructed using identical parameter values but different samples of Poisson-distributed noise, so that the constants of the equations were perturbed in such a way that the average change in the solution could be predicted. The sample variance of the a_ks thus obtained agreed with the error estimates returned by the analysis for both the linear and the nonlinear model, although other tests suggested that the former was approaching instability.

Even when a problem is quite well conditioned the iterative process may converge to the wrong solution because of roundoff errors - that is, errors which arise because the computer word length is finite (15 significant digits in the case of the CDC 6400 using floating point representation). Since roundoff error is cumulative, and the number of arithmetic operations required to invert a matrix increases rapidly with its size, it is desirable both to limit as far as possible the number of parameters to be determined and to exercise care in inverting the matrix. For this analysis the Gauss-Jordan method of inversion with full pivotal condensation was used (e.g. Martin, 1966), and the absence of roundoff error was confirmed by the fact that solutions obtained

with single or double precision arithmetic were identical.

4.4.4 Goodness of the Fit

A number of tests are available to decide whether or not a good fit has been obtained (Draper and Smith, 1966; Wolberg, 1967). If the model were linear and the fit good, the quantity Q should be described by a χ^2 distribution having a mean of ν and a variance of 2ν, where $\nu = n - N$ is the number of degrees of freedom of the problem. (For the remainder of this work Q and χ^2 will be taken to refer to the reduced values $Q_\nu = Q/\nu$ and $\chi^2_\nu = \chi^2/\nu$, so that the mean of their distributions should be one). However an analytic probability density function for Q cannot in general be obtained for the nonlinear case. A common practice, therefore, is to obtain the distribution empirically, by analysing a large number of data sets synthesized from the same parameter values but different samples of random noise. Frequently (and in fact for the model used here (Section 5.2)) the result is still close to the χ^2 curve, and the fit to individual records can then be evaluated using the integral χ^2 distribution. For example, for the present problem where $\nu = 94$, the 5% significance level for Q is 1.25. However, as can be seen from equation (4.20), this statistic does not differentiate between contributions due to the variance of the data points and departures of the fitting function from the correct form. It can also be strongly influenced by one or two aberrant data points, called outliers, which are not members of the parent population, and, when a poor fit is obtained, does not indicate the magnitude of the bias introduced in the parameter estimates.

Further information about the causes of high Q values can sometimes be obtained from a study of the residuals - the set of differences

$$\xi_j = Y_j - \hat{Y}_j \qquad \ldots(4.31)$$

(e.g. Anscombe and Tukey, 1963). They can be examined in graphical form as functions of the fitted values $\hat{Y}_j$, the independent variables $X_{j\ell}$ or the channel number j (a "time series" plot) to see whether they conform to the basic assumptions of the least squares procedure, namely that they be uncorrelated and normally distributed with zero mean. (In fact they should be slightly correlated since they have n-N degrees of freedom, but the effect is too small to be significant in graphical procedures when ν is large). If the scatter in the plot against $\hat{Y}_j$ does not have the appearance of becoming broader towards large $\hat{Y}_j$, the variance of the data is likely to have been correctly estimated, while evidence of structure in the other two plots can usually be attributed to inadequacies in the model.

The residuals can be tested for normal distribution by plotting their linearized cumulative distribution and comparing it with a number of samples of random normal deviates. This method does not, however, lend itself to the treatment of large quantities of data, and other tests such as the runs or Durbin-Watson tests are more useful. Both examine the residuals for serial correlation. The former compares the frequency of occurrence of sequences of values having the same sign with the result to be expected from random data. The quantity

$$r = \frac{m - \mu + \frac{1}{2}}{\sigma} \qquad \ldots(4.32)$$

where

$$\mu = \frac{2n_1 n_2}{n_1 + n_2} + 1 \qquad \ldots(4.33)$$

and

$$\sigma^2 = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 \, (n_1 + n_2 - 1)} \qquad \ldots(4.34)$$

and where m is the number of runs and there are n_1 positive and n_2 negative points (n_1, n_2 > 10), should be a unit normal deviate. Durbin

and Watson (1950 and 1951) have constructed a test which is based on the von Neumann statistic

$$d = \frac{\sum_{i=2}^{n} (\xi_i - \xi_{i-1})^2}{\sum_{i=1}^{n} \xi_i^2} \qquad \ldots(4.35)$$

but which is often inconclusive or at best approximate, although improved versions have been developed (Henshaw, 1966). The test is also strictly applicable only to linear models, so that for the present problem the same procedure was adopted as for Q - that is, an empirical distribution was obtained by analysing a large number of synthetic spectra. Both the r and d distributions are independent of the variance of the data adopted in the fit, and are particularly useful when, as in this case, structure could not easily be discerned in the residual plots. Finally the residuals can be used to reveal the presence of outliers, either by direct inspection of the time series plots or by testing for data points lying outside the chosen (e.g. 98%) confidence region about the fitted curve.

The goodness of fit may also be checked by comparing the solutions with theoretical or expected values. When these are not accurately known the presence of bias in the estimates is difficult to detect. The best that can be done is to examine the fit to synthetic data to which small but likely departures from the model have been added.

4.5 Functions and Constants Required for the Model

The functions characteristic of the spectrometer response which must be measured for use in the model are the instrument profile, the broadband filter profile, the wavelength dependence of the PM quantum efficiency, and a function representative of atmospheric water vapour

absorption. The last has already been considered in Section 4.3.3. As discussed in Section 2.2.2, $I(\lambda)$ is ideally found by recording the response of the spectrometer to a spectral line whose width is negligible compared to the instrument width. However the lack of a suitable source forced an alternative approach. Although it was possible to use the laser line and broaden the result by numerical means to simulate the desired function (Wilksch, 1975), neither the exact shape nor width of the latter is certain, depending as it does on the change in properties of the reflective coatings. It was considered better therefore, to use as a source a monochromator set to a bandwidth which compromised between the requirements of a high signal level and no broadening of the recorded function. The monochromator contribution was then removed by deconvolution using the discrete Fourier transform (Brigham, 1974, inter alia). The source profile width of 0.22 nm was sufficiently small compared with the instrument bandwidth (0.98 nm) that this procedure made little difference to the result. The numerical filtering inherent in the correction helped ensure that the errors in the resulting instrument profile were negligible compared with the statistical noise in the airglow record. In addition, since the instrument profile width changed somewhat from one end of the scan to the other (Section 3.4.1), it was necessary to broaden by interpolation each function describing an OH line according to its position in the spectrum.

Two measurements involving the prefilter are required. One of these, the relative transmission at the OH line wavelengths, was found using a monochromator, taking care that the filter temperature and field of view were as used in the FPI, and that the filter was uniformly illuminated and normal to the beam. The other is the response $[G * I]$ of the spectrometer, including the filter, to a spectrally flat source.

To obtain this the instrument was illuminated by a tungsten filament low brightness source used for calibrating photometers, scanned over the wavelength region of interest, and calibrated in absolute wavelength against a laser line. A small correction for the spectral intensity of the source had also to be made.

A comparison of the responses at different wavelengths of a number of photomultiplier tubes had shown that the quantum efficiency curve published by the manufacturer was unreliable. However, accurate measurements would have involved considerable effort, and since the expected wavelength dependence was in any case small, the tube was merely checked by comparing it with a PIN diode, the responsivity of which is inherently precise and repeatable.

In certain respects the precision with which the model had to be specified, and the performance required of the spectrometer, could best be determined by an analysis of a perturbed synthetic recorded spectrum. These include most importantly the linearity, parallelism and wavelength calibration (hereafter denoted $\Delta\lambda_Y$ nm per channel) of the scan, the separation stability of the etalon, and the wavelengths and number of significant OH lines in the data interval.

It was found that significant differences in Q ($\Delta Q \simeq 0.15$) occurred if the value adopted for $\Delta\lambda_Y$ or i, the average width of the instrument profile, were in error by as little as 0.4% and 1% respectively. This degree of stability in the effective piezoelectric strain constant could not be relied upon over a period much longer than 12 hours, and such small changes in i, if due to variations in parallelism, could not readily be measured by the alignment system. It was therefore necessary that both $\Delta\lambda_Y$ and i be recalculated for each night's data by minimizing Q with respect to each separately in several data blocks and averaging

the results. As discussed in Section 3.4.2, analyses of synthetic spectra also showed that it was necessary to optimize the positioning of the fitting functions to within 0.1 channels, and that this had to be done every two or three records. This also meant that the instrument profile had to be interpolated to this accuracy in the computer program. The variation of Q with respect to any of $\Delta\lambda_Y$, i or the position of the fitting function was not independent of the other two. Three dimensional contours of Q were therefore plotted to determine the best order in which to vary these parameters in a simple grid search to find their optimum values. This procedure involved no ambiguity because the OH lines themselves acted as calibration sources.

To determine whether or not a given OH line should be taken into account, a fit was constructed to a synthetic spectrum which had a known noise contribution to Q but to which the OH line in question had been added. It was established that apart from the P lines from $P_2(2)$ to $P_2(4)$ it was necessary to include the $P_1(4)$, Q(3) and Q(4) lines in the model, and that although severe attenuation of the $P_1(4)$ line would have a noticeable effect on the value of Q, it would not introduce a significant error in the rotational temperature obtained from a truncated 100 channel data interval (see following).

Using methods similar to the above it was found that, in respect to the relative positions of the OH lines, the maximum tolerable deviation from scan linearity was less than 0.05 nm, assuming that the nonlinearity had a quadratic form. Since it appeared that the initial portion of the scan was in error by more than this (Section 3.1.1.1), the first 28 channels were excluded from the analysis. The data and fit to a 128 channel spectrum and its truncated form are shown in Figure 4.1(a) and (b). The values of Q and rotational temperature are

2.33 and 177 K, and 1.02 and 197 K respectively.

The accuracy required of the OH line wavelengths depended on the prominence of the line in the spectrum, and was found to be about ±0.005 nm (±0.1 channels) for the $P_1(2)$ and $P_1(3)$ lines, and about ±0.02 nm for the $P_2(3)$ line. This result was of some importance because there is considerable spread, often exceeding 0.1 nm, in the relative positions of the lines obtained from wavelengths or spectroscopic constants published by other workers (Bass and Garvin, 1962; Krassovsky and Shefov, 1962; Kvifte, 1959; Meriwether, private communication; Wallace, 1960). At first relative wavelengths obtained by averaging these results were used, but it was found that Wallace's values gave a significantly better fit and they were therefore adopted, although their stated accuracy was only ±0.03 nm. However the separation of the $P_1(2)$ and $P_1(3)$ lines agreed with the averaged results to within ±0.004 nm. The wavelengths appear in the Appendix, together with the constants and equations from which they were calculated.

CHAPTER 5

RESULTS OF THE ANALYSIS AND OBSERVATIONS

5.1 Introduction

Problems involving least squares fitting of complex models must be examined with some care to ensure that the parameter estimates obtained are good approximations to the true values. This is particularly necessary in the present circumstances, where because a nonlinear model has been used many of the standard methods of evaluating the fit are not strictly valid, and where some of the physical characteristics of the problem cannot be exactly simulated. Much of this chapter will therefore be devoted to evaluating the performance of the analysis, first by means of goodness of fit tests and then by comparison of the estimates obtained from a number of nights of observation with expected values. A number of refinements which this evaluation showed to be necessary to the original model are described, and possible sources of error in the model are considered. Finally, under the assumption that the parameter estimates are accurate, the results are discussed in terms of the dynamic processes occurring at the OH emission height.

The observations referred to were made on fourteen nights spanning an interval of eighteen months, from 14 April 75 to 11 November 76. For the most part the observing conditions were as specified in the preceding chapters, and in particular the field of view remained free of cloud for the entire night, as nearly as could be ascertained. However in order to severely test the efficacy of the analysis, several nights on which the record was strongly contaminated by scattered moonlight or the Milky Way have also been included. The dates in question together with remarks on the prevailing observing conditions may be found in Table 5.1.

TABLE 5.1

Dates and remarks on observing conditions for the fourteen nights analysed

Date	Remarks
14 April 75	Zenith angle Z = 0°. Asimuth angle A = 0°. Milky Way in field of view at ∿ 0415 L.T.
6 June 75	Z = 0°, A = 0°. Milky Way in field of view at ∿ 0100 L.T.
30 Dec 75	Clear, dark night. Z = 55°, A = 180°. (South celestial pole).
2 Jan 76	Clear, dark night. Z = 55°, A = 180°. (South celestial pole).
16 Jan 76	Full moon. Z = 55°, A = 180°.
17 Jan 76	Full moon. Z = 55°, A = 180°.
3 Feb 76	As for 30 Dec.
4 Feb 76	As for 30 Dec.
29 March 76	As for 30 Dec.
30 March 76	As for 30 Dec.
22 June 76	As for 30 Dec.
24 Sept 76	As for 30 Dec.
25 Sept 76	As for 30 Dec.
11 Nov 76	Moon full on 6 Nov. Rises 2210 L.T. Z = 55°, A = 180°.

Note: L.T. = G.M.T. + 9.26 h

5.2 Results of Tests of the Goodness of Fit

The values of the quantity Q of equation (4.20) were relied upon as the primary indication of goodness of fit. Because the model used was nonlinear an expected Q distribution was obtained by analysing 590 synthetic spectra constructed as described in Section 2.3.3. The resulting histogram of Q was found by visual comparison to be very similar to a chi-square distribution. Since each night's data consisted of at least 30 individual 10 minute records, it was possible to assess the fit for the night as a whole by comparing the expected and experimental distributions. In general it was found that there was rather poor agreement, in that the mean $\overline{Q}$ of the latter was $\sim$0.3 units high on average, although the shapes of the histograms were similar. Values of $\overline{Q}$ for each night are shown in Table 5.2 in descending order of goodness of fit according to this statistic, and the distribution obtained for one night, 11 Nov. 76, is shown in Figure 5.1.

Residual plots of the type described in Section 4.4.4 were constructed for a random sample of analysed spectra. In general the residuals did not show a tendency to be more widely scattered at high or low values of $\hat{Y}$ when plotted against this quantity, thereby confirming the earlier observation of the signal count rate statistics (Section 3.3.5). Indications of nonrandom behaviour were found in the residual plots versus the fitting functions $X_{j\ell}$, but the information contained here was the same as that appearing in more concise form in the residual plots versus channel number j. Therefore only the last of the three representations will be discussed in the following.

With the exception of twilight records, for which the Q values were in any case very much inflated, it was in general only when all the residual plots for one night were summed that structure could clearly

TABLE 5.2

Mean values of the chi-square (Q), runs (r) and Durbin-Watson (d) distributions and the mean percentage of data points ($\bar{n}$) exceeding the 99% confidence limit. The nights are arranged in order of increasing $\bar{Q}$.

Night	$\bar{Q}$	$\bar{r}$	$\bar{d}$	$\bar{n}$
17 Jan	1.03	+0.11	2.02	0.83
16 Jan	1.18	-0.29	1.96	1.56
29 March	1.19	-0.31	1.91	1.52
11 Nov	1.21	-0.24	1.91	1.66
4 Feb	1.21	-0.59	1.77	1.60
30 Dec	1.23	-0.57	1.86	1.47
30 March	1.24	-0.27	1.89	1.29
25 Sept	1.30	-0.79	1.79	2.14
3 Feb	1.33	-0.32	1.89	2.27
22 June	1.35	-0.93	1.70	3.13
24 Sept	1.36	-0.39	1.82	2.25
14 April	1.39	-1.04	1.69	2.70
2 Jan	1.50	-1.36	1.60	3.13
6 June	1.77	-1.94	1.50	4.85
Mean	1.30	-0.64	1.81	2.17
Synthetic	1.04	+0.40	2.05	0.94

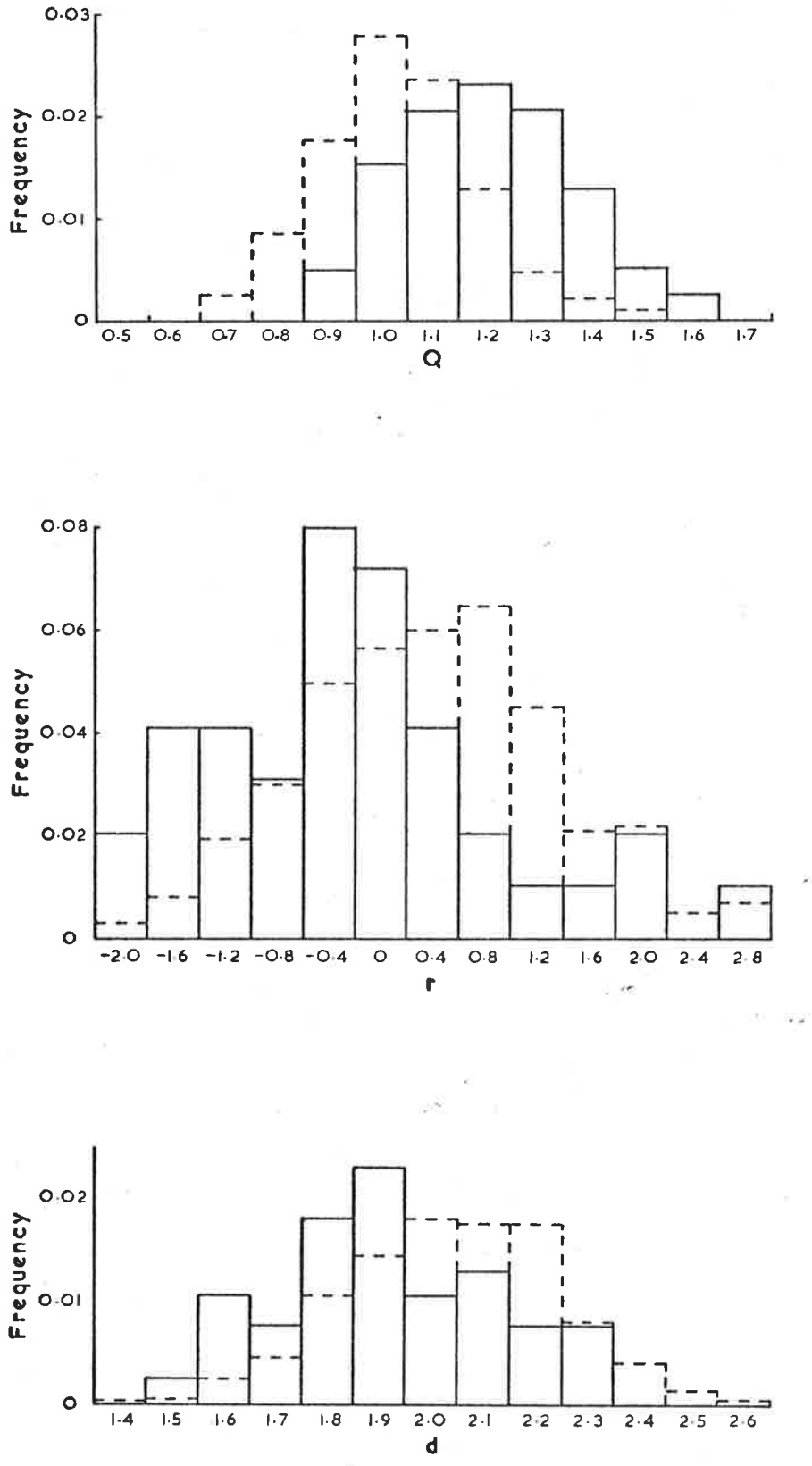


Figure 5.1 Histograms of the expected (dashed) and experimental (full) distributions of the chi-square (Q), runs (r) and Durbin-Watson (d) statistics for 11 Nov 1976. The distributions are normalized to unit area.

be seen. This structure was common to all but two of the nights studied, but its shape did not necessarily bear any physical resemblance to the airglow spectrum, and could not be directly used to improve the model. It may also be noted that in view of the approximations necessary in the model a noticeable error must be expected to accumulate after summing a sufficient number of records.

As mentioned in Section 3.3.6, it was the practice during the initial editing of the data to remove large noise spikes to which the detection circuitry was susceptible. The magnitude of these points was usually such that they could be readily identified by a simple visual inspection of low resolution (line printer) plots of the raw spectra. However because there was no essential lower bound to their amplitude this procedure could not remove all such events - a matter of some importance since a single small outlier could increase Q by 0.1 units. A second search for outliers was therefore carried out during the analysis by specifying for each data point a confidence limit, based on the assumption that the data came from a normally distributed parent population, and comparing this limit with the magnitude of the residual. The check revealed that an average of 2.2 points per hundred lay beyond the 99% limit. The mean rejection rate per night is given in Table 5.2 as the quantity $\bar{n}$.

It was by no means clear, however, that noise spikes were responsible for this result. The number of points within the 90-95% and 98-99% confidence regions was similarly higher than expected, suggesting that these points formed part of the main residual distribution rather than an extended tail. In addition they were found to fall equally frequently above and below the fitted curve, whereas had they been noise spikes they should have occurred preponderately above. Finally, although

measurements of the signal statistics were not made on the nights processed, there was no evidence in the dark current records of events of the magnitude and frequency required to inflate the Q values to the degree observed. An alternative explanation for the abnormally broad distribution of residuals is that the model is in error. It would therefore be expected that the frequency distribution of points beyond the 95% confidence limit versus channel number would show peaks corresponding to the peaks in the curve of the summed residuals, and in fact this was observed for most nights.

It was nevertheless found desirable to incorporate some form of additional discrimination against noise spikes. A confidence limit of 99.9% or 4 standard deviations was chosen as the rejection criterion for points above the fitted curve, and a new fit was made to the corrected data. For convenience the correction consisted of setting the number of counts in the outlier's channel equal to the original fitted value - a procedure which did not distort the parameter estimates.

Further quantitative measures of the residual behaviour were provided by the runs (r) and Durbin-Watson (d) tests (Section 4.4.4). As for the Q statistic a comparison was made between histograms obtained from real and synthetic data. The latter for the runs test was found to be approximately Gaussian in shape with a variance of 0.9 and a mean of 0.4, rather than zero as expected - a difference apparently caused to a large extent by the random number generator used. The distributions of r from the data were predominately Gaussian also, but were on average centred at -0.6 and had a variance of unity. Well-shaped histograms were likewise obtained for the Durbin-Watson statistic, but again the average value $\bar{d} = 1.8$ for all nights was rather less than the average of 2.1 obtained from synthetic data. The average values of r and d

for each night and histograms for 11 Nov 76 are shown in Table 5.2 and Figure 5.1.

As can be seen from Table 5.2 there is a pronounced tendency for the r and d values to decrease and n to increase as Q increases, consistent with the high Q values being caused by some error in the model. The flatness of the plots of the residuals versus $\hat{Y}$, the independent measurements of the variance of the signal count rates and the foregoing discussion of the occurrence of outliers all suggest that there is little or no contribution to Q due to non-normally distributed data. The question of the nature of the discrepancy in the model will be pursued in the following section.

5.3 Bias in the Parameter Estimates

Since the tests discussed in the previous section showed that the model adopted was imperfect, it was desirable to obtain some indication both of the magnitude of any bias in the parameter estimates - particularly rotational temperature and branch intensity - and of any variability of this bias not caused by the signal statistics. In regard to the first question, perturbed synthetic spectra were analysed to find the temperature change associated with an increase in Q from 1.0 to 1.3 (the average obtained from the fourteen nights considered). Of interest also was the magnitude of the perturbation required. A summary of the more significant results is given below.

To increase Q as required the relative wavelength of the $P_1(2)$ line had to be increased by about 0.02 nm, resulting in a temperature change of $\sim$ -10K, whereas for the $P_2(3)$ and $P_1(3)$ lines the variation was about 0.03 nm and +4K, and 0.04 nm and +10K, respectively. A quadratic departure from scan linearity of 0.05 nm in the sense expected of the

etalon (Section 4.5) was required and changed the temperature by ∿ +10K. By increasing the amplitudes of the $P_1(4)$ and $P_2(4)$ lines to simulate a non-Boltzmann distribution of the emitting population it was found that a temperature difference of 170K, as measured by the pairs of lines [P(2), P(3)] and [P(2), P(4)], changed the estimated temperature by +5K. Variations in the magnitude of the fixed background slope from -5% to -40% changed the estimate by +9K. (The degree of tolerance evinced in this example reflects the similarity of the role played in the model by the sloping background and water vapour absorption functions). A spectrum having W = 1.5, when analysed using a model which did not include a water vapour function, produced a temperature change of +11K. (The average value of W for all nights was ∿ 1.2). A number of additional lines, including the [O II] doublets at λ731.9 nm and λ733.0 nm and a mercury line at λ734.6 nm, were also modelled. The magnitude of their effect depended on their position relative to the OH lines. For example, while a line at λ734.6 nm and 4% of the amplitude of the $P_1(3)$ line changed Q by 0.3 units and the temperature by +10K, a similar line at λ731.9 changed Q by only 0.03 units and the temperature by -20K.

Since it is not certain which if any of these modifications simulate the actual defect in the model, it is clear that as they stand these findings do not support the conclusion that any systematic temperature error is small. However taken together with the results for the OH parameters and an assessment of the probable sources of error (see the following sections) they suggest that any bias would not greatly exceed 10K.

It is unlikely that any discrepancy in the model would contribute a fixed amount to the value of Q in the face of varying characteristics of the airglow. This is supported by the observation that the variances

of the distributions of Q are generally greater than the expected variance of 0.022, although the average for nine of the fourteen nights was only slightly larger at 0.026. It would therefore be expected that any systematic temperature error should also vary. However an inspection of the plots of Q, temperature and OH and background intensity versus time revealed no apparent correlation between Q and any of the other quantities. According to this, then, the variation of any bias in temperature is of the order of or less than the random component of the temperature error.

5.4 Presentation and Assessment of the Results

The preceding two sections have shown that the nonlinear model was imperfect, but the degree of error in the parameter estimates which this entailed has not been ascertained. In this section the results obtained from an analysis of the fourteen nights will be presented under the broad categories of the absolute and mean values of the parameter estimates, the nocturnal behaviour of the OH intensity and temperatures, correlations between parameters on a given night and periodic variations. Discussion of these results will at this stage be primarily directed towards establishing their validity by comparison with expected values and behaviour.

5.4.1 Mean and Absolute Values

The seasonal behaviour obtained for the rotational temperature T_R is shown in Figure 5.2. Each point represents the average for a night, excluding the four or five records closest to twilight. Also shown is the seasonal variation of the neutral atmospheric temperature at 80 km and 35° S latitude, as given by the CIRA (1972). Corrections for the effects of a broad emitting region (Section 1.3.2) would produce better

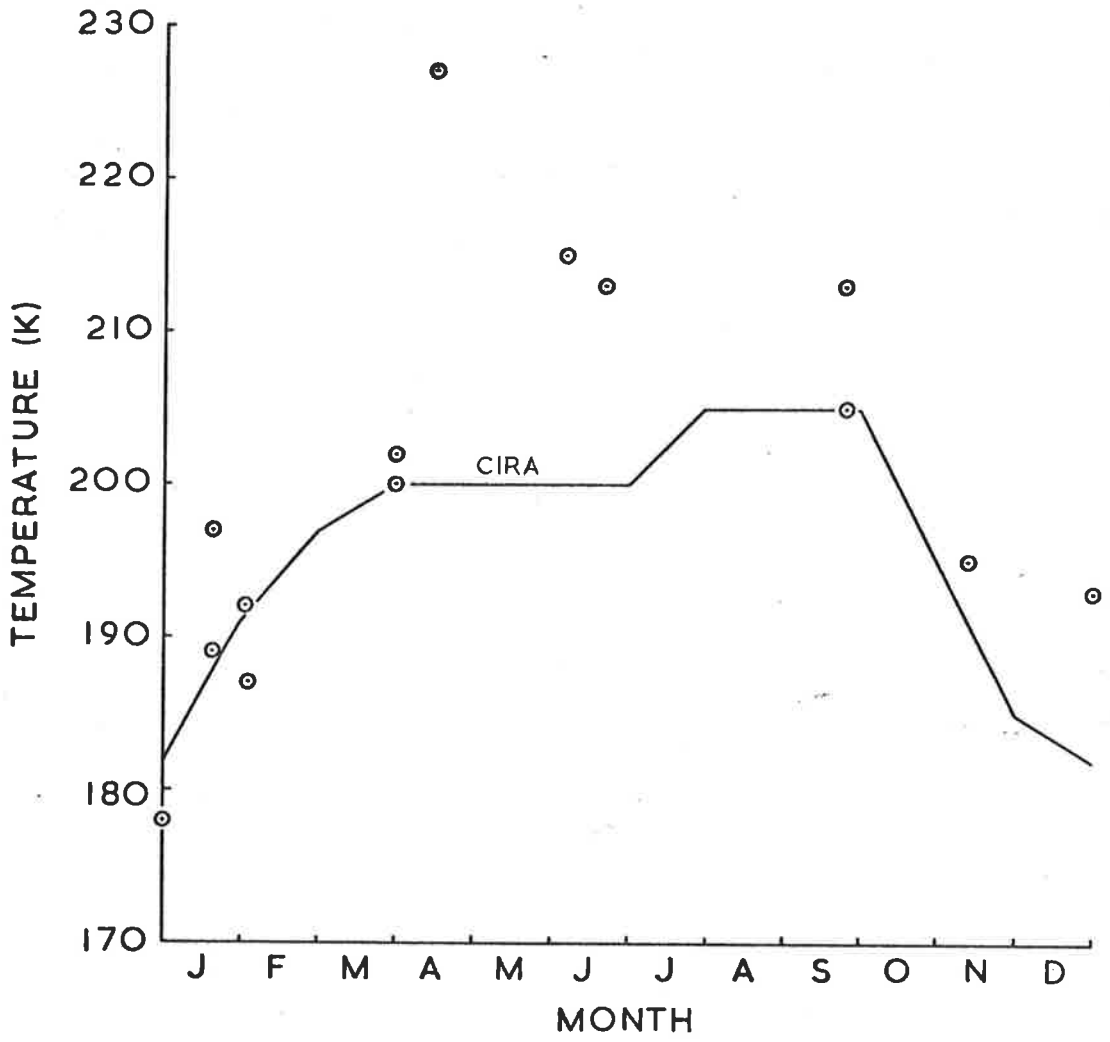


Figure 5.2 Seasonal behaviour of the measured rotational temperature, and the CIRA (1972) average monthly temperatures at 80 km.

agreement between the observations and the CIRA curve. However the absolute values of T_R are difficult to verify closely, bearing in mind the variability of the OH layer height shown by rocket measurements, and the degree of bias cannot be assessed to within less than ±10K. Possible height variations and the small sample size reduce the significance of the apparently excessive winter temperatures found. In general the results were consistent with several rocket and theoretical height distribution curves (summarized by Vallance Jones, 1973) as well as with values and seasonal behaviour discovered by a number of other workers (e.g. Shefov, 1969a; Harrison et al., 1971). Nevertheless it should be emphasized that the temperatures recorded here are not responsive to any non-Boltzmann behaviour of the OH lines with $K \gtrsim 4$, as has been shown in Section 5.3.

Further evidence that T_R was largely unbiased was provided, as outlined in Section 1.3.2, by the result for the doublet ratio temperature T_D, which was calculated from the estimates for the P_1 and P_2 branch relative intensities according to equation (1.9). The mean values of T_R and T_D for each night, again excluding the twilight records, are presented in Table 5.3, together with the number of spectra used in the average. There was generally good agreement between the two, as was found by Harrison and Kendall (1973a). However there were no evident anomalies in the acquisition or analysis of the data which were common to both 14 April and 24 Sept for example, and which might explain the larger differences obtained for these nights.

There are two other parameters the absolute values of which could be determined by the analysis and used to check the model. These are the effective water vapour absorber amount W and the magnitude of the background slope. In fact because of the similarity of the two functions

TABLE 5.3

Nightly mean values of the OH rotational (T_R) and doublet ratio (T_D) temperatures and the number of spectra used in the average.

Date	T_R	T_D	Number of Spectra Averaged
14 April	230±3	244±3	38
6 June	214±2	222±3	27
30 Dec	193±1	186±1	37
2 Jan	180±1	176±1	33
16 Jan	198±1	196±1	39
17 Jan	190±2	185±2	30
3 Feb	191±2	186±2	39
4 Feb	188±1	184±1	34
29 March	204±1	193±2	48
30 March	200±1	191±1	45
22 June	214±1	215±2	49
24 Sept	213±1	192±1	47
25 Sept	205±2	198±1	50
11 Nov	198±1	193±1	35
Mean	201±4	197±4	551

involved, the latter (that corresponding to the parameter a_{k+1} of equation (4.5) was usually fixed at -5% of the flat background component. This choice was consistent both with the measurements of Gadsen and Marovich (1973) and with the average obtained for a random sample of spectra analysed using the variable slope model. As shown in Section 5.3 the exact magnitude of a_{k+1} is of little consequence for the background levels usually encountered. However for high levels - i.e. on 16 and 17 Jan and 11 Nov when there was a full moon - both W and a_{k+1} were well defined and both had to be variables of the fit. The average slope was then $\sim$ -1.6% - a change in the sense expected since the solar spectrum has a +2% slope in this region of the spectrum.

The average values of W from the analysis were found to be consistently lower than those obtained from the Bureau of Meteorology records. A ready explanation for this is that the Mt. Torrens observatory was some 40 km distant and 600 m higher than the site of the radiosonde flights - a significant difference since most of the water vapour absorption occurs within 2 km of the ground. However a small model error must also have contributed, because on several very dry nights the analysis would, unless restrained, have returned small negative values of W. Despite these discrepancies it was found that the two measures of W were usually highly correlated, as shown in Figure 5.3.

5.4.2 Nocturnal Variations

Characterization of the behaviour of the OH intensity was difficult both because of the limited number of nights processed and because of the degree of variation observed from night to night. Comparisons with the results of other workers are hindered by an apparent dependence on latitude, season and upper vibrational level (Shefov, 1969 a,b, 1971).

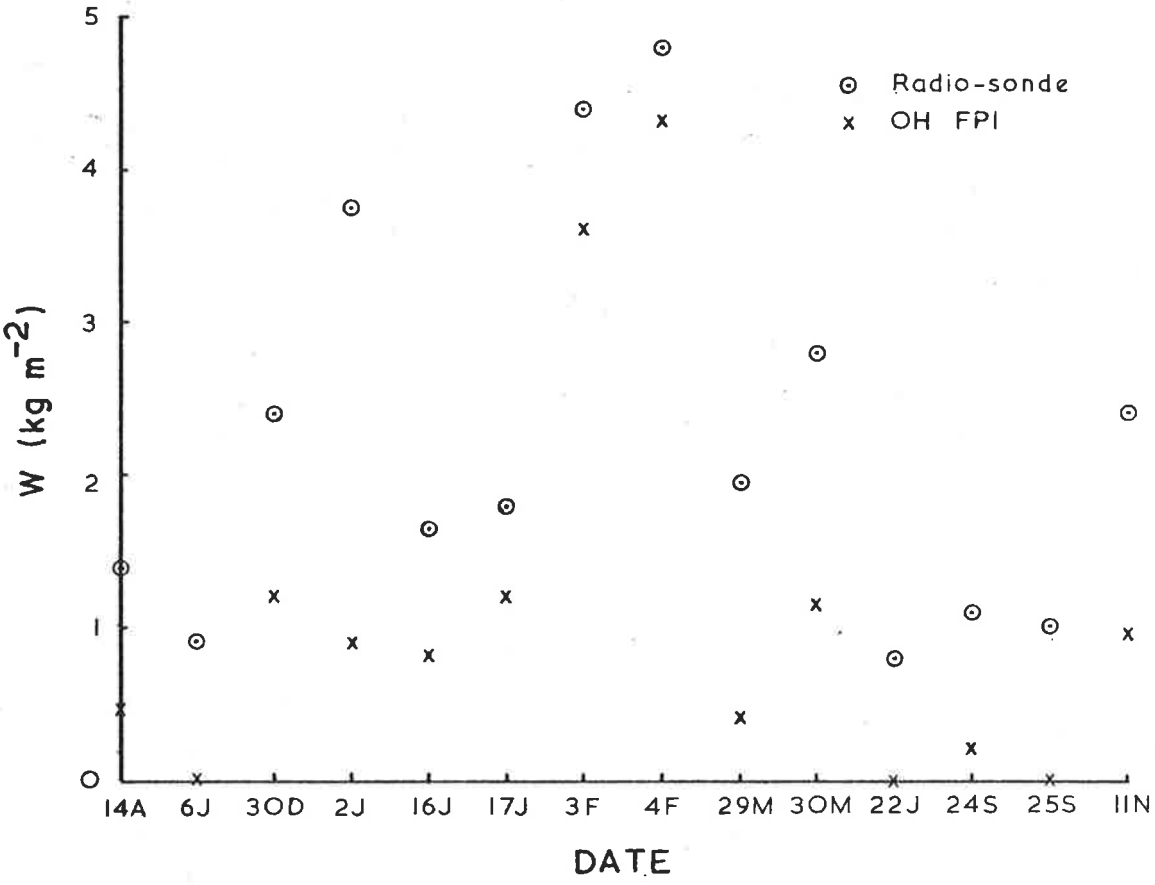


Figure 5.3 Average value of the equivalent sea level water vapour absorber amount (W) from the analysis and from radio-sonde measurements for each night.

Nevertheless it was possible to discern in the FPI results the basic forms of diurnal behaviour found by Wiens and Weill (1973) for the (9 - 3) band at latitude 33° S. That is, most of the nights from March to June showed a pronounced minimum prior to local midnight, while for many of the remainder the intensity either decreased from dusk until the middle of the night and then stabilized or simply decreased steadily. The magnitude of the variation seen was also similar, ranging from about 25% of the nightly mean in summer to about 100% in winter. The absolute intensities could not be compared however, as an absolute calibration of the FPI was not carried out on every night. There was also general agreement with the results of Dick (1972) and Takahashi et al. (1974), and on five nights (of which 24 Sept was one) there was a small peak in the record (Section 5.5) which may correspond to the local midnight maximum observed by Dick et al. (1970). The results as a whole did not agree with the predictions of the aeronomic models of Shimazaki and Laird (1970), Hesstvedt (1970) and others, that the intensity should decrease steadily from dusk till dawn.

The intensity records for three nights, 6 June, 24 Sept and 11 Nov, plotted as the percentage change in intensity of the $P_1(3)$ line relative to the mean for the night and as a function of local time, are illustrated in Figure 5.4(a), (b) and (c). There was no necessity to convert to total band intensity since for a constant emitting population a change in temperature of 70 K changes the intensity of this line by less than 3%. Also shown in Figure 5.5 is the average nightly behaviour found by fitting a fifth order polynomial to all the intensity data, after adjusting each night's values to the same mean and to time relative to sunset at 80 km.

There are few systematic studies of the nocturnal variation of T_R

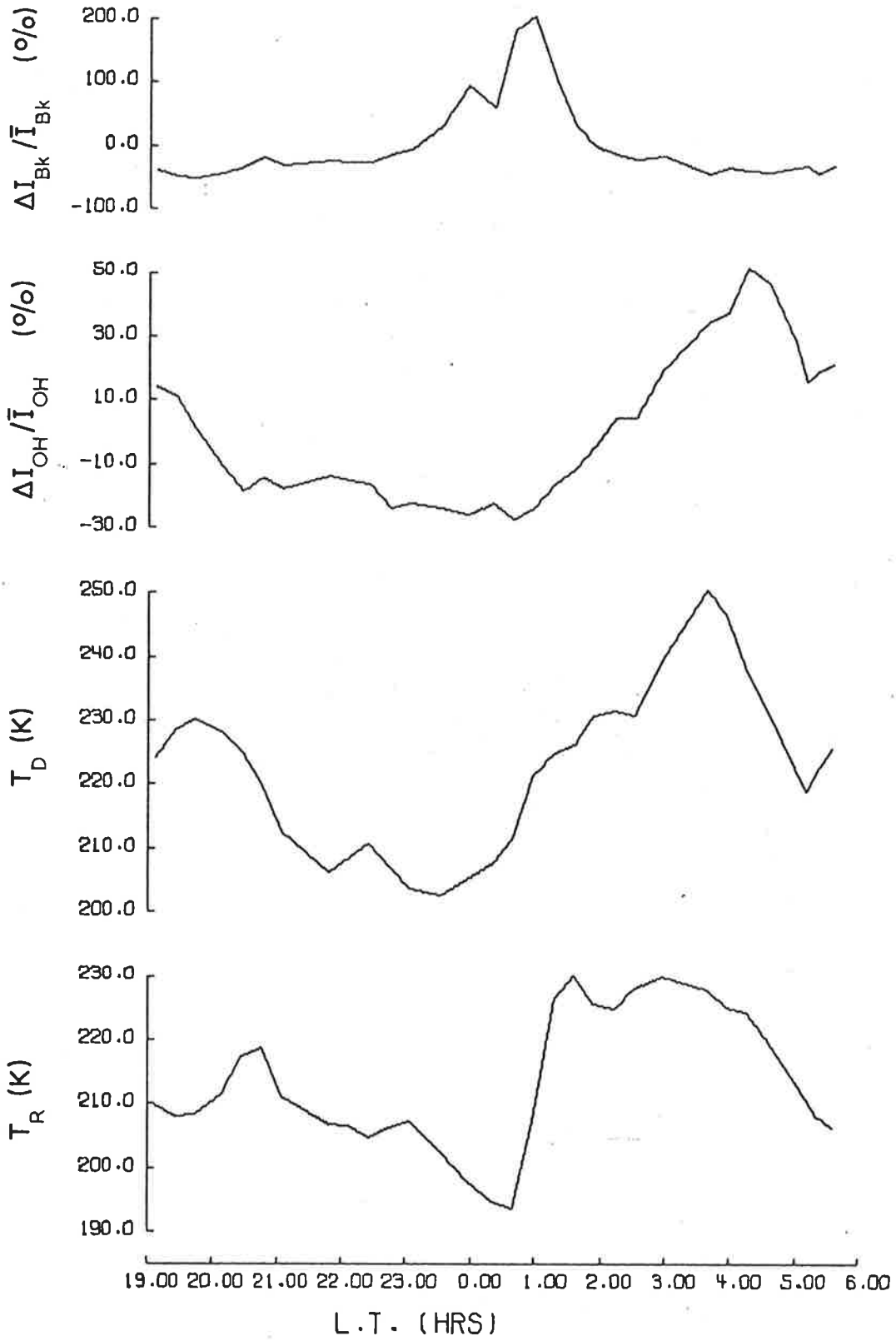


Figure 5.4 (a) Nocturnal variation of the rotational (T_R) and doublet ratio temperature (T_D), $P_1(3)$ line intensity (100 $\Delta I_{OH}/\bar{I}_{OH}$) and background intensity (100 $\Delta I_{Bk}/\bar{I}_{Bk}$) as a function of local time for 6 June 75. Twilight records have been omitted.

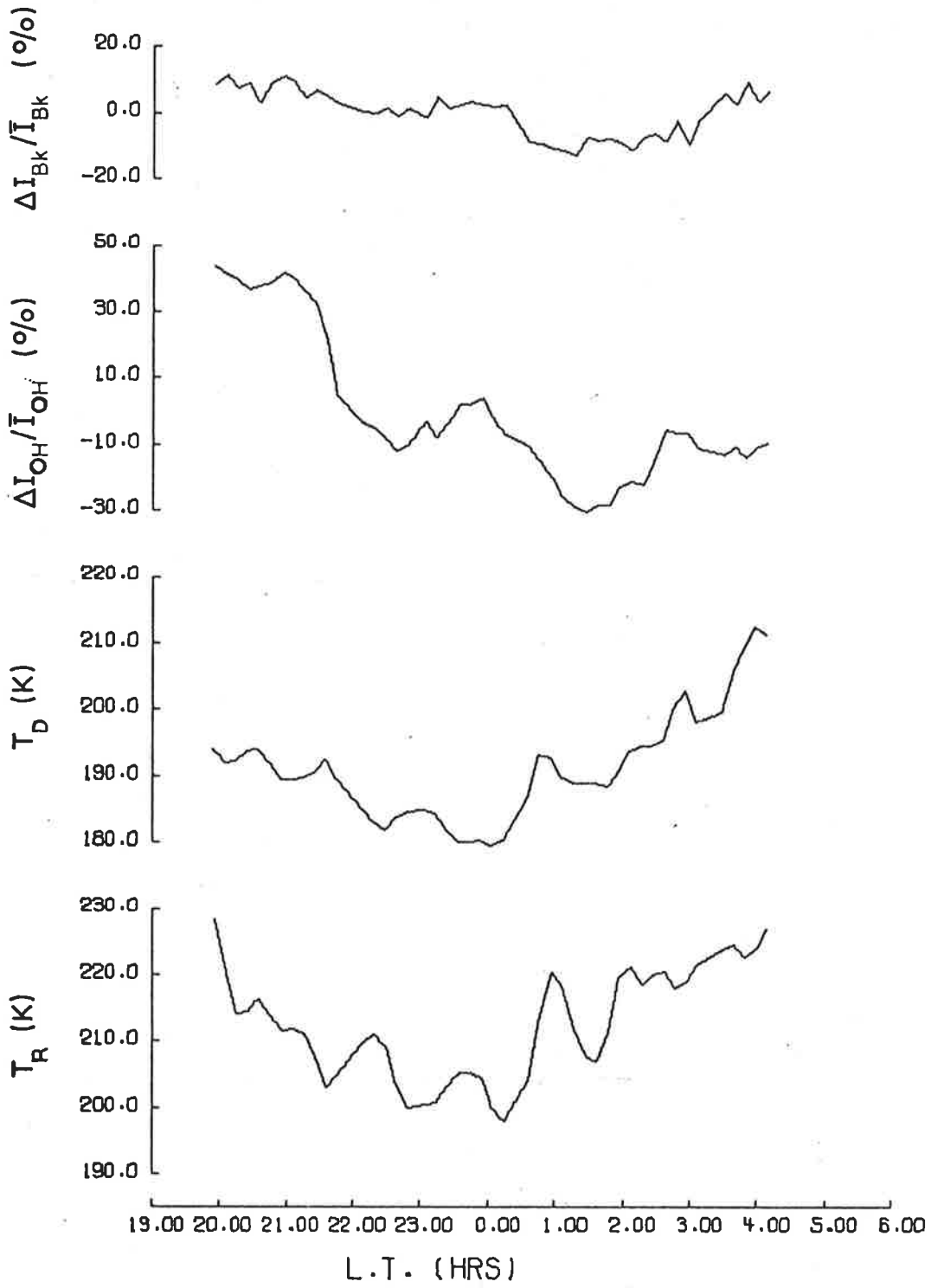


Figure 5.4 (b) As for Figure 5.4 (a) and 24 Sept 76

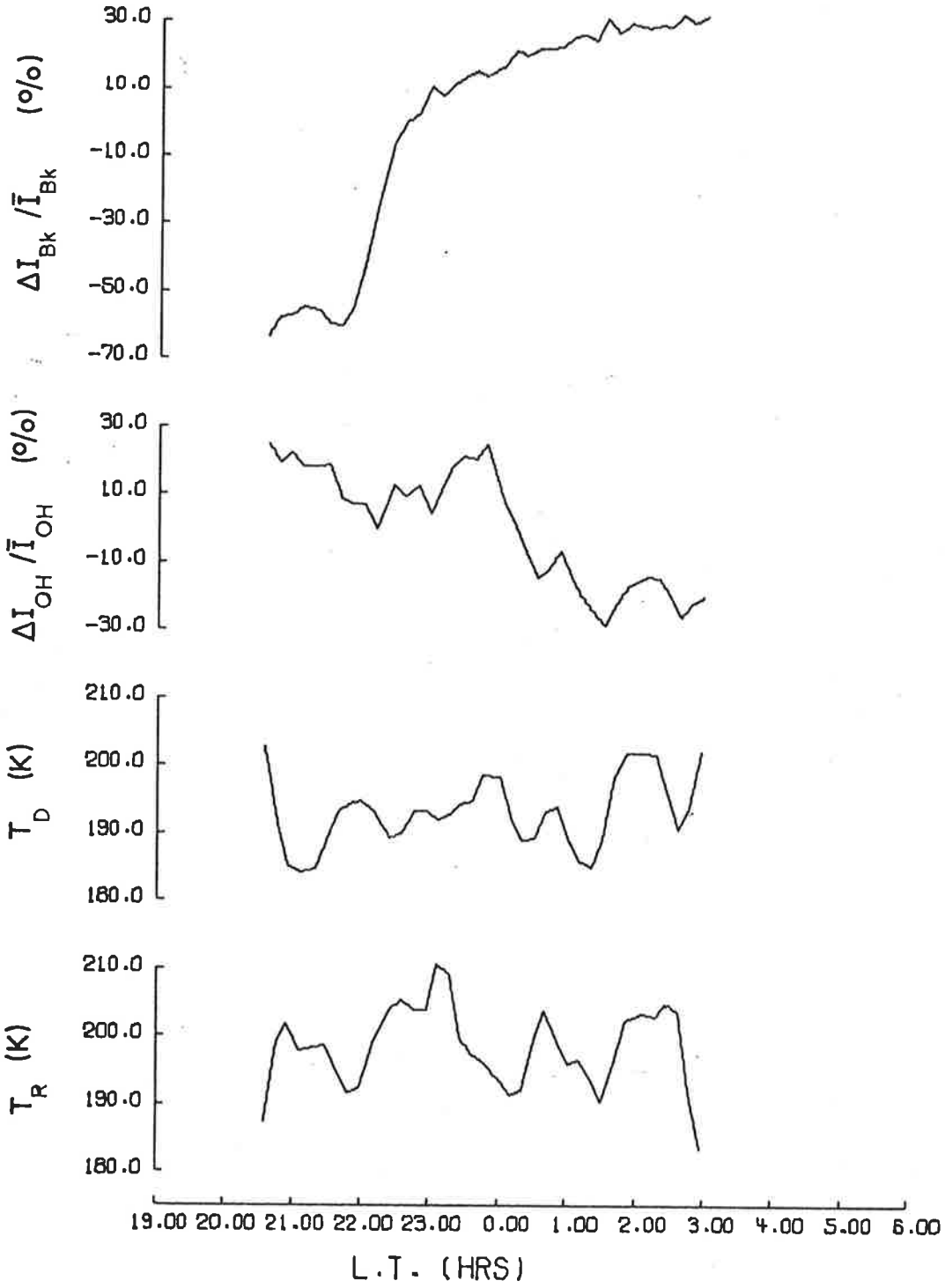


Figure 5.4 (c) As for Figure 5.4 (a) and 11 Nov 76

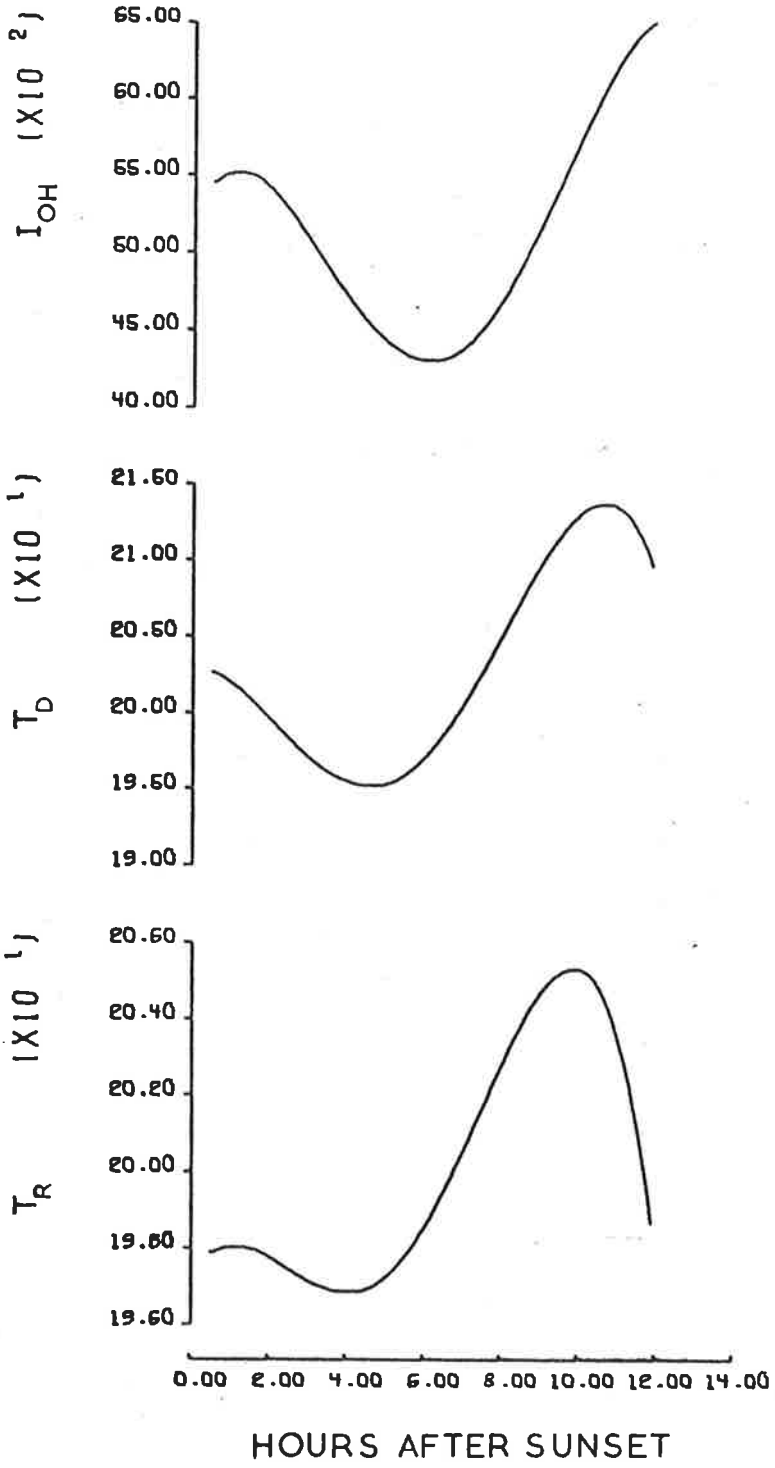


Figure 5.5 Smoothed mean nightly variation of the rotational (T_R) and doublet ratio (T_D) temperatures and the $P_1(3)$ line intensity (I).

with which the FPI measurements may usefully be compared. For the (8 - 3) band the results of Shefov (1971) and Dick (1972) suggest that on average there is a minimum in the vicinity of midnight, with an overall excursion of between 10 K and 40 K, while Visconti et al. (1971) found a fairly steady decrease throughout the night. Takahashi et al. (1974) found only small temperature changes and no systematic trend, as did Harrison et al. (1971) for the (4 - 1) band. In contrast the (8 - 3) band results of Sivjee et al. (1972) show a pronounced midnight maximum, the excursion sometimes being as large as 120 K.

A selection of the FPI results for T_R and T_D is shown in Figure 5.4. Since the estimated errors of these quantities were typically ±5 K and ±9 K respectively, the plots were smoothed by averaging over adjacent values with a binomial distribution spanning three points, thereby reducing the errors to about ±3 K and ±5 K. A similar procedure was not required for the intensities, which had a rather smaller error (usually < 1%) and exhibited much larger long term changes. Also shown in Figure 5.5 is the average nightly behaviour, found by fitting a fifth order polynomial to the massed data as described earlier. Although no particular allowance has been made in the mass plot for possible changes in the time of minimum temperature, its character was much the same as that observed by Dick. However the overall excursion of T_R was about a factor of three smaller. This was not particularly evident when the nights were examined individually, primarily because there was always considerable structure present. The magnitude of the overall temperature change during the night (both of T_R and T_D) ranged from 15 K to 50 K. On only one night, 14 April, was there a clear midnight temperature maximum, and this was accompanied by a pronounced intensity minimum.

There are several points of interest in the variation displayed

in Figure 5.5. Firstly, because measurements taken close to twilight were suspect (Section 5.4.3), and the reliability in the fitted polynomial decreases towards the ends of the interval, the drop in T_R at the end of the night might be regarded with reservation. Nevertheless the decrease was still present in curves drawn at one standard deviation from the original fitted polynomial, and was also present in T_D which, as discussed in Section 5.4.3, would rather be expected to increase were the effects of the twilight records present. Secondly, there was a lag in the time of minimum of the intensity and T_D relative to T_R, amounting to about two hours for the intensity and half an hour for T_D. There was also a statistically less significant lag in the time of the morning maximum of T_D. Thirdly, the amplitude of the average change in T_D was twice that of T_R.

In view of the variety of the behaviour seen by other workers, both the temperature and intensity variations found here appear to be not unreasonable. The question of the sensitivity of the analysis to real temperature changes will be pursued in Section 5.4.4.

5.4.3 Correlations between the Parameters

An examination of the parameter estimates for evidence of correlation is of considerable importance in that the transfer of counts from one component of the fitting function to another is a likely effect of a model error. Of primary concern is the question of correlations involving the total number of background counts for a spectrum since (a) this proved the most difficult quantity to model and as such was likely to be a source of error, (b) its accurate representation could be critical to the derived temperatures and (c) its behaviour was expected to be independent of the OH intensity and temperature.

The total background for several nights is shown in Figure 5.4. Two of these nights illustrate the severe conditions referred to earlier. On 6 June observations were made in the zenith, and the Milky Way crossed the field of view at approximately 0100 hours L.T. On 11 Nov the moon was about to enter its third quarter and rose (at 0 km) at 2210 L.T. Although the atmospheric water vapour content was not high on these occasions, there was no evidence that the background level influenced the intensity or temperature estimates. (The evidence from 6 June for the absence of an effect due to starlight was more convincing when taken in conjunction with the results of 14 April). The results for the nights of full moon, 16 and 17 Jan, when the background count rate was twice that of the $P_1(3)$ line and fairly constant, likewise showed no identifiable anomalies. Normally, however, the background was much less bright, contributing about one quarter of the $P_1(3)$ line count rate (i.e. about 15 R nm^{-1}) and changing by less than 40% during the night. Its effect on the other parameters should therefore have been negligible, at least when W was not too large. Even then there may be no serious error, as the results for 3 and 4 Feb, including the statistics $\overline{Q}$, $\overline{r}$ and $\overline{d}$, did not appear abnormal. Despite this it cannot be concluded that the background representation is exact - merely that the bias in the intensity and temperature estimates is smaller than their estimated random errors.

When the twilight spectra were examined a rather different picture emerged. It was found that as the background increased there was a steep decrease in T_R and increase in Q, T_D and W, but no systematic change in OH intensity. The trend continued until the OH line structure could no longer be distinguished in the recorded spectrum, and the analysis became incapable of finding a solution. Initially this behaviour suggested that the water vapour absorption function was wrong, and that

because it was derived using the sun as a source it was modelling the solar spectrum rather than atmospheric absorption. However the discovery that bright moonlight did not affect the fit prompted a search for an alternative explanation.

It appeared that the only spectral features likely to contribute at this time were the [O II] doublets (Section 5.3), and they were accordingly included in the model, with relative amplitudes as suggested by Dalgarno and McElroy (1965). The estimate of their intensity given by the analysis was found to increase steadily with decreasing solar depression angle δ, from a negative value at $\delta \simeq 19^\circ$ to about 10% of the $P_1(3)$ line intensity ($\sim$ 4R) at $\delta \simeq 14^\circ$, and with the change in the morning delayed relative to that in the evening. This result was in general agreement with the predictions of Dalgarno and McElroy (1966), but differed from the measurements of Carlson and Suzuki (1974) in that a significant contribution (i.e. > 0.5R) from these lines did not appear to be present during the night. With this alteration to the model the values of Q and W were reduced somewhat, but the trend of the temperatures with time was reversed. Because of the fairly close coincidence of the [O II] lines with the $P_1(2)$ and $P_1(3)$ OH lines, quite a small additional model error would have been sufficient to produce such an effect, but the nature of this error was not able to be determined, and the twilight spectra were therefore excluded from further consideration.

A search for correlations between the parameters describing the OH emission itself is much less useful for discovering model errors, since a degree of interdependence is to be expected. However a comparison of the behaviour of intensity and temperature can shed light on the dynamic processes occurring at the emission altitude. The FPI

results for each night were examined for dependence of the form

$$I = A \exp(B/T) \qquad \ldots(5.1)$$

and

$$I = C\, T^{n} \quad . \qquad \ldots(5.2)$$

These functions are suggested by the classical expression for the bimolecular rate constant

$$k \;\alpha\; T^{\frac{1}{2}} \exp(-E_A/RT) \quad . \qquad \ldots(5.3)$$

If E_A, the activation energy for the reaction, is not too small the exponential term dominates, giving equation (5.1), whereas if it is negligible equation (5.2) may be more appropriate. The behaviour to be expected of the ozone-hydrogen reaction is uncertain. Kaufman (1964), on the basis of the value obtained by Phillips and Schiff (1962) for the rate constant at 300 K, has suggested that E_A is less than 1.4 kcal mole^{-1}, Nicolet (1970) assumes E_A to be negligible, and other workers have adopted Bortner and Kummler's (1970) figure of B = -500 K for equation (5.1).

A degree of correlation between the intensity and both T_R and T_D was found. According to Student's t test the null hypothesis (corresponding to a parent distribution having a linear correlation coefficient of zero) could be rejected at the 10% level for about two thirds of the nights considered. One form of temperature dependence did not give noticeably better results than the other. The average values of the constant B were found to be -180 ± 60 K and -210 ± 60 K for T_R and T_D respectively, while n was on average 1.0 ± 0.3 for both T_R and T_D. The two temperatures were in general fairly well correlated, with only one night not being rejected by the significance test.

Several other workers have obtained results with which these findings may be compared. Harrison et al. (1971) obtained partial correlation with temperature for mean nightly intensities, but not for variations during the night. Takahashi et al. (1974) did find shorter term (∿ 3 hour averages) correlations and on two nights could relate I and T_R by equation (5.1) with B ≃ -500 K. Visconti et al. (1971) have also plotted ln I versus T^{-1} for a number of nights and claim to have found a real dependence of this form. Their average value of B for six nights was about -1000 K. Krassovsky et al. (1961) have likewise investigated the temperature dependence and obtained B = -580 K. In addition Krassovsky (1972) and Shagaev (1974), using equation (5.2), have reported an average value of n = 1.4 for a night, and instantaneous values as high as n = 5. The high time resolution results published by Shagaev show much stronger correlation than can be seen in the FPI data, the single night's record obtained by Meriwether (1975) or most of the examples presented by Shefov (1972). It therefore appears that this behaviour is rather exceptional, and that in general there is a weak dependence of intensity on rotational temperature both from night to night and during the night. There are no studies of short term relationships between T_R and T_D available for comparison.

5.4.4 Periodic Variations

Many of the records of I, T_R and T_D showed evidence of periodic behaviour, but this was usually of rather small amplitude and obscured by the irregular behaviour of the OH emission and the large variance of the measurements, particularly in the case of the temperatures. The power spectral density function (PSDF) was therefore calculated for each of the above parameters in a manner similar to that recommended by Bendat and Piersol (1971), using a mixed radix fast Fourier transform

algorithm. Because of the brevity of the time series involved, detrending was done by fitting a cubic polynomial by least squares methods rather than by applying a recursive high pass digital filter or by a linear detrend followed by truncation in the frequency domain. This procedure best preserved the information about low frequencies which were actually present in many of the records, but which because of the low resolution involved could not readily be distinguished from the trend. Significant errors were sometimes introduced, however, so that features with periods longer than 120 minutes had to be verified by inspection of the smoothed time series. Errors were also present at the highest frequencies because of segmentation of the records, which was brought about by the necessity to check the dark current and FPI alignment at intervals during the night. Fortunately this proved not to be a problem, since a feature was present in this part of the spectrum on only one occasion.

The usefulness of the PSDF in identifying spectral features was severely limited by the low statistical accuracy of the individual spectral estimates. For a Gaussian time series the relative error is approximately unity, while for the OH temperature results, where the estimated amplitudes of the wavelike phenomena were often less than twice the error of the measurement, the situation was little better. The usual methods of improving the accuracy, either by averaging over adjacent spectral estimates or by summing the spectra obtained from a segmented time series, are of little benefit because of the narrowness of the features typically encountered and the significant loss of resolution suffered, especially at low frequencies. By modelling the real data as the sum of a number of pure sinusoids and Gaussian noise an approximate criterion for identifying statistically significant peaks in the PSDF was obtained - namely that a peak should exceed by

a factor of three the average power density of band limited white noise, the variance of which was equal to that of the parameter (I, T_R or T_D) concerned.

The amount of power present in some low frequency narrow features necessitated the use of a Parzen apodizing function, of the type described by Otnes and Enochson (1972), to suppress leakage. This reduced the spectral resolution to between 7.5×10^{-5} Hz and 1.8×10^{-4} Hz. The Nyquist frequency was $\sim 1.3 \times 10^{-3}$ Hz for most nights, corresponding to the maximum sampling rate available (twice that used for the results as discussed so far, because the cutoff frequency for gravity waves at 85 km is about 3.5×10^{-3} Hz).

The PSDFs obtained for three nights, 3 Feb, 24 Sept and 25 Sept are shown in Figures 5.6 (a), (b) and (c), and are fairly representative. The units of the spectral ordinates are K^2 Hz^{-1} and $(\text{intensity})^2$ Hz^{-1}, where the units of intensity remain arbitrary. The approximate significance level adopted, calculated as outlined above, is indicated by the arrow.

The intensity spectra show a characteristic strong feature, which often contained more than 70% of the total power, at frequencies less than 1.4×10^{-4} Hz. That this is not an artefact of the detrending procedure on all occasions is clearly seen in the time series plot for 24 Sept (Figure 5.4(b)). On five nights a similar feature containing a smaller proportion of the power was also obtained for T_R and/or T_D, but the resolution was too low to establish whether or not this represented a true correlation. At higher frequencies significant peaks in either the intensity or T_R were usually obtained in conjunction with features (not always significant) in the other quantity. An example of this is found in Figure 5.6(a). However there were also several

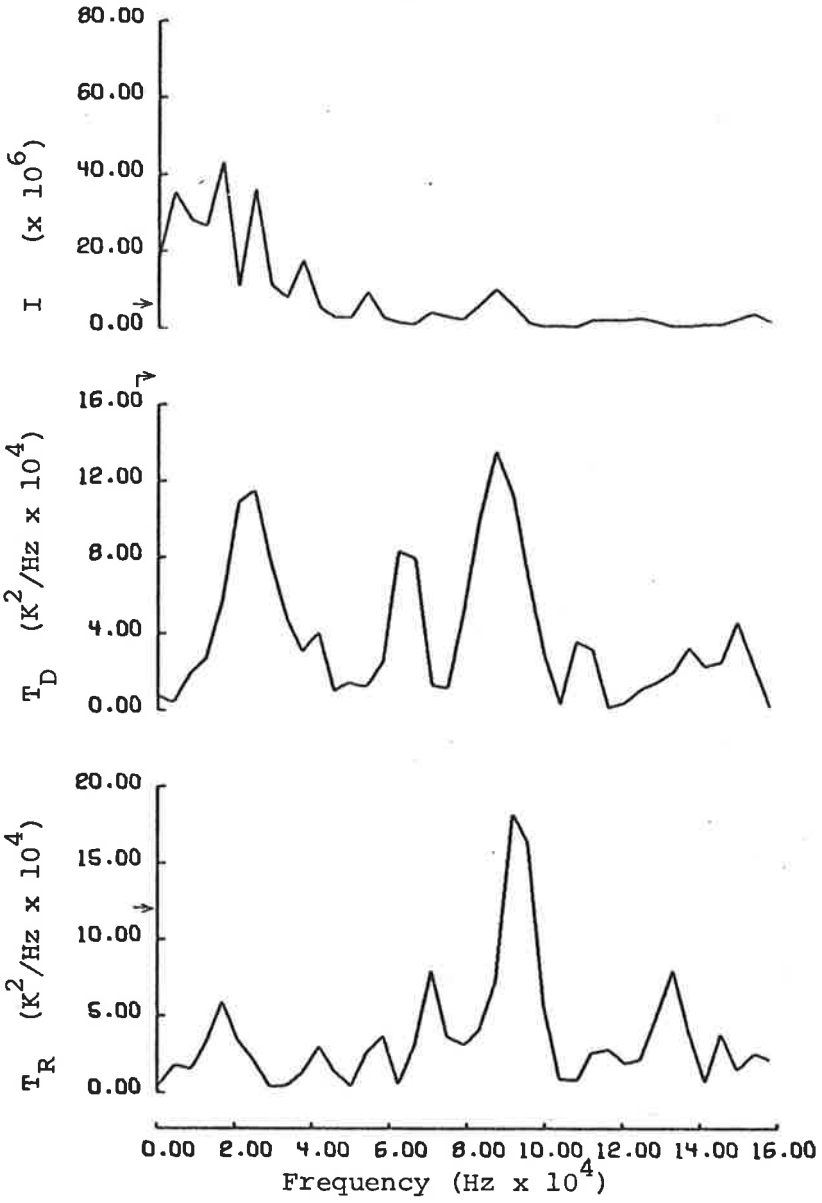


Figure 5.6 (a) Power spectral density functions of rotational (T_R) and doublet ratio (T_D) temperatures and the $P_1(3)$ line intensity (I) for 3 Feb 76. The arrows indicate the level above which the power was considered to be statistically significant.

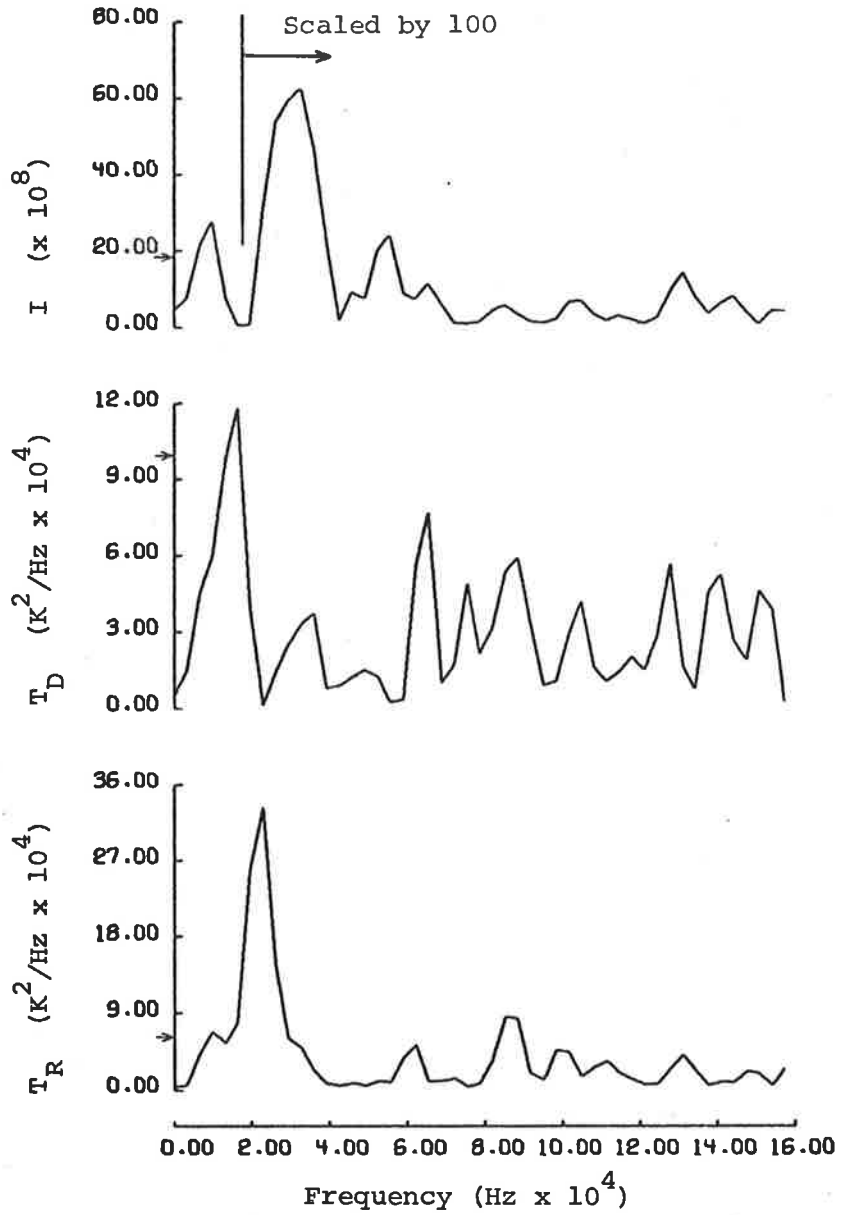


Figure 5.6 (b) As for Figure 5.6 (a) and 24 Sept 76

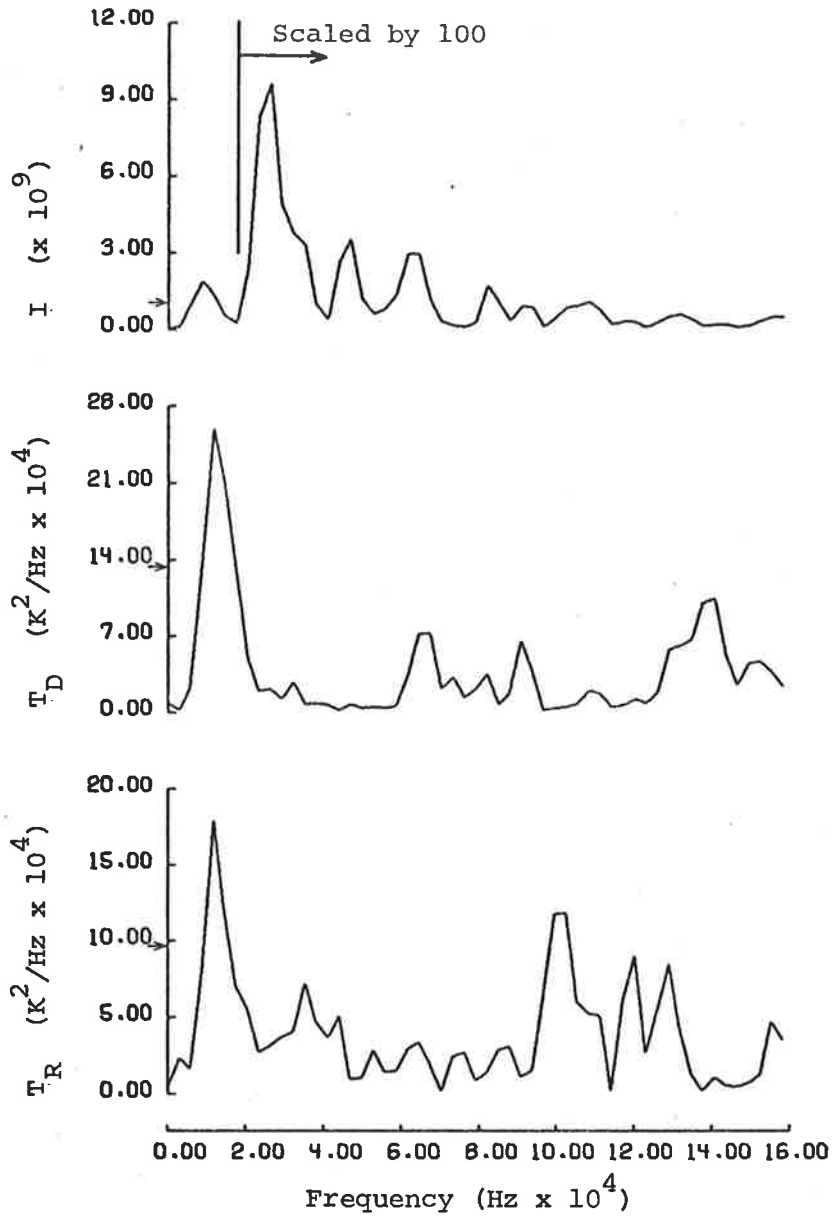


Figure 5.6 (c) As for Figure 5.6 (a) and 25 Sept 76

occasions when such was not the case. On 24 Sept a large oscillation in T_R with a period of $\sim$ 80 minutes had no obvious counterpart in the spectrum of the intensity, while on 25 Sept at a period of $\sim$ 65 minutes the reverse was the case. The amplitudes of the variations corresponding to the features in the intensity and T_R spectra ranged from 2K to 8K and 1% to 30% of the mean nightly intensity, with power generally concentrated towards lower frequencies.

Except for 11 Nov the uncertainties associated with T_D were too large for the PSDF to be useful in identifying periodic behaviour. However it was found that the mean nightly power was on all but one occasion lower in spectra for T_D than for T_R. Averaged over eight nights this quantity was 63 ± 4 and 75 ± 8 K^2 respectively. Since the fitting routine gives estimated uncertainties twice as large for T_D as for T_R, and since a large proportion of the measured power is due to noise, the difference in the average power is significant.

Periodic or quasiperiodic phenomena have been observed in the OH airglow by several other workers. The present results were not inconsistent with the measurements for a single night presented by Meriwether (1975), in which trains of wavelike small amplitude fluctuations with periods of $\sim$ 20 minutes were evident. On the other hand on none of the nights were there observed temperature variations of the magnitude or clear periodic character which can be seen in the events described by Armstrong (1975), even when allowance was made for the different spectroscopic constants used (Section 1.3.2). The same was true of most of the temperature records presented by Krassovsky and Shagaev (1974), which showed excursions in excess of 60 K in the space of several minutes. The FPI results most closely resemble the behaviour they designate as quiet.

These findings and those of Section 5.4.2 suggest that the least squares fitting procedure may be somewhat insensitive to real temperature variations. However such a conclusion is inconsistent with the comparatively low values found for the coefficients $|B|$ and n of equations (5.1) and (5.2), unless the model error which produces an underestimate of a temperature increase also produces a P_1 branch intensity decrease. Since the background is almost immune to OH parameter variations and since the P_1 and P_2 branch temperatures are set equal, it may be expected that this will result in an increase in the P_2 branch intensity and therefore an inflation of T_D (equation (1.9)). As shown in Figure 5.5, the amplitude of the average nightly behaviour of T_D was in fact twice that of T_R. However neither the mean nightly value of T_D nor the coefficient relating T_D and T_R was greater than expected. In addition the ranges of variation of the smoothed records of T_R and T_D taken on a nightly basis were equal on average. A more likely effect is that the P_1 branch intensity would be enhanced for large rotational temperatures, but in either event there was no obviously correlated change in intensity on the two occasions when T_R increased sharply by $\sim$ 30 K (e.g. Figure 5.4(a)). It is therefore very difficult to see how model errors could reduce the magnitude of the temperature variations by more than a few degrees.

5.4.5 Summary of Model Performance

In the preceding sections the results have been evaluated primarily with a view to discovering behaviour attributed to the model used in the least squares analysis. The principal findings and conclusions may be summarized as follows:

(a) There exists an error in the representation of the recorded spectrum.

(b) The mean nightly rotational temperatures obtained and their seasonal behaviour were in agreement with accepted neutral atmosphere temperatures at the assumed altitude of the emission.

(c) The rotational and doublet ratio temperatures were almost equal on average and were quite strongly correlated.

(d) The average nightly behaviour of W was well correlated with radiosonde data.

(e) The average nightly intensity and temperature variations were similar to those obtained by several other workers.

(f) The background intensity appeared not to influence other parameter estimates.

(g) A weak correlation between intensity and rotational temperature found by other workers was confirmed, although smaller average values for the coefficients $|B|$ or n of equations (5.1) and (5.2) were obtained.

(h) Short term quasiperiodic variations similar to those appearing in other results were also seen, though the magnitudes of the temperature fluctuations were generally smaller.

(i) Spectra recorded close to twilight were poorly modelled, and had to be discarded.

The nature of the model error remained unclear. Most of the simulations of likely errors produced an increase in the rotational temperature of only a few degrees, and changes in this bias caused by real variations in the intensity or temperature would have been even smaller. The most serious perturbation was caused by the [O II] lines, and these were apparently too weak to be modelled other than at twilight. It seems probable that in fact the error was the cumulative effect of

the approximations contained in the component functions of the model. It is concluded that the results of the analysis probably estimated the parent population values with an almost constant bias of less than ∿ 10 K. Henceforth it will also be assumed that the rotational temperature is a good measure of the ambient atmospheric temperature.

5.5 Dynamic Processes and the OH Emission

Understanding of the observed nocturnal OH temperature and intensity variation is at present incomplete. Existing aeronomic models predict that following an initial sharp increase just after dusk, the intensity should gradually decrease till dawn. At the same time the altitude of the emission layer is expected to rise steadily as ozone and hydrogen atoms are depleted at its base by reaction (1.4) and replenished at the top by downward transport of atomic oxygen, which forms ozone via

$$O + O_2 + M \rightarrow O_3 + M \qquad \ldots(5.4)$$

and recycles hydrogen atoms via reaction (1.5). The temperature behaviour should in this case be determined by the position of the layer relative to the mesopause. According to the model of Shimazaki and Laird (1970) and the CIRA (1972), the maximum temperature change to be expected at 35° latitude would be ∿ 10 K, in summer. On the basis of the midnight maxima in both I and T_R observed by them, Sivjee et al. (1972) have suggested an alternative dynamic model of OH production in which a rise in the emission height from 85 km to 100 km during the night is followed by a descent until dawn. Neither proposal is in agreement with the majority of the observations. Gattinger (1971) and Moreels et al. (1974 and 1976) have suggested that the increases in intensity are produced by an increase in the eddy diffusion coefficient

during the night, caused possibly by instabilities in tidal and/or gravity wave fields. Lindzen (1968) has calculated that this could occur in the vicinity of local midnight and persist for several hours, at least at low latitudes. Rocket measurements have also shown that the turbulent energy dissipation rate probably changes significantly during the night (Lloyd et al., 1972), and evidence for wave breaking has been found in power spectra of the [OI] λ557.7 nm intensity (Freund, 1976).

The calculations of Shimazaki and Laird (1970) give an emission peak some 5 km lower for an aeronomic model which incorporates eddy diffusion than for one which does not. In addition only two of the several rocket determinations of the layer height - Baker and Waddoups (1967) with 97 km, and Harrison (1970) with 95 km - yield values significantly above the mesopause height of ∿ 85 km. Therefore if tidal instabilities were to influence the OH emission in the manner suggested above, the following behaviour might be expected. Early in the night the layer should rise and the temperature and intensity decrease. Then as mixing increases the layer should descend and the temperature increase, followed by a recovery in the intensity as the increasing atomic oxygen concentration arrests and reverses the initial rapid loss of ozone.

This is the general form of the behaviour seen in the averaged FPI results and categorized by Dick (1972) as being typical. It may also be noted that the mechanism is broadly consistent with his atypical forms. The first of these, the very similar variations seen in the FPI results for 14 April, and the behaviour found by Sivjee et al. (1972) could represent an early reduction of turbulence, leading to a raising of the OH layer and a recovery of the temperature prior to dawn. The

second is much as would be expected if the OH layer were initially situated above the mesopause, as occasionally happens according to the rocket results just mentioned. However the magnitude (30 - 40 K) of the temperature changes observed in winter, when the temperature gradients near the mesopause are small, would require particularly large height changes and low emission altitudes.

Whether or not such a mechanism operates, the temperature and intensity should be responsive to atmospheric thermal and gravitational tides (Lindzen, 1967; Lindzen and Hong, (1974)). The mean diurnal behaviour of the atmosphere reflects a combination of several solar diurnal and semidiurnal modes (and a smaller lunar contribution), the relative importance of which are influenced by such factors as latitude, altitude and season. For the situation of interest here the diurnal temperature tide is expected to have an amplitude of $\sim$ 20 K and a maximum at 0600 hours L.T., and to be very sensitive to the tropospheric temperature structure. The semidiurnal tide should have an amplitude and time of maximum varying from $\sim$ 0.2 K and 1950 h at equinox to $\sim$ 2 K and 2130 h in summer and 10 K and 2330 h in winter. Generally applicable results for the density variation have apparently not been calculated, but measurements and calculations have been made for the equinox and 18° N (Kent et al., 1972). Here a density amplitude of $\sim$ 3% was found, and from the latitude distribution of temperature (Chapman and Lindzen, 1970) it might reasonably be expected to be rather greater at 35°.

These predictions are in moderate agreement both with the FPI results and those of Dick, insofar as the maximum temperatures and intensities occur close to dawn, as expected of the solar diurnal tide, while the secondary maximum found early in the night resembles the

semidiurnal component. The temperature amplitudes are also of the same order, bearing in mind that no account has been taken of a possible seasonal change in phase in Figure 5.5. However the density perturbations required to produce the observed intensity changes might be somewhat larger than the tide. Closer comparison is not possible at present, owing to the uncertainties of the theory and the paucity of the data. For example quite large discrepancies have been found between the theory and rocket measurements of the temperature made at 5° latitude (Groves, 1976). Comparison will also be hindered by the fact that the amplitude and phase of the tide are highly oscillatory functions of altitude in the region of the OH emission.

There are two other points to be made in connection with the nocturnal behaviour. The first is that according to either of the processes just outlined there is no simple relationship to be expected between the mean nightly temperature and intensity, although the intensity amplitude is probably augmented by a temperature dependent rate constant for OH production (see following). Secondly, because the observations are of the (8 - 3) band they are less likely to be greatly affected by quenching by atomic oxygen (see however the discussion following equation (5.14)) or by production processes other than the ozone-hydrogen reaction. Quenching principally affects the lower vibrational levels which have longer lifetimes, while additional production is thought to be likely only via reaction (1.8), which populates states with $v'' \lesssim 6$. On the other hand the results of Harrison et al. (1971), who used the (4 - 1) and (5 - 2) bands, may be influenced by these processes.

It is clear from the results of the Russian workers (Krassovsky and Shagaev, 1974 a and b), who have used spaced receivers, and of Moreels

and Herse (1977) who have used photographic methods, that the OH airglow is influenced by internal gravity waves. Their dominant period at 85 km is expected to be > 2 hours, as in fact was observed in the power spectra of the FPI results. The surprising aspect of these results was the occasional lack of coincidence of features in the spectra for I and T_R. The relationship to be expected was therefore considered more closely.

OH molecules are formed in the level v = 8 both directly via the ozone-hydrogen reaction and by radiative cascade from the ninth level. Loss processes are radiative transitions to lower levels and quenching in two body reactions with a number r of species which may be denoted X_i. The equilibrium condition can be described by

$$\alpha_8 \, k [O_3] [H] + A_{98} [OH_9] = \sum_{v=0}^{7} A_{8v} [OH_8] + \sum_{i=1}^{r} k_i [X_i] [OH_8] \qquad \ldots(5.5)$$

where $A_{v'v''}$ is the Einstein transition probability as used in Chapter 1 (omitting the →)

$[X_i]$ denotes the concentration of the species X_i. When X = OH the subscript denotes the vibrational level of excitation

k, k_i are the rate coefficients for reaction (1.4) and for the quenching reactions with the species X_i respectively

α_v is the fraction of the OH molecules formed by reaction (1.4) in the level v.

The number of OH molecules emitting into the (8 - 3) band per second from a volume V defined by the spectrometer field of view will be

$$I_{83} = V \, A_{83} [OH_8] \, . \qquad \ldots(5.6)$$

If the emission layer is subject to adiabatic perturbations (such as a

gravity wave) which obey the relation

$$T V^{\gamma-1} = C = \text{constant} \qquad \ldots(5.7)$$

and if the rate constants of the reactions of equation (5.5) have a temperature dependence of the form

$$k_i = k'_i \, T^{n_i} \qquad \ldots(5.8)$$

then (5.5) to (5.8) can be combined to give

$$I_{83} = A_{83} \frac{C \, \alpha_8 k'(O_3)(H) T^{n + (\gamma-1)^{-1}} + A_{98}(OH_9)}{\sum_v A_{8v} + \sum_i C k'_i (X_i) \, T^{n_i + (\gamma-1)^{-1}}} \qquad \ldots(5.9)$$

Here

$$[X_i] = (X_i) \, V^{-1} \qquad \ldots(5.10)$$

The quantity $[OH_9]$ can be evaluated using the equilibrium relation for the ninth level, namely

$$\alpha_9 \, k [O_3] [H] = \sum_{v=0}^{8} A_{9v} [OH_9] + \sum_i k_i [X_i] [OH_9] \qquad \ldots(5.11)$$

From equation (5.10) a relationship of the form

$$\frac{dI}{I} = \eta \frac{dT}{T} \qquad \ldots(5.12)$$

can then be derived. The expression for η can be simplified considerably by assuming that all the n_i for the r quenching reactions are equal - an approximation of little significance since in practice only one reaction need be considered. The result for η is

$$\eta = \frac{n + (\gamma-1)^{-1}}{1 + \dfrac{\alpha_9 A_{98}}{\alpha_8 \left(\sum_v A_{9v} + \sum_i k_i [X_i] \right)}} - \frac{n_x + (\gamma-1)^{-1}}{1 + \dfrac{\sum_v A_{8v}}{\sum_i k_i [X_i]}} \qquad \ldots(5.13)$$

The range of values which η can adopt can now be estimated with the help of the following results. The Einstein coefficients obtained by Murphy (1971) give $\sum_v A_{9v} = 15.8$, $\sum_v A_{8v} = 12.3$ and $A_{98} = 2.1$. In the 80 - 90 km region only atomic oxygen is important as a quenching agent, and has a concentration of between 10^{10} and 8×10^{11} cm^{-3} (on the basis of measurements by Evans and Llewellyn, 1973, and aeronomic model estimates) and a rate coefficient of $3 \times 10^{-12} T^{n_o}$ cm^3 s^{-1} with $0 < n_o < 1$ (Kaufman, 1964). A conservative value for n and n_o of 0.5 will be used here. From the measurements of Charters et al. (1971) and Fiocco and Visconti (1974) a value of 1 : 1.5 will be taken for the ratio $\alpha_9 : \alpha_8$. The ratio of specific heats, γ, is 1.4.

When quenching is virtually absent ($[O] < 10^{10}$ cm^{-3})

$$\eta = \frac{n + 2.5}{1 + 0.13 \frac{\alpha_9}{\alpha_8}} \simeq 2.5 \qquad \ldots(5.14)$$

for small ΔT, and is a maximum. It should increase slightly with decreasing altitude as $[O]$ decreases and $\alpha_9 : \alpha_8$ is reduced according to the pressure dependence found by Charters et al. A minimum value for η should occur for very strong quenching, when

$$\eta \simeq n - n_o \; . \qquad \ldots(5.15)$$

More realistically, if $[O] = 8 \times 10^{11}$ cm^{-3}, $\eta = 0.6$. Thus the power spectra for I and T_R will be very sensitive both to the amount of quenching and the temperature dependence of the ozone-hydrogen and hydroxyl-oxygen reactions. In view of the uncertainties in the latter it is possible that η could be as high as 5, corresponding to no quenching and n = 1.7, or as low as 0, if for example $[O] = 8 \times 10^{11}$ cm^{-3}, n = 0 and $n_o = 0.7$.

One of the examiners of this thesis has proposed that, whether or not reaction (1.8) contributes significantly to $[OH^*]$, there is significant difficulty in explaining the population of the lower vibrational levels. He suggests that by considering vibrational cascade quenching of OH^* by O_2 and N_2 the matter may be resolved, and that in fact the interaction of OH with O is insignificant. (Llewellyn, E.J. et al, 1977 - Papers in press.) The latter conclusion is certainly supported by the new values obtained by Streit and Johnston (1976) for the rate coefficients of the reactions $O_2 + OH^*$ and $N_2 + OH^*$. Using concentrations at 85 km obtained from CIRA 1972 and these values, as well as the figure obtained for atomic oxygen by Kaufman (1964), the quantity $k_i[X_i]$ evaluates approximately to 7.0, 7.0 and 18.4 s^{-1} when $X_i = O, O_2$ and N_2 respectively. (This also assumes a $T^{\frac{1}{2}}$ temperature dependence for the quenching rate coefficients.) Nevertheless, the general conclusion drawn in Section 5.5 - namely that the quantity η of equation (5.12) is strongly dependent on the temperature dependence of the reaction rates - remains unaltered. It should also be noted that if, in fact, vibrational cascade quenching is important, an additional production term of the form $\Sigma\beta_i k_i[X_i]\,[OH_9]$ must be included in the left hand side of equation (5.5). The quantity β represents the fraction of the molecules which are quenched to the 8^{th} level.

These calculation were compared with the power spectra illustrated in Figure 5.6. $\frac{\Delta T_R}{T_R}$ for the large feature on 24 Sept was 3 ± 1% and $\frac{\Delta I}{I}$ was < 1 ± 1%, so that on this occasion η was < 1. For the peak at 2.6×10^{-4} Hz on 25 Sept, $\frac{\Delta I}{I}$ = 4 ± 1% and $\frac{T_R}{T_R}$ < 0.4 ± 1%, giving η > 2. The peaks at 8.8×10^{-4} Hz on 3 Feb gave η = 0.5 ± 0.9 and other results were similar. However a much larger value for η (> 13) was found for the large ∿ 200 minute period oscillation appearing in the intensity record from 24 Sept. This suggests that the subsidiary maximum occurring near midnight (and also on four other nights, including 11 Nov) represents changes in local concentrations by a process such as an increase in the mixing ratio, rather than by adiabatic compression associated with a gravity wave.

Finally two matters in relation to the behaviour of T_D bear comment. The first is that both the lag in the nightly average variation of T_D relative to T_R and the smaller value obtained for the average nightly power of T_D suggest the operation of a process with a long time constant. This might, for example, be a relatively inefficient redistribution of energy between $^2\pi_{3/2}$ and $^2\pi_{1/2}$ states which have initial relative populations corresponding to a Boltzmann temperature close to T_R. The second is that whereas to date the larger average nightly excursion of T_D has been taken as suggestive of a bias in the estimation of T_R, it may well be a real effect characteristic of the manner in which the spin doublet states are excited.

CHAPTER 6

CONCLUDING REMARKS

In this thesis the construction of a small Fabry-Perot interferometer and its use in the investigation of the nocturnal and short term variations of the hydroxyl (8 - 3) band airglow has been described. The results obtained have been examined both to establish the absence of errors induced by the methods employed to analyse the low resolution spectra, and in terms of the dynamic processes of the upper atmosphere.

Since the intention was that the OH rotational temperatures should be obtained as rapidly as possible, consideration was given to optimizing the parameters of the spectrometer, particularly the operating bandwidth, in terms of the shape of the airglow spectrum, and to maximizing the efficiency of light detection by a suitable choice of detector, cooling and the use of a quantum efficiency enhancement device. In order to minimize expense and complexity, the spectrometer was designed to operate without servo-control of the plate separation or parallelism. Investigations of the behaviour of several types and configurations of piezoelectric material were therefore undertaken. They showed that the requirements of a parallelism stability of $\lambda/250$ for several hours and a linearity of 1% over a third of an order scan could just be met by using PZT-4 ceramic tubes and strictly periodic driving waveforms. This degree of linearity was, however, insufficient to prevent a variation of apparent instrument width across the interval scanned.

The nonlinear least squares estimation technique used to obtain relative intensities and spectroscopic temperatures from the raw data have been described, with particular attention being paid to the

problem of determining the effects of water vapour absorption on the background spectrum. Some of the properties of the algorithm used to solve the estimation problem have been investigated to ensure that it was both convergent and efficient in this context. Methods by which the goodness of fit may be assessed, and means of obtaining estimates for the uncertainties of the parameters of the fit have also been described.

Observations from fourteen clear nights spanning eighteen months, from 14 April 75 to 11 November 76, were analysed, and provided rotational and doublet ratio temperatures and relative intensities of the OH band and background at intervals of 5 or 10 minutes throughout the night, and with an accuracy of ±3 K to ±9 K. The behaviour of the analysis with respect to perturbed synthetic data was investigated, the residuals of the fit were subject to statistical tests and the parameter estimates were compared with known or expected values. On this basis it was concluded that there existed an error in the function used to model the recorded spectrum, but that, with the exception of measurements made close to twilight and a possible reduced sensitivity to real temperature fluctuations, the results were substantially unbiased.

The mean nightly rotational temperature was found to have values and a seasonal dependence consistent with an emission layer height of about 80-90 km. Its mean nightly variation and that of the intensity consisted of a decrease from dusk until the middle of the night, followed by a larger increase towards dawn, with the temperature beginning to decrease just prior to dawn. The intensity minimum lagged that of the temperature by about two hours. The data were too sparse to detect any seasonal trend in this behaviour. These

characteristics were considered with respect to two possible causal mechanisms - that of an increase in the eddy diffusion coefficient during the night, and atmospheric tides. Both appeared to be substantially compatible with the observed covariation, but a resolution of the problem awaits confirmation of a systematic variation of the mixing coefficient, further calculations of tidal effects and a greater body of OH data of this type. In addition to this average behaviour a small midnight peak in intensity, not obviously attributable to tides or gravity waves, was observed on five of the fourteen nights.

Periodicities evident in both the intensity and temperature records were examined by calculating their power spectra. Despite the low accuracy involved it appeared that on several occasions spectral peaks in the temperature PSDF were not present in the intensity and vice versa. It was shown that, depending on the degree of quenching of the vibrationally excited OH molecule and the temperature sensitivity of the reaction rate coefficients for quenching and production, this was to be expected. Power was concentrated at periods greater than two hours, as expected for gravity waves at the emission altitude.

The mean nightly value and variation of the doublet ratio temperature were found to correspond quite closely to that of the rotational temperature. However its diurnal peak-to-peak variation was twice as great, both its minimum and pre-dawn maximum lagged slightly, and its spectrum contained significantly less power. The reason for such behaviour is not clear, but is suggestive of a process of collisional relaxation between spin doublet states in violation of the $\Delta S \neq 0$ quantum selection rule.

Given the presence of a water vapour line coincident with the $P_1(4)$ OH line, the spectrometer could have been improved in a number of respects. Most importantly, since its operating range was reduced from the design figure of 9.8 nm to 6.0 nm, the width of the prefilter could have been made correspondingly smaller and the reflectance of the plate coatings reduced to match the new finesse requirements. To do so would produce a considerable improvement in light gathering power, since both τ_A and the ratio N_D/N_R would be increased, and would reduce light leakage arising from the relatively large transmission between orders. At the same time the opportunity could be taken to shift the edge of the filter from the region of the $P_2(2)$ line, where it caused problems in the fit. Had the $P_1(4)$ line not been blocked other modifications would have been desirable. For a full 9.8 nm scan some correction for the nonlinearity of the piezoelectric elements would have been necessary, and the averaging time would have suffered because of the need to scan at a uniform rate. However the present results suggest that there is little reason to scan beyond the $P_2(4)$ line, particularly because of the loss of temporal resolution which this would entail.

The principal drawback of the technique employed here lay in the enforced use of a low spectral resolution. The processing of the rather large quantities of raw data involved was very labour-intensive and expensive of computer time. More importantly, however, as with all such estimation problems, departures from the postulated model, and any bias in the parameter estimates when the fit was less than perfect, were extremely difficult to isolate and identify. This problem was compounded by the inability of the analysis to distinguish between components of the model which are similar in form, as in the case of the water vapour function and the slope of the airglow

continuum. The problem is not unique to the nonlinear fit, and in fact some experimentation with the linear model, which could be used if water vapour effects were absent and which would save considerably on computer time, suggested that an increase in the spectrometer resolution would then be necessary.

On the basis of the foregoing remarks, it is considered that a better method of obtaining much the same information from the (8 - 3) band would be by means of a multiple filter photometer. Two or three period dielectric filters offer the advantages of better rejection away from the centre wavelength, higher peak transmission and, for three periods, a flat-topped profile which reduces the positioning accuracy required. The high light gathering power of the FPI would be improved, construction and operation would be simpler and the quantity of raw data and the computation required very much less. If used with three channels the instrument would also be well suited to the measurement of gravity wave velocities by the spaced receiver method. Rotational temperature variations with an accuracy of about 1% could be obtained every one minute, using a single 5% quantum efficiency detector, three 48 mm diameter three period filters, a 10° field of view at the filter and the $P_1(2)$ and $P_1(3)$ lines of the (8 - 3) band. The effects of water vapour absorption would of course still be suffered. However calculations based on the absorption profile used in the FPI model show that on average this would only cause an overestimation of the temperature of between 5 K and 8 K, corresponding to the likely range of variation of W during the night.

Apart from the question of instrumentation, the present work could be extended in a number of respects. There is need for a greater body of data on the average nightly behaviour of the OH temperature,

and for both this and the intensity variation to be more closely compared with the predictions of tidal theory. Better quality power spectra may enable some measure to be made of changes in the turbulent energy dissipation rate during the night, and could shed further light on the relationship between intensity and temperature and the role played by quenching agents. Spaced receiver observations of the type already begun by the Russian workers would be particularly valuable in conjunction with similar measurements of the [OI] 557.7 nm and 630.0 nm lines.

APPENDIX

Values of $F_{i,vJ}$ in cm^{-1} (where i = 1, 2 correspond to the $^2\pi_{3/2}$ and $^2\pi_{1/2}$ states) required for the (8 - 3) band are as follows:

	$^2\pi_{3/2}$		$^2\pi_{1/2}$	
K	v = 8	v = 3	v = 8	v = 3
1	-47.81	-42.04	83.59	86.72
2	12.00	32.87	125.20	140.45
3	95.99	138.42	194.24	229.43
4	204.30	274.94	290.33	353.02
5	337.00	442.67	413.04	510.59

They were calculated using the formulae of Hill and Van Vleck (1928):

$$F_{1,vK} = B_v\{(K + 1)^2 - 1 - \tfrac{1}{2}[4(K + 1)^2 + Y_v(Y_v - 4)]^{\frac{1}{2}}\} - D_v K^2(K + 1)^2$$

$$F_{2,vK} = B_v\{K^2 - 1 + \tfrac{1}{2}[4K^2 + Y_v(Y_v - 4)]^{\frac{1}{2}}\} - D_v K^2(K + 1)^2$$

and the spectroscopic constants given by Herman and Hornbeck (1953) and Wallace (1960):

	v = 8	v = 3
B_v	12.91	16.414
Y_v	-10.95	-8.568
D_v	0.0018	0.0018

The wavelengths λ, Einstein coefficients A and line strengths S for the lines of interest are:

Line	J'	λ_{Air} (nm)	A	S
$Q_1(2)$	$\frac{5}{2}$	728.44	0.329	2.30
$Q_1(3)$	$\frac{7}{2}$	729.59	0.135	1.40
$Q_1(4)$	$\frac{9}{2}$	731.09	0.072	0.98

$P_1(2)$	$\frac{3}{2}$	731.63	0.234	1.62
$P_1(3)$	$\frac{5}{2}$	734.09	0.298	2.90
$P_1(4)$	$\frac{7}{2}$	736.93	0.321	4.04
$P_1(5)$	$\frac{9}{2}$	740.17	0.335	5.14
$P_2(2)$	$\frac{1}{2}$	730.35	0.387	1.29
$P_2(3)$	$\frac{3}{2}$	732.89	0.360	2.34
$P_2(4)$	$\frac{5}{2}$	735.83	0.354	3.37
$P_2(5)$	$\frac{7}{2}$	739.18	0.355	4.37

The wavelengths were calculated using the above term values $F_{i,vK}$ and the expression

$$\lambda_{vac} = [F_{i,v'K'} - F_{i,v"K"} + G_o(v') - G_o(v")]^{-1}$$

where from Herman and Hornbeck and Wallace

$$G_o(3) = 10213.83 \text{ cm}^{-1}$$

$$G_o(8) = 23958.9 \text{ cm}^{-1}$$

REFERENCES

Aarons, B.D. and Manuel, J. (1969). Enhancement of photomultiplier sensitivity by an optical method. *J. of Phys. E : Sci. Instrum.* 2, *734-735.*

Armstrong, E.B. (1975). The influence of a gravity wave on the airglow hydroxyl rotational temperature at night. *J. Atmos. Terr. Phys.* 37, *1585-1591.*

Arndt, R.A. and MacGregor, M.H. (1966). Nucleon-nucleon phase shift analyses by Chi-squared minimization. *Methods in Computational Physics* 6, *253-296.*

Anscombe, F.J. and Tukey, J.W. (1963). The examination and analysis of residuals. *Technometrics* 5, *141-160.*

Baker, D.J., Steed, A.J. and Stair, A.T. (1973). Ground observations of resolved hydroxyl ($\Delta v = 2$) airglow. *J. Geophys. Res.* 78, *8859-8863.*

Baker, D.J. and Waddoups, R.O. (1967). Rocket measurements of mid-latitude night airglow emissions. *J. Geophys. Res.* 72, *4881-4883.*

Bard, Y. (1970). Comparison of gradient methods for the solution of nonlinear parameter estimation problems. *SIAM J. Numer. Anal.* 7, *157-186.*

Bass, A.M. and Garvin, D. (1962). Analysis of the hydroxyl radical vibration rotation spectrum between 3900 Å and 11500 Å. *J. Mol. Spectrosc.* 9, *114-123.*

Bates, B., Conway, J.K., Courts, G.R., McKeith, C.D. and McKeith, N.E. (1971). A stable, high finesse scanning Fabry-Perot interferometer with piezoelectric transducers. *J. of Phys. E, Sci. Instrum.* 4, *899-901.*

Bates, D.R. and Moiseiwitsch, B.L. (1956). Origin of the Meinel hydroxyl system in the night airglow. *J. Atmos. Terr. Phys.* 8, *305-308.*

Bates, D.R. and Nicolet, M. (1950). Atmospheric hydrogen. *Publ. Ast. Soc. Pacific* 62, *106-110.*

Bendat, J.S. and Piersol, A.G. (1971). 'Measurement and Analysis of Random Data', (J. Wiley and Sons, New York).

Benedict, W.S. and Plyler, E.K. (1954). High resolution spectra of hydrocarbon flames in the infrared. ('Energy Transfer in Hot Gases') *U.S. Nat. Bureau Stds. Circ. 523, 57-73.*

Benedict, W.S., Plyler, E.K. and Humphreys, C.J. (1952). The emission spectrum of OH from 1.4 to 1.7 μ. *J. Chem. Phys. 21, 398-402.*

Bevington, P.R. (1969). 'Data Reduction and Error Analysis for the Physical Sciences', (McGraw-Hill, New York).

Blackwell, D.E., Ingham, M.F. and Rundle, H.N. (1960). The night sky spectrum λλ5000-6500 Å. *Astrophys. J. 131, 15-24.*

Born, M. and Wolf, E. (1975). 'Principles of Optics', (Pergamon Press, Oxford).

Bortner, M.H. and Kummler, R.H. (1970). *D.A.S.A. Report No. 2407.*

Bower, A.R.D. (1974). A Fabry-Perot interferometer and its application to measurement of thermospheric temperatures and winds. Ph.D. Thesis, University of Adelaide.

Bracewell, R.N. (1965). 'The Fourier Transform and its Applications', (McGraw-Hill, New York).

Breig, E.L. (1970). Secondary production processes for the hydroxyl atmospheric airglow. *Planet. Space. Sci. 18, 1271-1274.*

Brigham, E. Oran (1974). 'The Fast Fourier Transform', (Prentice-Hall, Inc., Englewood Cliffs, N.J.)

Broadfoot, A.L. and Kendall, K.R. (1968). The airglow spectrum, 3100-10,000 A. *J. Geophys. Res. 73, 426-428.*

Carlson, R.W. and Suzuki, K. (1974). Observation of O^+ ($^2P^0$ - $^2D^0$) λ7,319 Å emissions in the twilight and night airglow. *Nature 248, 400-401.*

Chabbal, R. (1953). *J. Rech. Cent. Natl. Rech. Sci. Lab. Bellevue (Paris) 5, No. 24, 138-186.* English Translation: Research on the best conditions for using a Fabry-Perot photo-electric spectrometer. A.E.R.E. Lib./Trans. 778, Harwell, U.K.

Chabbal, R. (1958). Finesse limite d'un Fabry-Perot formé de lames imparfaites. *J. Phys. Rad. 19, 295-300.*

Chamberlain, J.W. (1961). 'Physics of the Aurora and Airglow', (Academic Press, N.Y. and London).

Chapman, S. and Lindzen, R.S. (1970). 'Atmospheric Tides - Thermal and Gravitational', (Reidel Publishing Co., Dordrecht - Holland).

Charters, P.E., Macdonald, R.G. and Polanyi, J.C. (1971). Formation of vibrationally excited OH by the reaction H + O_3. *Appl. Opt.* **10**, *1747-1754*.

C.I.R.A. (1972). COSPAR International Reference Atmosphere 1972, (North-Holland Publishing Co., Amsterdam).

Clarke, J.H.R., Norman, M.A. and Borsay, F.L. (1975). A high performance Fabry-Perot interferometer for Rayleigh and Raman scattering studies. *J. of Phys. E : Sci. Instrum.* **8**, *144-146*.

Cocks, T.A. (1977). Studies of the atomic oxygen dayglow. Ph.D. Thesis, University of Adelaide.

Conte, S.D. (1965). 'Elementary Numerical Analysis - An Algorithmic Approach', (McGraw-Hill, New York).

Curcio, J.A., Drummeter, L.F. and Knestrick, G.L. (1964). An atlas of the absorption spectrum of the lower atmosphere from 5400 Å to 8520 Å. *Appl. Opt.* **3**, *1401-1409*.

Dalgarno, A. and McElroy, M.B. (1965). The fluorescence of solar ionizing radiation. *Planet. Space Sci.* **13**, *947-957*.

Dalgarno, A. and McElroy, M.B. (1966). Twilight effects of solar ionizing radiation. *Planet. Space Sci.* **14**, *1321-1329*.

Dick, K.A. (1972). Short time-interval spectrometric hydroxyl emission studies. *Ann. Géophys.* **28**, *149-153*.

Dick, K.A., Sivjee, G.G. and Crosswhite, H.M. (1970). Aircraft airglow intensity measurements: Variations in OH and OI (5577). *Planet. Space Sci.* **18**, *887-894*.

d'Incan, J., Effantin, C. and Roux, F. (1971). Intensites absolues et forces d'oscillateur de quelques raies des bandes de vibration-rotation (1-0) et (2-1) du radical OH. *J. Quant. Spectrosc. Radiat. Transf.* **11**, *1215-1224*.

Draper, N.R. and Smith, H. (1966). 'Applied Regression Analysis', (J. Wiley and Sons, New York).

Drummond, D. and Gallagher, A. (1973). A low resolution scanning multiple Fabry-Perot spectrometer. *Rev. Sci. Instrum.* 44, 396-399.

Durbin, J. and Watson, G.S. (1950). Testing for serial correlation in least squares regression. I. *Biometrica 37, 409-428.*

Durbin, J. and Watson, G.S. (1951). Testing for serial correlation in least squares regression. II. *Biometrica 38, 159-178.*

Eather, R.H. and Reasoner, D.L. (1969). Spectrophotometry of faint light sources with a tilting-filter photometer. *Appl. Opt. 8, 227-242.*

Evans, W.F.J. and Llewellyn, E.J. (1973). Atomic hydrogen concentrations in the mesosphere and the hydroxyl emissions. *J. Geophys. Res. 78, 323-326.*

Fiocco, G. and Visconti, G. (1974). Simultaneous measurements of the OH* nightglow in the (9-4), (8-3) and (5-1) bands and effects of quenching. *J. Atmos. Terr. Phys. 36, 583-590.*

Fiocco, G., Visconti, G., Congeduti, F. (1970). Nocturnal variation of the intensity and rotational temperature of the OH (8,3) band in the airglow. *Nature 228, 1079-1080.*

Freund, J.T. (1976). Structure in the 557.7 nm oxygen airglow. Ph.D. Thesis, University of Adelaide.

Gadsen, M. and Marovich, E. (1973). The nightglow continuum. *J. Atmos. Terr. Phys. 35, 1601-1614.*

Gattinger, R.L. (1971). Interpretation of airglow in terms of excitation mechanisms. 'The Radiating Atmosphere', (Ed. B.M. McCormac), 51-63. (Reidel Publishing Co., Dordrecht-Holland).

Gattinger, R.L. and Vallance Jones, A. (1973). Observation and interpretation of hydroxyl airglow emissions. 'Physics and Chemistry of Upper Atmospheres', (ed. B.M. McCormac), 184-192. (Reidel Publishing Co., Dordrecht-Holland).

Groves, G.V. (1976). Rocket studies of atmospheric tides. *Proc. Roy. Soc. Lond. A351, 437-469.*

Gunter, W.D. Jr., Grant, G.R. and Shaw, S.A. (1970). Optical devices to increase photocathode quantum efficiency. *Appl. Opt. 9, 251-257.*

Harrison, A.W. (1969). The night airglow spectrum 1.02 μ - 1.13 μ. *Planet. Space Sci.* 17, 173-179.

Harrison, A.W. (1970). Altitude profile of airglow hydroxyl emission. *Can. J. Phys.* 48, 2231-2234.

Harrison, A.W., Evans, W.F.J. and Llewellyn, E.J. (1971). Study of the (4-1) and (5-2) hydroxyl bands in the night airglow. *Can. J. Phys.* 49, 2509-2517.

Harrison, A.W. and Kendall, D.J.W. (1973a). Airglow hydroxyl doublet ratio temperatures. *J. Geophys. Rev.* 78, 4697-4701.

Harrison, A.W. and Kendall, D.J.W. (1973b). Airglow hydroxyl intensity measurements 0.6 - 2.3 μ. *Planet. Space Sci.* 21, 1731-1741.

Henshaw, R.C. (1966). Testing single equation least squares regression models for autocorrelated disturbances. *Econometrica* 34, 646-660.

Herman, R.C. and Hornbeck, G.A. (1953). Vibration-rotation bands of OH. *Astrophys. J.* 118, 214-227.

Hernandez, G. (1966). Analytical description of a Fabry-Perot photoelectric spectrometer. *Appl. Opt.* 5, 1745-1748.

Hernandez, G. (1970a). Analytical description of a Fabry-Perot photoelectric spectrometer 2: Numerical results. *Appl. Opt.* 9, 1591-1596.

Hernandez, G. (1970b). A high luminosity spectrometer for night airglow studies. *Appl. Opt.* 9, 1225-1227.

Hernandez, G. and Mills, O.A. (1973). Feedback stabilized Fabry-Perot interferometer. *Appl. Opt.* 12, 126-130.

Herzberg, G. (1950). 'Molecular Spectra and Molecular Structure. I. Spectra of Diatomic Molecules.' (van Nostrand, New York).

Herzberg, G. (1951). The atmospheres of planets. *J. Roy. Ast. Soc. Canada* 45, 100-123.

Hesstvedt, E. (1970). A theoretical study of the diurnal variation of hydroxyl emission. *J. Geophys. Res.* 75, 2337-2339.

Hill, E. and van Vleck, J.H. (1928). On the quantum mechanics of the rotational distortion of multiplets in molecular spectra. *Phys. Rev.* 32, 250-272.

Hines, C.O. (1960). Internal atmospheric gravity waves at ionospheric heights. *Can. J. Phys.* **38**, *1441-1481.*

Hirschberg, J.G., Fried, W.I., Hazelton, L. Jr. and Wouters, A. (1971). Multiplex Fabry-Perot interferometer. *Appl. Opt.* **10**, *1979-1980.*

Hirschfeld, T. (1968). Improvements in photomultipliers with total internal reflection sensitivity enhancement. *Appl. Opt.* **7**, *443-449.*

Horlick, G. (1972). Resolution enhancement of line emission spectra by deconvolution. *Applied Spectroscopy* **26**, *395-399.*

Hunt, B.G. (1971). A diffusive-photochemical study of the mesosphere and lower thermosphere and the associated conservation mechanisms. *J. Atmos. Terr. Phys.* **33**, *1869-1892.*

Jackson, D.A. and Pike, E.R. (1968). An automatic scanning Fabry-Perot etalon using multi-channel digital data storage. *J. of Phys. E : Sci. Instrum.* **1**, *394-396.*

Jacquinot, P. (1954). The luminosity of spectrometers with prisms, gratings, or Fabry-Perot etalons. *J. Opt. Soc. Am.* **44**, *761-765.*

Jacquinot, P. (1960). New developments in interference spectroscopy. *Rep. Prog. Phys.* **23**, *267-312.*

Jacquinot, P. and Dufour, Ch. (1948). Optical conditions in the use of photo-electric cells in spectrographs and interferometers. *J. Rech. Cent. Natl. Rech. Sci. Lab. Bellevue (Paris) No. 6, 91-103.*

Jonas, M. and Alon, Y. (1971). Dependence of signal-to-noise ratio on operating voltage in photomultipliers. *Appl. Opt.* **10**, *2436-2438.*

Jones, A. (1970). Spiral - A new algorithm for non-linear parameter estimation using least squares. *The Computer Journal* **13**, *301-308.*

Jones, D.P. (1976). Photomultiplier sensitivity variation with angle of incidence on the photocathode. *Appl. Opt.* **15**, *910-914.*

Kaufman, F. (1964). Aeronomic reactions involving hydrogen. A review of recent laboratory studies. *Ann. Géophys.* **20**, *106-114.*

Kent, G.S., Keenlislide, W., Sandford, M.C.W. and Wright, R.W.H. (1972). Laser radar observations of atmospheric tides in the 70-100 km height region. *J. Atmos. Terr. Phys.* **34**, *373-386.*

Kieffaber, L.M. (1974). OH airglow fluctuations at 1.65 and 2.15 μ. *J. Atmos. Terr. Phys.* 36, 1087-1091.

Kieffaber, L.M. and Peterson, A.W. (1973). Structure and fluctuations in the OH airglow at 1.65 μ. *J. Atmos. Terr. Phys.* 35, 2013-2022.

Krassovsky, V.I. (1956). Infra-red night airglow as a manifestation of the process of oxygen recombination. 'The Airglow and the Aurorae', (ed. Armstrong, E.B. and Dalgarno, A.), 193-196. (Pergamon Press, London and N.Y.)

Krassovsky, V.I. (1972). Infrasonic variations of OH emission in the upper atmosphere. *Ann. Géophys.* 28, 739-746.

Krassovsky, V.I., Shagaev, M.V. (1974a). Inhomogeneities and wavelike variations of the rotational temperature of atmospheric hydroxyl. *Planet. Space Sci.* 22, 1334-1337.

Krassovsky, V.I., Shagaev, M.V. (1974b). Optical method of recording acoustic or gravity waves in the upper atmosphere. *J. Atmos. Terr. Phys.* 36, 373-375.

Krassovsky, V.I. and Shefov, N.N. (1962). Atlas of the airglow spectrum - 3000 to 12000 Å. *Planet. Space Sci.* 9, 883-915.

Krassovsky, V.I., Shefov, N.N. and Yarin, V.I. (1961). On the OH airglow. *J. Atmos. Terr. Phys.* 21, 46-53.

Kvifte, G. (1959). Nightglow observations at Å̊s during the I.G.Y. *Geofysiske Publikasjoner* 20, 1-19.

Leś, F. and Leś, Z. (1968). An electrostrictively scanning Fabry-Perot interferometer. *Acta Physica Polonica* 34, 715-719.

Lindzen, R.S. (1967). Thermally driven diurnal tide in the atmosphere. *Roy. Met. Soc. Quart. J.* 93, 18-42.

Lindzen, R.S. (1968). The application of classical atmospheric tidal theory. *Proc. Roy. Soc.* A303, 299-316.

Lindzen, R.S. and Hong, S. (1974). Effects of mean winds and horizontal temperature gradients on solar and lunar semidiurnal tides in the atmosphere. *J. Atmos. Sci.* 31, 1421-1446.

Lloyd, K.H., Low, C.H., McAvaney, B.J., Rees, D. and Roper, R.G. (1972). Thermospheric observations combining chemical seeding and ground-based techniques - I. Winds, turbulence and the parameters of the neutral atmosphere. *Planet. Space Sci.* 20, 761-789.

McClatchey, R.A., Fenn, R.W., Selby, J.E.A., Volz, F.E. and Garing, J.S. (1972). Optical properties of the atmosphere (third edition). *AFCRL-72-0497, Environmental Research Papers, No. 411, Air Force Cambridge Research Laboratories, Massachusetts.*

Mack, J.E., McNutt, D.P., Roesler, F.L. and Chabbal, R. (1963). The PEPSIOS purely interferometric high-resolution scanning spectrometer. 1. The pilot model. *Appl. Opt. 2, 873-885.*

McKeith, N.E., Smith, R.W. and Whiteford, J.R. (1976). An interferometric technique for the electromechanical calibration of piezoelectric transducers. *J. of Phys. E : Sci. Instrum. 9, 15-18.*

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math. 11, 431-441.*

Martin, D.W. (1966). Linear algebraic equations. 'Numerical Analysis: An Introduction', (ed. Walsh, J.), 11-26. (Academic Press, Lond.)

Meinel, A.B. (1950). OH emission bands in the spectrum of the night sky. II. *Astrophys. J. 112, 120-130.*

Meriwether, J.W. (1975). High latitude airglow observations of correlated short-term fluctuations in the hydroxyl Meinel (8-3) band intensity and rotational temperature. *Planet. Space Sci. 23, 1211-1221.*

Mies, F.H. (1974). Calculated vibrational transition probabilities of OH ($X^2\Pi$). *J. Mol. Spect. 53, 150-188.*

Moore, C., Minnaert, M. and Houtgast, J. (1966). The Solar Spectrum 2935 Å to 8770 Å. *National Bureau of Standards Monograph 61.*

Moreels, G., Chahrokhi, D. and Blamont, J.E. (1976). OH emission intensity measurements during the 1969 NASA airborne auroral expedition. *J. Geophys. Res. 81, 5467-5478.*

Moreels, G., Gattinger, R.L. and Vallance Jones, A. (1974). Comments on the paper: Diurnal, annual and solar cycle variations of hydroxyl and sodium nightglow intensities in the Europe-Africa sector. *Planet. Space Sci. 22, 344-346.*

Moreels, G. and Herse, M. (1977). Photographic evidence of waves around the 85 km level. *Planet. Space Sci. 25, 265-273.*

Murphy, R.E. (1971). Infrared emission of OH in the fundamental and first overtone bands. *J. Chem. Phys.* 54, *4852-4859.*

Murray, R.B. and Manning, J.J. (1960). Response of end-window photomultiplier tubes as a function of temperature. *I.E.E.E. Trans. Nucl. Sci. NS-7, 80-86.*

Neiswander, R.S. and Plews, G.S. (1975). Low-noise extended-frequency response with cooled silicon photo-diodes. *Appl. Opt.* 14, *2720-2726.*

Nicholls, D.C. (1971). Hydroxyl rotational temperatures in the airglow. M.Sc. Thesis, University of Saskatchewan.

Nicholls, D.C., Evans, W.F.J. and Llewellyn, E.J. (1972). Collisional relaxation and rotational intensity distributions in spectra of aeronomic interest. *J. Quant. Spectrosc. Radiat. Transf.* 12, *549-558.*

Nicolet, M. (1970). Ozone and hydrogen reactions. *Ann. Géophys.* 26, *531-546.*

Nilson, J.A. and Shepherd, G.G. (1961). Upper atmospheric temperatures from Doppler line widths - 1. Some preliminary measurements on OI 5577 Å in aurora. *Planet. Space Sci.* 5, *299-306.*

Noxon, J.F. (1964). The latitude dependence of OH rotational temperature in the night airglow. *J. Geophys. Res.* 69, *4087-4092.*

Otnes, R.K. and Enochson, L. (1972). 'Digital Time Series Analysis', (J. Wiley and Sons, New York).

Packer, D.M. (1961). Altitudes of the night airglow radiations. *Ann. Géophys.* 17, *67-75.*

Phillips, L.F. and Schiff, H.I. (1962). Mass spectrometric studies of atomic reactions. III. Reactions of hydrogen atoms with nitrogen dioxide and with ozone. *J. Chem. Phys.* 37, *1233-1238.*

Potter, A.E. Jr., Coltharp, R.N. and Worley, S.D. (1971). Mean radiative lifetime of vibrationally excited (v = 9) hydroxyl. Rate of the reaction of vibrationally excited hydroxyl (v = 9) with ozone. *J. Chem. Phys.* 54, *992-996.*

Powell, M.J.D. (1965). A method for minimizing a sum of squares of non-linear functions without calculating derivatives. *The Computer Journal* 7, *303-307.*

Ramsay, J.V. (1966). Automatic control of the spacing of Fabry-Perot interferometers. *Appl. Opt. 5, 1297-1301.*

Ramsay, J.V. (1969). Abberations of Fabry-Perot interferometers when used as filters. *Appl. Opt. 8, 569-573.*

Ramsay, J.V. and Mugridge, E.G.V. (1962). Barium titanate ceramics for fine-movement control. *J. Sci. Instrum. 39, 636-637.*

Reay, N.K., Ring, J. and Scadden, R.J. (1974). A tunable Fabry-Perot filter for the visible. *J. of Phys. E : Sci. Instrum. 7, 673-677.*

Roig, J. (1958). Balayage thermique des anneaux de Perot et Fabry. *J. Phys. Rad. 19, 284-289.*

Sacconi, A. (1970). Linearità e ripetibilità della caratteristica tensione-spostamento di trasduttori piezoelettrici. *Alta Frequenza 9, 797-800.*

Schaeffer, R.C. (1970). Mid-latitude airglow and auroral phenomena. Ph.D. Thesis, University of Adelaide.

Selby, J.E.A. and McClatchey, R.A. (1975). Atmospheric transmittance from 0.25 to 28.5 μ: Computer code LOWTRAN 3. *AFCRL-75-0255, Environmental Research Papers, No. 513, Air Force Cambridge Research Laboratories, Massachusetts.*

Shagaev, M.V. (1974). Fast variations of hydroxyl night airglow emission. *J. Atmos. Terr. Phys. 36, 367-371.*

Shefov, N.N. (1961). On determination of the rotational temperature of OH bands. *Spektral'nyye, Elektrofotometricheskiye i Radiolakatsionnyye Issledovaniya Polyarnykh Siyaniy i Svecheniya Nochnogo Neba No. 5, 5-9.* English translation in NASA-TT-F-106, pp. 1-7, (1962).

Shefov, N.N. (1969a). Hydroxyl emission in the upper atmosphere - I. The behaviour during a solar cycle, seasons and geomagnetic disturbances. *Planet. Space Sci. 17, 797-813.*

Shefov, N.N. (1969b). Hydroxyl emission in the upper atmosphere - II. Effects of a sunlit atmosphere. *Planet. Space Sci. 17, 1629-1639.*

Shefov, N.N. (1971). Hydroxyl emission of the upper atmosphere - III. Diurnal variations. *Planet. Space Sci. 19, 129-136.*

Shefov, N.N. (1972). Hydroxyl emission. *Ann. Géophys.* 28, *137-143.*

Shemansky, D.E. and Vallance Jones, A. (1961). New measurements of the night airglow spectrum in the 1.5 μ region. *J. Atmos. Terr. Phys.* 22, *166-175.*

Shepherd, G.G. (1960). A Fabry-Perot spectrometer for auroral and airglow observations. *Can. J. Phys.* 38, *1560-1569.*

Shepherd, G.G. (1967). Applications of the Fabry-Perot spectrometer to upper atmospheric spectroscopy. *J. de Phys.* 28, *301-307.*

Shepherd, G.G., Lake, C.W., Miller, J.R. and Cogger, L.L. (1965). A spatial spectral scanning technique for the Fabry-Perot spectrometer. *Appl. Opt.* 4, *267-272.*

Shimazaki, T. and Laird, A.R. (1970). A model calculation of the diurnal variation in minor neutral constituents in the mesosphere and lower thermosphere including transport effects. *J. Geophys. Res.* 75, *3221-3235.*

Sivjee, G.G., Dick, K.A. and Feldman, P.D. (1972). Temporal variations in night-time hydroxyl rotational temperature. *Planet. Space Sci.* 20, *261-269.*

Slater, P.N., Betz, H.T. and Henderson, G. (1965). A new design of a scanning Fabry-Perot interferometer. *Jap. J. Appl. Phys.* 4, *Supplement 1, 440-444.*

Strobel, D.F. (1972). Minor neutral constituents in the mesosphere and lower thermosphere. *Radio Science* 7, *1-21.*

Tabata, T. and Ito, R. (1975). Effective treatment of the interpolation factor in Marquardt's nonlinear least-squares fit algorithm. *The Computer Journal* 18, *250-251.*

Takahashi, H., Clemesha, B.R. and Sahai, Y. (1974). Nightglow OH (8-3) band intensities and rotational temperature at 23°S. *Planet. Space Sci.* 22, *1323-1329.*

Taranova, O.G. (1967). An investigation of space-time characteristics of hydroxyl emission. (English translation of title.) *Spektral'nyye, Elektrofotometricheskiye i Radiolakatsionnyye Issledovaniya Polyarnykh Siyaniy i Svecheniya Nochnogo Neba No. 13, 13-21. (Aurorae and Airglow* 13, *13-21.)*

Tarasova, T.M. (1963). Night-sky emission-line intensity distribution with respect to height. *Space Res.* 3, 162-172.

Taylor, J.H. and Yates, H.W. (1957). Atmospheric transmission in the infrared. *J. Opt. Soc. Am.* 47, 223-226.

Vallance Jones, A. (1973). The infrared spectrum of the airglow. *Space Sci. Rev.* 15, 355-400.

Velculescu, V.G. (1970). On the production of excited hydroxyl radicals in the H + O_3 - atomic flame. *Z. Physik* 237, 69-74.

Visconti, G., Congeduti, F. and Fiocco, G. (1971). Fluctuations in the intensity and excitation temperature in the OH airglow (8-3) band. 'The Radiating Atmosphere', (ed. B.M. McCormac), 82-89. (Reidel Publishing Co., Dordrecht-Holland).

Wallace, L. (1960). The constants of the $^2\Pi$ - $^2\Pi$ OH bands. *Astrophys. J.* 132, 894-897.

Wallace, L. (1961). Seasonal variation and interpretation of the OH rotational temperature of the airglow. *J. Atmos. Terr. Phys.* 20, 85-93.

Wallace, L. (1962). The OH nightglow emission. *J. Atmos. Sci.* 19, 1-16.

Wiens, R.H. and Weill, G. (1973). Diurnal, annual and solar cycle variations of hydroxyl and sodium airglow intensities in the Europe-Africa sector. *Planet. Space Sci.* 21, 1011-1027.

Wilkinson, J.H. (1967). The solution of ill-conditioned linear equations. 'Mathematical Methods for Digital Computers II', (ed. Ralston, A. and Wilf, H.S.), 65-94. (John Wiley and Sons, New York).

Wilksch, P.A. (1975). Measurement of thermospheric temperatures and winds using a Fabry-Perot spectrometer. Ph.D. Thesis, University of Adelaide.

Wolberg, J.R. (1967). 'Prediction Analysis', (Van Nostrand; Princeton, New Jersey).

Worley, S.D., Coltharp, R.N. and Potter, A.E. (1972). Rates of interaction of vibrationally excited hydroxyl (v = 9) with diatomic and small polyatomic molecules. *J. Phys. Chem.* 76, 1511-1514.

Yates, H.W. and Taylor, J.H. (1960). Infrared transmission of the atmosphere. *N.R. Report 5453, U.S. Naval Research Lab., Washington, D.C.*

Young, A.T. (1969). Photometric error analysis. IX: Optimum use of photomultipliers. *Appl. Opt. 8, 2431-2447.*