General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

UNIVERSITY OF ILLINOIS
URBANA

(NASA-CR-157770) THEORETICAL AND EXPERIMENTAL STUDIES OF THE ATMOSPHERIC SODIUM LAYER (Illinois Univ.) 263 p HC A12/MF A01
CSCL 04A
N78-33653
Unclas
G3/46 35790

AERONOMY REPORT NO. 79

THEORETICAL AND EXPERIMENTAL STUDIES OF THE ATMOSPHERIC SODIUM LAYER

by
E. S. Richter
C. F. Sechrist, Jr.

May 1, 1978

Library of Congress ISSN 0568-0581

NOV 1978
RECEIVED
NASA STI FACILITY

National Aeronautics and Space Administration
Grant NGR 14-005-181
National Science Foundation
Grant ATM 77-00933

Aeronomy Laboratory
Department of Electrical Engineering
University of Illinois
Urbana, Illinois

UILU-ENG-78-2501

AERONOMY REPORT

NO. 79

THEORETICAL AND EXPERIMENTAL STUDIES OF THE ATMOSPHERIC SODIUM LAYER

by

E. S. Richter
C. F. Sechrist, Jr.

May 1, 1978

Supported by
National Aeronautics and
Space Administration
Grant NGR 14-005-181
National Science Foundation
Grant ATM-77-00933

Aeronomy Laboratory
Department of Electrical Engineering
University of Illinois
Urbana, Illinois

PRECEDING PAGE BLANK NOT FILMED

ABSTRACT

Atmospheric atomic sodium has been observed over Urbana using a laser radar system tuned to 589 nm. The layer typically has a peak concentration of 10^9-10^{10} m^{-3} at 90 km and is about 10 km thick. The photocount data are processed using a digital filter to obtain continuous estimates of the sodium concentration versus altitude. Wave-like structures in the sodium layer have been observed with typical wavelengths of 3-15 km and downward phase velocities of 1-3 m s^{-1}. On one occasion there was evidence for the presence of a standing wave in the layer. During two successive evenings the bottomside of the layer was observed to undulate with a period of about 2½ hours. In agreement with worldwide measurements the layer is observed to broaden through the night.

A meteor ablation-cluster ion theory of sodium is developed and found to be in good agreement with existing atmospheric observations as well as laboratory measurements of rate constants. The Na^+ chemistry is connected to the Na chemistry by cluster ion reactions involving $Na^+ \cdot N_2$, $Na^+ \cdot CO_2$, and $Na^+ \cdot H_2O$. The chemistry includes switching and dissociative recombination of cluster ions but does not include their photodestruction. The chemistry is then used in conjunction with a meteor ablation source and a downward sweeping of ionization caused by an interaction of the neutral wind profile with the geomagnetic lines of force (corkscrew mechanism) to explain many features of the sodium layer. Sodium is removed from the mesosphere in the form of higher order cluster ions which eventually undergo sedimentation to the lower atmosphere.

PRECEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure

Figure

Figure

Figure

Figure

Figure

Figure

1. INTRODUCTION

1.1 *Past Observations and Results*

Since the discovery of atmospheric sodium emissions in the late thirties, several investigators have measured the features of the sodium layer and theorized about its origin and variation. A review of this activity is presented in the publications by *Brown* [1973] and *Kvifte* [1973]. In spite of the varied approaches in both ground-based and rocket-borne experiments to learn more about the sodium layer, many of the reasons for its origin and behavior remain elusive because important correlative measurements have not been made. Yet knowledge of the physics and chemistry of atmospheric sodium continues to be important for the understanding of the aeronomy of the middle atmosphere. In particular the chemistry of sodium is intimately connected to such important constituents as atomic oxygen and ozone as well as electron density. Furthermore, it is felt that sodium may be used as a tracer for atmospheric motions in the 80-100 km region.

Prior to 1970, sodium measurements were largely restricted to measuring the total flux of resonantly scattered sunlight, giving only the total column density of sodium. Rocket measurements [*Hunten and Wallace*, 1967; *Donahue and Meier*, 1967] were the first to reveal the existence of sharp upper and lower boundaries for the layer. Since the initial lidar measurements of sodium [*Bowman et al.*, 1969], nighttime monitoring has established several important features of the layer. These measurements combine the altitude resolution of rocket experiments with the continuous coverage ability of previous ground-based measurements.

1.2 *Lidar Observations and Results*

Remote sensing of the sodium layer using a dye laser tuned to one of the sodium resonance lines, usually 589.0 nm, has produced detailed nighttime profiles. In general it has been found that atomic sodium consistently resides in the upper part of what is referred to as the middle atmosphere in the altitude range of 80-100 km. Its distribution is rather sharply peaked at about 90 km, with a maximum number density of 10^9 - 10^{10} m^{-3}. The scale height at the top of the layer is about 3-4 km, much smaller than that of the surrounding neutral atmosphere; and the lower portion of the layer falls off quite rapidly with an order of magnitude drop in about 4 km.

However, it is the desire to understand the variations of the layer which has stirred so much interest in the sodium problem. Specifically there is a winter enhancement at mid-latitudes of about a factor of two and a lowering of the altitude of the layer bottomside. This well-defined seasonal variability is illustrated by Figure 1.1 [*Gibson and Sandford*, 1971] and also appears in the sodium column abundance shown in Figure 1.2 [*Megie*, 1975]. Although ground-based [*Gadsden and Purdy*, 1970] and rocket-borne [*Donahue and Meier*, 1967] dayglow measurements have indicated a diurnal variation, laser radar measurements have indicated no such variation [*Gibson and Sandford*, 1972]. The nighttime sodium concentration shows little systematic variation during the night [*Gibson and Sandford*, 1971] although *Hake et al.* [1972] have noticed an increase in the sodium layer during a period of the Geminids meteor shower activity. Detailed structure was observed in the sodium layer by *Kirchhoff and Clemesha* [1973] and they attributed this structure to atmospheric tidal oscillations. *Megie* [1975] reported on 80 nights of observations from October 1973 to July 1975 and speculated that dynamics of the mesopause region were responsible for the time-dependent stratification of the layer. Using a steerable sodium lidar, *Thomas et al.* [1976, 1977] claim to have observed lateral movement of sections of the sodium layer.

ORIGINAL PAGE IS
OF POOR QUALITY

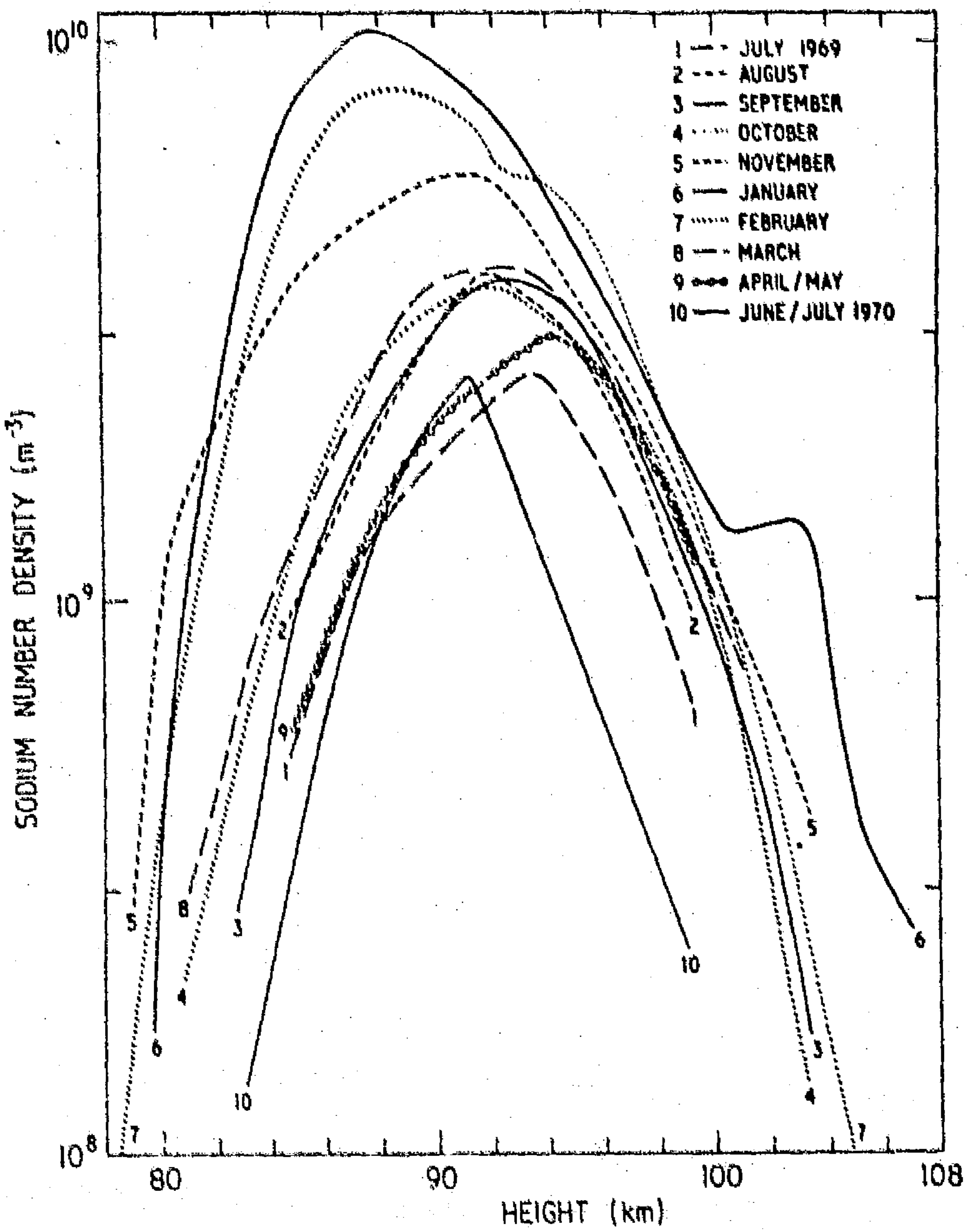


Figure 1.1 Seasonal variation of sodium concentration profiles [*Gibson and Sandford*, 1971].

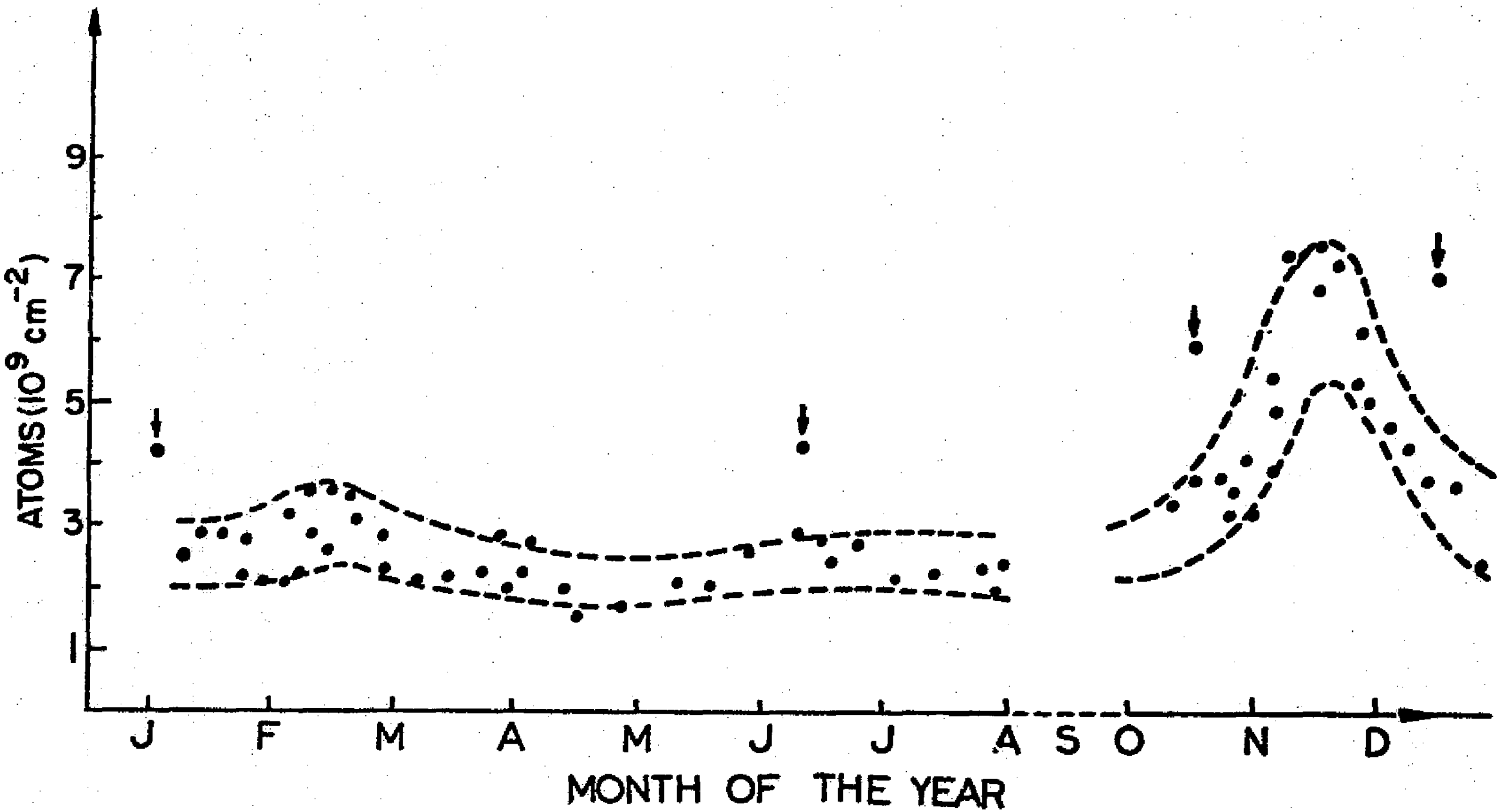


Figure 1.2 Seasonal variation of total sodium abundance [*Megie*, 1975].

Hence, a wealth of information about the sodium layer has been provided by ground-based lidars, but as yet there is no theoretical study which can simultaneously explain all of the observed features.

1.3 *Theories of the Sodium Layer and Problems*

Detailed theories of the sodium layer have been presented by *Hunten and Wallace* [1967], *Donahue and Meier* [1967], *Hanson and Donaldson* [1967], *Gadsden* [1968], *Hunten* [1971], *Mitra* [1973], and *Fiocco and Visconti* [1973, 1974]. A problem common to all of these theories is that none of them can explain all of the observed features of the sodium layer. It is the attempt to explain these observations which has prompted the theoretical investigation of this work, and this theory is briefly summarized below.

Gadsden [1970] made the interesting suggestion that sodium atoms ablated from meteors are ionized in the *E* region and the long-lived Na^+ ions move down to the base of the *E* region by the corkscrew mechanism [*Chimonas and Axford*, 1968]. Following their transport to the 80-100 km region, the Na^+ ions are converted to neutral Na through cluster ion reactions. In this way the dumping of ions by the corkscrew mechanism maintains the steep topside gradient of the sodium layer in the presence of vertical eddy and molecular diffusion. Within the sodium layer in the 80-100 km region, chemical quasi-equilibrium exists with the predominant forms of sodium being Na, NaO, Na^+, and $Na^+ \cdot H_2O$. It is questionable whether other forms of sodium compounds such as NaO_2, and NaO^+ play an important part in the chemistry. Nevertheless it may be shown that sodium concentration is proportional to both electron density and atomic oxygen. Hence the sharp bottomside falloff of sodium is probably a result of the steep electron and oxygen density gradients below 90 km. Below 85 km sodium is removed from the layer, possibly in the form of higher order water clusters which undergo sedimentation to the lower mesosphere. The details of this meteor-ion theory are presented in Chapter 7.

1.4 *Objectives and Scope of This Study*

The scope of this work has included an experimental as well as theoretical investigation of the sodium layer. As part of this work a sodium lidar system was designed, constructed, and operated at the Aeronomy Laboratory Field Station. Although the published system characteristics of many groups are rather sparse, some comparison between lasers is possible and these characteristics are tabulated in Table 1.1. A similar comparison between receivers is impossible because of lack of data.

As indicated in Table 1.1, the Urbana lidar is quite adequate for the first scientific measurements. In addition, it should be noted that the higher energy lasers employ an oscillator-amplifier configuration and an amplifier would be an excellent addition for future correlative measurements. The theoretical work was developed simultaneously and conclusions were drawn about the experimental observations.

A brief review of terminology and principles involved with the study of sodium vapor is presented in Chapter 2. Chapter 3 provides a detailed discussion of the upper atmospheric lidar application which is appropriate to this study.

Important concepts about dye lasers and a description of the Urbana laser follows in Chapter 4. The Urbana receiving system and routine operating techniques are discussed in Chapter 5.

The experimental results are presented in Chapter 6. The feasibility study for the first correlative measurements with the Urbana meteor radar facility is also included in this chapter.

Chapter 7 develops the meteor-ion theory. This study involves a mechanistic approach to the problem. That is, individual dynamical and chemical

TABLE 1.1

Summary of dye lasers used for sodium lidar measurements

Institution	Wavelength	Energy/Shot	Repetition Rate	Linewidth
Radio and Space Research Station, ENGLAND [*Bowman et al.*, 1969]	589.6 nm	4 mJ	0.2 s^{-1}	0.15 nm
GCA Corporation Technology Division, Bedford, Massachusetts [*Schuler et al.*, 1971]	----- -----	600 mJ 200 mJ	--- ---	0.1-0.2 nm 12 pm
SRC Radio and Space Research Station, ENGLAND [*Gibson and Sandford*, 1972]	589.0	200 mJ	2 s^{-1}	3 pm
Service d'Aeronomie du CNRS, FRANCE [*Blamont et al.*, 1972]	589.0 nm 589.0 nm	125 mJ 300 mJ*	0.017-0.05 s^{-1} 0.017-0.05 s^{-1}	15 pm 13 pm
European Space Research Institute, ITALY [*Magyar and Schneider-Muntau*, 1972]	589.0 nm	600 mJ*	---	10 pm
Stanford Research Institute, Menlo Park, California [*Hake et al.*, 1972]	589.0 nm	500 mJ*	.25 s^{-1}	<5 pm
Instituto de Pesquisas Espaciais, BRAZIL [*Kirchhoff and Clemesha*, 1973]	589.0 nm	5 mJ	0.055 s^{-1}	5 pm
Tohoku University, JAPAN [*Aruga et al.*, 1974]	-----	2 mJ	1 s^{-1}	2 pm
Service d'Aeronomie du CNRS, FRANCE [*Loth and Megie*, 1974]	589.0 nm 589.0 nm	400 mJ 1.2 J*	1 s^{-1} 1 s^{-1}	10 pm 10 pm
University of Illinois, Urbana, Illinois	589.0 nm	100 mJ	.5 s^{-1}	10 pm

*Denotes an oscillator-amplifier configuration.

ORIGINAL PAGE IS
OF POOR QUALITY

mechanisms are investigated by numerical methods and the results are interpreted in the light of experimental findings.

Chapter 8 summarizes the experimental and theoretical implications of this study. Future areas for research are defined and further correlative measurements are suggested.

2. ATOMIC SODIUM VAPOR CHARACTERISTICS AND SPECTROSCOPIC NOTATION

2.1 *Basic Transitions*

Sodium is the eleventh element of the periodic table and exists at mesospheric heights in the atomic gaseous phase. Being an alkali atom, its structure consists of one valence electron and two inner shells of non-radiating electrons. A brief review of spectroscopic notation applicable to the sodium atom is helpful in understanding some basic concepts important to this lidar experiment.

It is necessary to designate a number of spatial quantum numbers to describe the complicated spectrum of sodium. Among these we include the principal or total quantum number, n, and the angular momentum quantum number, l. The principal quantum number takes on integral values $(1, 2, \ldots, n, \ldots)$ and denotes the orbit or shell in which the electron is moving. For a given value of n, l may take on values $0, 1, 2, \ldots n-1$, and characterizes the orbital angular momentum of the electron. The states associated with the l values are sometimes referred to as sub-shells. In spectroscopic notation it is common to designate these sub-shells by the letters, s, p, d, f, (alternatively S, P, D, F) corresponding to the values $l = 0, 1, 2, 3$, respectively. The notation grew out of optical spectroscopy where sets of spectral lines were characterized as sharp, principal, diffuse, and fine. Using this notation the electronic ground state configuration for sodium may be written as $1s^2 2s^2 2p^6 3s^1$. The number preceding a letter refers to the total quantum number of state, n, while the superscripts denote the number of electrons in each subshell.

Sodium, like all alkali atoms, exhibits a series of doublets rather than single lines in its spectrum. To describe this phenomenon, we must introduce a new quantum number, j. This total angular momentum quantum number for extranuclear electrons is obtained by the expression:

$$j = \begin{cases} l + s & l \neq 0 \\ 1/2 & l = 0 \end{cases} \tag{2.1}$$

where s is the spin angular momentum number and takes on values $s \pm 1/2$. Note that the $l = 0$ state is not split. Hence, all states except the S state are actually double.

We may then label our states as shown in Table 2.1 where j appears as a subscript in the entries and the superscript 2 denotes a doublet state. The selection rules for allowed transitions become $\Delta j = \pm 1, 0$ with the transition $j = 0 \rightarrow j = 0$ not allowed and $\Delta j = \pm 1, 0$ with $l = 0 \rightarrow l = 0$ forbidden. Therefore, the two D-line transitions of interest to our problem are:

$$3^2S_{1/2} \rightarrow 3^2P_{1/2} \qquad \lambda = 589.6 \text{ nm } (D_1 \text{ line}) \tag{2.2}$$

$$3^2S_{1/2} \rightarrow 3^2P_{3/2} \qquad \lambda = 589.0 \text{ nm } (D_2 \text{ line}) \tag{2.3}$$

These lines appear in the energy level diagram shown in Figure 2.1.

2.2 *Spectral Properties of High Resolution Lines*

2.2.1 *Hyperfine structure.* If one observes the line spectrum with an apparatus of high resolving power, one will find a fine grained or hyperfine structure to the lines. The hyperfine structure is qualitatively explained by assuming the nucleus possesses an intrinsic angular momentum and a corresponding magnetic moment. We assign another quantum number, i, to denote this momentum. The quantum number i differs from one type of atom to the next. For sodium $i = 3/2$. A new quantity, f, called the hyperfine quantum number is defined to represent the total angular momentum and may take values $f = j + i$, $j + i - 1$, $j + i - 2, \ldots, |j - i|$. If $j \geq i$, there are $2i + 1$ values for f (D_2 line). If $i \geq j$, there are $2j + 1$ values for f (D_1 line). The selection rules that apply to f are $\Delta f = \pm 1, 0$ but $f = 0 \rightarrow f = 0$ is forbidden.

Table 2.1 Labeling of Doublet States

j l	1/2	3/2	5/2	7/2
0	$^2S_{1/2}$			
1	$^2P_{1/2}$	$^2P_{3/2}$		
2		$^2D_{3/2}$	$^2D_{5/2}$	
3			$^2F_{5/2}$	$^2F_{7/2}$

ORIGINAL PAGE IS
OF POOR QUALITY

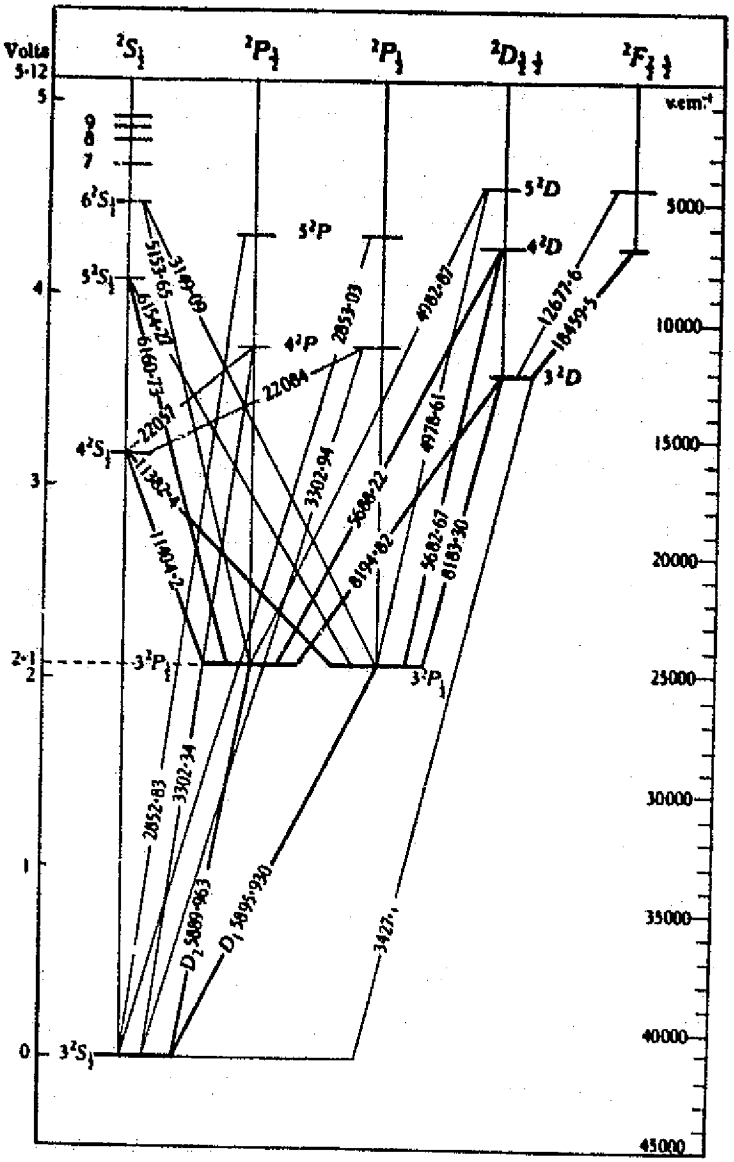


Figure 2.1 Energy levels of sodium [*Mitchell and Zemansky*, 1971].

We observe splitting at the $^2S_{1/2}$, $^2P_{1/2}$, and $^2P_{3/2}$ levels and because the same interval rule holds as did for Δj, the hypermultiplet will have a similar appearance to an ordinary multiplet. Hence, each D-line has two groups of fine structure lines associated with it and we designate these groups as D_{1a}, D_{1b}, D_{2a}, and D_{2b}. The splitting applicable to the D_2 line is shown in Figure 2.2.

2.2.2 *Zeeman effect.* Because the atom has a magnetic dipole moment, the splitting of the hyperfine lines can be observed when the atom is placed in an external magnetic field (*Zeeman effect*). We may estimate the effect of the earth's magnetic field on this splitting by the expression [*White*, 1934]:

$$\Delta\sigma = m_j g \left(\frac{eB}{4\pi\ mc}\right) \qquad \text{(SI units)} \qquad (2.4)$$

where m_j is the magnetic quantum number and denotes the sublevel of hyperfine structure and g is the Landé g-factor. The quantum number, m_j, takes values differing by one from $-j$ to $+j$. Landé g-values depend on the quantum numbers j and l and if known, are usually included in tables of atomic energy states [e.g., *Kuhn*, 1969, p. 291]. The values for m_j and the product $m_j \cdot g$ for our case are listed in Figure 2.2. The quantity enclosed by parentheses is called the Larmor frequency. Quantities indicated have their usual meaning, namely:

e = electronic charge
B = magnetic field
m = mass of electron
c = speed of light

If we assume the earth has a magnetic field of 6 × 10^{-5} Tesla (.6 Gauss) at our latitude, then we calculate:

$$\Delta\sigma = 5 \times 10^{-3}\ \mathrm{m}^{-1} \qquad (2.5)$$

or

$$\Delta\nu = c\Delta\sigma = 1.5 \times 10^{6}\ \mathrm{s}^{-1} \qquad (2.6)$$

ORIGINAL PAGE IS OF POOR QUALITY

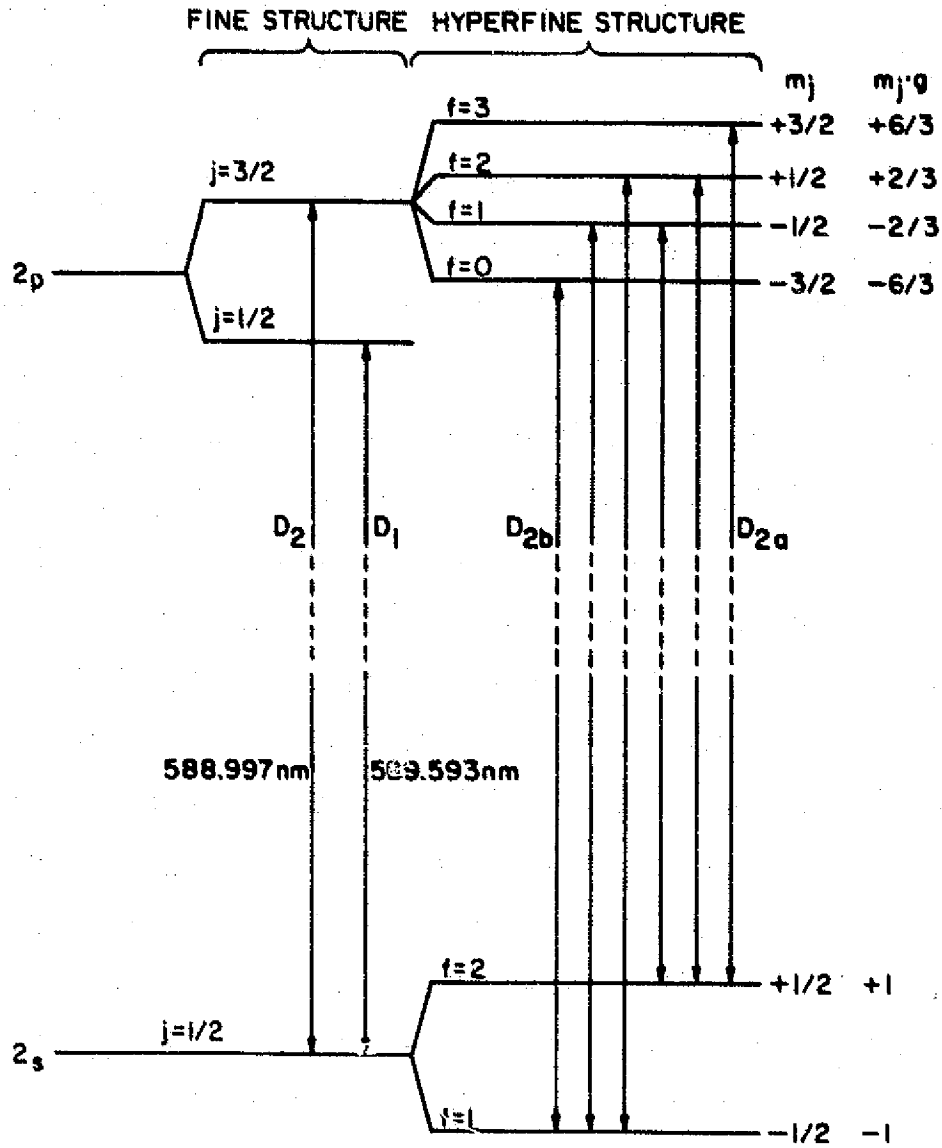


Figure 2.2 Zeeman effect for the sodium D_2 line. Values for m_f and $m_f \cdot g$ are taken from Herzberg [1945].

for $m_j \cdot g = 2$. This is more than 2 orders of magnitude smaller than the separation of the hyperfine lines. Hence, it is appropriate to neglect the splitting of the line due to the earth's magnetic field.

2.3 *Line Intensities*

Another principle, important to the sodium spectrum, relates the relative intensities of the lines within a given multiplet. The statement, known as the sum rule of Ornstein, Burgers, and Dorgelo, is as follows: *The sum of the intensities of all the lines of a multiplet which belong to the same initial or final state is proportional to the statistical weight* $2j + 1$ *of the initial or final state, respectively.* Hence, if we let I represent the intensity of the line, then:

$$\frac{I_{D_2}}{I_{D_1}} = \frac{2\ (3/2) + 1}{2\ (1/2) + 1} = \frac{2}{1} \tag{2.7}$$

Hence, it is preferable to tune the system to the D_2 line for a twofold increase in returns.

Furthermore, this rule applies to the hyperfine structure if we replace j by f. For example, the ratio of the sum of intensities with ${}^2P_{3/2}f = 3$ are to those with ${}^2P_{3/2}f = 2$ as $(2 \times 3 + 1)/(2 \times 2 + 1) = 7/5$. Referring to Figure 2.3 we note that the only line from the $f = 3$ state $({}^2P_{3/2}f = 3 \rightarrow {}^2S_{1/2}f = 2)$ has a relative strength of 14. There are two lines coming from the $f = 2$ state $({}^2P_{3/2}f = 2 \rightarrow {}^2S_{1/2}f = 1$ and ${}^2P_{3/2}f = 2 \rightarrow {}^2S_{1/2}f = 2)$ both having a relative strength of 5. The ratio of the intensities is therefore $14/(5 + 5) = 7/5$ in agreement with the sum rule.

All of the relative intensities may be obtained by this rule. The line spectrum as shown in Figure 2.3 would be that of sodium at 0°K. However, Doppler broadening is dominant for higher temperatures, and this tends to merge

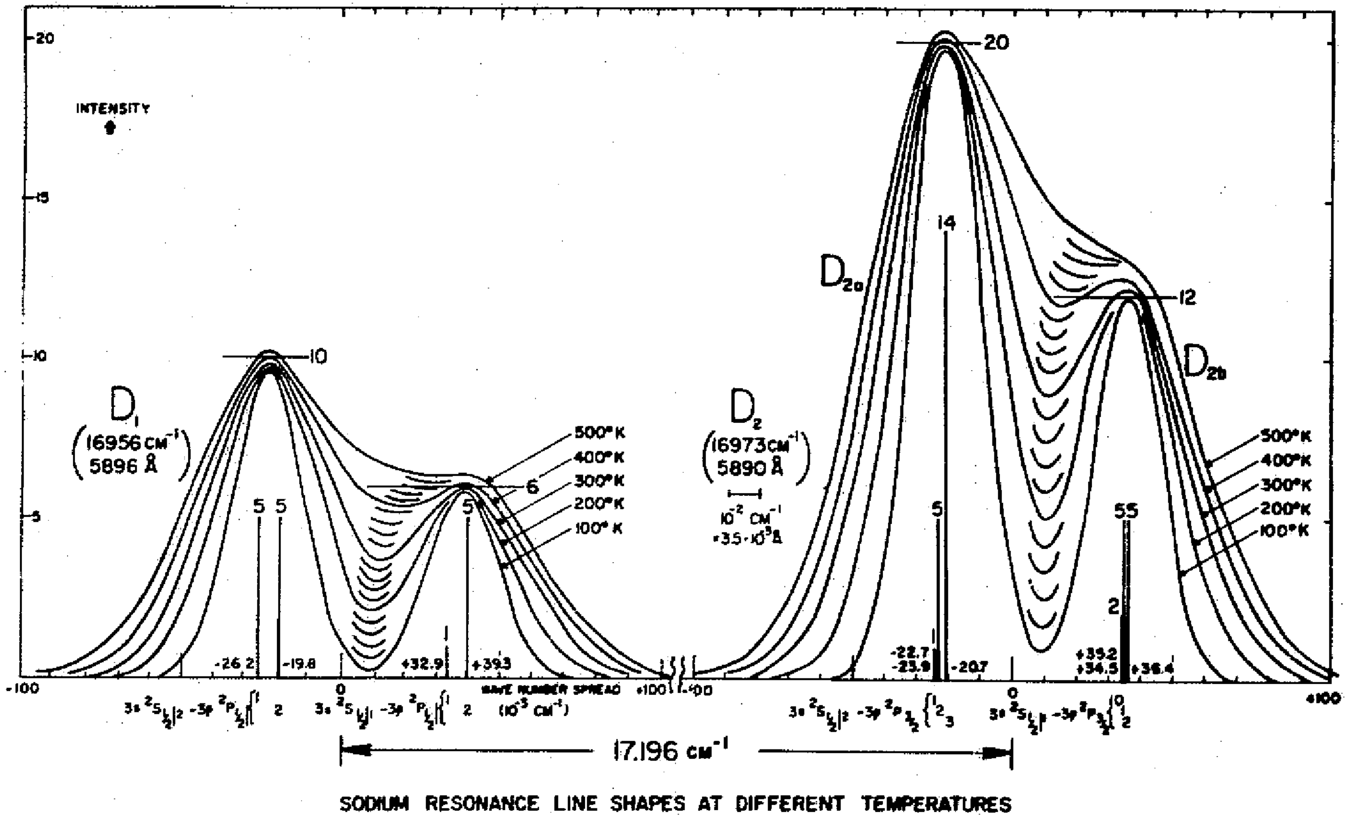


Figure 2.3 Line profiles for D_1 and D_2 emission at various temperatures [*Chamberlain,* 1961].

ORIGINAL PAGE IS OF POOR QUALITY

the lines into a continuous absorption profile as also shown in Figure 2.3. The profile for 200°K is suitable for mesospheric heights corresponding to the sodium layer.

2.4 *Absorption Coefficient*

2.4.1 *Natural and Doppler broadening.* There are several processes that contribute to the formation of an absorption line of a gas. For mesospheric heights we need only consider the processes of natural broadening and Doppler broadening.

Natural broadening results from the finite lifetime of the excited state. For sodium the lifetime of the $^2P_{3/2}$ state is $\tau = 16$ ns [*Mitchell and Zemansky*, 1971]. Hence, we calculate:

$$\Delta\nu_n = \frac{1}{2\pi\tau} = 10^7 \text{ Hz} \qquad (2.8)$$

We will see that this is negligible compared to the effect of Doppler broadening.

Neglecting natural damping, the expression for Doppler broadening of a line is given by:

$$\Delta\nu_D = \frac{2\sqrt{2R \ \ln 2}}{c} \quad \nu_0 \sqrt{\frac{T}{M}} \qquad (2.9)$$

where

R = universal gas constant

c = speed of light

ν_0 = center line frequency ($\nu_0 = c/\lambda_0$)

T = absolute temperature of the gas

M = molecular weight

Taking $\lambda_0 = 589$ nm and $T = 200$°K, we calculate for sodium:

$$\Delta\nu_D = 1.07 \times 10^9 \text{ Hz} \qquad (2.10)$$

Clearly we have $\Delta\nu_D >> \Delta\nu_N$.

2.4.2 *Maximum absorption coefficient,* k_0. We will follow closely the excellent work by *Mitchell and Zemansky* [1971] in developing the expression for the absorption coefficient. If we consider only Doppler broadening, the absorption coefficient of a gas is given by:

$$k_\nu = k_0 \exp - \left[\frac{2\sqrt{\ln 2} \quad (\nu - \nu_0)}{\Delta\nu_D} \right]^2 \tag{2.11}$$

where k_0 represents the maximum absorption coefficient. Integrating equation (2.11) we obtain:

$$\int_0^\infty k_\nu d\nu = \frac{1}{2}\sqrt{\frac{\pi}{\ln 2}} \; k_0 \Delta\nu_D \tag{2.12}$$

Two alternate forms for this expression exist due to Ladenburg:

$$\int_0^\infty k_\nu d\nu = \begin{cases} \dfrac{\lambda_0^2}{8\pi} \; \dfrac{g_2}{g_1} \; \dfrac{\rho}{\tau} & \text{(2.13a)} \\[2ex] \dfrac{e^2}{4\varepsilon_0 mc} \; f \; \rho \qquad \text{(SI units)} & \text{(2.13b)} \end{cases}$$

where

g_1, g_2 = statistical weights of the normal and excited states respectively

ρ = number density of sodium

τ = lifetime of the atom in the excited state

f = oscillator strength or spectral line f-value

e = electronic charge

ε_0 = permittivity of free space

m = mass of an electron

c = speed of light

If we define the scattering cross section, σ_0, at the peak of the line by the expression $k_0 = \sigma_0 \rho$, then by combining equations (2.12) and (2.13) we may solve for σ_0 and obtain:

$$\sigma_0 = \begin{cases} \dfrac{2}{\Delta\nu_D} \quad \dfrac{\ln 2}{\pi} \quad \dfrac{\lambda_0^2}{8\pi} \quad \dfrac{g_2}{g_1} \quad \dfrac{1}{\tau} & \qquad (2.14a) \\[2ex] \dfrac{2}{\Delta\nu_D} \quad \dfrac{\ln 2}{\pi} \quad \dfrac{e^2}{4\epsilon_0 mc} \; f \quad \text{(SI units)} & \qquad (2.14b) \end{cases}$$

Taking f = .67 [*Mitchell and Zemansky*, 1971] we calculate:

$$\sigma_0 = 1.55 \times 10^{-15} \text{ m}^2 \qquad (2.15)$$

2.4.3 *Weighted strengths and positions of D_{2a} and D_{2b} lines.* Since the D_2 line on a hyperfine structure level consists of 6 separate lines, we may write a more exact expression for k_ν. Assuming all hyperfine levels of a resonance level have the same lifetime, we express k_ν in sum form as:

$$k_\nu = \sum_{i=1}^{6} k_{0i} \exp - \left[\frac{2\sqrt{\ln 2} \quad (\nu - \nu_0)}{\Delta\nu_{Di}} \right]^2 \qquad (2.16)$$

where the sum is over the 6 hyperfine lines.

However, for this study, it was found sufficient to use just two lines designated as the D_{2a} and D_{2b} lines. The positions of these lines are taken as the weighted averages of the appropriate hyperfine lines. Referring to Figure 2.3 we obtain a wave number spacing for the two lines of:

$$\frac{14(-2.07) + 5(-2.27) + 1(-2.39)}{20} = -2.14 \text{ m}^{-1} \qquad \begin{matrix} (2.17a) \\ (D_{2a} \text{ line}) \end{matrix}$$

and

$$\frac{2(3.45) + 5(3.52) + 5(3.64)}{12} = 3.56 \text{ m}^{-1} \qquad (2.17b)\ (D_{2b} \text{ line})$$

The cross section values for the D_{2a} and D_{2b} lines are taken to be the following:

$$\sigma_{0a} = \frac{20}{32}\sigma_0 = 9.5 \times 10^{-16} \text{ m}^2 \qquad (2.18a)\ (D_{2a} \text{ line})$$

$$\sigma_{0b} = \frac{12}{32}\sigma_0 = 5.7 \times 10^{-16} \text{ m}^2 \qquad (2.18b)\ (D_{2b} \text{ line})$$

3. SODIUM LIDAR PRINCIPLES

3.1 *Introduction*

Lidar (*l*ight *d*etection *a*nd *r*anging) systems operate in much the same manner as conventional radar systems. An alternative name for lidar, laser radar, is used to emphasize the similarities. Essentially only the operating frequency has been raised.

However to understand the operation of a lidar system it is necessary to cover several topics pertinent to optical atmospheric transmission as well as to laser characteristics. In Section 3.2 we will discuss the different types of optical scattering which occur in the atmosphere. Sections 3.3 and 3.4 will develop the lidar equation for application to the sodium layer. Section 3.5 will present the technique of system calibration. A design calculation for system analysis will be presented in Section 3.6. And finally a check on the possibility of line saturation will be given in Section 3.7.

3.2 *Scattering Processes*

A laser beam passing through a homogeneous atmosphere of thickness z is attenuated and Bouguer's law applies as follows:

$$I = I_0 \exp - [\beta z] \tag{3.1}$$

where I_0 is the incident intensity and β is the extinction coefficient having units m^{-1}. In general the expression is more complicated since the atmosphere is not homogeneous and hence $\beta = \beta(z)$ and equation (3.1) becomes

$$I = I_0 \exp - \left[\int_0^z \beta(z)dz \right] \tag{3.2}$$

The quantity β is usually the sum of several components. Lower altitude lidar returns clearly indicate that the main contribution comes from Rayleigh (molecular) scattering. Occasionally layers of dust or high altitude clouds may cause deviations from molecular scattering. Mie scattering, or scattering from aerosols, is responsible for this effect. Both Rayleigh and Mie scattering are elastic processes for which excitation can take place at any frequency.

On the other hand, several inelastic scattering processes which are wavelength specific may be used to distinguish constituents. Raman scattering occurs when the incident light excites vibrational or rotational transitions of molecules. Re-emission takes place on either side of the incident frequency and the shift characterizes the molecule involved. The efficiency of the process may be enhanced by as much as a factor of 10^6 if the incident radiation is close to an electronic transition. The process is sometimes referred to as resonant Raman scattering. In either event the process plays a minor role in the value of the extinction coefficient for this experiment. Fluorescence involves the absorption of the incident radiation in one of the absorption bands of an atom or molecule and the re-emission in a longer wavelength band of the spectrum. The current literature draws little distinction between the fluorescence and resonance processes. Resonance scattering results from the excitation of electronic transitions. In this case the re-emitted light is the same frequency as the incident light and occurs after a period of time corresponding to the lifetime of the excited state (1-20 ns). Since the process is very efficient, it may be necessary to calculate the attenuation of the beam as it passes through the sodium layer to compute accurately the densities at the top of the layer. Hence resonance scattering may contribute to the extinction coefficient. The final contribution to the extinction coefficient comes from absorption. In this case the incident energy is not scattered but rather converted to a non-optical form of energy. A calculation is made later to estimate absorption by ozone.

The extinction coefficient is proportional to the density, ρ, of the scattering medium; and often we express the extinction coefficient for a particular process in terms of a scattering cross section, σ, by the relation

$$\beta = \rho\sigma \tag{3.3}$$

The scattering cross section is essentially the cross sectional area that the particle presents to the incident beam for scattering. As a guide the scattering cross sections for these various processes are compared in Table 3.1. It is apparent from Table 3.1 that resonance scattering may be used to detect very small densities of a constituent since its scattering cross section may be 10 orders of magnitude greater than the other processes involved.

3.3 *Lidar Equation*

3.3.1 *Formulation.* The University of Illinois lidar is a pulsed system operating in the monostatic mode. The laser is tuned to a resonance line of sodium and the received backscatter signal is processed to yield information about the sodium density profile. This translation of the received signal into a density profile requires a knowledge of the interaction of the laser pulse with the atmosphere. The system configuration is illustrated in Figure 3.1. The laser transmitter and receiver are separated slightly ($\sim$25 m) but the separation is small compared to the altitude of the sodium layer ($\sim$90 km) and can therefore be neglected. The solid angle of the receiver, Ω_R, is determined by the field stop iris and the diffraction limit of the Fresnel objective lens. The Fresnel lens limits the minimum half angle field of view to about 1 mrad which is comparable to the laser diver-

Table 3.1 Typical Values for Scattering Cross Sections

Process	Cross section (m^2)
Mie	10^{-14} - 10^{-31}
Raman	10^{-33} - 10^{-34}
Rayleigh	10^{-32}
Resonance-Fluorescence	10^{-16} - 10^{-23}

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

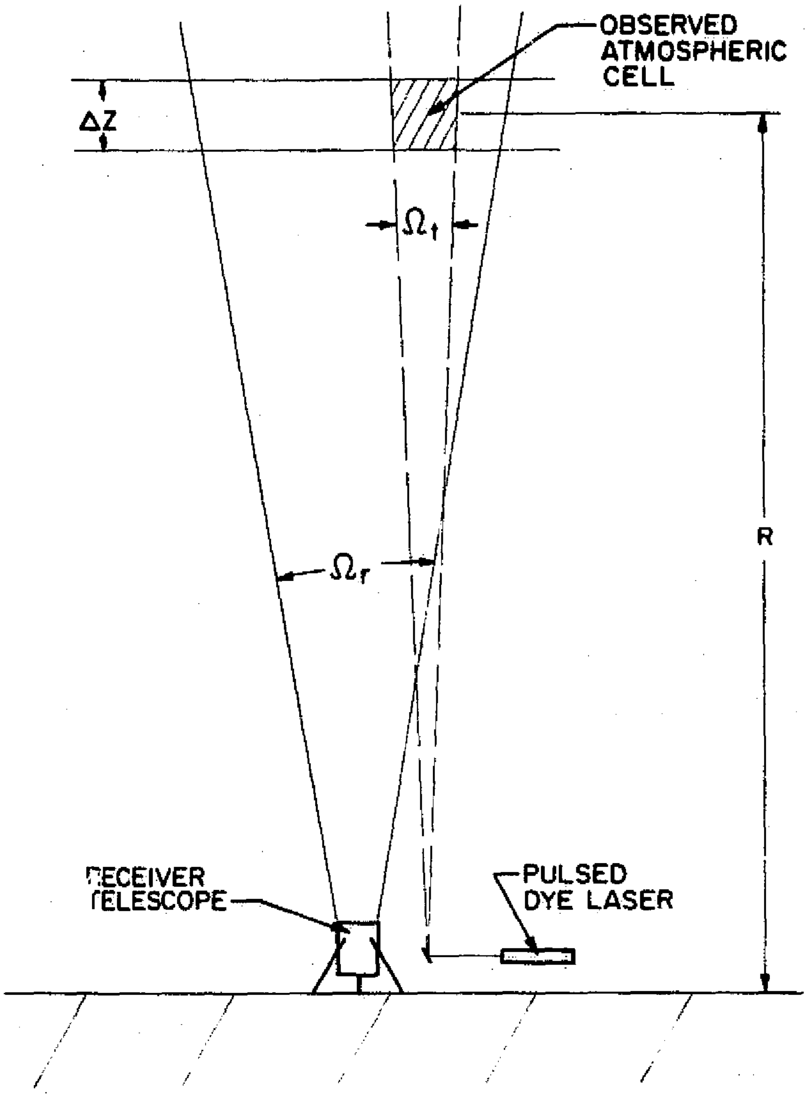


Figure 3.1 Monostatic lidar system.

gence. Consequently the field stop iris is opened slightly so that the entire scattering volume is illuminated by the laser ($\Omega_r > \Omega_t$).

The laser radar equation indicates the system and atmospheric parameters which contribute to the measured signal. Since the received signal is small, it is necessary to operate in a photon counting mode. A few remarks concerning the derivation of the lidar equation for this high altitude application will be helpful. Excellent references for the following are *Northend et al.* [1966], *Derr* [1972], and *Megie* [1976].

We assume that the laser emits an expected value of N_t photons per pulse with a frequency distribution $g(\nu)$. The quantity $g(\nu)$ represents the fraction of photons emitted between ν and $\nu + d\nu$ and by definition we have $\int_0^\infty g(\nu)\, d\nu = 1$. For simplicity we do not consider the temporal variation of the pulse at this time but rather only assume that the duration of the pulse is less than $2\Delta z/c$ where Δz is the range bin size.

The energy per photon is $h\nu$, where h is Planck's constant and ν is the frequency of the emitted photon. If P and τ represent the power and duration respectively of the transmitted pulse, then we may write

$$N_t = \frac{P\tau}{h\nu} \tag{3.4}$$

This expression is useful for approximate calculations; however, the frequency dependence of the pulse, $g(\nu)$, is ultimately needed and hence a spectral analysis of the laser pulse must be performed.

Consider the photon flux incident on the bottom of the layer which occurs at altitude z_0 (Figure 3.2). If we lump all of the atmospheric attenuation which results from elastic scattering (Rayleigh and Mie scattering) and absorption into a parameter T, then we may express this flux as

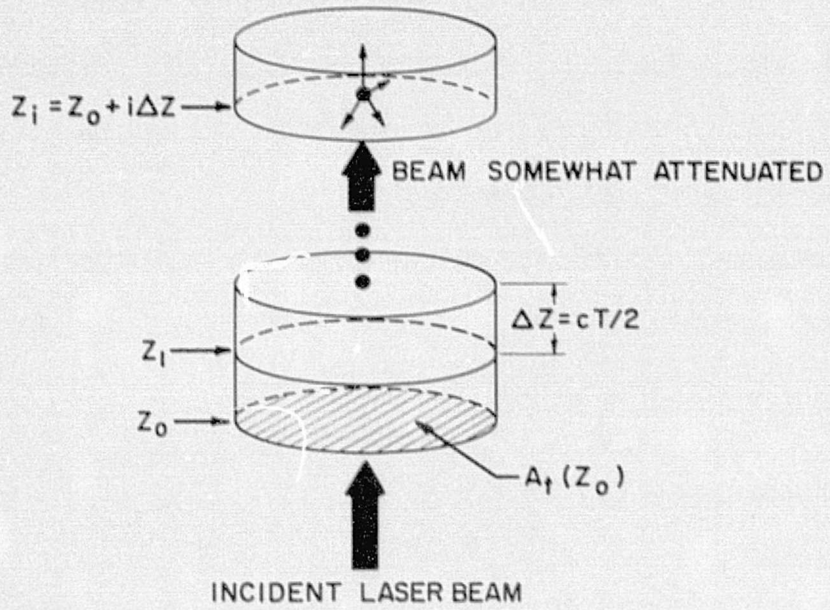


Figure 3.2 Range cell convention.

$$\phi_\nu(z_0) = \frac{N_t g(\nu) T}{A_t(z_0)} \tag{3.5}$$

where $A_t(z)$ is the surface illuminated by the laser beam at altitude z. This area is given by the expression

$$A_t(z) = \Omega_t\, z^2 = \pi(\alpha_L z)^2 \tag{3.6}$$

where α_L is the angular halfwidth or divergence angle of the laser beam. For $\alpha_L = 1$ mrad the beam has a radius of about 80-100 m at the altitude of the sodium layer and so $A_t(z_0) \equiv 20000\ \text{m}^2$.

As the beam passes through the sodium layer, it is also attenuated by sodium D line resonance scattering. Hence at altitude $z > z_0$, the flux is given by

$$\phi_\nu(z) = \frac{N_t g(\nu) T}{A_t(z)} \ \exp - [\tau_\nu(z)] \tag{3.7}$$

where τ_ν is the optical depth of sodium given by

$$\tau_\nu(z) = \int_{z_0}^{z} \sigma_\nu \rho(s) ds \tag{3.8}$$

In expression (3.8) σ_ν is the absorption cross section for sodium and $\rho(z)$ is the density of sodium at altitude z. (Note that $k_\nu = \sigma_\nu \rho$). The cross section σ_ν is not constant with altitude as it depends on temperature.

We may approximate the scattering atmospheric volume by a cylinder since the divergence angle for the laser is small. Hence the number of photons re-emitted per unit area from an infinitesimally small thickness, dz, is given by

$$dN(\nu, z) = \sigma_\nu \rho(z) \phi_\nu(z) dz \tag{3.9}$$

We are assuming in equation (3.9) that natural damping is more important than deactivation by collisions with atomic or molecular constituents. This is reasonable since the collisional frequency at 80 km is about 10^5 s^{-1} and the natural lifetime of the excited state is 16 ns. In addition we may assume that the re-emitted photon has the same frequency as the incident photon.

If the area of the receiver is A_r, then the receiver solid angle as seen from the atmospheric cell is given by

$$\Omega_r = \frac{A_r}{s^2} \tag{3.10}$$

Assuming that the scattering from the sodium layer is isotropic, the number of received photons from the partition of the sodium layer ds thick is

$$dN_r(\nu, s) = dN(\nu, s)A_t(s)\ \frac{A_r}{4\pi s^2}\ T \exp - [\tau_\nu(s)] \tag{3.11}$$

where the factor of 1/4π converts the isotropic scatterer to a steradian basis.

Combining equations (3.7), (3.9) and (3.11) we may write the lidar equation for our case. If we define

$$s_i = s_0 + i\Delta s \qquad\qquad i = \text{integer} \tag{3.12}$$

then the expected value of the number of photoelectrons, $N_i(\nu)$, counted during the interval of time corresponding to the returns from the ith range bin is given by

$$N_i(\nu) = N_t g(\nu)\ \frac{A_r}{4\pi}\ T^2\eta \int_{s_i}^{s_{i+1}} \sigma_\nu\ \frac{\rho(s)}{s^2} \exp - [2\tau_\nu(s)]ds \tag{3.13}$$

where η is the receiver efficiency.

3.3.2 *Approximations.* A number of approximations may be used to simplify the integration in (3.13). The factor of $1/z^2$ may be approximated by the constant $1/(z_i + \Delta z/2)$. This is equivalent to assuming that $1/z^2$ changes slowly over distances on the order of the range bin size, and the approximation introduces a maximum range error of $\Delta z/2z$. For $\Delta z = 1.5$ km this means an error 1% at 80 km and 0.75% at 100 km.

The exponential factor in (3.13) may also be approximated by the first two terms of its Taylor's series expansion. If we take the column density of sodium to be 10^{14} m^{-2} and let the scattering cross section be 9×10^{-16} m^2 which corresponds to the peak of the D_2 line, then

$$\tau_\nu \cong (10^{14}\ \mathrm{m}^{-2})(9 \times 10^{-16}\ \mathrm{m}^2) = 0.09 \qquad (3.14$$

Hence the error incurred by replacing $\exp - [2\tau_\nu(z)]$ by $1 - 2\tau\nu(z)$ is less than 2%.

The absorption cross section σ_ν, varies with altitude since it is temperature dependent. This is because the expression for k_ν is inversely proportional to $\Delta_{\nu D}$ and hence inversely proportional to $\sqrt{T}$ (see equations 2.9 and 2.11). The relative error that results from neglecting temperature variations depends on the frequency ν of the incident radiation. For example at $\nu = \nu_0$, the dependence of the exponential factor in σ_ν on $\Delta\nu_D$ is eliminated. The temperature profile at this altitude can be highly variable and exhibit a wavelike structure at times. Temperatures can vary between 150-250°K. For $T = 200 \pm 50$°K, σ_ν varies by at most $\mp$ 15%. However when the laser linewidth is greater than the width of the absorption line, the effective scattering cross section is independent of temperature (see Section 3.4.2). For this reason σ_ν will be assumed to be constant with altitude.

It is appropriate to discuss the consequences of neglecting the time response of the laser. If $N_t(t)$ represents the photon time response of the laser pulse, then the total intensity received from the passage of the laser pulse through the atmosphere is given by

$$\int_0^\infty F(s)N_t(t - \frac{2s}{c})\, ds \tag{3.15}$$

where F is a fraction which depends on lidar parameters such as number density of scattering medium, cross section of scatterer, range, geometry, and efficiency factors.

To obtain the number of counts in a bin, the signal processing equipment counts the number of photoelectrons emitted from the cathode of the photomultiplier tube in an interval of duration $T = 2\Delta s/c$. Hence for a particular bin, the number of counts recorded is

$$\int_{t-T}^{t} d\tau \int_{-\infty}^{\infty} F(s)N_t(t - \frac{2s}{c})\, ds \tag{3.16}$$

where the lower limit on the s integrand has been extended to $-\infty$ by using the convention that $F(s) = 0$ for s negative. It is convenient to rewrite (3.16) in terms of a receiver gating function G,

$$G\left(\frac{t}{T}\right) = \begin{cases} 1 & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

Then expression (3.16) becomes

$$\int_{-\infty}^{\infty} F(s)ds \int_{-\infty}^{\infty} N_t(\tau - t_d)G\left(\frac{t - \tau}{T}\right) d\tau \tag{3.18}$$

where $t_d = 2s/c$ is the delay between the transmitted and received pulse. Recalling that $F(s)$ is proportional to the density of the scattering medium, $\rho(s)$, we find that (3.18) is a relationship involving the two-fold convolution of the density, transmitter pulse, and receiver gating function.

If the receiver gate is larger than the pulse width, the system resolution is approximately $\Delta s = cT/2$. The characteristic resolution is the value which results from the convolving of the laser pulse with the gating function and is the sum of the range bin size and spatial length of pulse. We may estimate the error incurred if we neglect this convolution. An approximation to $N_t(t)$ based on photographs of the time response of the laser pulse (Figure 3.3) is

$$N_t(t) = \begin{cases} 2a^2t \exp - [at]^2 & t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

where $N_t(t)$ is normalized, $a \cong 1/3$ and t is in μs (Figure 3.4). The full width at half maximum is about 3.5 μs which agrees with experimental values.

If we neglect the translation t_d, then we may consider the convolution to be of the form

$$v{*}w = \int_{-\infty}^{\infty} v(t)w(t - \tau)d\tau \tag{3.20}$$

where

$$v(t) = \begin{cases} 2a^2t \exp - [at]^2 & t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.21}$$

$$w(t) = \begin{cases} 1 & 0 \leq t \leq T \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

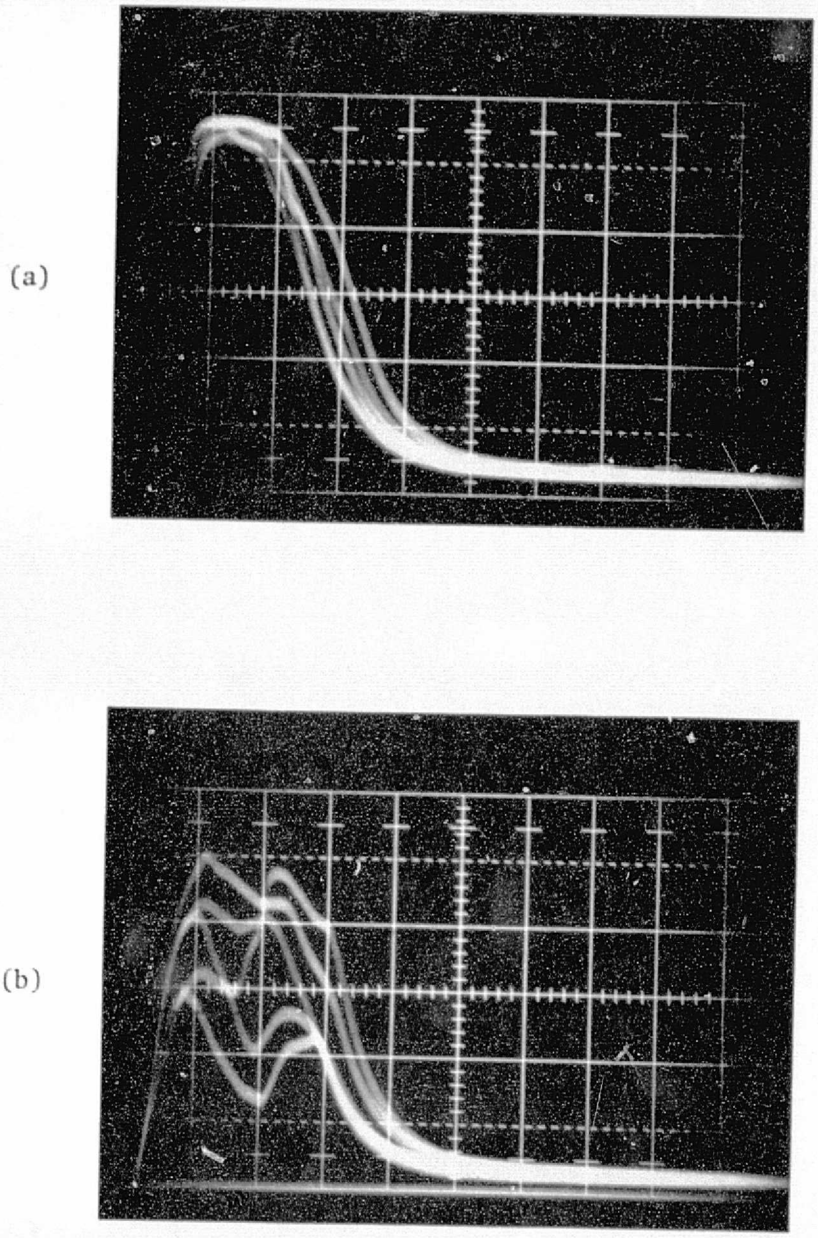
or convenience we subdivide the problem into three cases.

Case 1: $\tau < 0$. In this case the two functions do not overlap and their product is zero. Hence

$$v{*}w = 0 \qquad t < 0 \tag{3.23}$$

Case 2: $0 \leq t \leq T$. The functions partially overlap but only over the range $0 \leq \tau \leq t$. Hence

$$v{*}w = \int_0^t 2a^2\tau \exp - [a\tau]^2 \, d\tau \tag{3.24a}$$



ORIGINAL
OF POOR QUALITY

Figure 3.3 (a) Five superimposed laser pulses as seen by a sodium cell monitor. Horizontal scale is 2 μs/div. (b) Five superimposed laser pulses shown as tuning drifts to the edge of the D_2 line. Horizontal scale is 2 μs/div.

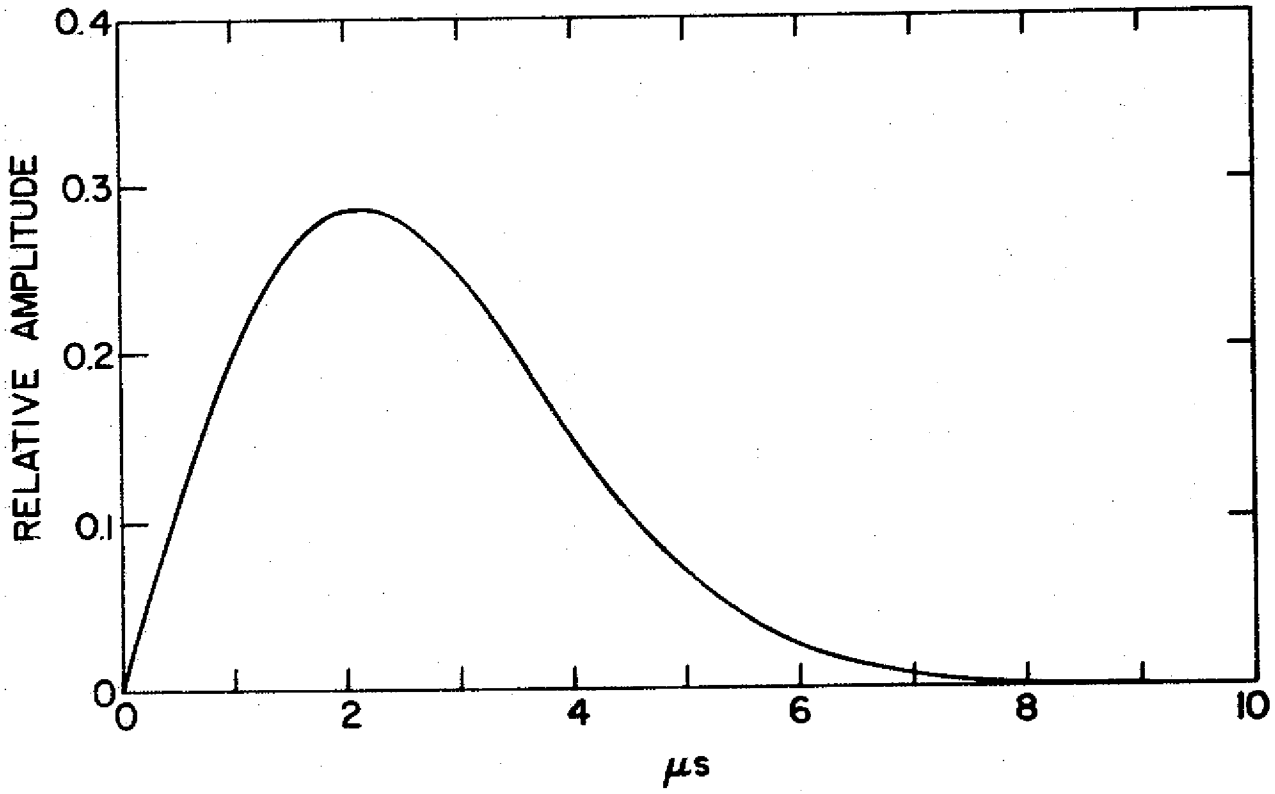


Figure 3.4 Approximate profile of the laser pulse time response, $N_t(t)$.

This integral may be solved by making the substitution $z = (a\tau)^2$ and we obtain

$$v{*}w = 1 - \exp - [at]^2 \qquad 0 \leq t \leq T \tag{3.24b}$$

Case 3: $t \geq T$. In this situation the functions fully overlap and thus

$$\begin{aligned} v{*}w &= \int_{t-T}^{T} 2a^2\tau \exp - [a\tau]^2 \, d\tau \\ &= \exp - [a^2(t - T)^2] - \exp - [at]^2 \qquad t \geq T \end{aligned} \tag{3.25}$$

This convolution is plotted in Figure 3.5. Note that when we consider the convolution of the laser pulse and the gating function, there is some overlap of signals from adjacent range bins.

A estimate may be made on the effect of neglecting this overlap. The sodium layer can fall by over a factor of 10 in 4 km on the bottomside and has a scale height of 3-4 km at the topside. This means that adjacent 1.5 km range bins at the bottomside may differ by as much as a factor of 4 and by a factor of 2 at the topside. Referring to Figure 3.6, we see that the total counts received during the sampling corresponding to bin i consists of a tail from bin $i - 1$ and the counts received from bin i while the laser pulse is entirely within the $i^{\underline{th}}$ range bin. This return is to be compared to the total number of counts received form the $i^{\underline{th}}$ bin. Referring to Figure 3.6b, if we assume that the density of scatters in range bin $i - 1$ is a factor of 4 smaller than that of bin i, then we form

$$\frac{\text{\# counts received during interval of duration } T}{\text{\# counts emitted from range bin } i} = \frac{A_1 + 4A_2}{4(A_2 + A_3)} = 0.80 \tag{3.26}$$

where we are integrating the expressions derived in equations (3.24b) and (3.25) to determine A_2 and A_1 respectively. Obviously $A_1 = A_3$. Alternatively

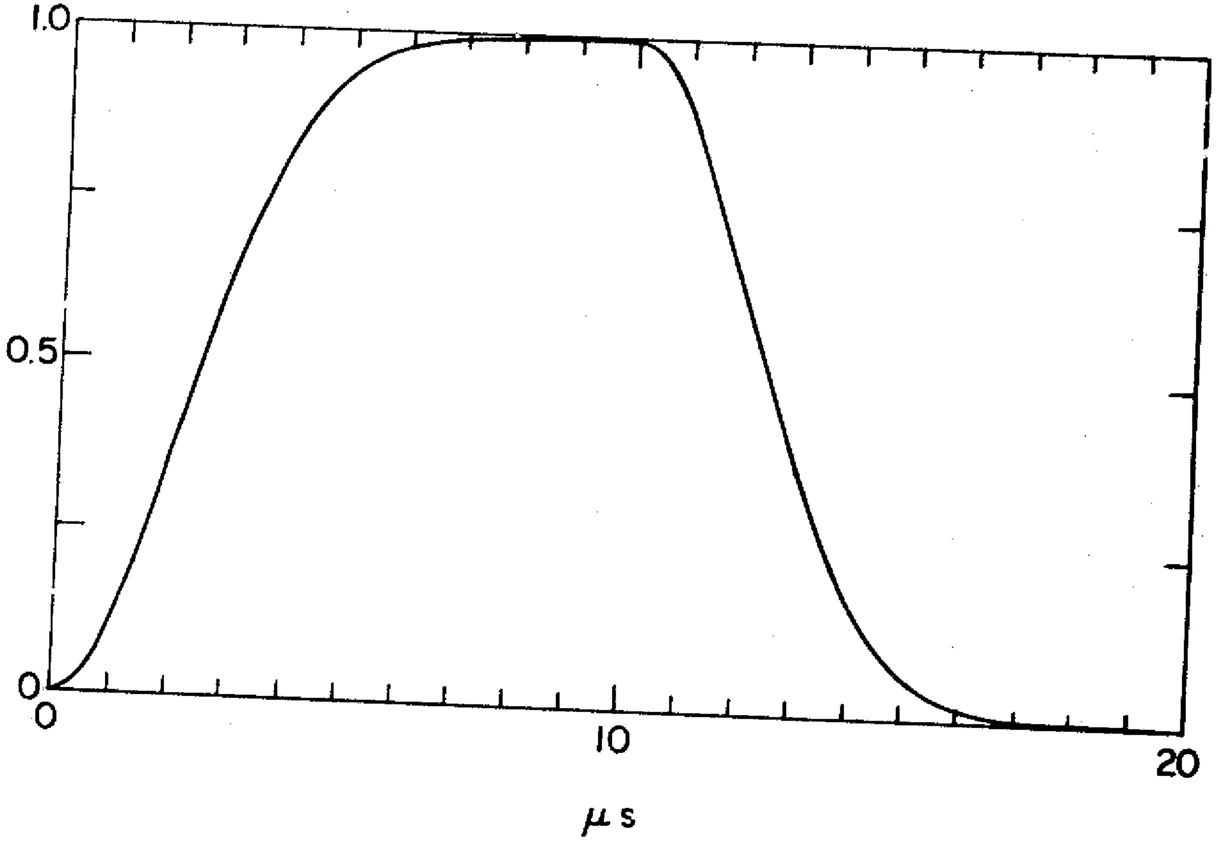


Figure 3.5 Convolution of the laser pulse with the receiver gating function.

ORIGINAL PAGE IS
OF POOR QUALITY

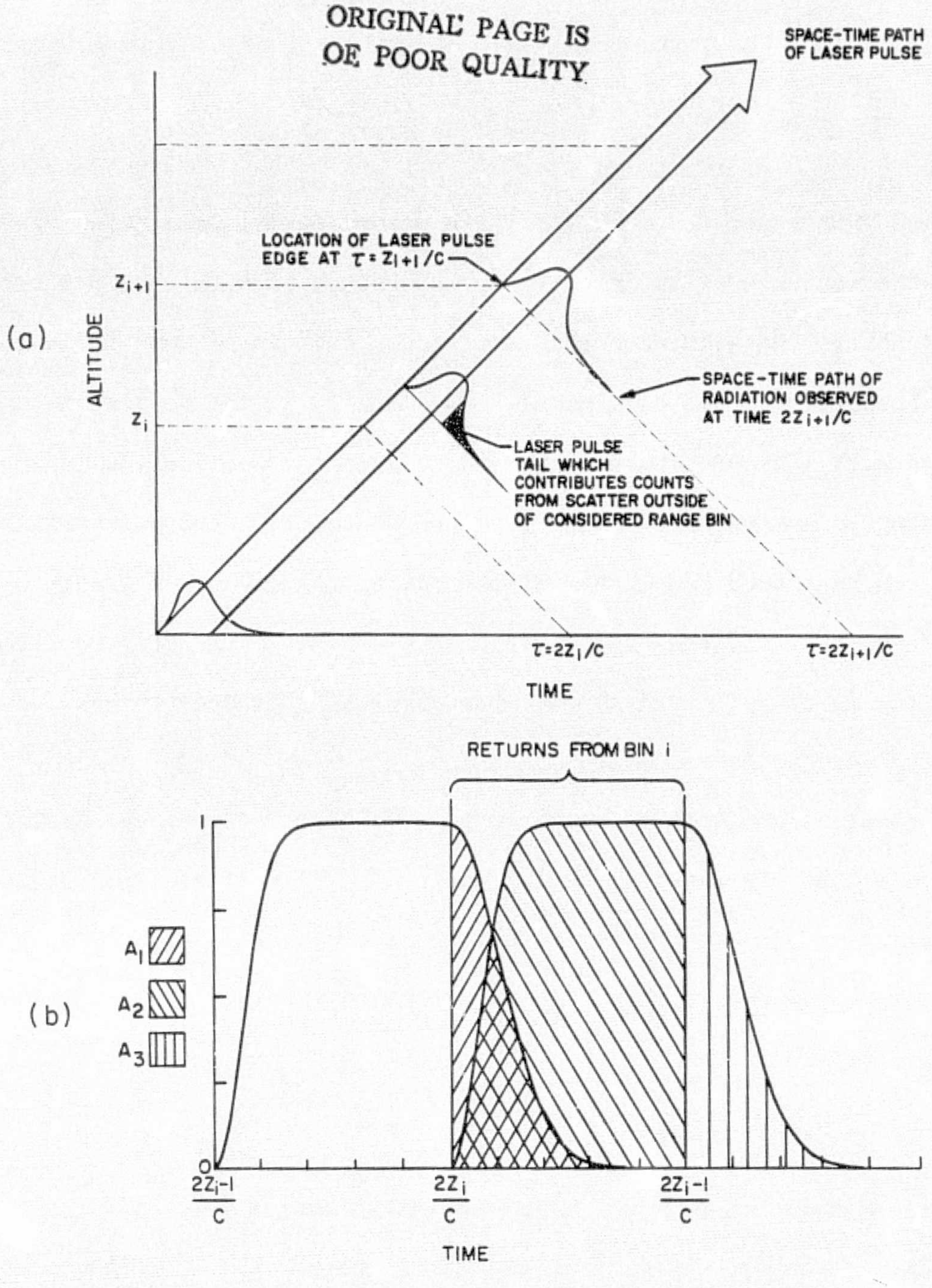


Figure 3.6 (a) Space-time diagram of propagating laser pulse; (b) overlap of scattering from adjacent range bins caused by convolution of laser pulse with receiving gating function.

for the case when the i^{th} bin is a factor of 2 less dense than bin $i - 1$, the ratio becomes

$$\frac{2A_1 + A_2}{A_2 + A_3} = 1.27 \tag{3.27}$$

We see then that under the most unfavorable conditions where steep gradients occur in the sodium profile, returns could produce as much as a 30% error in some range bins. However when adjacent range bins are of comparable number density, the error should be minimal. In summary, if the density is increasing with altitude, then the measured gradient will be less than the true gradient. Alternatively if the sodium density is decreasing with altitude, the measured gradient will be greater than the true gradient. These effects would be enhanced if the range bin size were reduced without reducing the laser pulse width. Then deconvolution procedures would need to be introduced to solve (3.18).

For completeness we note that the minimum resolution of the system is determined by the lifetime, τ, of the excited state. If we allow 4τ seconds for the excited atoms to return to the ground state, which corresponds to 92% of the excited atoms, then the minimum range resolution for a monostatic lidar configuration is

$$\Delta s_{\text{min}} = \frac{4\tau c}{2} = 9.6 \text{ m} \tag{3.28}$$

3.3.3 *Simplified expression for returns from a range bin.* Applying the approximations discussed in Section 3.3.2 to equation (3.13) we obtain

$$N_i(\nu) = \frac{N_t g(\nu) A_r T^2 \eta \sigma_\nu}{4\pi[s_i + \Delta s/2]^2} \int_{s_i}^{s_{i+1}} \rho(s) \exp - [2\tau_\nu(s)] ds \tag{3.29}$$

Because the system receives data in the form of range bins having thickness Δz, it is natural to replace the continuous density profile, $\rho(z)$, by a discrete profile, $\bar{\rho}_i$, which represents the average density in the $i^{\underline{\text{th}}}$ range bin. Let us compute N_i for the case $i = 0$. If we expand $\tau_\nu(z)$ as defined in (3.8) by assuming no variation of σ_ν with altitude and consider the discrete profile mentioned above, we obtain

$$N_0(\nu) = N_t g(\nu) \frac{A_r T^2 \eta \sigma_\nu}{4\pi[z_i + \Delta z/2]^2} \int_{z_0}^{z_1} \bar{\rho}_0 \exp - [2\bar{\rho}_0 \sigma_\nu (z - z_0)] dz \tag{3.30}$$

We may use the Taylor's series expansion for the exponential factor as discussed in Section 3.3.2 and perform the integration as indicated below

$$\sigma_\nu \int_{z_0}^{z_1} \bar{\rho}_0 \exp - [2\bar{\rho}_0 \sigma_\nu (z - z_0)] dz = 1/2 - 1/2 \exp - [2\bar{\rho}_0 \sigma_\nu \Delta z] \approx \bar{\rho}_0 \sigma_\nu \Delta z \tag{3.31}$$

To eliminate the frequency dependence of $N_0(\nu)$ we must integrate over all ν. Substituting the approximation (3.31) into (3.30) and performing the integration results in

$$N_0 = N_t \frac{A_r T^2 \eta \bar{\rho}_0 \Delta z}{4\pi[z_i + \Delta z/2]^2} \int_0^\infty \sigma_\nu g(\nu) d\nu \tag{3.32}$$

This quantity represents the expected number of counts registered in the $i = 0$ bin.

To calculate the number of photons received from the $i^{\underline{\text{th}}}$ bin, we must account for the absorption of the laser line as it passes through bins, j, $j < i$. Let us assume that the laser linewidth is a Gaussian given by

$$g(\nu) = G_0 \exp - [\omega/\alpha]^2 \qquad (3.33)$$

where

$$\omega = 2\,\frac{\sqrt{\ln 2}}{\Delta\nu_D}\,(\nu - \nu_0)$$

$$G_0 = \frac{2}{\Delta\nu_L}\sqrt{\frac{\ln 2}{\pi}}$$

$$\alpha = \Delta\nu_L/\Delta\nu_D$$

$\Delta\nu_D$, $\Delta\nu_L$ = absorption line Doppler width and laser linewidth respectively.

If we express the absorption coefficient as $k_\nu = k_0 \exp[-\omega^2]$, then the absorption incurred by the beam passing through the ith range bin is given by

$$A = \frac{\int_{-\infty}^{\infty} \{1 - \exp[-k_0 l \exp(-\omega^2)]\} \exp[-(\omega/\alpha)^2]d\omega}{\int_{-\infty}^{\infty} \exp[-(\omega/\alpha)^2]d\omega} \qquad (3.34)$$

A series expansion exists for the quotient [*Mitchell and Zemansky*, 1971] and is given by

$$A = \frac{k_0 l}{\sqrt{1+\alpha^2}} - \frac{(k_0 l)^2}{2!\sqrt{1+2\alpha^2}} + \cdots + \frac{(-1)^n (k_0 l)^n}{n!\sqrt{1+n\alpha^2}} + \cdots \qquad (3.35)$$

We may eliminate the higher order terms if the value of $k_0 l$ is very small. To estimate its value we take l = 1.5 km which corresponds to the thickness of a range bin. The value of k_0 is given by $\sigma_0\rho$ as indicated in equation (3.3). If we take the maximum cross section for the D_{2a} line of $\sigma_{0a} = 9.5 \times 10^{-16}$ m^2 and a typical maximum density for sodium of $\rho = 5 \times 10^9$ m^{-3} we obtain $k_0 l = 7.1 \times 10^{-3}$. Hence A may be approximated by the first term of the series.

If the laser linewidth is $\Delta\lambda_L = 10$ pm, then $\Delta\nu_L = \frac{c}{\lambda^2}\Delta\lambda_L = 8.6 \times 10^9$ Hz for $\lambda = 589$ nm. The quantity $\Delta\nu_D$ was calculated in equation (2.10). Hence $\alpha = (8.6 \times 10^9 \text{ Hz})/(1.07 \times 10^9 \text{ Hz}) = 8.06$. Therefore

$$A \approx \frac{k_0 l}{8.06} \qquad (\Delta\lambda_L = 10 \text{ pm}, T = 200°K) \tag{3.36}$$

To calculate the flux impinging on the i^{th} bin, ϕ_ν^i, we account for the absorption in each range bin j, $j < i$, and write

$$\phi_\nu^i = \phi_\nu(z_i)\ T_{Na} \tag{3.37}$$

$$T_{Na} = \prod_{j=0}^{i-1} \left(1 - \frac{\sigma_0 \bar{\rho}_j\ \Delta z}{8.06}\right) \tag{3.38}$$

Often an upper bound on the estimate of the transmission through the layer is given by

$$T_{Na} \approx \exp -[k_0 l] = \exp\left[-\sigma_0 \sum_{j=0}^{i-1} \bar{\rho}_j\ \Delta z\right] \qquad (\Delta\nu_L \ll \Delta\nu_D) \tag{3.39}$$

An estimate on the effect of this transmission on the sodium density calculated is given in Section 3.5.3. The final expression for the expected number of counts received from the i^{th} bin is given by

$$N_i = \frac{N_t A_r T^2 T_{Na}^2 \eta \bar{\rho}_i \Delta z}{4\pi[z_i + \Delta z/2]^2} \qquad \int_0^\infty \sigma_\nu\ g(\nu)\ d\nu \tag{3.40}$$

where T_{Na} is given by either expression (3.38) or (3.39).

3.4 *Effective Scattering Cross Section*, σ_{eff}

3.4.1 *Definition*. The expression for the expected number of photons received from the i^{th} range bin (3.40) contains an integral of the product of the laser line and the cross section of sodium. We may rewrite this integral as

$$\int_0^\infty \sigma_\nu\ N_t\ g(\nu)\ d\nu = \sigma_{eff}\ \ N_t \tag{3.41}$$

where N_t represents the total number of photons emitted in one laser pulse and σ_{eff} is the effective scattering cross section. Hence σ_{eff} is defined by the equation

$$\sigma_{eff} = \frac{\int_0^\infty \sigma_\nu \, N_t \, g(\nu) \, d\nu}{\int_0^\infty N_t \, g(\nu) \, d\nu} = \int_0^\infty \sigma_\nu \, g(\nu) d\nu \qquad (3.42)$$

3.4.2 *Approximate values for large and small laser linewidths.* As a first approximation, let us assume that the laser pulse is square in the frequency domain with width $\Delta\nu_L << \Delta\nu_D$. Employing the simple formula for the integrated cross section (see equation (2.13b))

$$\int_0^\infty \sigma_\nu \, d\nu = \frac{e^2}{4\varepsilon_0 m_e c} \, f \qquad \text{(SI units)} \qquad (3.43)$$

and taking $g(\nu)$ to be constant over the width of the D_2 line we obtain

$$\sigma_{eff} = \frac{e^2}{4\varepsilon_0 m_e c} \, \frac{f}{\Delta\nu_L} \qquad \Delta\nu_L >> \Delta\nu_D, \; \Delta\nu < \nu_{D_2} - \nu_{D_1} \qquad (3.44)$$

The value of this expression is appropriate not only when the conditions specified in equation (3.44) are satisfied but also when the laser spectrum may cover different sections of the D_2 line on a shot to shot basis because of temperature variations of the laser tuning elements. In the latter case we assume that $g(\nu)$ is effectively constant over the D_2 line when averaged over many shots.

Notice that as the laser linewidth is narrowed so that it becomes much smaller than the resonance line, σ_{eff} increases to a maximum

$$\sigma_{eff} = \frac{e^2}{4\varepsilon_0 m_e c} \, \frac{f}{\Delta\nu_D} \qquad \Delta\nu_D >> \Delta\nu_L \qquad (3.45)$$

where again $g(\nu)$ is uniform over its frequency distribution.

3.4.3 σ_{eff} *for a Gaussian laser frequency profile.* To compute σ_{eff} more realistically, we will assume that $g(\nu)$ is a Gaussian profile given by

$$g(\nu) = \frac{2}{\Delta\nu_L}\sqrt{\frac{\ln 2}{\pi}} \exp - [\frac{2\sqrt{\ln 2}}{\Delta\nu_L} (\nu - \nu_0)]^2 \tag{3.46}$$

where ν_0 = center frequency of the laser line

$\Delta\nu_L$ = laser line halfwidth

Since σ_ν is taken to be the sum of two Gaussians, one centered about the D_{2b} and the other about the D_{2a} line as described in Section 2.4.3, an analytical solution to (3.41) exists and is derived below.

We may express σ_{eff} as

$$\sigma_{eff} = \int_{-\infty}^{\infty} \left\{ \sigma_{0a} \exp - \left[\frac{(\nu - \nu_{Da})^2}{2\sigma_D^2} \right] + \sigma_{0b} \exp - \left[\frac{(\nu - \nu_{Db})^2}{2\sigma_D^2} \right] \right\} \cdot \left\{ \frac{1}{\sqrt{2\pi\sigma_L^2}} \exp - \left[\frac{(\nu - \nu_0)^2}{2\sigma_L^2} \right] \right\} d\nu \tag{3.47}$$

where

$\sigma_D^2 = \frac{\Delta\nu_D^2}{8 \ln 2}$ = variance of Doppler broadened profile

$\sigma_L^2 = \frac{\Delta\nu_L^2}{8 \ln 2}$ = variance of laser line

ν_{Da}, ν_{Db} = center frequency of D_{2a} and D_{2b} line, respectively.

The integral of (3.47) may be divided into two similar integrals and we indicate the solution to one of these integrals below. The portion of (3.47) which involves the D_{2a} line may be written as

$$\int_{-\infty}^{\infty} \frac{\sigma_{0a}}{\sqrt{2\pi\sigma_L^2}} \exp - \left[\frac{(\nu - \nu_{Da})^2}{2\sigma_D^2} + \frac{(\nu - \nu_0)^2}{2\sigma_L^2}\right] d\nu \tag{3.48}$$

Changing variables by letting $u = \nu - \nu_{Da}$ and defining $\nu_a = \nu_{Da} - \nu_0$, (3.48) may be rewritten as

$$\int_{-\infty}^{\infty} \frac{\sigma_{0a}}{\sqrt{2\pi\sigma_L^2}} \exp - \left\{ \frac{1}{2\sigma_D^2\sigma_L^2} \left[u^2\sigma_L^2 + (u + \nu_a)^2\,\sigma_D^2\right]\right\} du \tag{3.49}$$

By completing the square of the quantity enclosed by square brackets we obtain a Gaussian with variance

$$\frac{\sigma_D^2\;\sigma_L^2}{\sigma_D^2 + \sigma_L^2} \text{ and mean } - \frac{\sigma_D^2\nu_a}{\sigma_D^2 + \sigma_L^2}\,.$$ Hence (3.49) may be written as

$$\frac{\sigma_{0a}}{\sqrt{2\pi\sigma_L^2}} \sqrt{\frac{2\pi\sigma_D^2\sigma_L^2}{\sigma_D^2 + \sigma_L^2}} \exp - \frac{\nu_a^2}{2(\sigma_D^2 + \sigma_L^2)} \tag{3.50}$$

A similar expression may be obtained for the D_{2b} line contribution. Substituting the appropriate values for σ_D, σ_ν and ν_a we obtain the final expression

$$\sigma_{eff} = \frac{\Delta\nu_D}{\sqrt{\Delta\nu_D^2 + \Delta\nu_L^2}} \left\{ \sigma_{0a} \exp - \left[\frac{4\ \ln 2(\nu_{Da} - \nu_0)^2}{\Delta\nu_D^2 + \Delta\nu_L^2}\right] + \sigma_{0b} \exp - \left[\frac{4\ \ln 2(\nu_{Db} - \nu_0)^2}{\Delta\nu_D^2 + \Delta\nu_L^2}\right]\right\} \tag{3.51}$$

Values of σ_{eff} for various laser linewidths are plotted in Figure 3.7. Note that if the laser is tuned to the D_{2a} line rather than the center of the D_2 line, an increase in σ_{eff} is obtained for small $\Delta\lambda_L$.

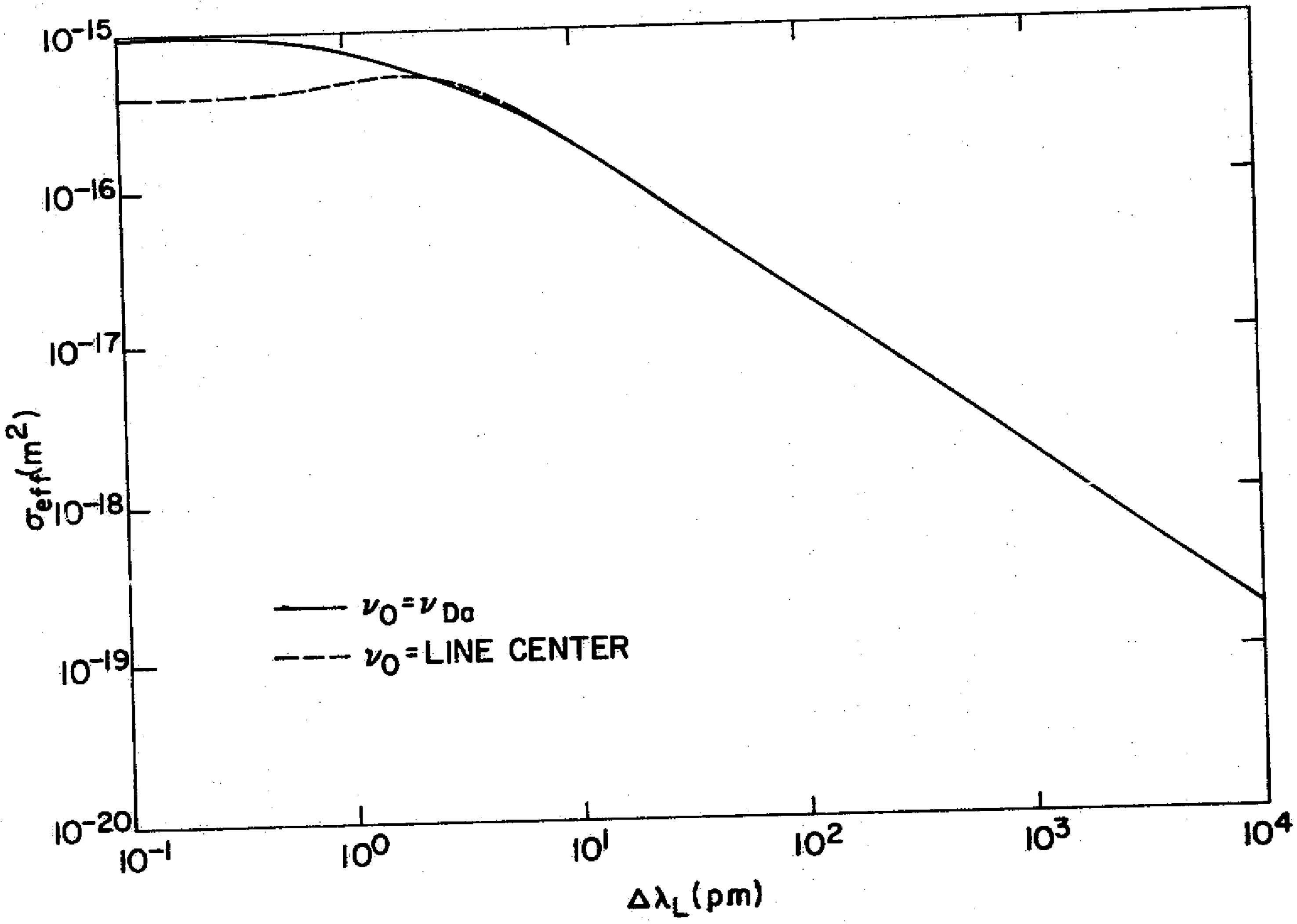


Figure 3.7 Values of σ_{eff} as a function of laser linewidth.

We may now compare the value of σ_{eff} obtained by the approximation (3.44) with the more exact expression (3.51) for $\Delta\nu_L >> \Delta\nu_D$. Substituting the expression for the scattering cross section of the D_{2b} line (2.14b) and using the relative weights of the D_{2a} and D_{2a} lines as given in (2.18), equation (3.51) becomes

$$\sigma_{eff} = 2\sqrt{\frac{\ln 2}{\pi}}\ \frac{e^2}{4\varepsilon_0 m_e c}\ \frac{f}{\Delta\nu_L}\left\{ \frac{20}{32} \exp - \left[\frac{4\ \ln 2(\nu_{Da} - \nu_0)^2}{\Delta\nu_L{}^2}\right] + \frac{12}{32}\ \exp - \left[\frac{4\ \ln 2(\nu_{Db} - \nu_0)^2}{\Delta\nu_L{}^2}\right]\right\} \tag{3.52}$$

The Doppler width of the line is $\Delta\nu_D = 1.07 \times 10^9\ \text{s}^{-1}$ as given in equation (2.10). The linewidth of the laser is calculated by $\Delta\nu_L = c\Delta\lambda_L/\lambda_0{}^2$ For $\Delta\lambda_L = 10$ pm, $\Delta\nu_L = 8.65 \times 10^9\ \text{s}^{-1}$. If we assume the laser emits at line center, namely $\sigma = 1697340\ \text{m}^{-1}$, and take ν_{Da} and ν_{Db} to be positioned relative to ν_0 as calculated in (2.17) then we obtain from equation (3.52)

$$\sigma_{eff} = 1.82 \times 10^{-16}\ \text{m}^2 \qquad (\Delta\lambda_L = 10\ \text{pm}) \tag{3.53}$$

This value may be compared to that by applying (3.44). In this case employing equation (3.44)

$$\sigma_{eff} \cong 2.05 \times 10^{-16}\ \text{m}^2 \qquad (\Delta\lambda_L = 10\ \text{pm}) \tag{3.54}$$

Hence we find that the approximate formula (3.44) gives a value about 12% higher than the expression obtained by more realistically modeling the laser frequency distribution.

3.4.4 *Sensitivity of* σ_{eff} *to linewidth and stability.* It is worthwhile to study the changes in σ_{eff} for values of $\Delta\nu_L$ comparable to $\Delta\nu_D$ since this is the regime in which the laser may be operated in the future. In Figure

3.7 we showed the increase in σ_{eff} as the spectral width of the laser was narrowed and noted an enhancement in σ_{eff} for small $\Delta\lambda_L$ if $\nu_0 = \nu_{Da}$. For a more realistic presentation, one must realize that tuning elements have at least a 30% insertion loss. If we assume an etalon narrows the laser linewidth by a factor of 10, then we should experience a 30% decrease in pulse energy for every factor of 10 decrease in linewidth. Hence in reality although σ_{eff} would increase by roughly two orders of magnitude with the use of two etalons, the laser energy would drop simultaneously by about one order of magnitude resulting in only an order of magnitude increase in returns.

From expression (3.51) we also note that σ_{eff} is sensitive to the center line of the laser. This effect is displayed in Figure 3.8. The series of plots show that although σ_{eff} may increase as the laser linewidth is narrowed, laser instability could cancel all the benefits of a narrow linewidth. By using both Figures 3.7 and 3.8, it was decided that a linewidth of 10 pm was optimum for attaining the highest value of σ_{eff} given the frequency stability of the laser. Certainly when the stability of the laser is increased, then a more narrow linewidth may be considered.

3.5 *System Calibration*

3.5.1 *Normalization by Rayleigh scattering returns.* Referring to equation (3.40) one may observe that it is not necessary to measure T^2, A_r, η, and N_t individually to obtain a value for $\bar{\rho}_i$. Rather, the sodium returns may be normalized by returns from the 30-40 km region, assuming returns in this region are consistent with Rayleigh scattering. This may be established if there is little evidence of aerosol layers or high altitude clouds giving rise to Mie scattering. The signal strength of returns from this altitude will determine the produce $N_t A_r T^2 \eta$.

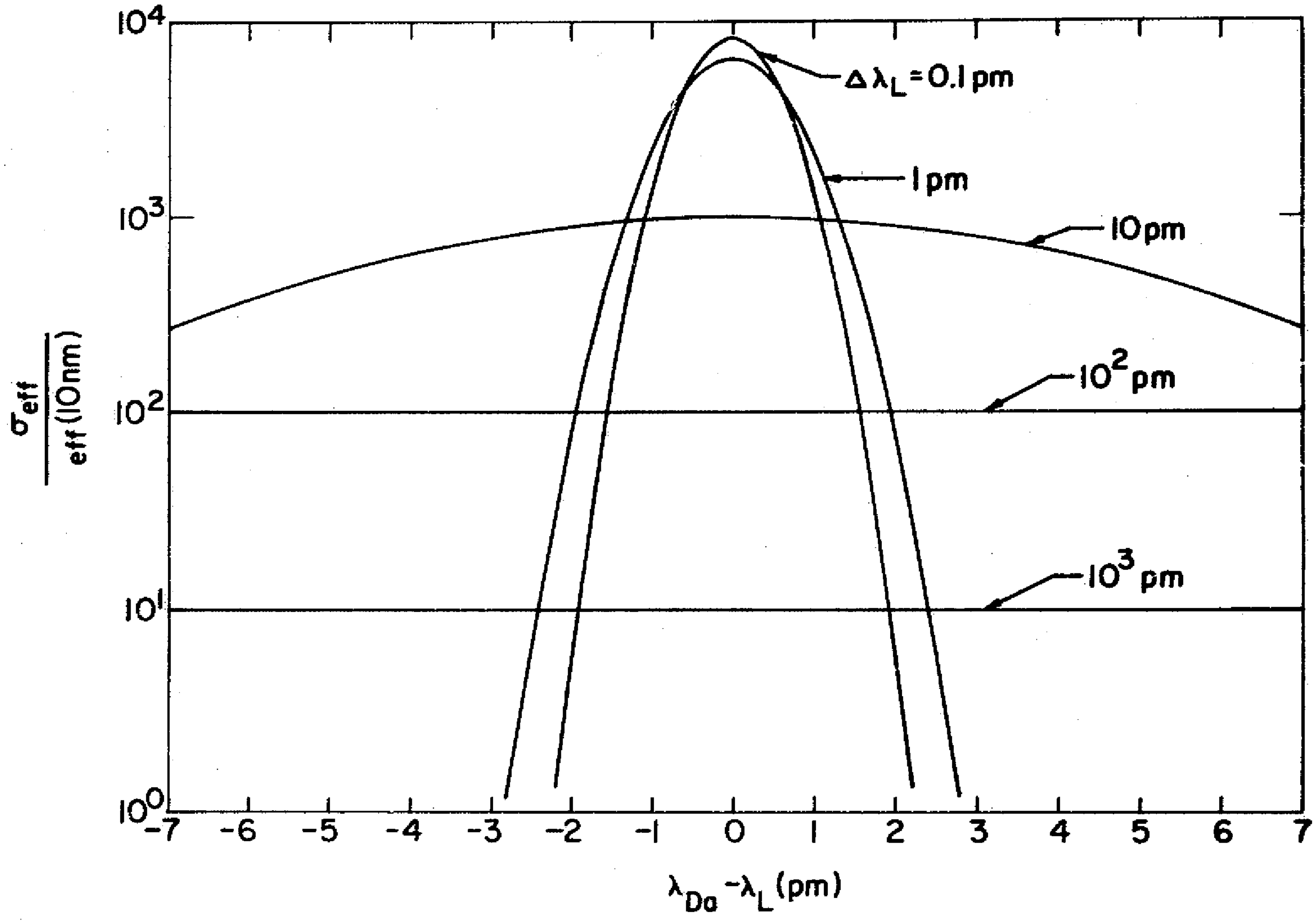


Figure 3.8 Variation of σ_{eff} versus the deviation of the laser line from the center of the D_{2a} line.

Since Rayleigh scattering is frequency non-specific, we may dispense with the integral in (3.40) and write the expression for the number of counts received from a range bin in the 30-40 km region as

$$N_R = \frac{N_t A_r T_R^{2} n \bar{\rho}_R \Delta z \sigma_R(\pi)}{4\pi z_R^{2}} \tag{3.55}$$

where T_R is the atmosphere transmission to the altitude z_R and $\bar{\rho}_R$ is the average density of the atmosphere in the range bin of thickness Δz centered around z_R. The quantity $\sigma_R(\pi)$ is the classical Rayleigh backscatter cross section.

Taking the ratio of equation (3.40) and (3.55) and solving for $\bar{\rho}_i$ we obtain

$$\bar{\rho}_i = \frac{\sigma_R(\pi)(z_i + \Delta z/2)^2}{\sigma_{eff}\, z_R^{2}} \; \frac{N_i}{N_R} \; \bar{\rho}_R \; \frac{T_{Na}^{2} T_R^{2}}{T^2} \tag{3.56}$$

The ratio T_R/T is discussed in Section 3.5.3. The absorption due to sodium which appears as T_{Na} is calculated in Section 3.5.4. The number density, $\bar{\rho}_R$, of the atmosphere range cell used for normalization is taken to be $\bar{\rho}_R = 3.69 \times 10^{23}\ \mathrm{m}^{-3}$ at $z_R = 30$ km [*CIRA*, 1972].

3.5.2 *Rayleigh scattering cross section.* The laser is only weakly polarized and hence we may assume that the incident light on the atmospheric range bins is unpolarized (natural) light. The angular Rayleigh scattering cross section is then given by

$$\sigma_R(\theta) = \sigma \cdot f \cdot \psi(\theta) \tag{3.57}$$

where σ = classical Rayleigh cross section for isotropic molecules

f = factor which accounts for the anisotropy of the polarizability of air molecules

$\psi(\theta)$ = angular scattering function

The classical Rayleigh cross section is given by

$$\sigma = \frac{32\pi^3}{3\lambda^4} \frac{(n_R-1)^2}{\rho_R^2} \tag{3.58}$$

where

n_R = refractive index of air at the altitude of interest

λ = wavelength of incident light

ρ_R = atmospheric number density

Often the quantity enclosed by parentheses will be approximated by

$$(n_R - 1)^2 \simeq \frac{1}{4} (n_R^2 - 1)^2 \qquad (n_R - 1 \text{ small}) \tag{3.59}$$

yielding the expression

$$\sigma = \frac{8\pi^3}{3\lambda^4} \frac{(n_R^2 - 1)^2}{\rho_R^2} \tag{3.60}$$

However the approximation (3.59) under the assumptions of Rayleigh scattering is not more correct [*Van De Hulst*, 1957] and we will use expression (3.58) for the classical Rayleigh cross section rather than (3.60).

The factor f is given for unpolarized light by

$$f = \frac{6 + 3\Delta}{6 - 7\Delta} \tag{3.61}$$

where Δ is the depolarization factor ($\Delta \simeq .035$) [*Penndorf*, 1957]. The scattering is not isotropic and assuming isotropic air molecules the angular scattering function is given for unpolarized light as [*Johnson*, 1953]

$$\psi_r(\theta) = \frac{3}{8\pi} \left[\frac{\cos^2\theta + 1}{2}\right] = \frac{\rho(\cos\theta)}{4\pi} \tag{3.62}$$

where $\rho(\cos\theta)$ is the Rayleigh phase function and $\psi_r(\theta)$ converts σ to a steradian basis. In this case the phase function is normalized to unity and represents the conservative case of perfect scattering [*Chandrasekhar*, 1960]; absorption is not treated in this approach. A more exact value for $\rho(\cos\theta)$ is given by [*Penndorf*, 1957]

$$\rho(\cos\theta) = 0.7629\ (1 + 0.932\ \cos^2\theta) \tag{3.63}$$

and will be used for this work.

Therefore for backscatter $\theta = \pi$, and $\psi_r(\pi) = \frac{3}{8\pi}\ (0.983)$. Note that $f \cong 1.06$ and hence we may make the approximation

$$\psi_r(\pi)f = \frac{3}{8\pi}\ (1.04) \approx \frac{3}{8\pi} \tag{3.64}$$

Taking the above into consideration, the Rayleigh backscatter cross section then becomes

$$\sigma_r(\pi) = \frac{4\pi^2}{\lambda^4}\ \frac{(n_R - 1)^2}{\rho_R{}^2} \tag{3.65}$$

The units of σ_r are $\mathrm{cm}^2\ \mathrm{sr}^{-1}$. To obtain $\sigma_R(\pi)$, which is used in equation (3.56), $\sigma_r(\pi)$ must be multiplied by 4π. The Rayleigh backscatter cross section then becomes

$$\sigma_R(\pi) = \frac{16\pi^3}{\lambda^4}\ \frac{(n_R - 1)^2}{\rho_R{}^2} \tag{3.66}$$

The index of refraction shows a variation with density and temperature given by [*Penndorf*, 1957]

$$(n_R - 1) = (n_S - 1)\ \left(\frac{1 + \alpha t_S}{1 + \alpha t_R}\right) \frac{p_R}{p_S} \tag{3.67}$$

where n_S = index of refraction for standard air where $(n_S - 1) \times 10^{+8}$ is calculated to be 27714 for λ = 589 nm

α = 0.00367

p_R, p_S = number density of air where Rayleigh scattering occurs and of standard air, respectively

t_R, t_S = temperature of Rayleigh scattering medium measured in °C and of standard air (15°), respectively.

The temperature at 30 km is about -40°C and assuming $\rho_S = 2.5 \times 10^{25}$ m^{-3} we calculated with the aid of equations (3.65) and (3.66) that

$$\sigma_R(\pi) = 5.2 \times 10^{-31} \text{ m}^2 \tag{3.68}$$

Notice that $\sigma_R(\pi)$ depends on selected definitions and "standard" adopted values for several parameters. As a result the value of $\sigma_R(\pi)$ adopted will differ between authors.

3.5.3 *Atmospheric transmission in the 30-90 km region.* Equation (3.56) may be simplified by noting $T_R \cong T$. We determine this by computing the effect of Rayleigh scattering and absorption on the laser beam between the altitudes z_R and $z_i + \Delta z/2$. For computation purposes we set z_R = 30 km and $z_i + \Delta z/2$ = 90 km.

The extinction coefficient due to Rayleigh scattering is given by

$$\beta_R = \int_{30 \text{ km}}^{90 \text{ km}} \sigma_R \rho_R(z) dz = 2 \times 10^{-4} \tag{3.69}$$

where we assume that σ_R is constant with altitude and of magnitude $\sigma_R = 7 \times 10^{-32}$ m^2. The values for $\rho_R(z)$ are those of the CIRA 1972 atmosphere.

The Chappuis absorption bands of ozone span the sodium D_2 line. Ozone has a maximum absorption cross section at 590 nm on the order of 5×10^{-25} m^2 [*Banks and Kockarts*, 1973]. If we assume a column density of 2×10^{22} m^{-2} [*Shimazaki and Ogawa*, 1974], then the extinction coefficient resulting from ozone absorption is

$$\beta_A = (5 \times 10^{-25} \text{ m}^2)(2 \times 10^{22} \text{ m}^{-2}) = 0.01 \tag{3.70}$$

With the above information we may now relate T and T_R by writing

$$T = T_R \exp - [\beta_R + \beta_A] = 0.99\, T_R \tag{3.71}$$

Hence it is reasonable to equate T and T_R. If we neglect the attenuation in the laser beam due to sodium absorption, then we may write (3.56) in its simplest form as

$$\bar{\rho}_i = \frac{\sigma_R(\pi)}{\sigma_{eff}} \frac{[z_i + \Delta z/2]^2}{z_R^2} \frac{N_i}{N_R} \bar{\rho}_R \qquad (3.72)$$

3.5.4 *Accuracy of the method.* The calibration method depends on demanding little deviation in returns from those expected by Rayleigh scattering in the 30-40 km region. Hence, if enhancements occur in this region due to aerosol scattering, calibration may be difficult if not impossible. In addition, the method assumes that we know the number density of the atmosphere in the 30-40 km region. Hence the calibration is only as accurate as the model atmosphere used; this accuracy is 10% typically.

Furthermore we assume a temperature for the 30-40 km region to calculate the Rayleigh scattering cross section. A variation in this temperature of ± 10°C results in a ∓ 4% change in returns which is not serious. More importantly, we find a serious variation in Rayleigh scattering cross section values in the literature. Because of the various definitions, values of σ_R differ by as much as a factor of 2 and one must be aware of this, prior to comparing data between groups.

Finally we determine the error if absorption in the sodium layer is neglected. In equation (3.14) we estimated the optical depth for a one way path through the layer of $\tau_\nu = 0.09$. Hence a two way path will have a transmission of

$$T_{Na}^2 = \exp - 2(0.09) = 0.84 \qquad (3.73)$$

and the top of the layer should be corrected by up to 16%. However, because the linewidth of the laser can be measured to only a factor of 2 or 3 and with the systematic errors discussed above, errors incurred by neglecting this attenuation are comparatively small. For a qualitative presentation therefore, equation (3.72) applies.

3.6 *Signal-to-Noise Calculation*

3.6.1 *Noise contributions.* In Section 3.3 we obtained an expression for the expected number of signal photons received from the scattering volume. In this section we will investigate the effect of noise on returns and calculate the signal-to-noise ratio for the system. The noise considered consists of background noise, detector noise, logic circuit generated noise, signal-induced noise, and self noise of the signal.

Background, or sky noise, is caused by airglow or heavenly bodies which may enter the field of view of the telescope. Scattered light from nearby cities would also fall into this category. Its effect is strong enough that sometimes it prevents accurate measurements from being taken. Detector noise, which includes dark noise, is present when photoelectrons are emitted from the photocathode of the photomultiplier tube and as a consequence may be minimized by cooling the photomultiplier tube. Data channel failures result in logic circuit generated noise. This noise is easily observable and software routines are used to identify the problem. It is not a prime source of noise. In addition the laser and receiver have been separated by a large distance to minimize interference problems. Signal-induced noise occurs when noise is generated after the irradiation of the photomultiplier tube. *Jones and Kent* [1974] suggest that the electrons are excited into trap levels which subsequently require less thermal energy for their removal. It has been shown that this effect occurs for low light levels and may be present for up to 200 μs after the signal is quenched [*Pettifer*, 1975]. Since the scattered laser intensity is of relatively long duration, the effect is cumulative. That is the $i^{\underline{\text{th}}}$ bin will have signal-induced noise contributions from the $i - 1$, $i - 2$, etc., bins. Our returns from the sodium layer occur approximately 600 μs after the firing of the laser and even with the cumulative effect, it seems doubtful that these returns are affected. However, since we normalize our returns on the basis of scattering from heights around 30 km, it is possible that these returns could be affected. We will assume that signal-induced noise is not a serious problem for times greater than 200 μs after the firing of the laser pulse. To verify this assumption one would need to run tests on the C31034A photomultiplier tube performance as different photomultiplier tubes vary by orders of magnitude in their amount of signal-induced noise [*Young*, 1976]. Finally self noise, or shot noise, results from statistical variations in the received signal; and to estimate the effect we will assume that the received photons obey Poisson statistics. Its effect will become evident below.

Under the above assumptions, the contribution from noise to the number of counts registered is proportional to the time the receiver gate is open ($2\Delta z/c$) and is given by an expression quite similar to that for signal photons. If we let N represent the number of noise counts registered in a range bin, then

$$N = \frac{2\Delta z}{c}\left\{\frac{\eta\ B_\lambda\ \Omega_r\ \Delta\lambda\ A_r}{h\nu} + N_D\right\} \tag{3.74}$$

where

B_λ = background brightness at the receiver
η = receiver efficiency
Ω_r = solid angle of the receiver

$\Delta\lambda$ = receiver bandwidth centered at wavelength $\lambda = c/\nu$

A_r = receiver area

$h\nu$ = energy of incident photon

N_D = detector dark count

and the other symbols have their usual meaning.

3.6.2 *Formulation of the* S/N *ratio.* Several expressions for the signal-to-noise ratio are used in practice and care must be taken to be aware of the particular one in use. For our purposes the received signal and noise are statistical in nature since we are in a photon counting mode. Hence an appropriate measure of the noise is the standard deviation of the received counts rather than the number N. If we assume Poisson statistics, this standard deviation is $\sqrt{N_i + N}$ where N_i is determined by (3.40). Hence the signal-to-noise ratio for a range bin on a per shot basis is given by the ratio of the mean number of received signal photons to the standard deviation of the mean number of received photons, namely

$$S/N = \frac{N_i}{\sqrt{N_i + N}} \tag{3.75}$$

Many authors will multiply N by a factor of 2 to account for the uncertainty in the noise count, N. Ideally if the noise is observed for a long time, N can be accurately determined.

We note that for x shots, the signal-to-noise ratio is given by

$$S/N = \frac{x\,N_i}{\sqrt{x(N_i + N)}} = \frac{\sqrt{x}\,N_i}{\sqrt{N_i + N}} \tag{3.76}$$

Hence in theory we can increase the signal-to-noise ratio indefinitely by integrating over a greater and greater number of shots. In practice one is reluctant to do this since over the time scales involved, the sodium profile may change and hence only an average profile rather than an instantaneous profile is obtained. We believe that integration times must be kept shorter than 1/2 hour to study the time variation of the layer. In applying (3.76) one must be aware of the appropriate approximations. The signal-to-noise ratio may be approximated by $S/N \cong \sqrt{x\ N_i}$ for the low noise condition. In this case a high repetition rate, low energy per pulse system might be suitable. However the system does not always operate under low noise conditions. For $N >> N_i$, $S/N \cong \sqrt{x\ N_i}\ /\ \sqrt{N}$ and a much longer integration time will be required to achieve an appropriate S/N ratio. In summary then (3.76) implies that the sodium lidar system should incorporate not only a high energy per pulse laser but also one with high average power.

3.6.3 *System calculation for the Urbana lidar*. It is instructive to calculate the signal-to-noise ratio for our system. Values for several of the parameters used in this calculation are given in Table 3.2. The receiver solid angle may be calculated using the telescope divergence half angle, α_R, by

$$\Omega_R = \frac{\pi(\alpha_R z)^2}{z^2} = \frac{\pi\alpha_R^{\ 2}}{4} \tag{3.77}$$

Hence $\Omega_R = 7 \times 10^{-6}$ sr. The detector dark count is taken from the specification sheet for the RCA 31034A photomultiplier tube (see Figure 5.4). The dark count is a function of temperature and for -20°C is $N_D = 50\ \mathrm{s}^{-1}$. Alternatively $N_D = 5 \times 10^{-4}/10\ \mu\mathrm{s}$. We use a value for $B_\lambda = 10^{-4}\ \mathrm{W\ m^{-2}\ sr^{-1}\ \mu m^{-1}}$ which is an appropriate value for scatter from city lights [*Stewart and Hopfield*, 1965]. With this information the background count, N_B may be calculated using (3.74) and has a value of $N_B = 5.9 \times 10^4\ \mathrm{s}^{-1}$ or alternatively 0.59/10 μs. For this example the system is background noise

Table 3.2 Parameters of the Urbana Sodium Laser Radar System

Wavelength	589 nm
Transmitted bandwidth	10 pm
Laser beam divergence (half angle)	$\alpha_T < 1$ mrad
Laser pulse energy	100 mJ = 3×10^{17} photons
Pulse duration	4 μs FWHM
Receiver gate time	10 μs (1.5 km range bin)
Receiver area	0.114 m^2
Receiver bandwidth	5 nm
Receiver efficiency	0.05
Receiver beam divergence (half angle)	$\alpha_R = 3$ mrad
Separation of receiver and transmitter	25 m
Repetition rate	0.5 s^{-1}

limited. Usually the experiment is done on very clear nights when N_B is more than an order of magnitude lower. Under these conditions the system is self-noise limited. For further computation we will take $N = 0.6$.

Substituting the appropriate values of Table 3.2 into equation (3.40) and neglecting the optical depth of the sodium layer we may calculate the mean number of photoelectron counts from the range bin corresponding to the peak of the sodium layer taken to be at $z = 90$ km and $\bar{\rho}_i = 5 \times 10^9$ m^{-3}. If we let $T^2 = 0.2$, then $N_i = 5$. Hence employing equation (3.75), the signal-to-noise ratio becomes $S/N = 2.1$ on a per shot basis. Hence in 400 shots, which corresponds to a 15 min interval, we would expect a 20-fold increase in the signal-to-noise ratio.

3.6.4 *Confidence interval.* More exactly we must assign a confidence interval to our results [*Hogg and Craig*, 1970]. We consider the experiment as one in which we seek to determine the mean number of counts, μ, returned from a range bin. Each shot produces a value, μ_i, which is close to μ and is the mean of a distribution with variance σ^2. Furthermore we assume that the process obeys a Poisson distribution, namely

$$P(X = k) = \frac{\mu^k \exp - \mu}{k!} \tag{3.78}$$

where k is an integral number of counts.

Under the assumption of low noise ($N < N_i$), the variance for the distribution is simply the square of the signal-to-noise ratio or the total number of received counts from the particular bin. Hence, if X is to be within 10% of μ, we write the probability of this event as

$$P(.9\mu < X < 1.1\mu) = \sum_{k=.9\mu}^{1.1\mu} \frac{\mu^k e^{-\mu}}{\mu!} = \sum_{k=.9N_i}^{1.1N_i} \frac{(N_i)^k \exp\text{-}N_i}{N_i!} \tag{3.79}$$

If we take $S/N = 2$ per shot, then in 25 shots the $S/N = 10$. Using (3.79), the probability of the event of being within 10% of 100 photoelectron counts is 0.7. Alternatively for 225 shots the $S/N = 30$. In this case we have a probability of 0.99 of being within 10% of the actual mean.

3.7 *Sodium Line Saturation*

3.7.1 *Statement of the problem.* For short duration laser pulses and a low divergent system, saturation effects may become important. Although it will become apparent that saturation is not a problem for our system, we discuss the topic as a reminder for future system improvements or studies of other constituents such as potassium [*Megie et al.*, 1978] where saturation effects may need to be considered.

Let N_1 and N_2 represent the population number densities of the ground state and excited state, respectively. The absorption coefficient, k_ν, is proportional to the population difference and is related to the effective cross section by

$$k_\nu = (N_1 - N_2)\, \sigma_{eff} \tag{3.80}$$

Hence for equal populations of the two states ($N_1 = N_2$), the material becomes transparent to the incident radiation. This condition of no absorption is called saturation. Obviously a change in the absorption coefficient will occur before saturation, and it is useful to determine if any effects are noted in our experiment.

3.7.2 *Solution to the rate equations.* We may write the rate equations for the excited state population as

$$\frac{dN_2}{dt} = -\frac{N_2}{\tau} - (N_2 - N_1)\, W_i \tag{3.81}$$

where

W_i = rate of induced emission or absorption

τ = lifetime of the excited species

The rate W_i may be written in terms of the incident intensity of the beam, $I(\nu,t)$, where I has units $W \cdot m^{-2}$. Specifically

$$W_i = \int_\nu k_\nu \frac{I(\nu,t)}{h\nu} d\nu \tag{3.82}$$

The quantity in the brackets is the incident photon flux in units per frequency unit per second per square meter. Hence

$$W_i = \frac{\sigma_{eff} N(t) T}{\pi(\alpha_L z)^2} \tag{3.83}$$

where

σ_{eff} = effective scattering cross section of sodium

$N(t)$ = photon time response of the laser beam

T = transmission of atmosphere

α_L = half angle divergence of the laser beam

z = altitude of range bin

We may rewrite (3.81) as

$$\frac{dN_2}{dt} = -\frac{N_2}{\tau} - (N_2 - N_1) \frac{\sigma_{eff} N(t) T}{\pi(\alpha_L z)^2} \tag{3.84}$$

Let us assume that the laser pulse is a square pulse of duration Δt_L, given analytically as

$$N(t) = \begin{cases} \dfrac{N_t}{\Delta t_L} & 0 < t < \Delta t_L \\ 0 & \text{otherwise} \end{cases} \tag{3.85}$$

where N_t represents the total number of photons emitted in the laser pulse. The solution to (3.84) is given by

$$N_2(t) = \begin{cases} \dfrac{\rho\bar{\tau}}{\tau_s}\ [1 - \exp - (t/\bar{\tau})] & t < \Delta t_2 \\ \dfrac{\rho\bar{\tau}}{\tau_s}\ [1 - \exp - (\Delta t_L/\bar{\tau})]\ \exp - \left[\dfrac{t - \Delta t_L}{\tau}\right] & t > \Delta t_L \end{cases} \tag{3.86}$$

where

$\rho = N_1 + N_2$ = number density in scattering volume

$$\tau_s = \frac{\pi(\alpha_L s)^2\ t_L}{\sigma_{eff}\ N_t T}$$

$$\frac{1}{\bar{\tau}} = \frac{1}{\tau} + \frac{2}{t_s}$$

3.7.3 *Saturation criterion.* At this point we may obtain an estimate for the condition of saturation. The maximum number of atoms which could be excited by a laser beam is given by

$$\rho_{\max} V = \rho V\ \frac{\Delta t_L}{\tau} \tag{3.87}$$

where

ρ = number density of sodium

V = scattering volume

Δt_L = laser pulse duration

τ = lifetime of the excited state

If each photon excited an atom, then under the most favorable conditions the total number of atoms excited would be given by

$$\#\text{excited atoms} = N_t T\ \left[\frac{\sigma_{eff}\ \rho V}{\pi(\alpha_L s)^2}\right] \tag{3.88}$$

where the quantity enclosed by parenthesis is the ratio of the maximum effective area that the medium presents for scattering divided by area of the laser beam at altitude z. Equating (3.87) and (3.88) we obtain a crude condition for saturation, namely

$$\left(\frac{N_t}{\alpha_L{}^2}\right) = \frac{\pi\ z^2}{T\sigma_{eff}}\ \frac{\Delta t_L}{\tau} \tag{3.89}$$

Equation (3.89) gives us a restraint on the laser parameters. In particular we note that the divergence effect is quadratic. Using the appropriate values from Table 3.2 and taking τ = 16 ns, we obtain the condition for z = 90 km that

$$\left(\frac{N_t}{\alpha_L{}^2}\right)_{\text{sat}} = 7.0 \times 10^{28} \tag{3.90}$$

For our system $(N_t/\alpha_L{}^2) = 3.0 \times 10^{23}$ hence saturation is obviously not a problem. In fact (3.90) would represent the condition of a laser with parameters 230 J pulse energy and 10^{-4} rad half angle divergence.

Although the above gives an approximation for saturation effects, we may explicitly obtain an expression which will indicate the deviation from expected returns for high photon flux conditions. Since stimulated emission is emitted in the same direction as the laser beam, only spontaneous emission contributes to the signal received at the ground. Neglecting any absorption of this spontaneous emission or any stimulated emission caused by it, the range bin volume re-emits $\pi(\alpha_L z)^2\ \Delta z\ N_2(t)/\tau$ photons per second. Hence the received number of photoelectrons from this range bin is given by

$$N_{R(\text{sat})} = \frac{A}{4\pi z^2}\ T\eta\ \frac{\pi(\alpha_L z)^2\ \Delta z}{\tau} \int_0^{\frac{2\Delta z}{c}} N_2(t)dt \tag{3.91}$$

where Δz is the range resolution and $t = 0$ corresponds to the bottom of the range bin. This equation is accurate for $\Delta t_L \ll 2\Delta z/c$, and in fact we may extend the integration to ∞ since $N_2(t)$ falls off rapidly before $t = 2\Delta z/c$. The integral in (3.91) becomes

$$\int_0^{\Delta t_L} \frac{\rho\bar{\tau}}{\tau_s} [1 - \exp - (t/\bar{\tau})]dt$$

$$+ \int_{\Delta t_L}^{\infty} \frac{\rho\bar{\tau}}{\tau_s} (1 - \exp - (\Delta t_L/\bar{\tau})] \exp - [(t - \Delta t_L)/\tau]dt$$

$$= \frac{\rho\bar{\tau}}{\tau_s} \{\Delta t_L + (\tau - \bar{\tau})[1 - \exp - (\Delta t_L/\bar{\tau})]\} \qquad (3.92)$$

Substituting (3.92) into (3.91), we may rewrite $N_{R(\text{sat})}$ as

$$N_{R(\text{sat})} = \frac{N_i}{1 + R} \left\{1 + \frac{\tau}{\Delta t_L} \frac{R}{1 + R} \left[1 - \exp - \left(\frac{\Delta t_L(1 + R)}{\tau}\right)\right]\right\} \qquad (3.93)$$

where

τ = lifetime of excited atom

Δt_L = laser pulse width

R = $2\tau/\tau_s$

N_i = number of counts received under the condition of no saturation

To observe the effect of beam divergence on saturation, we plot the ratio $N_{R(\text{sat})}/N_i$ in Figure 3.9 as a function of α_L and see that saturation effects depend on Δt_L. In general saturation effects will only become appreciable for $\alpha_L < 10^{-4}$ rad, and $\Delta t_L < 300$ ns.

Although equation (3.93) was derived for atmospheric applications, it may be used for determining saturation effects in the sodium cell monitor. These effects may be important if returns are to be normalized on the basis of of sodium cell scattering.

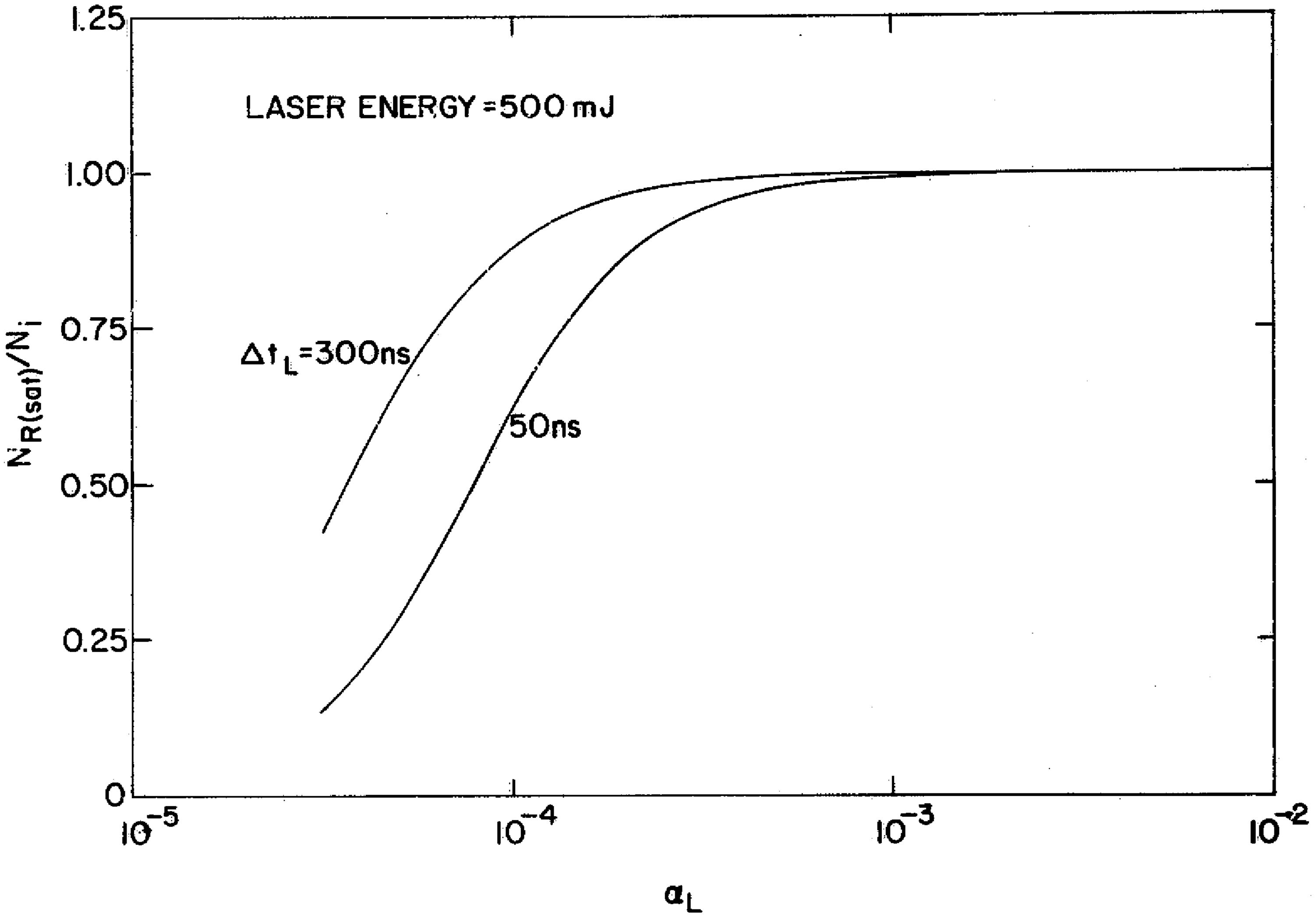


Figure 3.9 Ratio of $N_{R(sat)}/N_i$ for different values of the half angle laser divergence.

4. DYE LASER CHARACTERISTICS

4.1 *Principles of Dye Lasers*

4.1.1 *Lasing transitions*. The general properties of organic dye lasers have been reviewed in the literature [*Snavely*, 1969; *Schafer*, 1973; *Magyar*, 1974]. This section will present the information necessary for understanding the specific problems common to flashlamp-pumped dye lasers.

There are many ways in which the excited dye molecule may return to the ground state. Some of these paths are indicated in the energy level diagram of Figure 4.1; the electronic ground state of the molecule is S_0, and S_1 and S_2 represent the first and second excited singlet states. Transitions between singlet states are spin allowed and hence correspond to strong absorption bands. The molar extinction coefficient, ε, is a measure of the strength of an absorption band and is defined by the relation

$$\varepsilon = \frac{\ln (I_0/I)}{NL} \tag{4.1}$$

where

N = concentration of dye molecules

L = dye solution path length

I_0, I = incident and transmitted light intensity

The dye used for this experiment, Rhodamine 6G Perchlorate, has a maximum molar extinction coefficient of 11.2×10^4 l mol^{-1} cm^{-1}.

The lasing process begins with the excitation of the dye molecule from the lowest levels of the ground state S_0 to the higher vibrational-rotational levels of the S_1 state ($A \rightarrow b$ transition). The molecule decays in a non-radiative process to the lowest level of the S_1 state from which it returns to the higher levels of the S_0 state via stimulated emission ($B \rightarrow a$ transition).

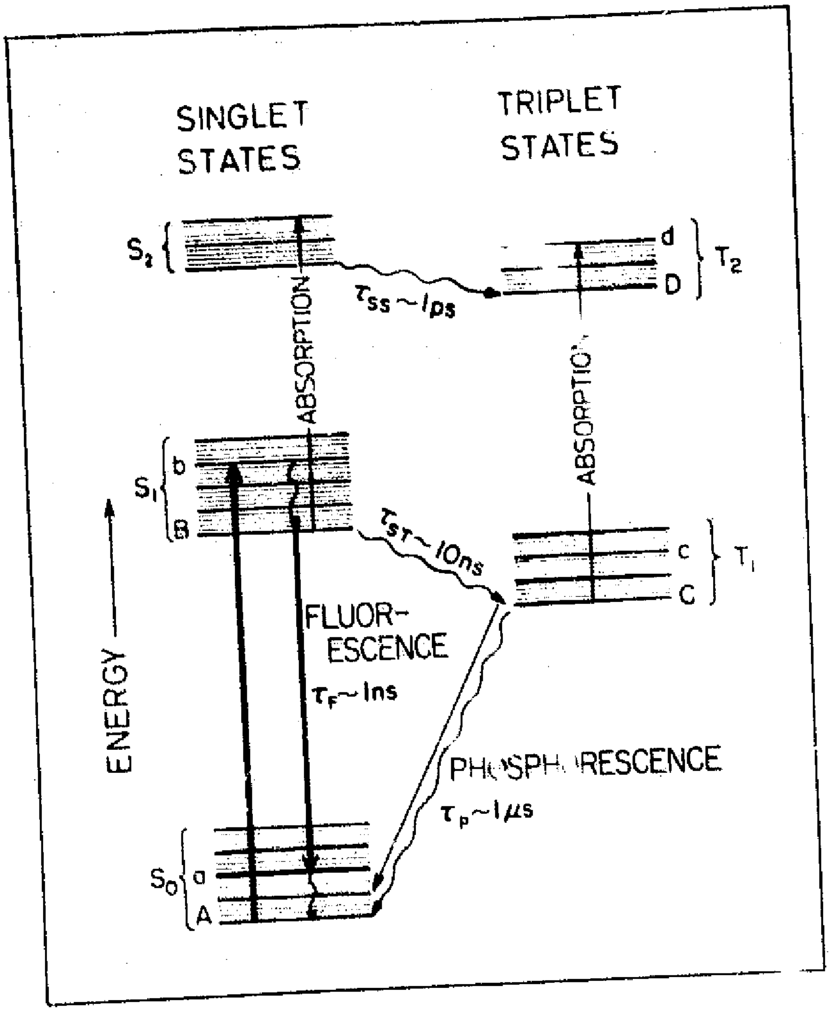


Figure 4.1 Eigenstates of a typical dye molecule with radiative (straight lines) and non-radiative (wavy lines) transitions. Excitation and lasing transitions are represented by the paths A→b and B→a respectively.

ORIGINAL PAGE IS
OF POOR QUALITY

4.1.2 *Loss processes and triplet formation.* To facilitate lasing, losses from the S_1 state cannot be excessive. The term, critical inversion, is used to describe the concentration of molecules in the S_1 excited state required to produce lasing. Some of the loss processes for flashlamp-pumped dye lasers are described below.

Spontaneous emission from the S_1 to S_0 state (B → a transition) is known as fluorescence. The time required to decay by this process is on the order of 1-5 ns. Also since the molecule decays non-radiatively prior to fluorescence, the fluorescence maximum is shifted to the longer wavelengths in the Anti-Stokes region. This shift is important because the unexcited dye will not absorb the spontaneous emission. In Figure 4.2 the absorption and fluorescence bands are drawn as practically continuous over a band of frequencies. Because the dye molecule has about 50 atoms and hence about 150 normal vibrations, collisional and electrostatic perturbations caused by the surrounding solvent molecules broaden these lines, resulting in a quasicontinuum of lines superimposed on each electronic state. The result is a mirror image of the absorption band displaced toward lower wave numbers by a reflection at the wave number of the purely electronic transition.

Intersystem crossing is another process by which molecules may be removed from the lasing action. In this case a molecule in the S_1 state may relax to a lower lying triplet state, T_1. The transition is forbidden but may be collisionally assisted. If it is a radiative process, then it is called phosphorescence. Since the time spent in the triplet state is relatively long, molecules which decay to this state are effectively trapped and removed from the lasing process.

ORIGINAL PAGE IS
OF POOR QUALITY

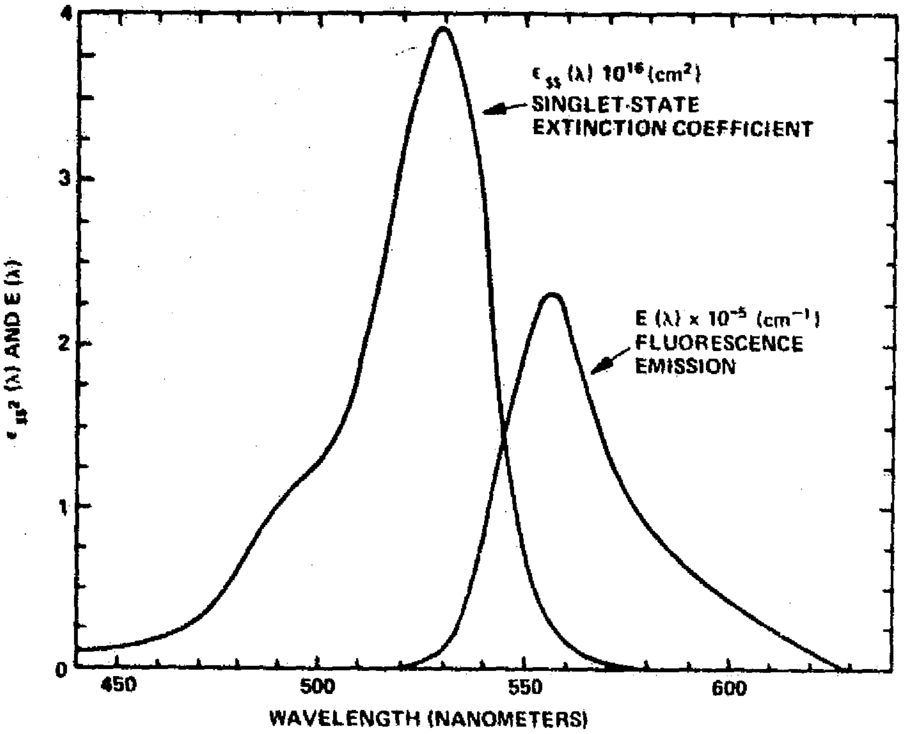


Figure 4.2 Singlet-state absorption and fluorescence spectra for a 10^{-4} molar ethanol solution of Rhodamine 6G [*Snavely*, 1969].

To compound the problem, transitions between triplet states are spin allowed. Hence these transitions can form absorption bands which may overlap the spontaneous singlet state fluorescence emission. Absorption between singlet states such as S_1 and S_2 is present but is not an important loss process, since the population in the S_1 state is relatively small. To minimize the effects of the triplet state formation, a pump source whose intensity rises rapidly with time is required. Flashlamp pumping is an accepted dye excitation method.

4.1.3 *Solvent and quenching agent.* The processes discussed so far have been related to the molecular structure of the dye molecule. However the environment of the molecule, i.e., solvent, is very important. For example dimerization of the molecule may provide competing quenching processes. The dimers usually have a strong absorption band at shorter wavelengths and an additional weaker band at the long wavelength side of the monomer band. One way to minimize dimerization, is to use a less polar solvent such as ethanol. For safety reasons the solvent was chosen to be half water and half ethanol [*Grant and Hawley*, 1975]. The equilibrium between monomers and dimers shifts toward the latter with increasing dye concentration and decreasing temperature; and at room temperatures, a concentration of 5×10^{-4} M Rhodamine 6 G may stop lasing action completely. An optimum concentration for the laser used in this study is 5×10^{-5} M.

Quenching agents may be added to the dye solution to reduce triplet formation [*Marling et al.*, 1970]. The quenching agent provides a lower lying triplet state than the dye molecule, facilitating a triplet-triplet decay which then effectively increases the intersystem crossing rate between the T_1 and S_0 states. The condensed phase meterial, 1, 3, 5, 7-cyclooctatetraene (COT) in concentrations of 5×10^{-3} M has performed quite adequately [*Pappalardo et al.*, 1970].

4.2 *Laser Mainframe*

4.2.1 *Layout*. The laser and associated monitoring equipment were designed and built to monitor the sodium layer as well as to have sufficient versatility to investigate transmitter options and modifications. Among the factors which determined the physical layout were financial constraints, possible inclusion of the transmitter into a mobile lidar system and space limitations. Although some design choices were not optimal, the resulting system is adequate for the first basic scientific experiments and has allowed us to define precisely the areas of development needed to accomplish future system improvements.

The flashlamp-pumped dye laser is shown in Figure 4.3. The optical bench consists of a 2.5 m I-beam (1) supported by two containers of sand (2) which serve to isolate acoustically the I-beam from the floor. Two plates of 1/4" steel were bolted to the I-beam. The top plate was tapped with 1/4" - 20 holes and the bottom plate became a spacing washer having clearance holes aligned with the holes of the top plate.

Effort was made to avoid mechanical vibrations from directly shaking the I-beam. As a result the flashlamp capacitor is mounted independently of the I-beam with its terminals straddling the I-beam. The flashlamp assembly is bolted to a frame which bridges the I-beam but does not touch it. The dye cell mount is clamped to the I-beam and does not touch the flashlamp assembly; all optics are similarly bolted to the I-beam. Tubing for the flashlamp cooling, dye flow, and vacuum systems is attached to the side of the I-beam for convenience.

Sampling of the laser beam is done by a light-activated computer trigger pulse generator, spectrometer, relative power indicator and sodium

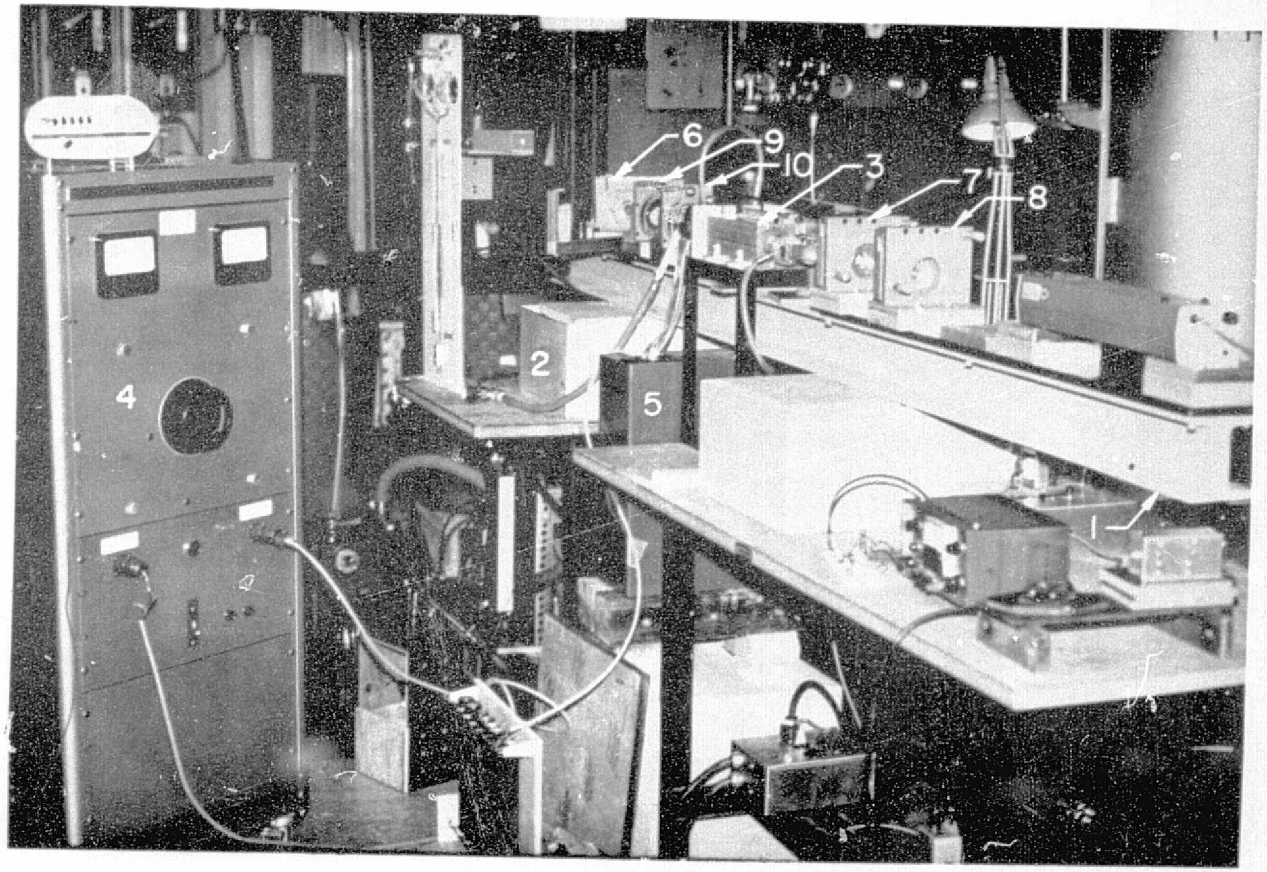


Figure 4.3 Flashlamp pump dye laser at the Aeronomy Laboratory Field Station.

ORIGINAL PAGE IS
OF POOR QUALITY

cell monitor. All of these items are placed on various tables surrounding the laser and beam splitters are used to direct the light into the appropriate directions. A pictorial display of the laser transmitter system is shown in Figure 4.4.

4.2.2 *Elliptical cavity*

4.2.2.1 *General consideration.* Most flashlamp-pumped dye lasers employ an elliptical cavity for focusing the excitation light onto the dye cell. The elliptical cavity is usually made from aluminum and the machined surface is polished mechanically or by hand. The cavity may be made on a lathe or milling machine by feeding the material at an angle to the cutting tool as indicated in Figure 4.5.

As a general rule, the eccentricity of the resulting cavity should be as small as possible [*Röss*, 1969]. The ray tracing diagrams of Figure 4.6 indicate the projection distortion that occurs for ellipsoids of high eccentricity. To minimize projection distortion, the dye cell diameter should be small compared to the semi-axis of the ellipse. Associated with this rule is the idea that the homogeneous illumination is limited to a small region which is dependent on the size of the source (Figure 4.7a). Assuming a two-dimensional analysis, we see that the boundaries of the homogeneously illuminated area, A, are given by

$$\frac{r_1}{r} = \frac{a - c}{a + c} = \frac{1 - e}{1 + e} \tag{4.2a}$$

and

$$\frac{r_2}{r} = \frac{a^2 - c^2}{a^2 + c^2} = \frac{1 - e^2}{1 + e^2} \tag{4.2b}$$

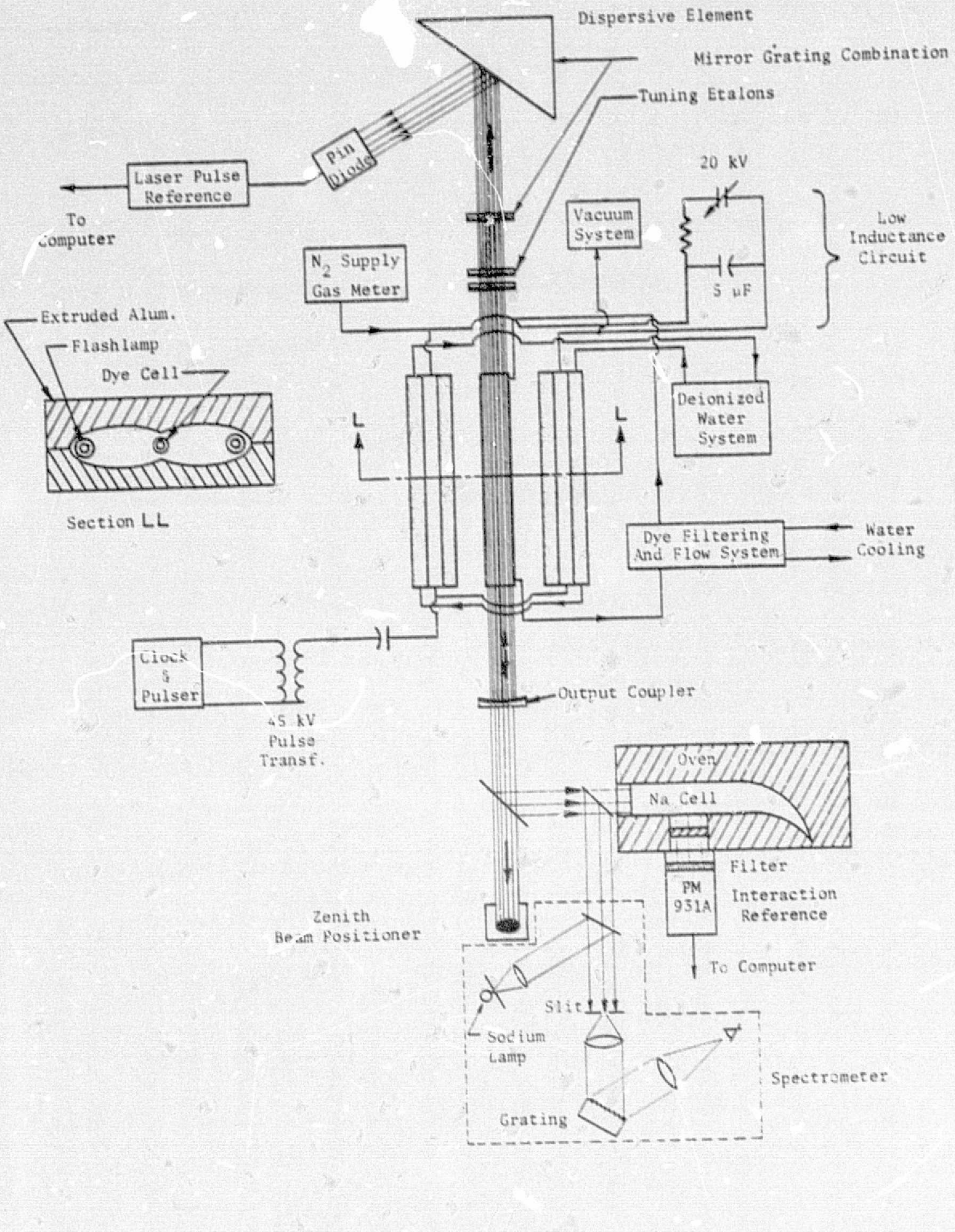


Figure 4.4 Laser transmitting system.

ORIGINAL
OF POOR QUALITY

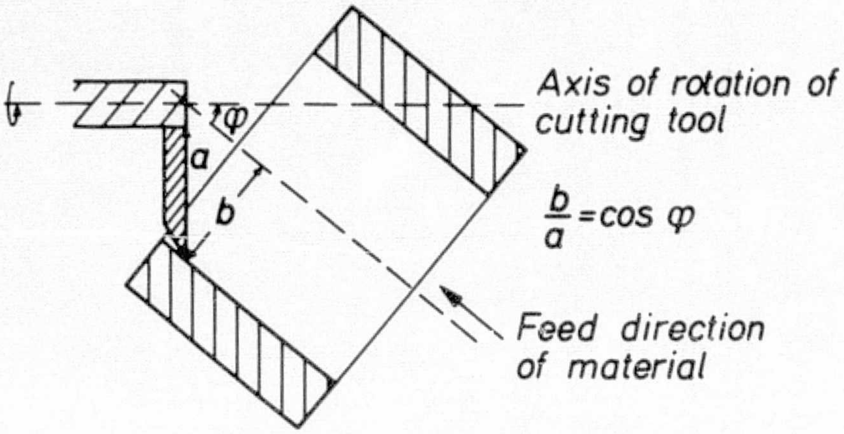
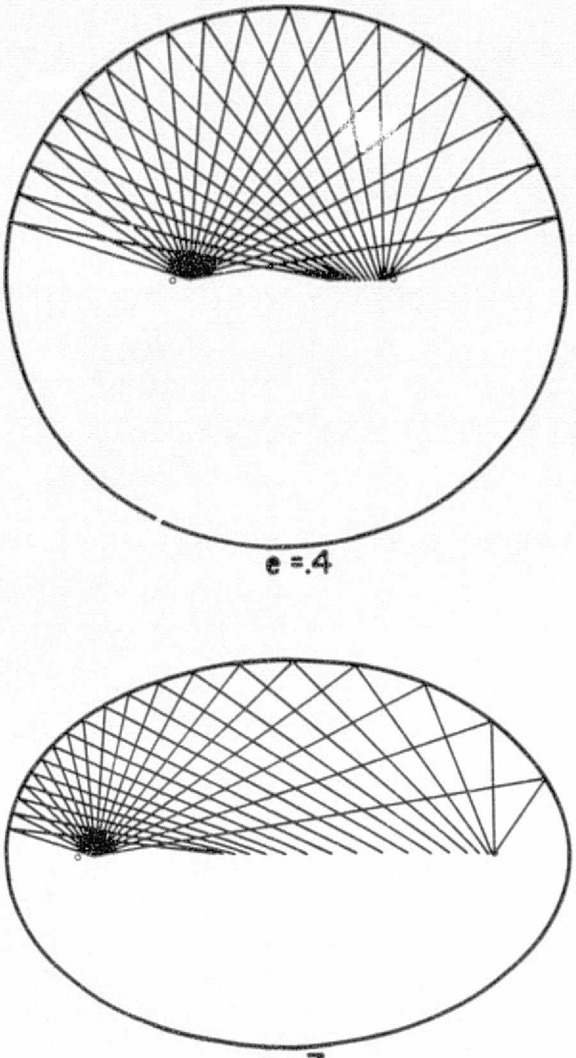


Figure 4.5 Production of an exactly elliptical cylinder on a lathe or milling machine; ϕ is the angle between the axis of rotation and the feed direction. With long cylinders, the mirrors are made in two halves because the axis of the cutting tool would otherwise interfere [*Röss*, 1969].



e =.4

e =.7

Figure 4.6 Projection distortion for an elliptical cavity. The major axis of each ellipse is the same and the source of rays is displaced equally.

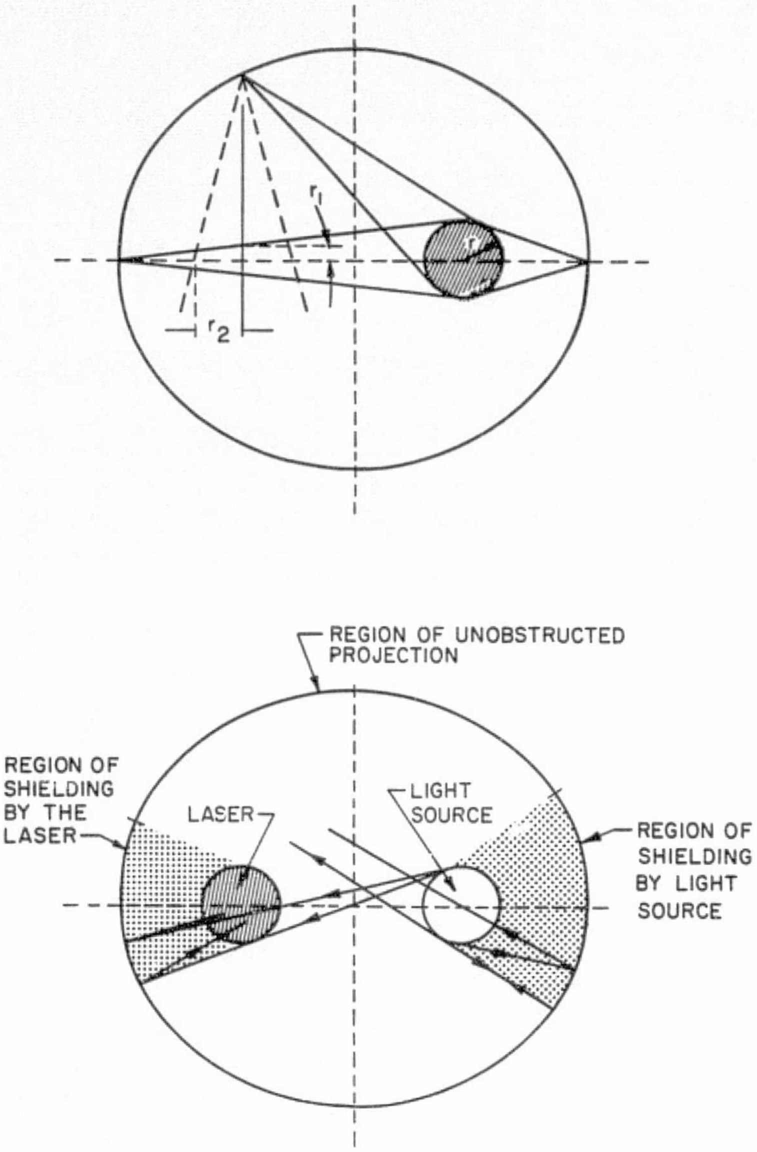


Figure 4.7 (a) Extreme cases of distorted reflection.
(b) Shadowing in elliptical cavity mirrors [*Röss*, 1969].

ORIGINAL PAGE IS
OF POOR QUALITY

where

$$e = \frac{c}{a} = \text{eccentricity of the ellipse}$$

We obtain the approximation

$$\frac{A}{\pi r^2} = \frac{\pi r_1 r_2}{\pi r^2} = \frac{(1 - e)^2}{(1 + e)^2} \tag{4.3}$$

Hence if $e = 0.4$, then the homogeneous illuminated area is about 30% of the source cross sectional area.

Several other properties of elliptical cavities become evident. The length of the cavity should be comparable to or larger than the dimension of the semi-axis. Hence multiple reflections resulting from end reflections are avoided. Direct light becomes more important as the flashlamp is brought closer to the dye cell; this occurs for elliptical cavities having small eccentricities. However under these conditions the dye cell would shade part of the cavity wall, resulting in an uneven excitation of the dye (Figure 4.7b). Furthermore when flashlamps are brought as close as 1-2 cm from the dye cell, flashlamp shock waves may create gradients in the index of refraction of the dye; the resulting spurious deflection of laser light passing through the dye cell then degrades tuning performance.

Several problems are often overcome by the use of multiple elliptical cavities. It is important to remember that the energy density produced in the dye cell region is dependent on the beam angle of the flashlamp. In any case the beam angle of all flashlamps adds up to a constant 2π. Hence on first principles, the efficiency of multiple elliptical cavities should decrease as the number of lamps is increased; the maximum attainable laser output should remain the same. This however is not observed experimentally.

What occurs is a larger homogeneous illumination area for multiple elliptical cavities. Furthermore gradient effects in the dye become more symmetrical as the number of ellipses increases, resulting in a more stable tuning arrangement. As a compromise between maintenance cost and improved laser output, most groups have chosen a double elliptical cavity. In the extreme case, a cavity having many elliptical components approaches the direct light situation and a helical or coaxial flashlamp would then provide greater symmetry.

4.2.2.2 *Conjoining of the vacuum, dye flow, and flashlamp cooling systems*. The elliptical cavity is central to the operation of the laser; and as a consequence, it becomes the interaction area for a number of laser subsystems. We discuss these subsystems in this section but postpone the construction details of the elliptical cavity until Section 4.2.3.1.

The flashlamps are an ablating version and as such must be constantly evacuated. The pump and vacuum system is extremely simple in design (Figure 4.8a). A Welch 1402B pump (140 l/min) is used to evacuate the system and dry nitrogen is leaked into the system through a needle valve. Pressure is monitored on the pump side of the flashlamps by a manometer; typical pressure for the lamps is in the 4-8 torr region depending on the life of the flashlamps. A trap consisting of a flask is placed between the manometer and the flashlamps to prevent water from entering the vacuum pump or manometer in the unlikely event that mixing occurs between the water cooling and vacuum systems where they are coaxial. In addition the trap collects much of the quartz dust before it reaches the manometer or pump.

The dye flow system consists of a magnetically coupled stainless steel pump (Micropump Corp. #12-60-303-374), 5 μm filter (Filterrite Corp. # U001A4UCE), dye cell and assorted Imperial Eastman *Poly Flo* tubing (Figure 4.8b). The emphasis of the design was to provide a fast dye flow,

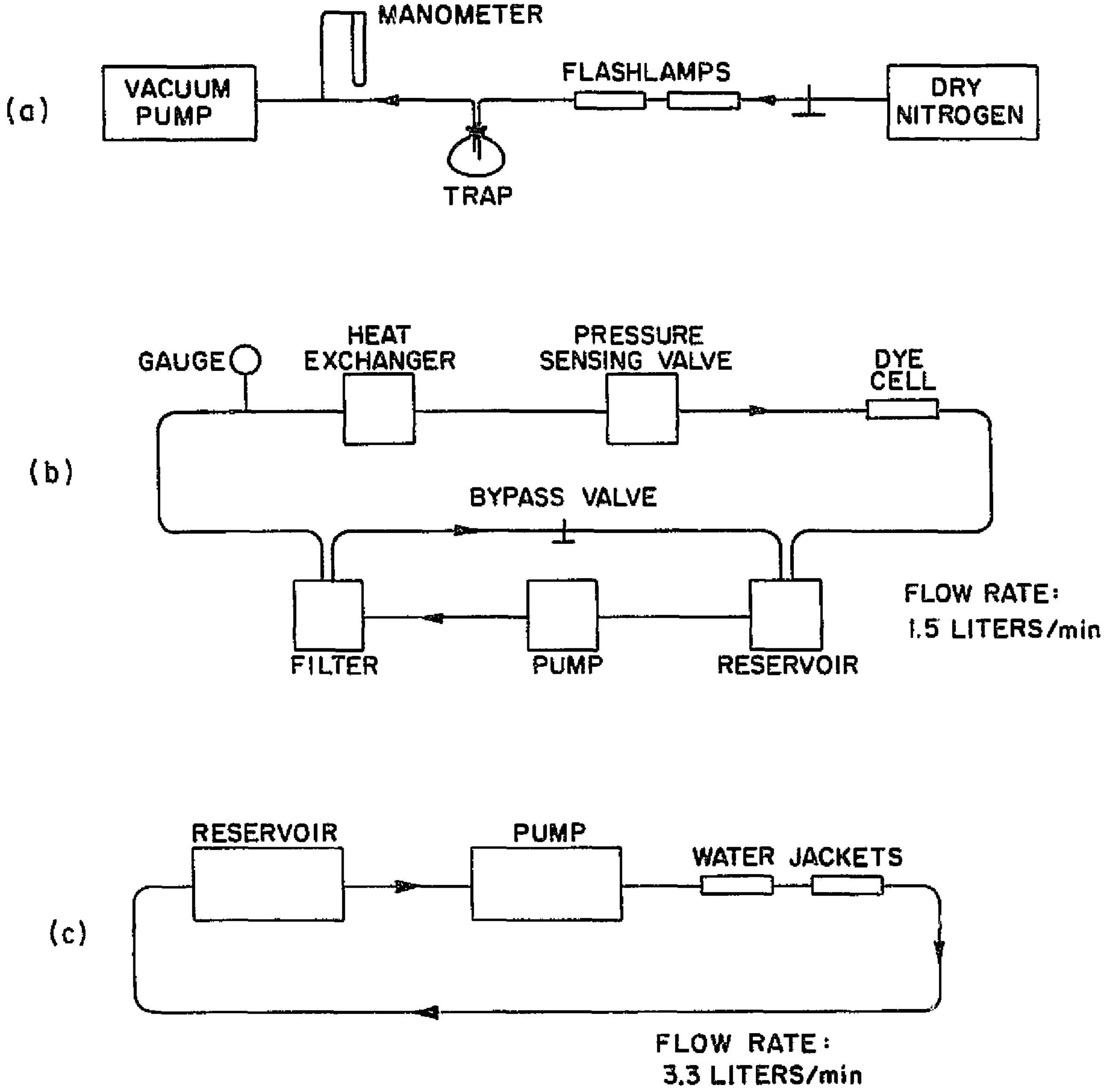


Figure 4.8 Laser subsystems (a) vacuum system, (b) dye flow system, (c) flashlamp cooling system.

to reduce bubbles to a minimum, to build the system from components impervious to a wide selection of solvents and to have the option of jet stream operation should laser pumping be employed. The pump system has a pressure-activated relay which will turn the pump off when a break occurs in the dye system. If a fire results from a ruptured dye flow system, this feature prevents more of the combustible solvent from being added to the fire. Large volumes of dye are needed to minimize the loss in power due to photolysis of the dye, and hence the system employs a stainless steel tub which increases the system capacity to about 20 liters. The dye cell consists of two detachable quartz windows (15 mm diameter × 10 mm thickness) and a 10 mm ID × 280 mm pyrex tube. These windows are sealed with O-rings as indicated in Figure 4.10b. Thick quartz windows were chosen since we observed etalon effects with thinner windows [E. S. Richter in *Edwards*, 1977]. With our present window holder arrangement, approximately 6 cm of dye is not excited while 26 cm length is excited. Hence some loss is incurred in the dye cell.

The flashlamp cooling system consists of an 80-liter reservoir of deionized water, magnetically-coupled pump to minimize bubble formation (Micropump #10-41-316), flashlamp water jacket, and assorted hosing (Figure 4.8c). Since the conductivity of the cooling water increases with laser use, the water is deionized prior to each run by using a Barnestead deionizer (DS041 Mixed Resin Cartridge). Since portions of the flashlamp electrodes are immersed in this cooling water, it is imperative that the resistivity of the water is kept high to prevent a short circuiting of the lamp discharge path by the water.

4.2.3 *Optical pump*

4.2.3.1 *Flashlamp mount and assembly*. The elliptical cavity assembly is patterned after that of *Gibson* [1972] and is shown in Figures 4.9 and 4.10. The double elliptical cavity was used originally in a ruby laser; and although it does not satisfy the criterion discussed in Section 4.2.2.1, it is adequate for first results. The eccentricity of each ellipse is 0.7 (major axis = 75 mm, minor axis = 50 mm) and the surface is marred in many spots from flashlamp breakage during operation. The flashlamps are held by lucite support pieces and are allowed to move slightly because of the O-ring seals. This freedom of motion is advantageous since the flashlamps undergo shock after each pulse. Lucite was chosen for the pieces which come in contact with water since lucite does not absorb water. All other pieces were fabricated from bakelite. Initially there was a problem with an arc which originated at the high voltage electrode and traveled down the cooling jacket until it completed its path to the grounded elliptical cavity. This was overcome by placing teflon tape around the water jacket at the appropriate location, and replacing the high-voltage end of the cavity with a bakelite-teflon assembly.

The flashlamps are a variation of the design of *Ferrar* [1969]. These ablating flashlamps generally radiate 15-20% of the electrical input energy into the 200-400 nm range and may be considered as 20,000-30,000°K blackbody radiators for this wavelength range. The lamps consist of a 5 mm ID × 10 mm OD × 280 mm quartz tube to which two brass electrodes are connected as shown in Figure 4.10 and 4.11. The electrodes are not symmetric because the walls of the elliptical cavity are of different thickness.

Brass was chosen as a suitable electrode material because brass tubing was readily available, easy to machine, and did not sputter excessively. The electrodes will clog eventually with glass deposits in spite of the nitrogen atmosphere, but the lamps are replaced well before the constriction is complete. A piece of metal tape (Scotch #5524) provides a mechanical

ORIGINAL PAGE IS
OF POOR QUALITY

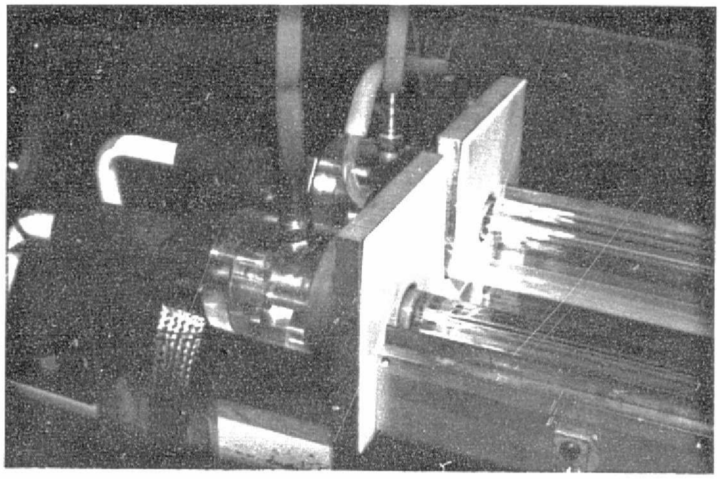


Figure 4.9 Photograph of the high voltage end of the elliptical cavity.

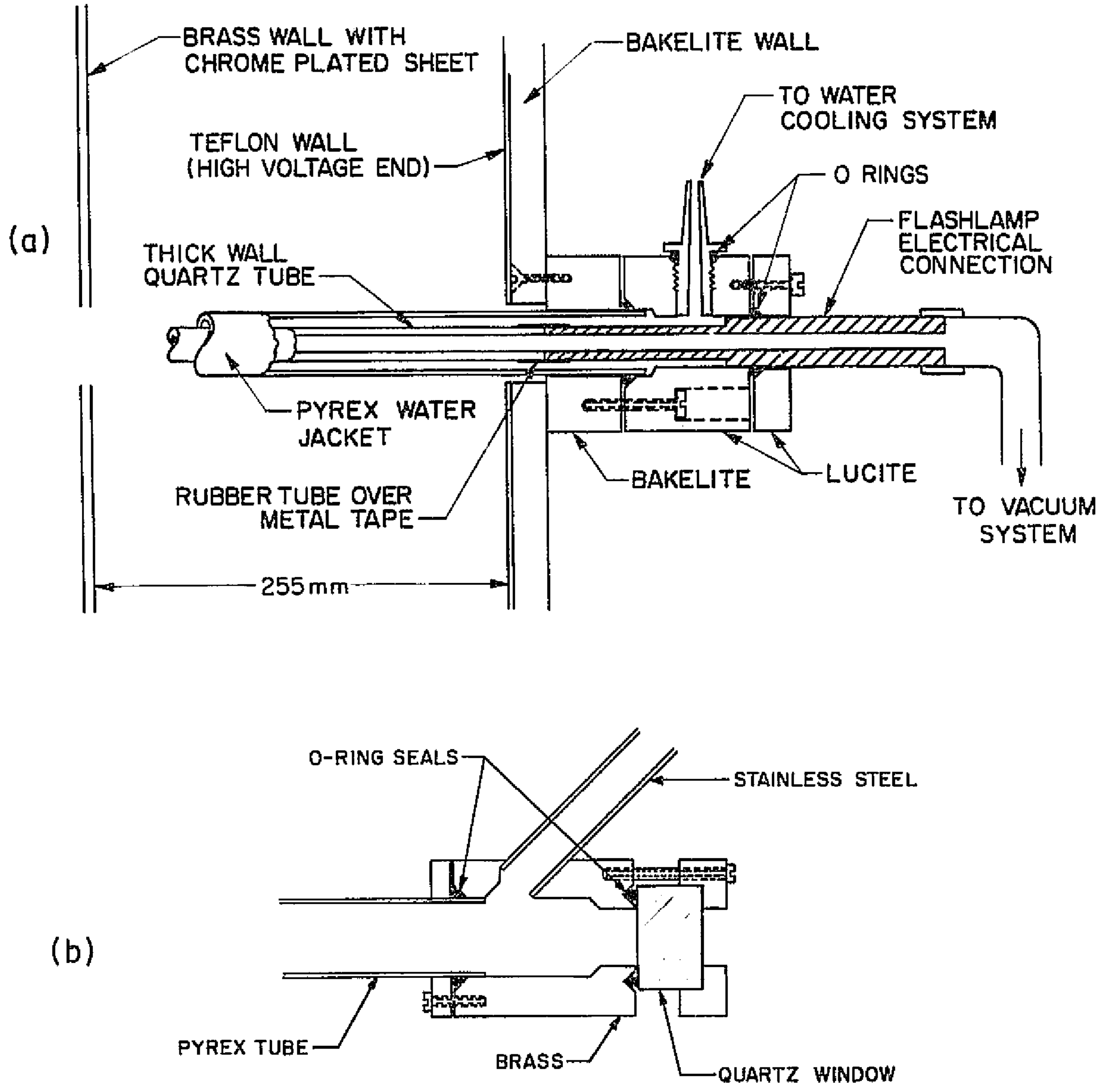


Figure 4.10 (a) Elliptical cavity cross section showing the seating of the flashlamp (1/2 scale); (b) Dye cell window holder (to scale).

ORIGINAL PAGE IS
OF POOR QUALITY

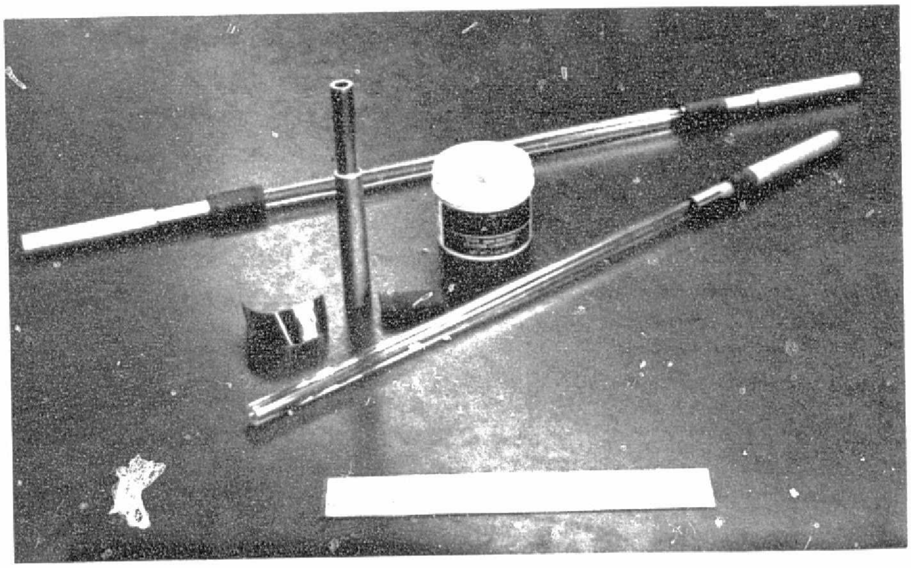


Figure 4.11 Flashlamp assembly.

seal between the quartz and brass and prevents a blast from exiting at this junction. Gum rubber tubing (5/16 ID) wetted with silicone grease (Dow Corning #970V) is placed over this metal tape to provide a vacuum seal. Lamp assemblies last about 10,000 shots and are entirely replaced at that time; laser design insured that flashlamps replacement did not upset the optical alignment.

4.2.3.2 *Triggering and discharge circuitry.* The series combination of flashlamps is powered by a 19 kV, 0.5 A power supply. This power supply charges a 5μF/24kV low-inductance capacitor (Maxwell Laboratories, Series C, #33017) to 18.5-19 kV. The capacitor was chosen using the criterion of *Hayworth* [1971]. Once the capacitor is charged to the desired voltage, a pulser ionizes the gas in one of the lamps causing both lamps to discharge the capacitor rapidly.

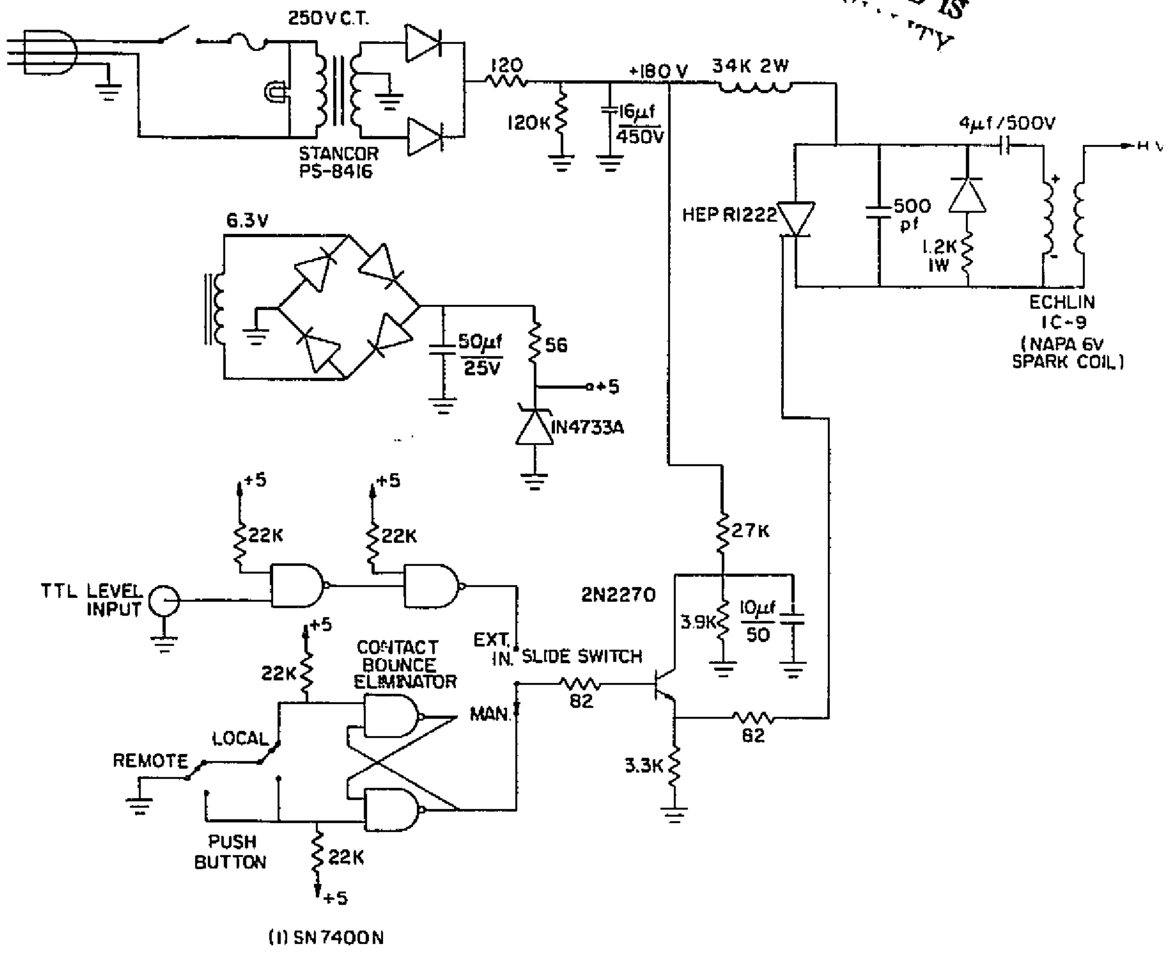
The pulser incorporates an SCR circuit which drives an automobile spark coil (Figure 4.12a). Inputs to the pulser may be either from a timing circuit which generates a TTL level pulse every 1-5 sec (variable) or from a manual input. The resulting 15 kV output pulse is passed through a 500pF/30kV D.C. blocking capacitor to the common junction of the flashlamps (Figure 4.12b). Generally some adjustment in lamp pressure is required to obtain consistent triggering; but once this is accomplished, the laser will trigger faithfully throughout a run.

4.3 *Tuning*

4.3.1 *Mirror-grating combination.* In the early stages of the project a grating was used for crude tuning of the dye laser. It was found that the grating burned easily and ways were sought to minimize this. The present system employs a mirror-grating combination (MGC) [*Bjorkholm et al.*, 1971] and uses a high damage threshold grating (PTR optics #UTF-26, 1800 g/mm).

ORIGINAL PAGE IS
OF POOR QUALITY

(a)



(b)

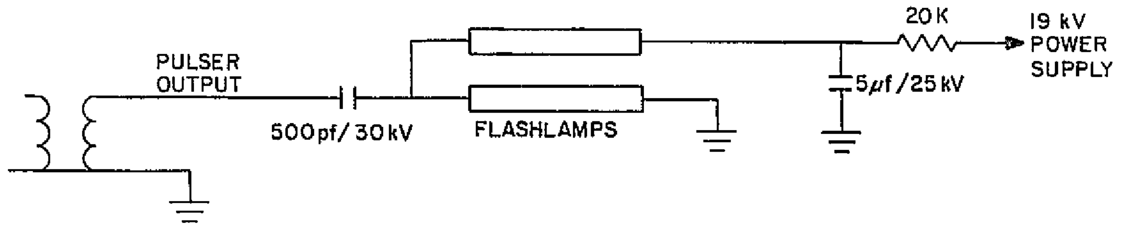


Figure 4.12 (a) Spark gap initiator; (b) flashlamp electrical circuit.

The grating is mounted in auto collimation (Littrow arrangement), and for this application the grating equation is expressed by

$$n\lambda = 2d \sin \theta \tag{4.4}$$

where

n = order of diffraction

λ = wavelength

d = groove spacing

θ = angle of diffraction

The incident beam having a center wavelength λ and linewidth $\Delta\lambda$ is diffracted into a multiplicity of discrete beams leaving the grating at different angles. Each of these beams is a spectral replica of the incident beam. Because the grating is blazed for the first order, most of the energy is concentrated in this order. The term echellette is sometimes used to describe a grating blazed for low orders ($n = 1$ or 2).

We chose a grating for coarse tuning because of its simplicity of use, availability, and relatively low cost. The grating was mounted such that all orders were parallel to the optical bench. This can be easily verified with a He-Ne laser by rotating the grating about an axis parallel to the laser beam until the laser beam is reflected back on itself for each order. Each order may be checked by rotating the grating about an axis perpendicular to the bench beam until the condition is met. Three strong orders are observed with our grating, corresponding to $n = 1, 0, -1$; the $n = 1$ order is easily resolved since the power in the $n = 1$ order is greater than the power in the $n = -1$ order. The order $n = 0$ of course corresponds to specular reflection.

Several problems exist with the use of gratings. Cleaning of gratings is virtually impossible; and in the dusty environment of the Aeronomy

Laboratory Field Station, if dust is not blown off the grating with dry nitrogen, the small particles will absorb light and the resulting heat will destroy the reflective surface. In addition the reflectivity of the aluminized surface is about 90% and hence the surface is constantly absorbing heat; oxidation of the surface compounds the problem. Eventually the metal surface is evaporated and localized heating causes a warping of the plastic backing of the replica grating. Thus the groove configuration is eventually destroyed and tuning suffers.

To minimize the grating incident energy, we adopted an MGC as shown schematically in Figure 4.13. The MGC may be considered as an etalon which has one mirror strongly reflective over a narrow wavelength range. Within this range the peak reflectivity of the MGC is given by

$$R_{\max} = \frac{\sqrt{R_1} + \sqrt{R_2}}{1 + \sqrt{R_1 R_2}} \tag{4.5}$$

which is the expression for an etalon in antiresonance. As Figure 4.14 indicates, the MGC can actually reduce cavity losses because the reflectivity of the MGC is higher than that of the grating.

In addition the MGC reduces the incident power on the grating by a factor, η, where

$$\eta = \frac{1 - R_1}{[1 + \sqrt{R_1 R_2}]^2} \tag{4.6}$$

For $R_1 = 0.44$ and $R_2 = 0.8$, equations (4.5) and (4.6) imply that $R_{\max} = 0.96$ and $\eta = 22\%$. These values are appropriate for the present tuning configuration. It was found for $R_1 > 0.5$ that the laser would lase off the partial reflector with the grating having no tuning effect.

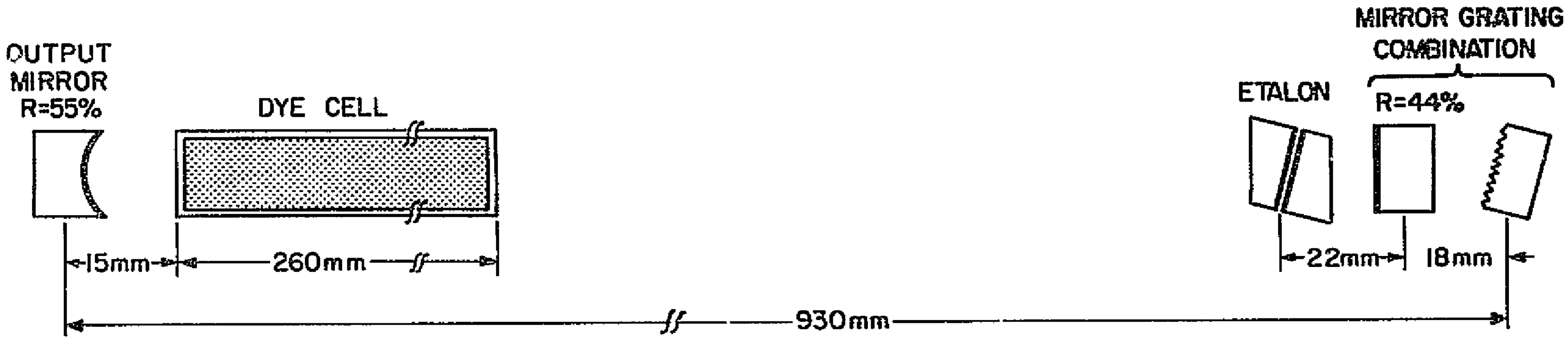


Figure 4.13 Optical cavity presently used in the dye laser.

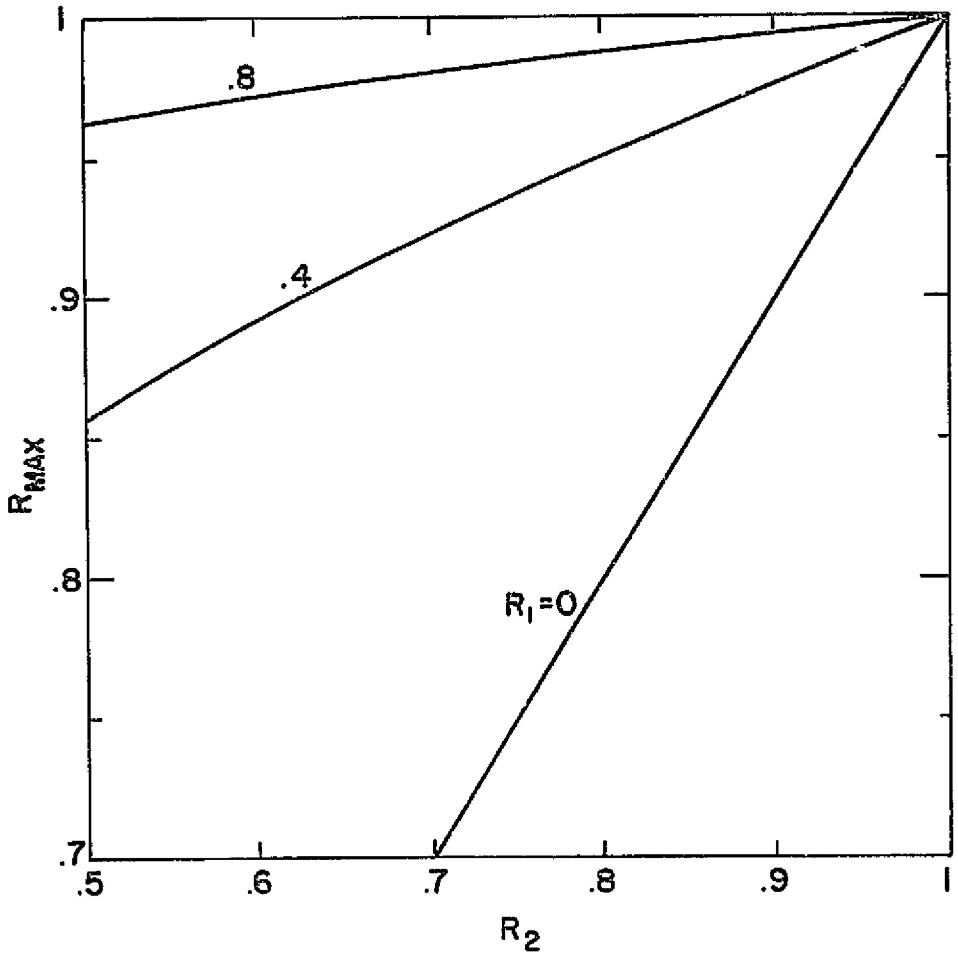


Figure 4.14 Curves of R_{max} for a mirror-grating combination (MGC) when considered as an etalon in antiresonance. The reflectivity of the grating is denoted by R_2 while R_1 represents the reflectivity of the mirror.

4.3.2 *Etalon tuning*. The use of etalons for tuning high energy dye lasers has become an accepted technique. An etalon is a Fabry-Perot interferometer whose gap is held fixed by pressing the interferometer plates against a rigid spacer. For our application the medium between the plates is air; and in the ensuing calculations, the index of refraction for air will be assumed to be 1.

The etalon may be considered as a filter which has maximum transmission at wavelengths λ_n satisfying

$$n\lambda_n = 2d \cos \theta \tag{4.7}$$

where

n = an integer

d = spacing between the plates

θ = angle of tilt of the etalon to the incident beam

The integer n is termed the order of the interference; and because the optical wavelength is much smaller than the physical separation of the plates, n is usually on the order of 10^4 to 10^5.

Suppose plane waves of equal intensity are incident over a range of angles, and the transmitted light is collected by a lens as shown in Figure 4.15. There will appear on the screen fringes of equal inclination in the focal plane of L, following loci of constant θ when the order of interference has integral values.

Minima will occur when n has half integral values. Considering two adjacent fringes for the same wavelength we write

$$n\lambda_1 = 2d \cos \theta_1 \tag{4.8}$$

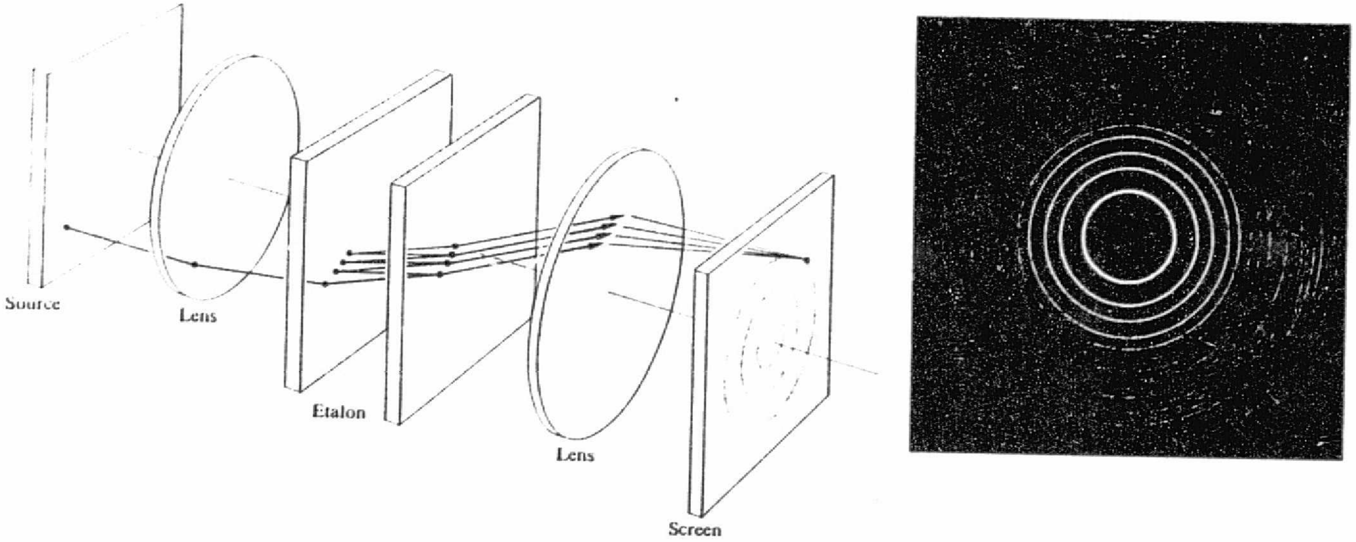


Figure 4.15 Fabry-Perot etalon [after *Hecht and Zajac*, 1975].

ORIGINAL PAGE IS
OF POOR QUALITY

and

$$(n - 1)\lambda_1 = 2d \cos \theta_2 \tag{4.9}$$

If we consider a spectral component λ_2 close to λ_1 such that this component in order n falls on the order $n - 1$ of λ_1, then for $\Delta\lambda = \lambda_1 - \lambda_2$

$$2d \cos \theta_2 = n(\lambda_1 - \Delta\lambda) \tag{4.10}$$

From equations (4.9) and (4.10) we find

$$\lambda_1 = n \, \Delta\lambda \tag{4.11}$$

Substituting the value for n obtained from (4.8) into (4.11) we obtain

$$\Delta\lambda = \frac{\lambda_1{}^2}{2d \cos \theta_1} \tag{4.12}$$

for θ_1 small. This is the wavelength interval between fringes of adjacent orders for the same wavelength. It is the spacing between successive maxima of transmission and is called the free spectral range (F.S.R.) of the etalon.

A measure of the sharpness of the fringes is given by the finesse, F, of the etalon. By definition the bandwidth of the fringe is given by

$$\text{B.W.} = \frac{\text{F.S.R.}}{F} \tag{4.13}$$

and the relation between these quantities is shown diagrammatically in Figure 4.16. The finesse can be thought of as the effective number of interfering beams involved in forming the interference fringes. For our particular case the major factors which limit the net finesse are mirror reflectivity and lack of parallelism and/or planeness of the mirrors. The net finesse may be calculated as

$$F^{-2} = \sum_i F_i{}^{-2} \tag{4.14}$$

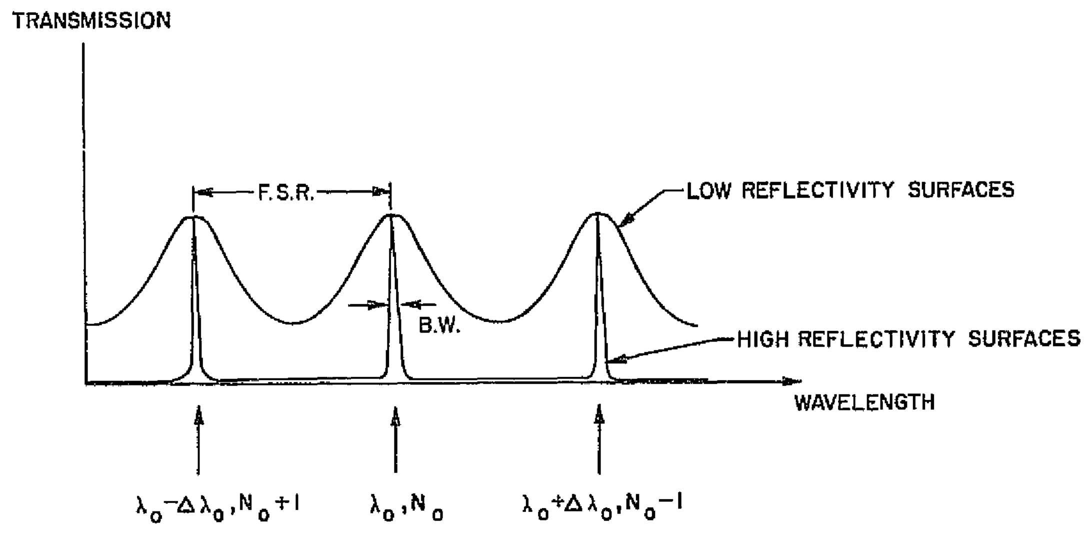


Figure 4.16 Transmission of an etalon as a function of wavelength and the relation between free spectral range and bandwidth.

ORIGINAL PAGE IS
OF POOR QUALITY

where the F_i represent the component finesses.

The reflective finesse is given by the expression

$$F_R = \frac{\pi\sqrt{R}}{1 - R} \tag{4.15}$$

where $R = \sqrt{R_1 R_2}$ is the geometric mean of the reflectivity of the two plates. If M is the fractional deviation from planeness across the mirror aperture, or the measure of the lack of parallelism of the mirrors within the aperture being used, the flatness of figure finesse is given by

$$F_F = M/2 \tag{4.16}$$

for a λ/M plate. This plate flatness is normally specified at 500 nm.

In the above discussion on bandwidth, the divergence of the laser has not been taken in account. If we consider only a plane parallel beam, the bandwidth of the lines transmitted by the etalon is given by

$$\text{B.W.} = \frac{\lambda}{nF} \tag{4.17}$$

However a contribution to the bandwidth also comes from the spread $\Delta\theta$ in the angle of the laser beam. If we take the derivative of (4.8) and consider only magnitudes of the quantities involved, we obtain

$$\frac{d\lambda}{\lambda} = \tan\theta \, d\theta \tag{4.18}$$

For θ small we may therefore write

$$\Delta\lambda = \lambda\theta\Delta\theta \tag{4.19}$$

To prevent light reflected from the etalon from feeding back through the gain medium, we require $\theta > \Delta\theta/2$. Hence for a value of $\Delta\theta = 5$ mrad, the theoretical minimum value of the bandwidth is 7.4 pm. Realistically since the laser is nonlinearly sensitive to losses, the linewidth will vary with the pump level and may be much smaller than the computed value.

Several practical choices should be made when specifying plates. The overall finesse of the etalon should not be excessively high. For flashlamp-pumped dye lasers the finesse of each etalon should be about 5. Standard interferometer plates have a flatness of λ/20; and if the plates are as large as 1.5" in diameter, then over the area of the laser beam the flatness specification will be better than λ/50 which is adequate. Plates should be wedged to avoid etalon effects from the outside surfaces, and these outside surfaces may be anti-reflection (A.R.) coated. Highly reflective plates will cause excessive walk-off losses because the plates are not perfectly flat. Reflectivities of 30% and 50% were found to be quite satisfactory, but 85% coatings were unsatisfactory unless the plates were exceptionally flat. Coatings should be able to withstand the energy of the laser beam; and since the etalon forms a resonant cavity, the energy density will be higher within the etalon than outside of it. We have found electron beam deposited dielectric coatings to be satisfactory but have burned some of the more traditional coatings.

Etalon holders were fabricated from invar and are shown in Figure 4.17. Spacers were made from either thin microscope slides or pieces of aluminum foil. For slightly larger gaps we understand that shafts of drills or bearing balls are quite acceptable. Temperature variation through a night required slight adjustments about every hour indicating an oven housing is perhaps ultimately needed. Day-to-day variations in temperature and/or pressure did cause sufficient drift that careful tuning was required prior to each run.

It is important to remember that although these etalons are crude, they do work and cost perhaps 1/3 of a suitable commercial unit. One plate rests on a machined ridge while 3 screws are used to press the plates together. A

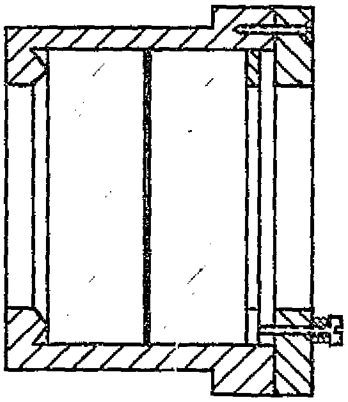


Figure 4.17 Etalon holder (to scale).

thin invar washer prevents direct contact between the screws and the plate.

For satisfactory performance it is important that the coated surface of the plates be parallel. By illuminating the etalon with a He-Ne laser and looking at the output on a screen, a train of dots will be seen. These result from multiple reflections from the misaligned plates. By adjusting the three screws, one can collapse these dots to a single spot.

For final alignment several methods exist. If one can illuminate the etalon with a large, collimated monochromatic beam, a couple of straight line fringes will be seen on the screen. Fine adjustment of the screws will yield even transmission across the aperture of the etalon indicating that the etalon is tuned to the wavelength of the incident beam.

However most of our etalons were aligned using an alternate technique. In this case the emission from a sodium lamp was passed through a sheet of ground (frosted) glass. If one views the light source through the etalon with the eye focused at infinity, a system of rings will be seen with the reflected image of the pupil of the eye in the center of the ring pattern. When the eye is moved in any direction, the ring pattern will also move while the eye pupil remains at the center of the pattern. If for example the ring pattern expands in size as the eye is moved to the right, the plates then are further apart at the right than at the left. Suitable adjustments of the alignment screws will result in an aligned etalon which is characterized by observing that the rings remain the same size as the eye is moved to any point in the field of view.

4.3.3 *Monitoring the spectral emission*

4.3.3.1 *Spectrometer.* A 5" divided circle spectrometer is used for monitoring the tuning operation, and a used laser grating having a dispersion of 0.5 nm/mrad @500 nm is used as the dispersive element. Part of the laser output and the focused output of an Osram sodium lamp are superimposed onto the spectrometer slits. In this way the sodium emission appears as a doublet, and the laser may be tuned to the vicinity of the D_2 line at which point the sodium cell monitor must be used.

The resolution of this spectrometer is such that the D lines appear to be spaced by about 2 mm. Hence since we know this spacing corresponds to 0.6 nm, we judge the linewidth of the laser to be about 0.2 nm when only an MGC is used. With an etalon we estimate the linewidth to be 10 pm. Because of the resolution of the present equipment, we feel these estimates are accurate to a factor of 3.

4.3.3.2 *Sodium resonance cell.* The sodium cell is simply designed, yet adequate for first measurements. A more elaborate cell is described in *Lory-Chanin* [1965]. The cell used in this work consists of a glass envelope and allows for either absorption or scattering measurements. Because the laser linewidth is much greater than the D_2 linewidth, the latter measurement has been used exclusively. Two sodium dispensers are spot-welded to three electrodes and assembly-mounted in the base of the cell. The cell's exhaust hole is then connected to a vacuum system and pumped down for about 3 days resulting in a pressure in the 10^{-6} - 10^{-5} torr range; the vacuum is limited by the pump system used. Outgassing of the glass may be accomplished by heating the cell while under evacuation. Finally the cell is sealed. Sodium is released by passing a current of a few amperes through either dispenser.

The cell is mounted in an oven built from fire bricks (Figure 4.18). A 1500 watt heater coil is placed under each window and the input power to the heaters is controlled by a variac. With this configuration the windows become the hottest part of the cell and hence sodium is less likely to condense on

ORIGINAL PAGE IS
OF POOR QUALITY

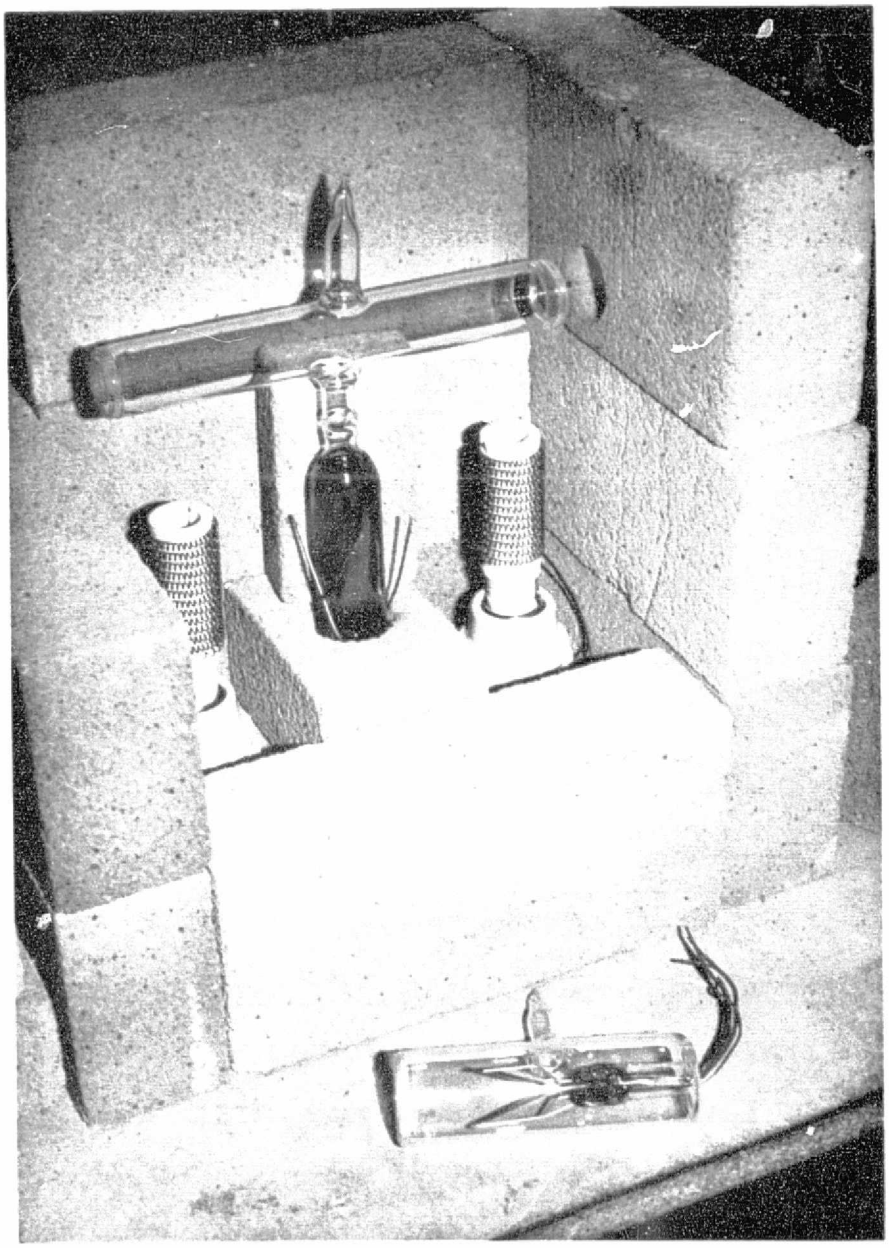


Figure 4.18 Sodium cell and oven.

the windows. Temperature is monitored by a mercury thermometer placed near the center of the cell and the equilibrium temperature is kept to 160 ± 1°C. This temperature allows for bright scattering without excessive secondary scattering.

It is instructive to calculate the temperature dependence of the scattering process to estimate the temperature control required. Since low pressures are involved, the line is mostly Doppler-broadened and equation (2.14a) applies, namely

$$k_0 = \frac{2}{\Delta\nu_D}\sqrt{\frac{\ln 2}{\pi}}\ \frac{\lambda_0^{\ 2}}{8\pi}\ \frac{g_1}{g_2}\ \frac{\rho}{\tau} \tag{4.20}$$

Employing the perfect gas law and taking $g_2/g_1 = 2$ one obtains

$$k_0 = 1.57 \times 10^{14}\ \frac{P}{T^{3/2}} \tag{4.21}$$

where k_0 has units m^{-1}, P has units of atmospheres, and T has units °K. We note then that for 160 ± 1°C, k_0 varies by less than 1%.

Equation (4.21) may also be applied to determining the length of the cell for $1/e$ attenuation. In this case $k_0\ l = 1$. If one knows the expected pressure for the evacuated cell, then by referring to the vapor pressure of sodium [*Sittig,* 1956] one may calculate l. Generally if the cell can be evacuated to 10^{-6} - 10^{-5} torr, significant attenuation will occur for $T > 150°C$. Practically a sodium lamp source may be used to calibrate the cell.

5. RECEIVER DESCRIPTION AND SYSTEM OPERATION

5.1 *Receiver Hardware*

The initial receiver incorporated the Aeronomy Laboratory Field Station PDP-15 computer. The first observations of sodium over Urbana were made with this system; and during the early experimental phase, several areas for system improvement were defined. This first lidar receiver is described in Appendix I. The present receiving system is based on a Hewlett-Packard 2114A computer and was designed for eventual inclusing in a mobile lidar system. In addition it affords more flexibility in a data collection.

The present receiving system consists of three basic subsections: a telescope with photomultiplier tube, a lidar system interface with memory, and a HP2114A minicomputer (Figure 5.1). Although a complete description of this system may be found in *Kinter* [1977], a general discussion of the system hardware and operation follow below.

The telescope is a 0.38 m (15 inch) diameter refractor which uses a plastic f/1.6 Fresnel lens as the objective. A major problem with a Fresnel lens is that collimated light is focused to a circle of least confusion which for this lens is 1 mm. This circle is larger than the diffraction limited spot size of a comparative perfect lens and hence the field of view for the telescope is larger than that of a telescope employing a more perfect objective lens. Yet, because the lens is so large, the Fresnel lens is judged more cost effective. A subassembly which contains a field stop iris, a collimating lens, an aperture stop iris, and a 5 nm interference filter may be removed from the telescope for precise alignment under laboratory conditions. The schematic layout of the telescope is shown in Figure 5.2.

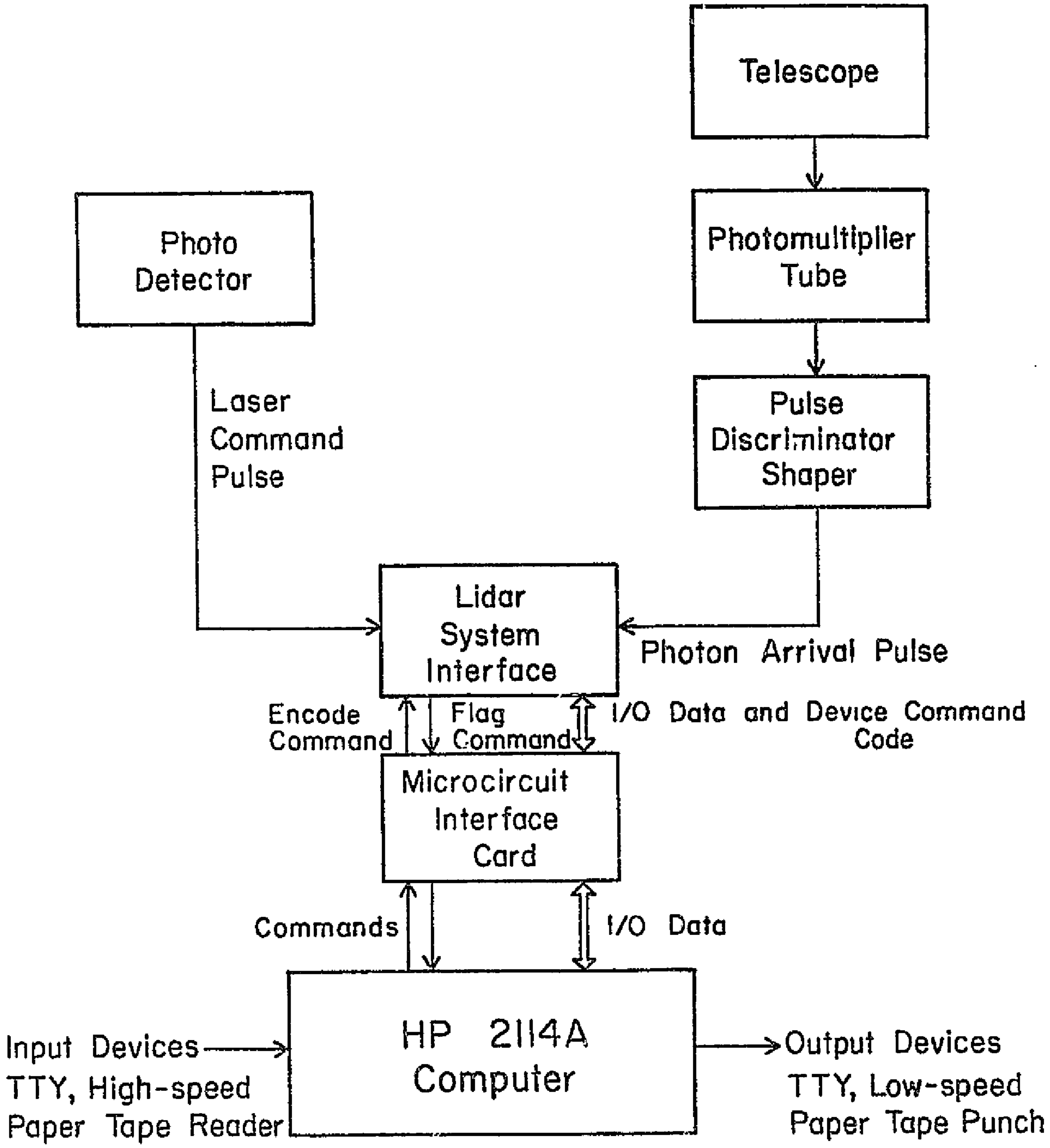


Figure 5.1 Lidar receiver with system interface.

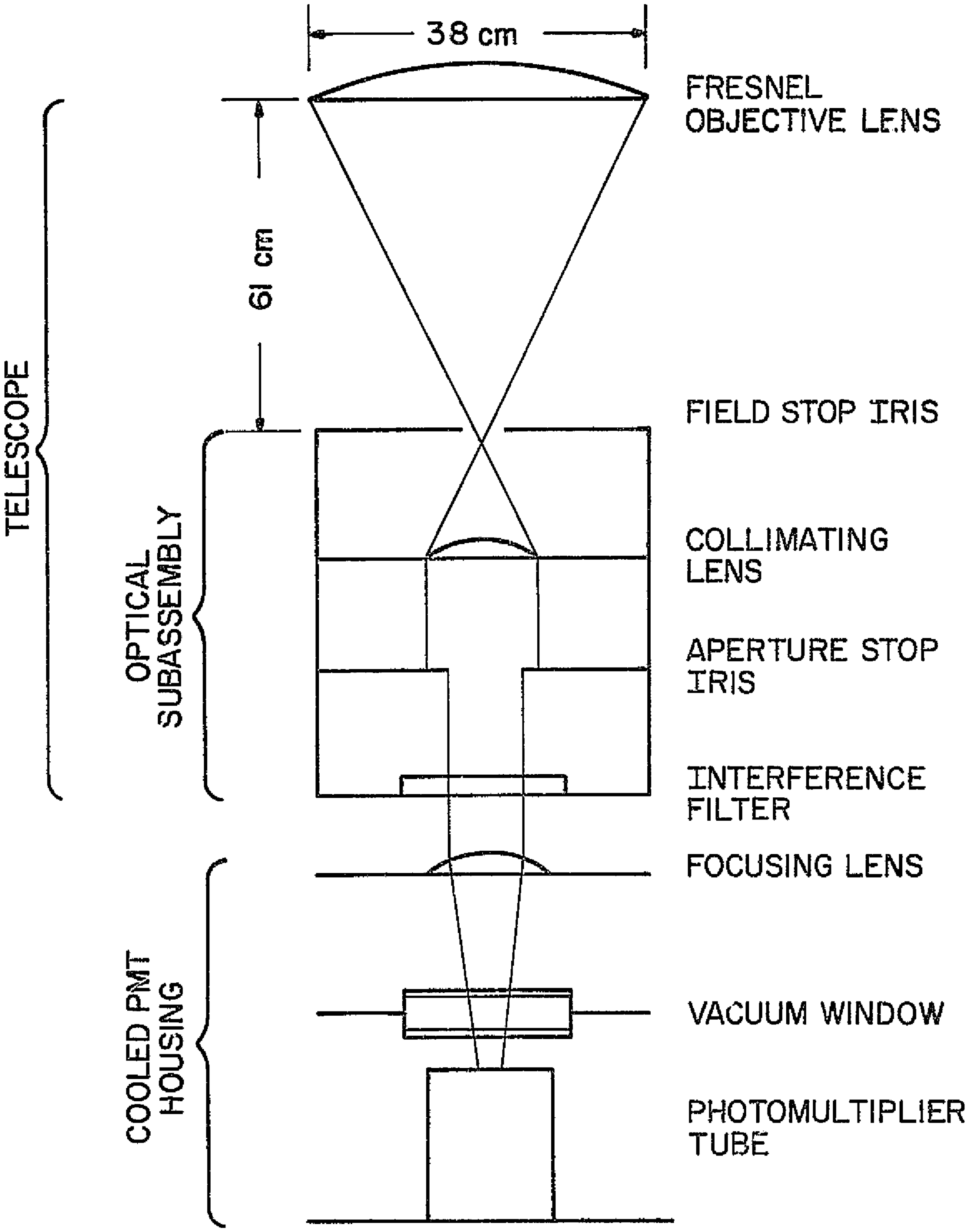


Figure 5.2 Telescope optical system.

In the photograph of Figure 5.3, the portion of the telescope (1) containing the subassembly can be clearly seen. A RCA C31034A/2A photomultiplier tube (PMT) mounts in an assembly beneath the telescope (2) and is cooled to -20° C. This is accomplished by inserting a heating coil into a dewar of liquid nitrogen (3) and circulating the resultant vapors around the PMT. A variac (4) is used to control the heat input and an ohmmeter (5) monitors the resistance of a calibrated thermistor mounted inside the PMT housing. The PMT views the received light through a double paned window to prevent frosting of the window.

The C31034A/2A PMT has a quantum efficiency of 0.2 at 589 nm as shown in Figure 5.4. Unfortunately the tube has a characteristic of losing this high quantum efficiency with age. Presumably impurities emitted from the inner dynodes may eventually coat the gallium arsenide photocathode chip and thus reduce the efficiency. From the experience of other groups, this degradation in quantum efficiency is proportional to use and can be accelerated if high anode currents frequently occur. Low light levels should be maintained under all conditions of operation or storage. Finally the rectangular geometry of the chip is not a perfect match to the circular geometry of the telescope but problems do not result from this mismatch.

Negative-going pulses from the photomultiplier tube are routed by a matched cable to an amplitude discriminator (University of Illinois Elementary Particle Physics Research, Trigger T100A) where a threshold level of 50-100 mV is appropriate for good signal discrimination. The discriminator eliminates a large number of the pulses which originate from the inner dynodes of the PMT. The output of the discriminator is a NIM standard ("ϕ" = 0V into 50 Ω, "1" = 0.8V into 50 Ω) and is fed through a matched line to the vicinity of the lidar interface unit where the signal is converted to TTL levels using

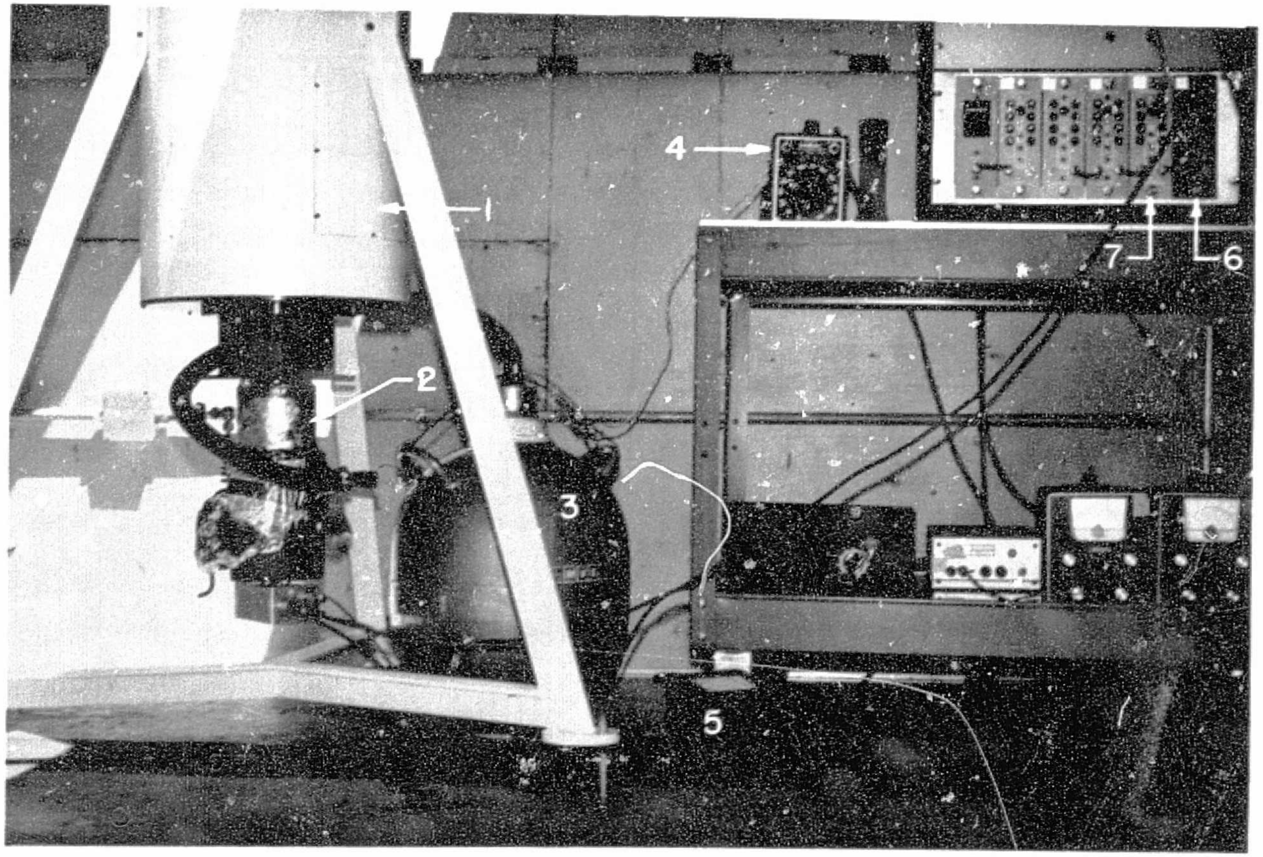
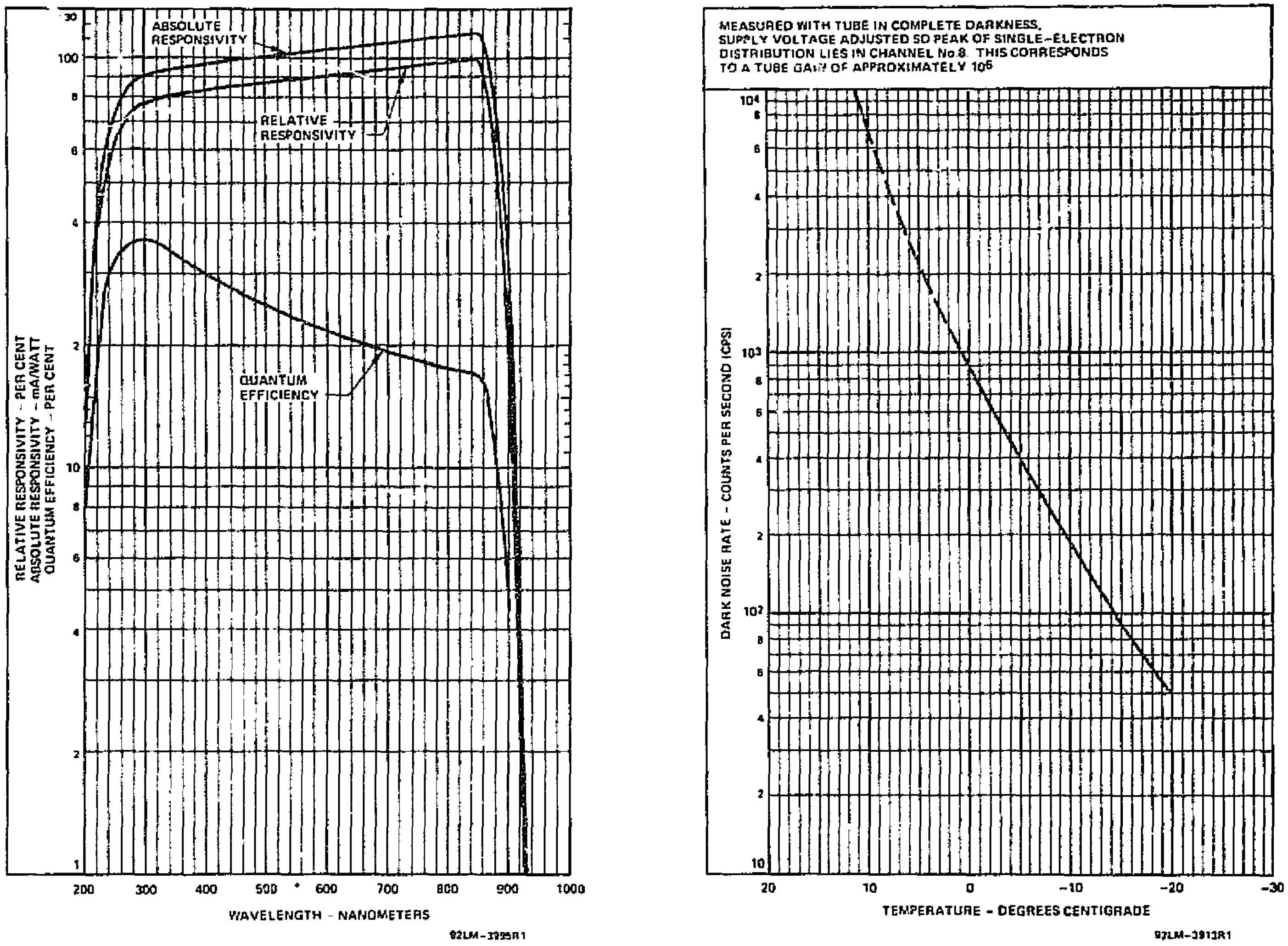


Figure 5.3 Lidar receiving system telescope and photon counters.

ORIGINAL PAGE IS
OF POOR QUALITY

...L PAGE IS
OF POOR QUALITY



MEASURED WITH TUBE IN COMPLETE DARKNESS, SUPPLY VOLTAGE ADJUSTED SO PEAK OF SINGLE-ELECTRON DISTRIBUTION LIES IN CHANNEL No.8. THIS CORRESPONDS TO A TUBE GAIN OF APPROXIMATELY 10^6

Figure 5.4 RCA C31034A photomultiplier characteristics.

ORIGINAL PAGE IS
OF POOR QUALITY

a fast ECL circuit. The pulses are then presented to the interface unit.

Referring to Figure 5.5, a brief summary of what transpires during data collection will now be given. The system interface includes a photon upcounter, processor, clock and random access memory (RAM) with the processor communicating with the computer and coordinating operation within the system interface.

When data are collected the following sequence of events takes place. An initialization pulse is given by the processor to prepare for data collection. The lidar software allows the operator to load into the computer the desired sampling rate and delay, and these values are transferred by the processor from the computer to the interface. The sampling rate determines how frequently the binary upcounter will be sampled and therefore determines the size of the range bin. The sampling period then is programmable and may take on values from 1 μs to 128 μs. The delay corresponds to the time difference between the firing of the laser and the initial sampling of the upcounter. The delay time is a multiple of the sampling period and may be from 1 to 4095 times the sampling period.

After this information is loaded, the system waits until the laser fires; the laser pulse triggers the system into operation. After the desired delay, the upcounter is sampled at the desired rate for 256 times. Each sample is routed to an addressed portion of the RAM. After this sequence the processor disables the clock signal and informs the computer that data collection has been completed. Between laser pulses the data are transferred from the RAM to the computer. Adjacent bins are then subtracted to yield the number of counts registered in each bin. This information is integrated with previous shots and the sequence continues until the desired number of shots is taken.

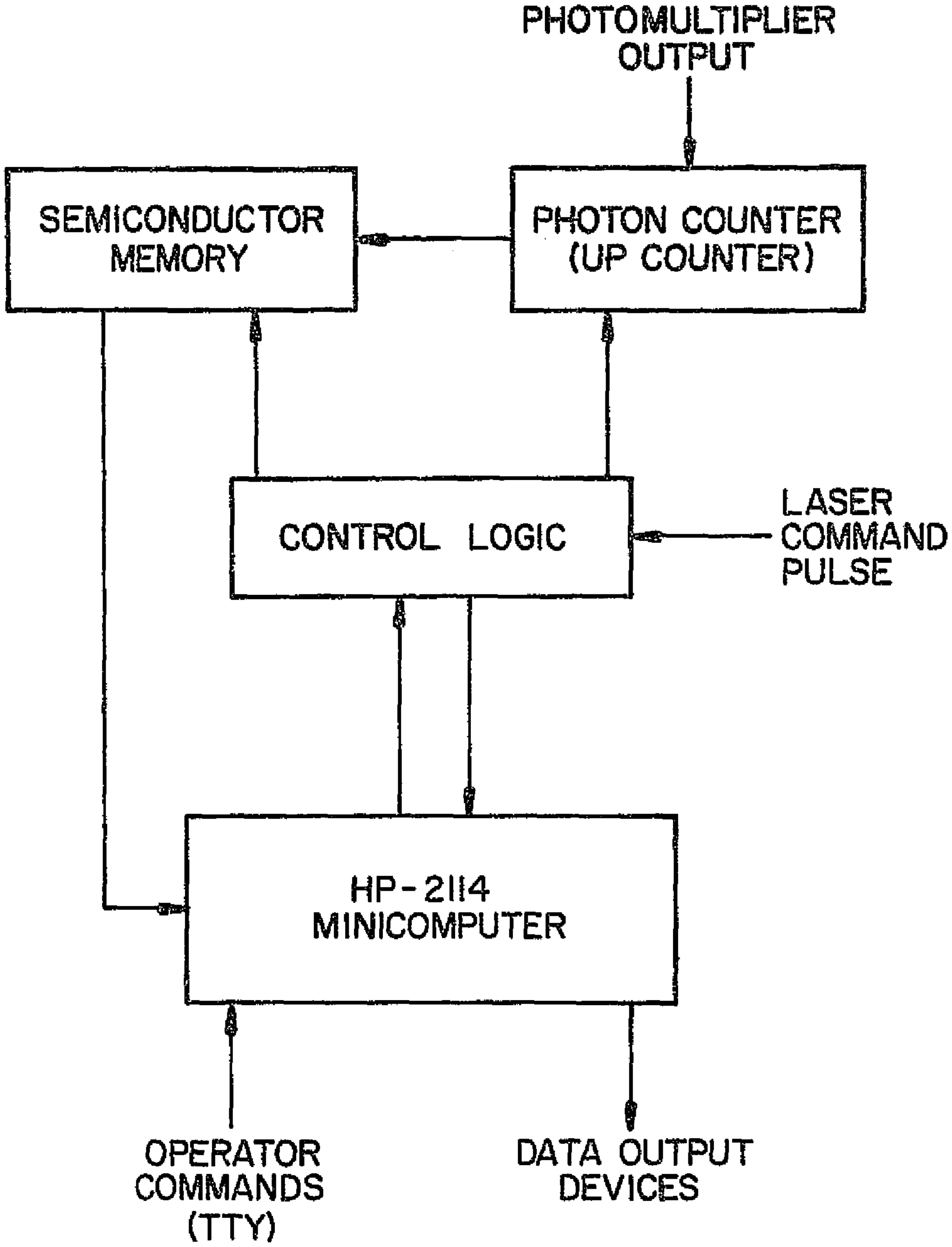


Figure 5.5 Block diagram of lidar system interface device and associated links to HP 2114A minicomputer.

Since direct memory access was not available in the HP2114A computer, the inclusion of the device's separate memory was necessary. While the maximum available memory of 8k of core in the HP2114A is a severe limitation, the options of variable sampling rate and delay provide a more flexible system over the initial system described in Appendix I. Furthermore, the receiving system may be easily included in a mobile lidar system.

5.2 *System Evaluation*

It is instructive to compare the improvement in system performance over the duration of this work. Two profiles are superimposed in Figure 5.6 which were taken about 1½ years apart. As noted in Chapter 3, we define the maximum signal-to-noise ratio as the number of "signal" counts obtained from the peak of the layer divided by the standard deviation of the count from the peak of the layer. Explicitly this is written as

$$S/N = \frac{N_i}{\sqrt{N_i + N}} \tag{5.1}$$

where

N_i = "signal" count from the peak of the layer

N = background noise count

Referring to Figure 5.6 the signal-to-noise ratio is calculated to be 4.8 for the June data while by November 1977, $S/N = 13$. More importantly, it should be emphasized that the 1976 data required the integration of 1181 shots while the 1977 data was obtained with only 246 shots. Taking into account the difference in integration time, the small change in range bin sizes, and the seasonal change in sodium density we find that the signal strength has been increased by a factor of 10.

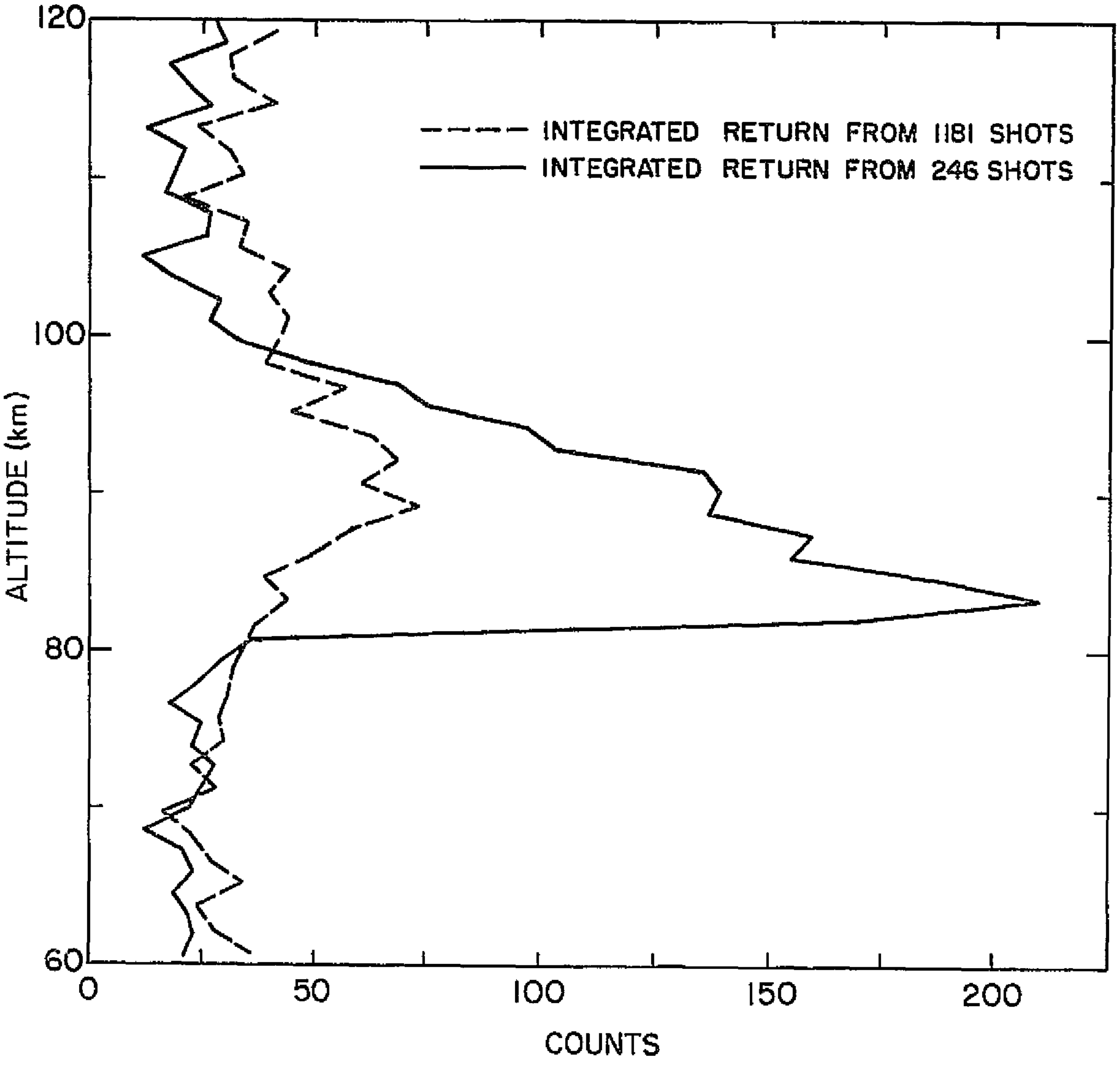


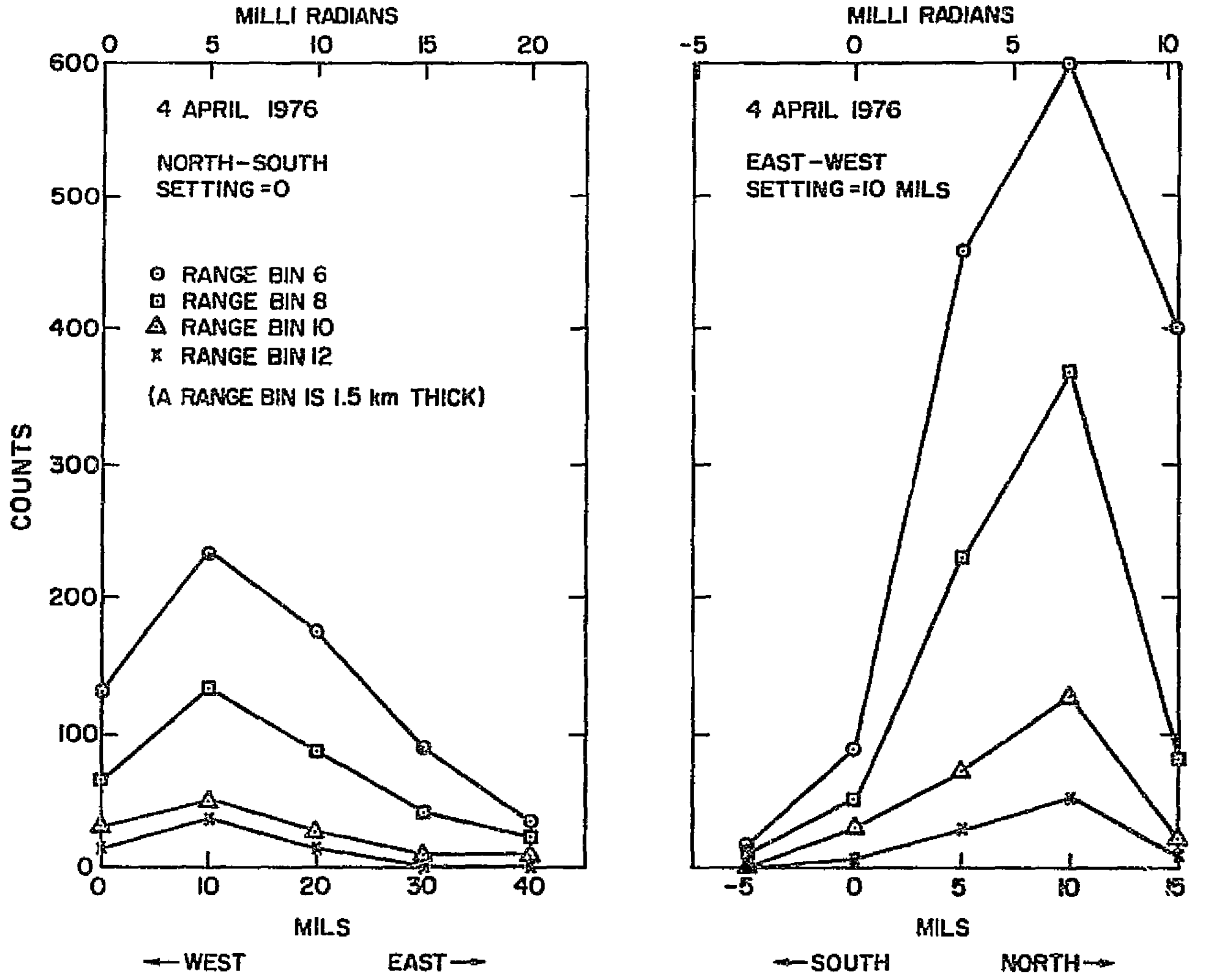
Figure 5.6 Comparison of profiles taken 1½ years apart reflecting system improvements. Dotted curve was from the evening of June 25-26, 1976. Solid curve is taken from the evening of Nove..ber 11-12, 1977. The November data are significant in themselves, since a primary peak centered at 90 km developed during the evening, indicating the variability of the layer.

5.3 *System Alignment*

Alignment of receiver and transmitter is crucial to the proper operation of the laser radar system and a rather simple approach to the problem has been used. Our telescope, once leveled, is stationary. We therefore align the system by steering the laser beam to seek a peak in returns at some altitude range bin in the 20-30 km region. Generally 20 shots are sufficient to give statistically significant returns from these altitudes. Micrometer heads are used to steer the laser beam in either an east-west or a north-south direction.

Figure 5.7 depicts the type of alignment returns we observe. In this particular case the final micrometer settings were taken relative to the arbitrary starting position as 10 mils east and 10 mils north. Alternatively, the angular movement from the starting position was approximately 5 mrad east and 7 mrad north. Finer adjustments of the micrometer heads would show a rounding of the curves near the peak. Improvements in the laser output have now allowed us to align at much higher altitudes than those indicated in Figure 5.7, although the procedure remains unchanged.

Lengthy alignment is performed only when some major laser optical alignment has occurred between runs. This alignment of transmitter and receive beams may take up to an hour to perform. However, once this alignment is achieved, the system requires only fine adjustments before future runs. This fine adjustment usually takes less than 15 minutes to accomplish.



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 5.7 Integrated return from 20 shot samples versus laser beam direction

6. EXPERIMENTAL RESULTS

6.1 *Introduction*

The initial years of this project were spent designing, constructing, and optimizing the sodium lidar facility. During this time the techniques for successful lidar measurements of the sodium layer were developed, and monitoring of the layer concentrated on obtaining one or two profiles for an entire evening. By October 1977 the system signal-to-noise ratio had been increased substantially and experimental efforts focused on studying the short-term variability of the layer. Emphasis was placed on observing relative structural changes in the layer rather than measuring the absolute value of the sodium concentration versus altitude. In conjunction with this effort, signal processing techniques were developed to study the apparent wave motion often noticed in the layer.

The theoretical work has indicated the need for several correlative measurements. Joint experiments with the Aeronomy Laboratory Field Station meteor radar facility and ionosonde are being planned. The signal processing techniques are briefly discussed in Section 6.2. Results from the nighttime monitoring of the sodium layer are presented in Section 6.3. Section 6.4 presents the results of a feasibility study for joint meteor radar and laser radar measurements. Section 6.5 describes the interesting data obtained from monitoring the pre-sunrise sodium layer and simultaneously sounding the mesosphere with a 2.66 MHz transmitter. And finally a brief conclusion is given in Section 6.6.

6.2 *Sodium Lidar Data Collection and Analysis*

6.2.1 *Introduction.* Our lidar system has been developed to measure the sodium concentration profile over Urbana, Illinois. Because low signal levels are encountered, the receiving system operates in the photon counting mode. During the past year J. R. Rowlett and C. S. Gardner have been investigating various techniques for processing photocount data to obtain estimates of the sodium profile. The usual procedure is to divide the time axis into disjoint intervals and to count the number of photons detected during each time interval. Using the lidar equation, the photocount in each interval or range bin can be related to the density and is then plotted versus time or distance as a histogram.

The main advantage of the histogram approach is the simplicity of the photon counting electronics. However, the technique only yields the density at discrete points and since the measurement is actually an integrated density, the range resolution is limited by the receiver gate time. In addition photon counting does not fully utilize all the available information in the signal. Useful information contained in the arrival times of the detected photons is suppressed by the histogram data. If the arrival times are known, it should be possible to obtain a more accurate estimate of the sodium profile. Unfortunately, because the arrival time measurement requires sophisticated and expensive electronics, it is seldom used in lidar applications.

Initially the signal processing work concentrated on the problem of estimating the sodium density from the histogram data using statistical techniques such as least-squares and minimum variance estimators. Particular attention was given to deconvolving the effects of the receiver gate time and laser pulse width. In addition it was necessary to determine the receiver gate time effect on the accuracy of the sodium density estimate.

A gate time of zero corresponds to an arrival time measuring receiver. Since histogram data can be generated from knowledge of the photon arrival times but not vice versa, it is expected that the density estimates become more accurate as the gate time is decreased. This work has led to the development of a promising signal processing technique based upon frequency domain filters which can be used to generate continuous density profiles from the histogram data. Furthermore it has been shown that under appropriate conditions, histogram data can yield density estimates whose accuracy is comparable to that obtained with arrival time data. A detailed presentation of this work is given by *Rowlett and Gardner* [1978].

In this chapter data are presented sometimes as a histogram and sometimes as filtered data. In the former case the counts registered in the $i^{\underline{th}}$ range bin are plotted at the center of the range bin; unless otherwise noted, the range bin size is 1.5 km (10 μs). Error bars are obtained by equating the square root of the received count to the standard deviation of the returns. The processed (filtered) data can easily be recognized because the filter cutoff frequency always appears with these data. A horizontal scale labeled "counts" implies that the data have not been corrected for the $1/z^2$ dependence mentioned in Chapter 3. In most cases these data have been taken with a 150 m (1 μs) range bin size; occasionally the data were taken with a 450 m (3 μs) range bin size, and these cases are clearly indicated.

6.2.2 *Signal processing.* In Section 3.3.2 it was mentioned that the lidar equation has the form of a two-fold convolution of the receiver gate function, the transmitter pulse, and the unknown sodium profile. The profile is "smeared" by the effects of the laser pulse shape and the receiver gate, and it is necessary to deconvolve these smearing effects to obtain an estimate of the sodium profile. However, the observed count also contains noise counts from background radiation and from "dark counts" generated by the photomultiplier tube in the absence of any signal. The presence of noise complicates the deconvolution process.

Deconvolution is an inherently ill-conditioned process, because the deconvolution filter is essentially an inverse filter to the original convolution. From the frequency domain viewpoint, the convolution process attenuates the high-frequency spectral components of profiles. The attenuated components will then be amplified in deconvolution, but the measured count will contain primarily noise at these frequencies. Amplifying the noise components produces an unstable solution. To reduce this instability we incorporate *a priori* information about the solution.

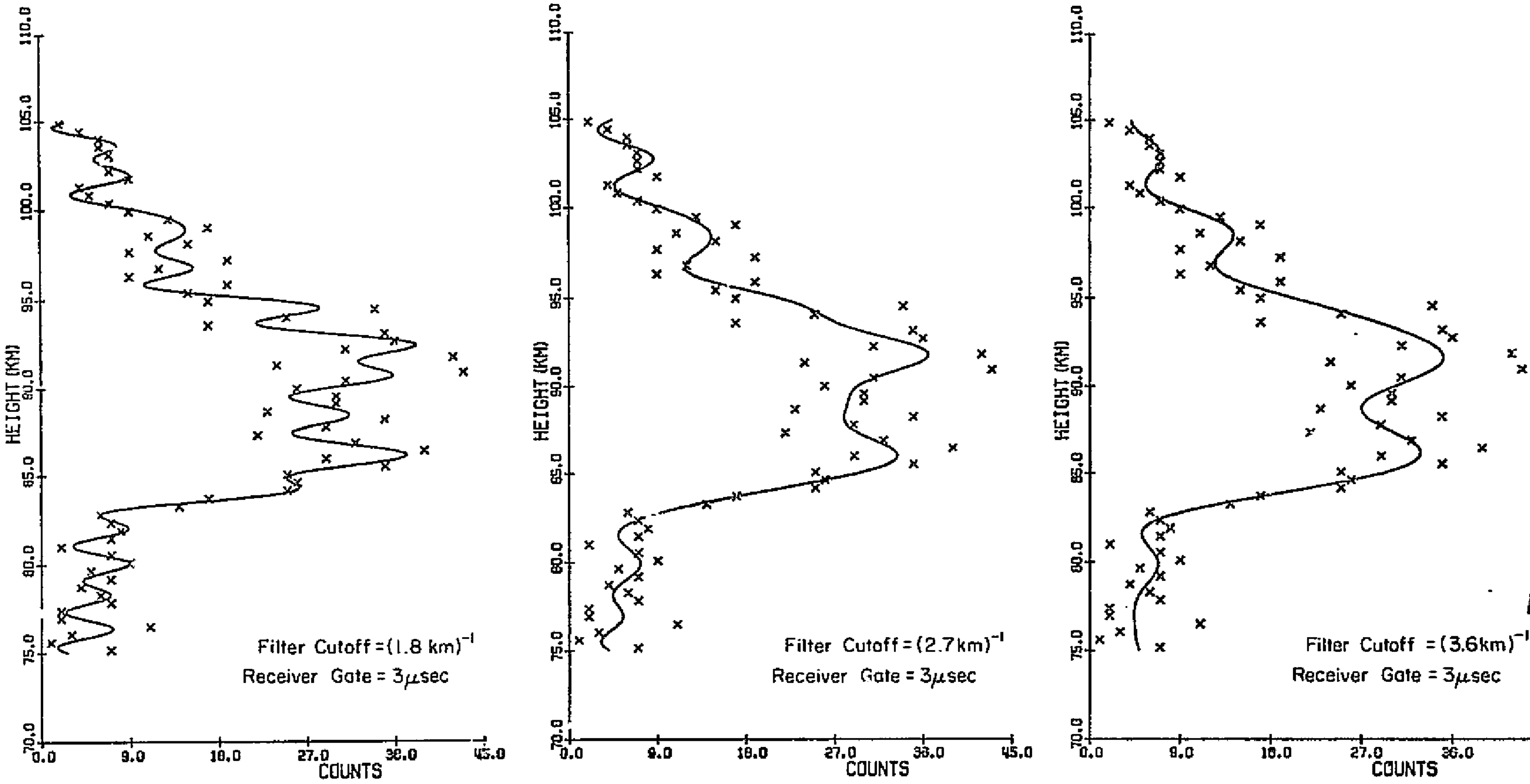
We can gain an intuitive idea of the effect of using *a priori* information by examining the deconvolution procedure from the frequency domain viewpoint. In the frequency domain, the lidar equation (3.18) becomes

$$E\{N_i(\omega)\} = N_t(\omega)G(\omega)D(\omega) \tag{7.5}$$

where $D(\omega)$ is the transform of $\rho(r)/r^2$. To recover an estimate $\hat{\rho}(r)$ of the profile, we should pass the data through a filter

$$H(\omega) = \frac{1}{N_t(\omega)G(\omega)} \tag{7.6}$$

This filter has poles at $\omega = \frac{\pm 2\pi k}{T}$, k an integer, $k \neq 0$ from the zeros of $G(\omega)$. However, if we assume that $D(\omega)$ is bandlimited to $\omega < \frac{2\pi}{2T}$, we can set $H(\omega) = 0$ for ω larger than the cutoff of $D(\omega)$. The physical meaning of assuming $D(\omega)$ is bandlimited to some ω_c is that $\hat{\rho}(r)$ has no sinusoidal component whose wavelength is less than $\frac{c\pi}{2\omega_c}$, where c = speed of light. We would choose the cutoff ω_c as small as possible to exclude as much noise while at the same time retaining as much information as possible. Figure 6.1 shows the deconvolution filter applied to the data collected on October 19, 1977 with three different filter cutoffs. As ω_c is decreased, the



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 6.1 Deconvolution filter applied to the data of October 19, 1977. The filter cutoff is varied from (1.8 km)$^{-1}$ to (3.6 km)$^{-1}$. The x's denote the original photon counts.

estimates become less noisy, but some information is also lost. Presumably, a middle ground can be reached where most of the information in the signal is retained while most of the noise is excluded. The present opinion is that a cutoff of $(2.7\ \text{km})^{-1}$ is near optimum for these data but ultimately the exact value will vary from one data set to another.

6.3 *The Nighttime Sodium Layer*

6.3.1 *The initial monitoring of the layer and system performance tests.*

The initial monitoring of the sodium layer in the spring of 1977 concentrated on obtaining one or two profiles for an entire night. The peak of the layer was consistently found to occur near 90 km. The layer was typically quite narrow and showed on occasion the characteristic bifurcation. Evidence of secondary peaks on the topside of the layer also existed. Several of these profiles are shown in Figures 6.2-6.5.

By October 1977 the system signal-to-noise had increased enough to enable us to obtain several profiles in a single evening. Such an example is shown in Figure 6.6. In this figure the Rayleigh scattering is also indicated to emphasize that lower altitude returns fall off exponentially and are not affected by Mie scattering. Furthermore this night of data indicated the variability of the sodium layer. The profiles in the latter part of the evening show a slight decrease in the altitude of the peak of the layer and a lowering of the bottomside. The layer was observed to broaden through the night. Each of these profiles represents the integrated return from a 520 shot sequence and indicate the improvement in signal-to-noise over the Spring 1977 profiles.

Several runs were hampered by haze or moonlight. Since returns from these evenings were usually noisy, time was spent on these night checking system performance. A standard test for a resonance scatter experiment is to

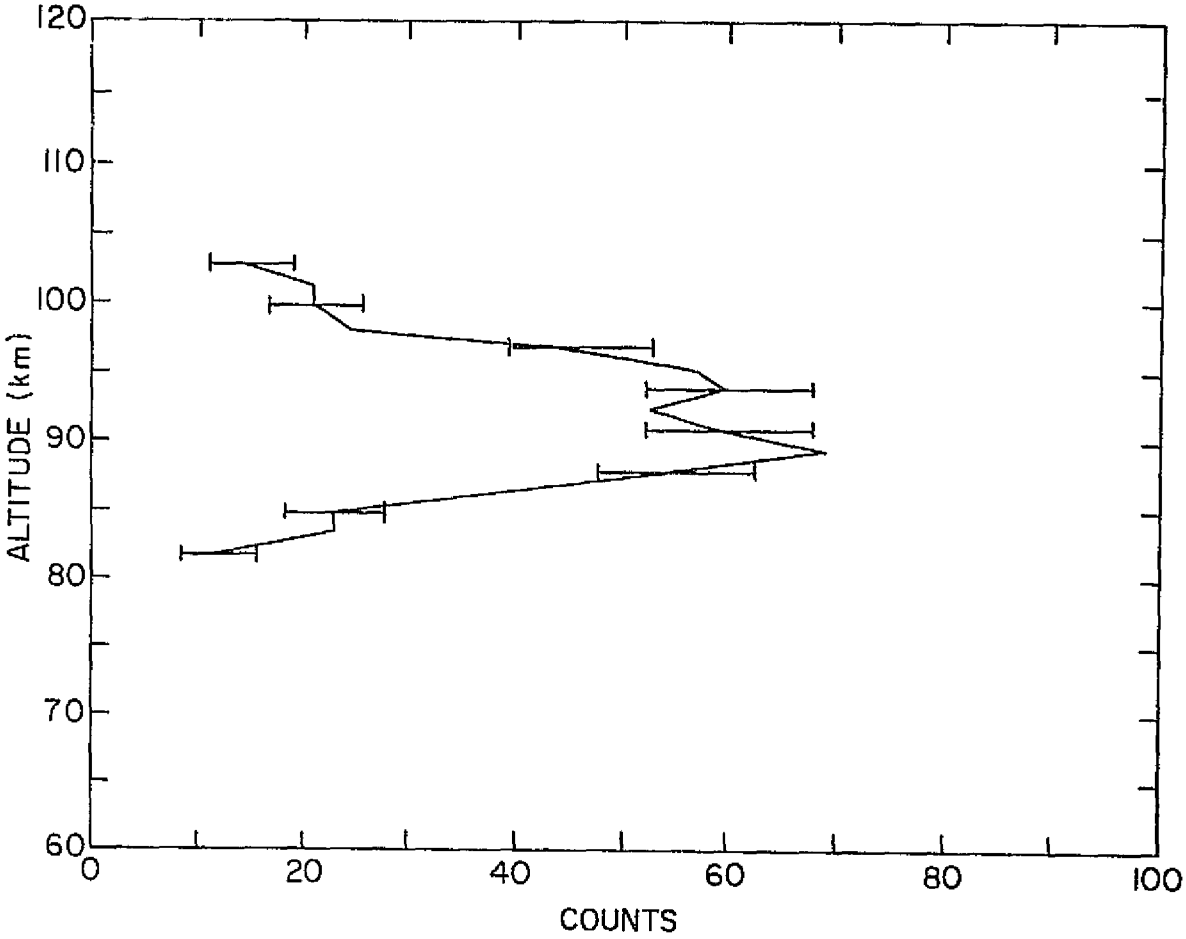


Figure 6.2 Integrated return from 2000 shots, March 23-24, 1977, 2211-0206 hr.

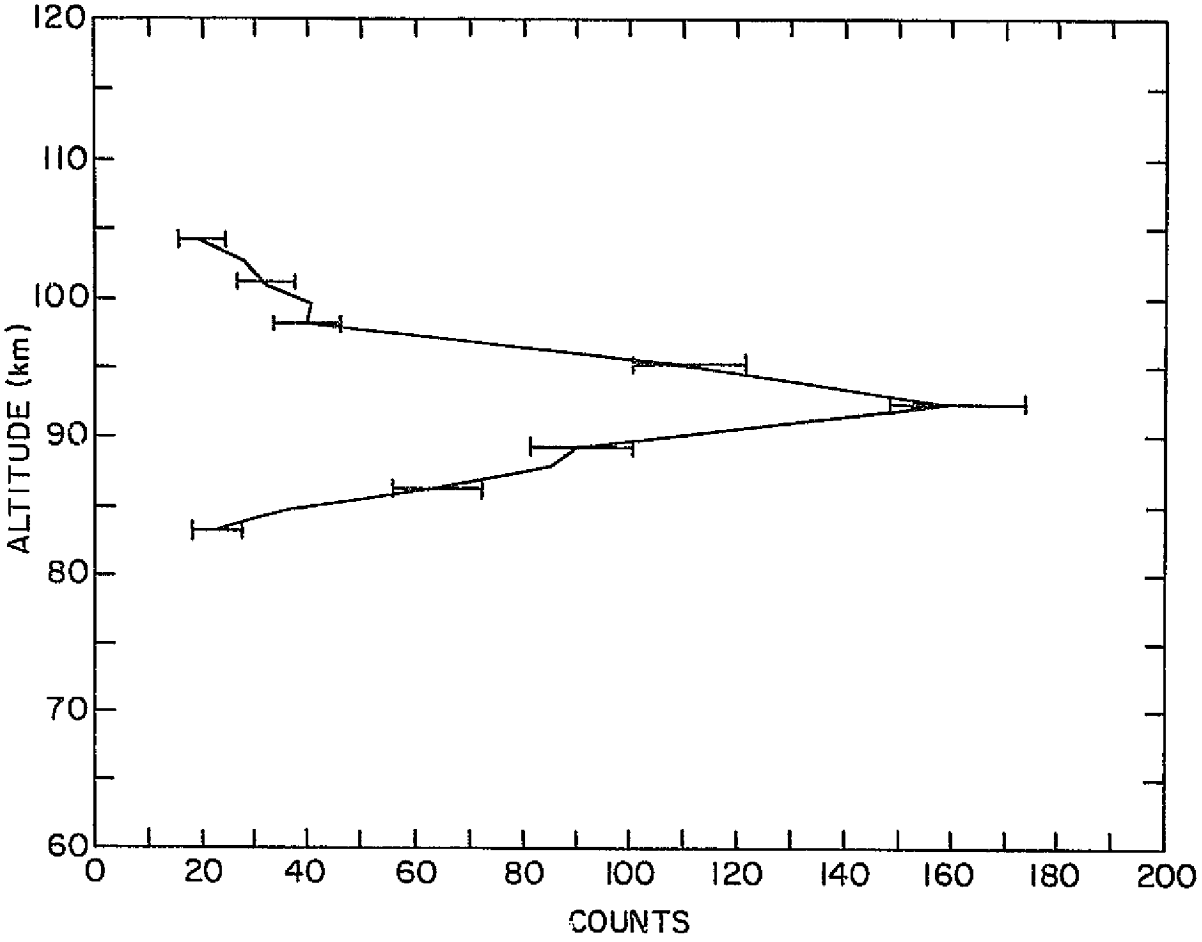


Figure 6.3 Integrated return from 3000 shots, March 24-25, 1977, 2130-0111 hr.

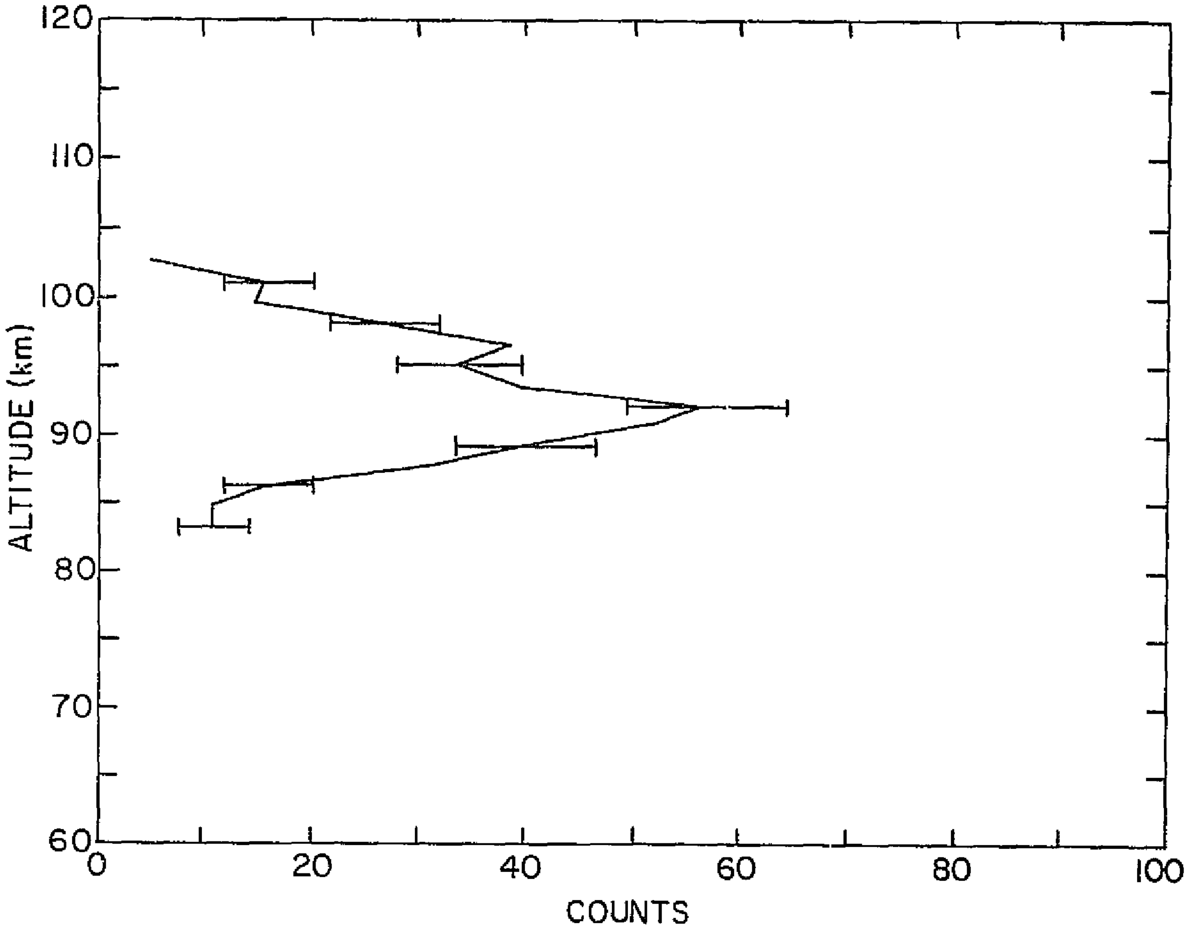


Figure 6.4 Integrated return from 1200 shots, March 26, 1977, 0201-0321 hr.

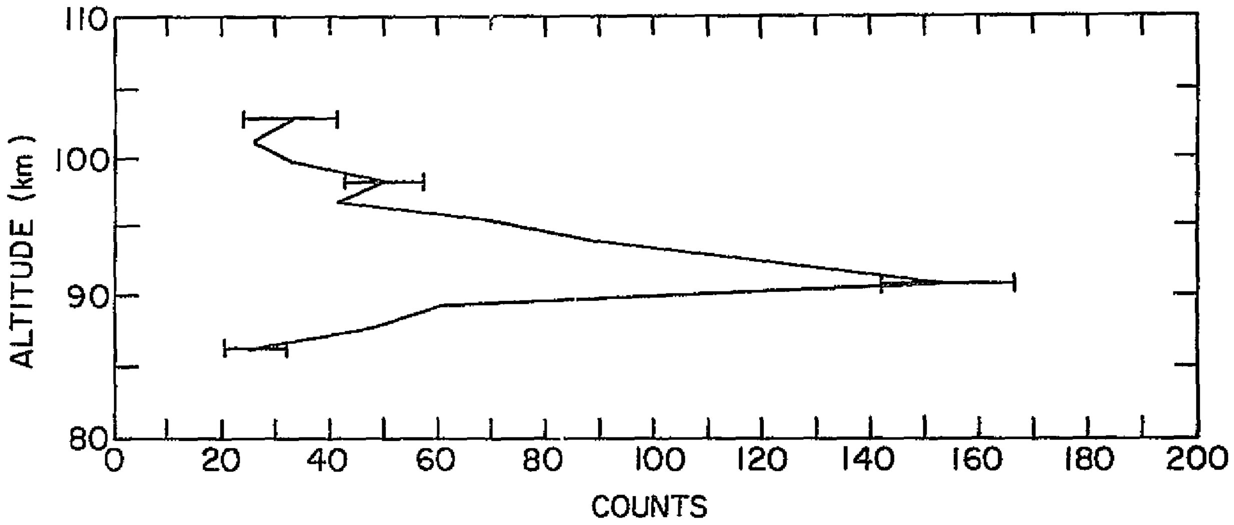


Figure 6.5 Integrated return from 2400 shots, April 10-11, 1977, 2335-0218 hr. Rayleigh scattering returns from 30 km were comparable to the resonance scattering returns from 90 km.

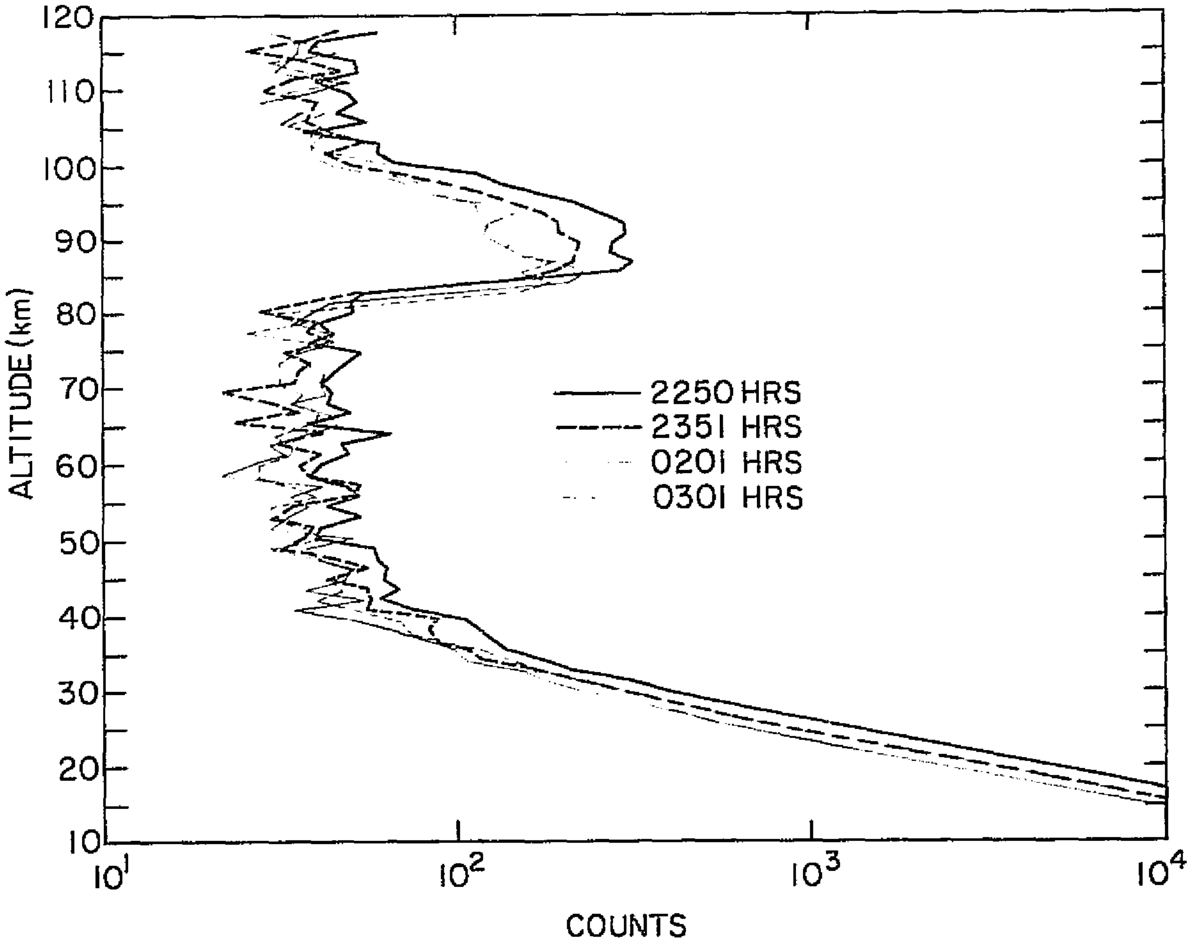


Figure 6.6 Four sets of data taken on the night of October 19-20, 1977. Starting times for the data collection of each set are indicated. Each profile represents an integration time of 20 minutes.

tune off the 589 nm resonance line and monitor the returns. The absence of returns in this case would indicate resonance scattering is the dominant process at 90 km. In Figure 6.7 the results of such a test are presented.

In most of this work the emphasis will be placed on observing relative changes in the layer. However it is important to verify the value of the peak concentration of the layer. Using the data of February 22, 1978 and applying equation (3.72) to the returns, the sodium profile is shown in Figure 6.8. The concentration scale is accurate to a factor of 3 which results from our inability to measure the linewidth of the laser precisely.

6.3.2 *Development of the layer structure with time--observations of wavelike structures and variations of the topside scale height.* Some of the more interesting lidar data we have collected was obtained during the late night and early morning hours of November 11 and 12, 1977, and the night of November 12. The data were processed using the unbiased estimate filter and plotted on a linear scale in Figures 6.9 and 6.10. The amplitude spectra of the raw data indicated that the signal usually drops into the noise at a spatial frequency somewhere between $(1.8\text{ km})^{-1}$ and $(2.7\text{ km})^{-1}$.

The raw data for each profile were obtained by firing the laser 250 times over a period of approximately 10 minutes. An additional 5 minutes is required by the computer to output the data on punched paper tape. Samples of the November 11 and 12 data are replotted in Figures 6.11 and 6.12 (solid lines) along with curves representing plus and minus one standard deviation about the estimated profiles (dashed lines). These plots are typical of the remaining data and illustrate the relative accuracy of the estimated sodium profiles.

An anamolous feature of these returns is that the peak of the layer remained at 85 km for much of the November 11-12 monitoring. On the basis

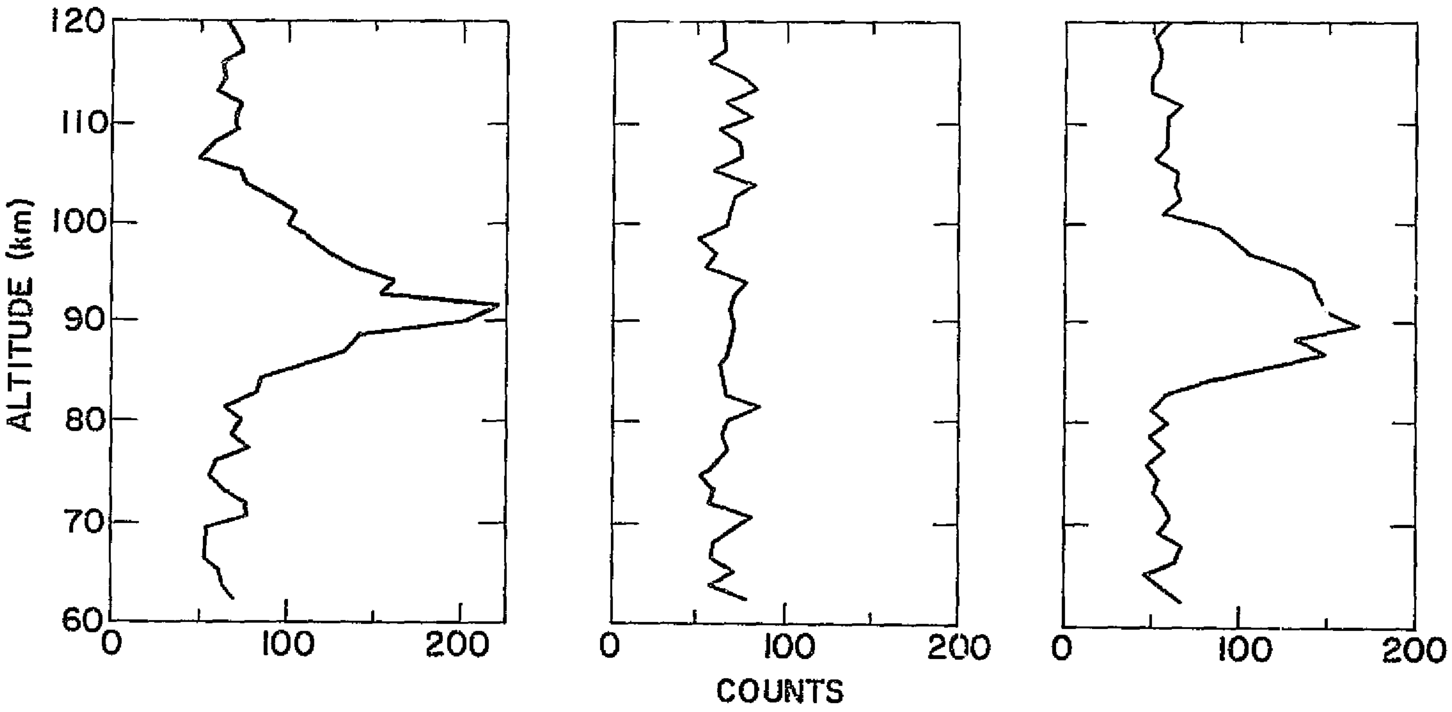


Figure 6.7 Returns showing the effect of laser tuning. Each graph is an integrated return from 260 shots on October 21, 1977. Starting time for the returns are from left to right 2135, 2157, and 2209 hrs. The center set of returns was taken when the laser was tuned to resonance scattering by sodium atoms.

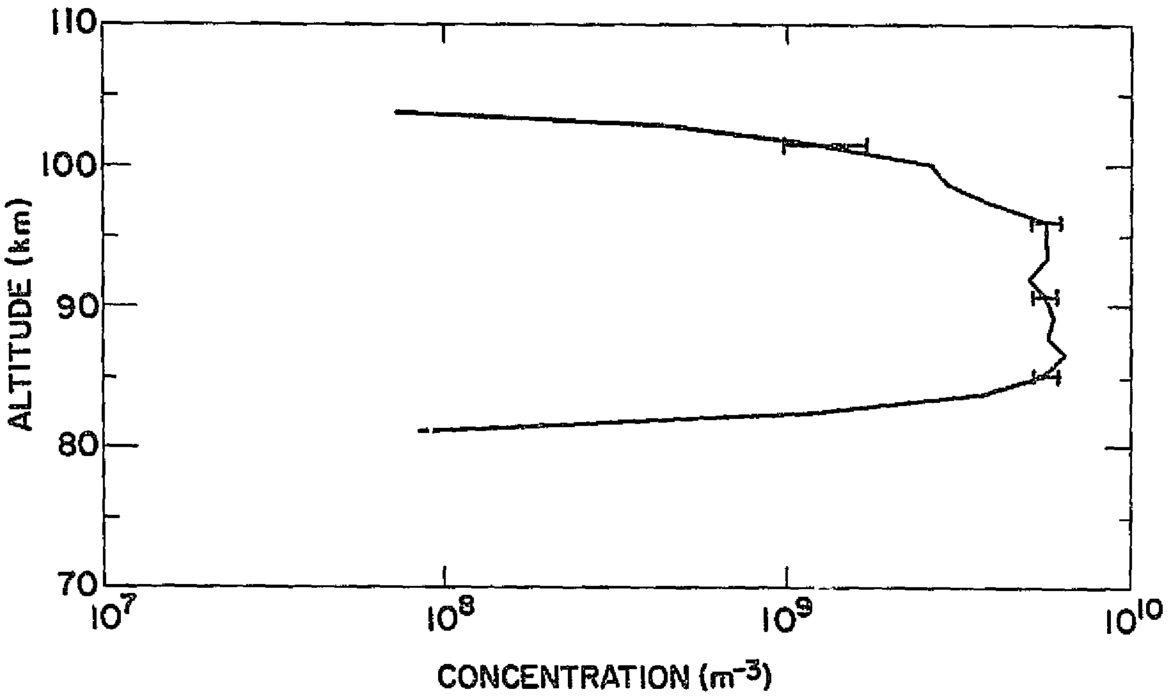


Figure 6.8 Sodium profile obtained on the night of February 22, 1978 resulting from the integration of returns from 300 shots. The concentration scale is accurate to a factor of 3 which results from the inaccuracy of the linewidth measurement.

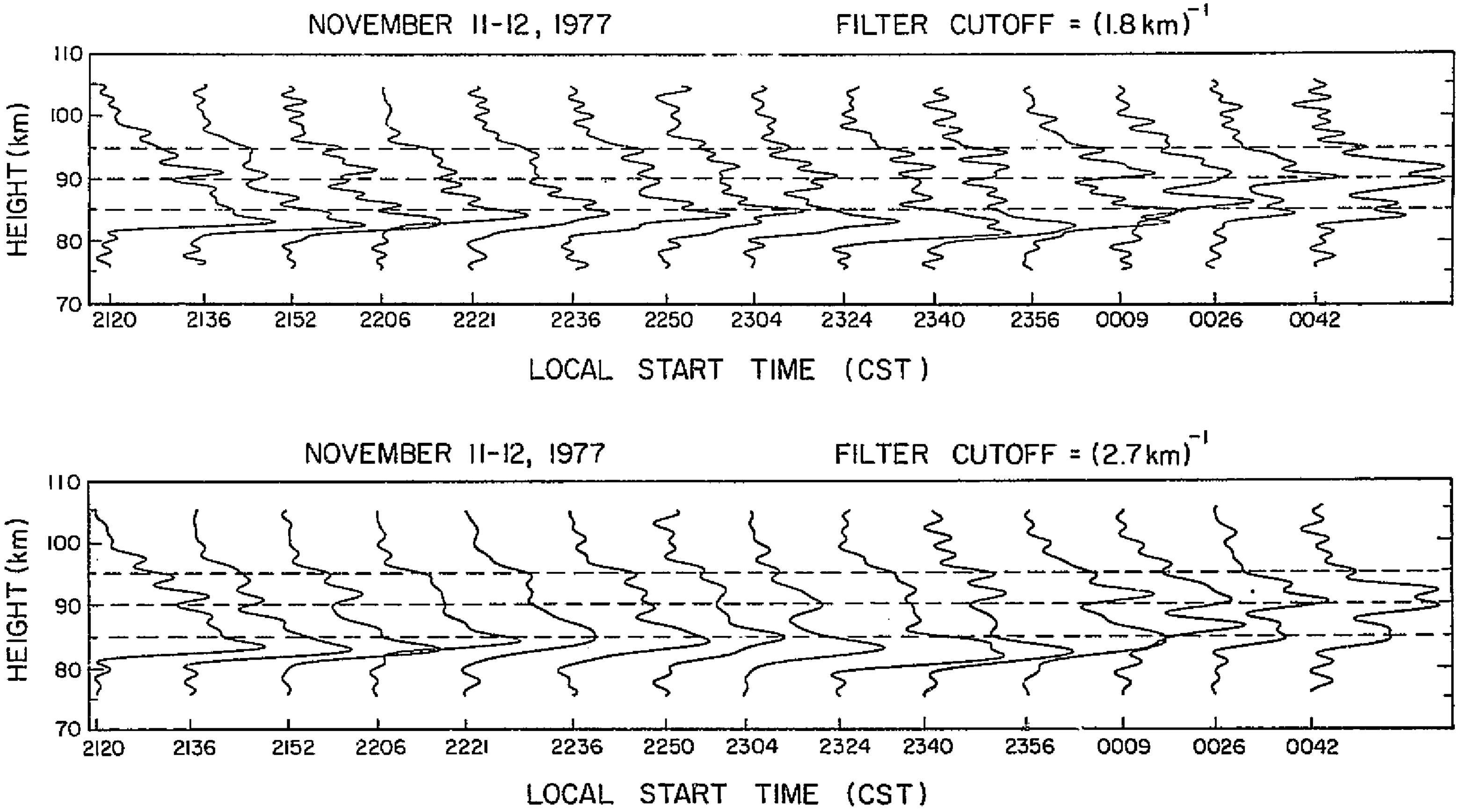


Figure 6.9 Time evolution of the sodium lidar data obtained during the late evening of November 11, 1977 and the early morning of November 12, 1977. The receiver gate time was 1 μsec.

ORIGINAL PAGE IS
OF POOR QUALITY

OF POOR QUALITY

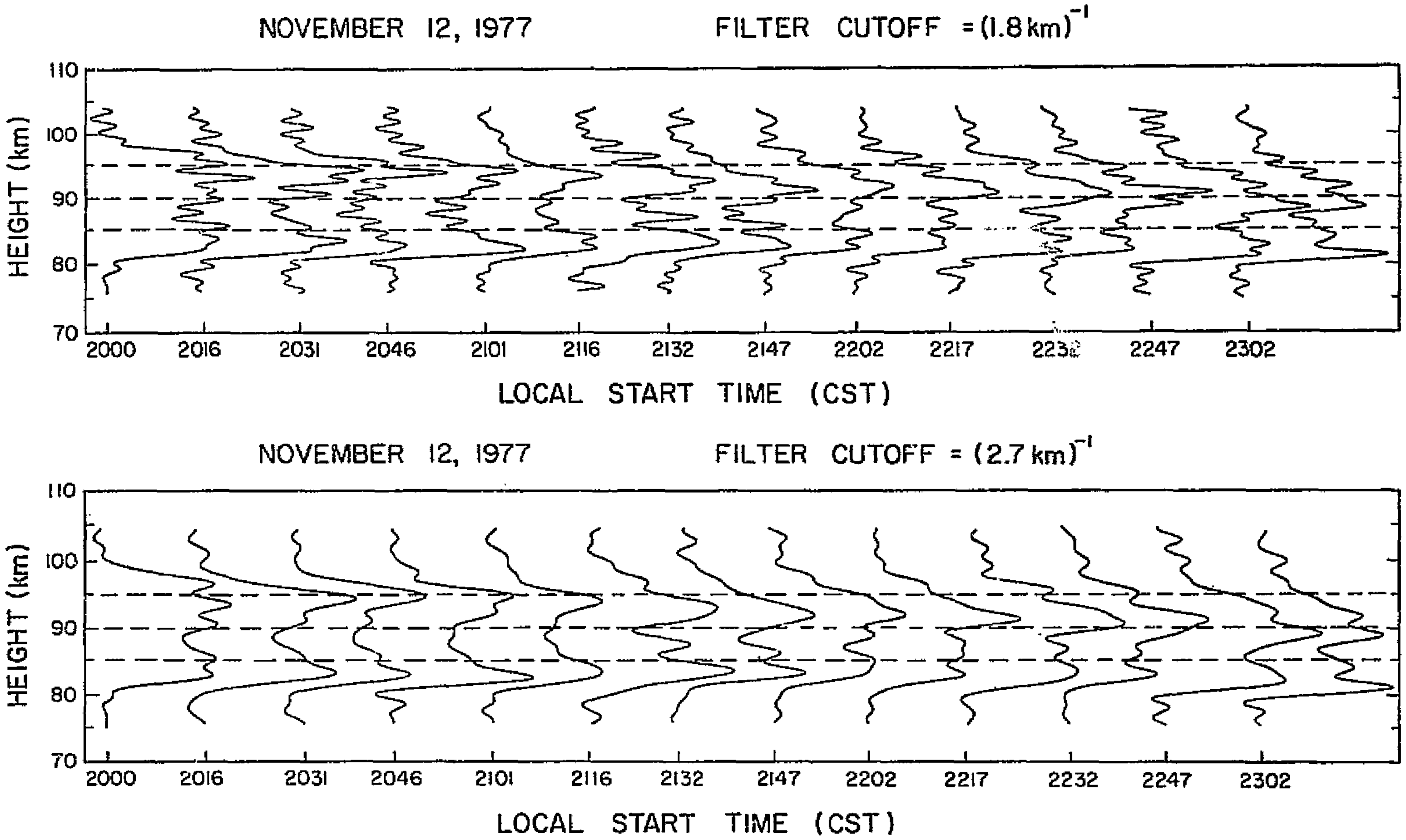
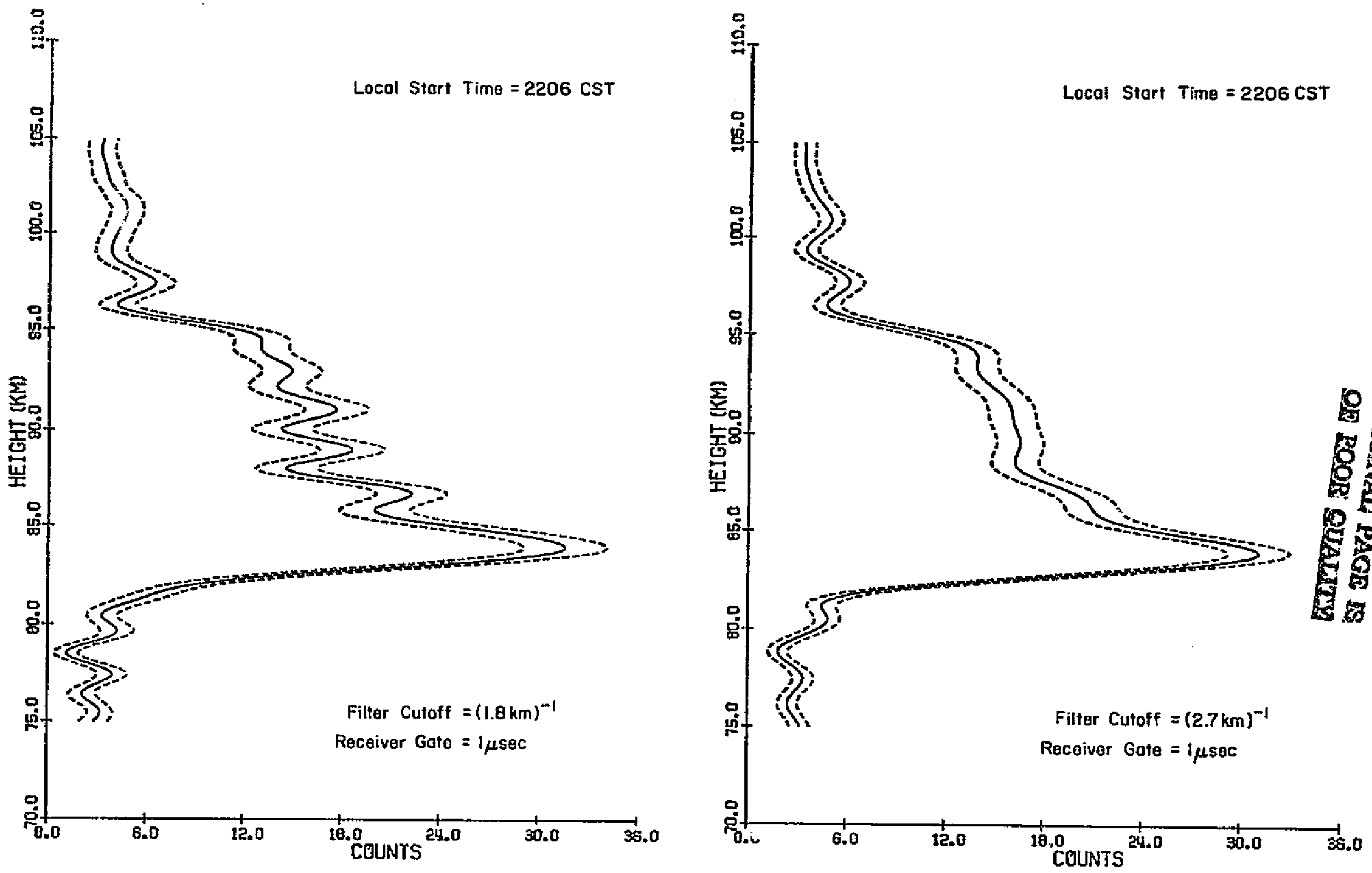


Figure 6.10 Time evolution of the sodium lidar data obtained during the evening of November 12, 1977. The receiver gate time was 1 μsec.



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 6.11 Samples of the November 11, 1977 data of Figure 6.9 (solid line) plotted along with curves representing plus and minus one standard deviation about the estimated profile (dashed lines).

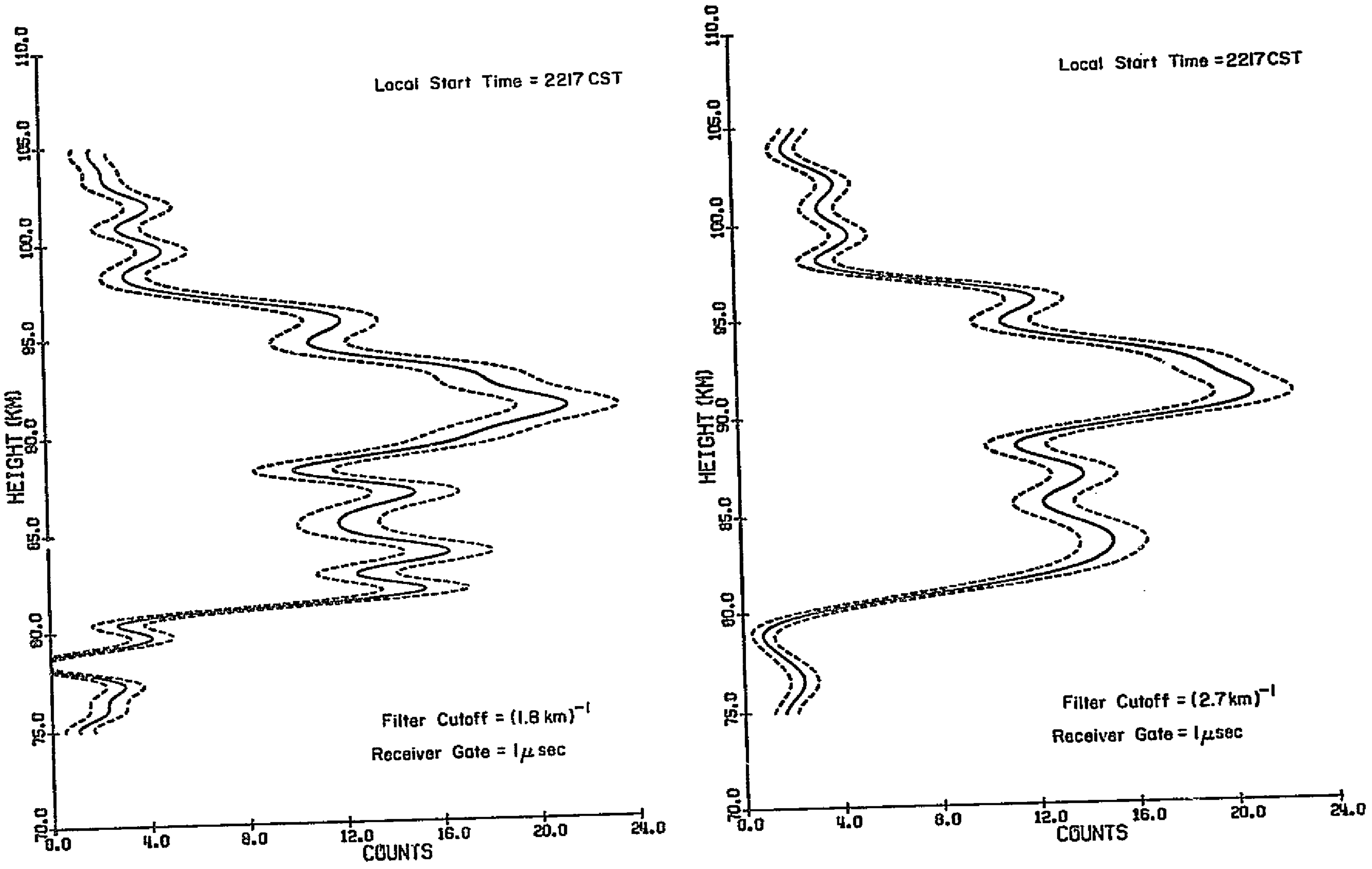


Figure 6.12 Samples of the November 12, 1977 data of Figure 6.10 (solid line) plotted along with curves representing plus and minus one standard deviation about the estimated profile (dashed lines).

of these data one wonders if the seasonal lowering of the peak of the layer as observed by *Gibson and Sandford* [1971] is caused by a systematic lowering or rather is the result of averaging several profiles, many of which have substantial contributions well below the average peak of the layer. The data also indicate a relation between the topside scale height and the altitude of the peak of the layer. The profile of 2120 hrs has a topside scale height of about 12 km; by 0026 hrs the scale height is closer to 5 km and the peak of the layer has changed from 83 km to 90 km. These data are in good agreement with the observations of *Megie and Blamont* [1977] which show a relation of the topside scale height to the altitude of the peak of the layer.

The November 12, 1977 data (Figure 6.10) provide an equally interesting aspect to the sodium layer. On this night the layer bifurcated, forming peaks at 83 km and 95 km. The peak at 95 km is then observed to move downward with time with a phase velocity of about 5 km/2 hr or 70 cm s^{-1}. This is the same order of magnitude for the downward transport of ionization by the wind shear mechanism as discussed in Chapter 7.

Another interesting feature of the time evolution of the sodium lidar data shown in Figures 6.9 and 6.10 is the change in the height of the bottomside throughout the night. In Figure 6.9 the height of the bottomside goes through a minimum in altitude at about 2324 hrs and a maximum near 2206 and 0026 hrs. On the night of November 12, 1977 the same phenomenon occurs but now the maximum height of the bottomside is at 2132 hrs while the minimum height occurs at 2302. In both cases the period of this oscillation appears to be near 2 1/2 hours. *Witt et al.* [1974] have suggested that the bottomside gradient of the sodium layer may undergo undulatory motions which result from dynamical processes and our observations appear to support their conclusion. A similar oscillation was noted by *Thomas et al.* [1976].

The data of January 3, 1978 indicated a highly variable layer (Figure 6.13). Wave structure seemed evident but a bulk movement as suggested in the November 12, 1977 data could not be found. It is possible that several waves are interacting in the region causing cancellation at one time and reinforcement at another. Indeed the amplitude spectra of these data indicate the possibility of many waves being simultaneously present (Figures 6.14 and 6.15).

It is well known that the wave structure in this region may be quite complicated. Evidence for the reflection of waves over Urbana exists [*Hess*, 1976] and standing waves are therefore possible. Such a standing wave in the lidar data would be typified by no vertical movement of the nulls and peaks of the wave structure but rather a change in amplitude with time at any set altitude. Using a notch filter to study only a portion of the spectrum of the returns, *Rowlett and Gardner* [1978] observed such a structure. The wave profile appears to return to its starting profile about every hour. The vertical wavelength of the wave is slightly greater than 5 km.

6.3.3 *Broadening of the layer with time*. Evidence for broadening of the layer with time has already been shown in Figure 6.6. The monitoring of the layer during the night of March 16-17, 1978 indicated less structural variation than on previous nights (Figure 6.16). However the layer did broaden during the night, starting at a full width at quarter maximum of 16.4 km (2231 hrs) and becoming 19.2 km wide at 0350 hrs. If the distribution topside scale height is taken to be the altitude increase from the peak of the layer at which the density is equal to $1/e$ of the peak density, then the distribution scale height increases from 8.5 km to 10.7 km in this time range for an average rate of change of 400 m hr^{-1}. The peak of the layer which started at 90.3 km fell by 1.7 km by the end of the observation period. This

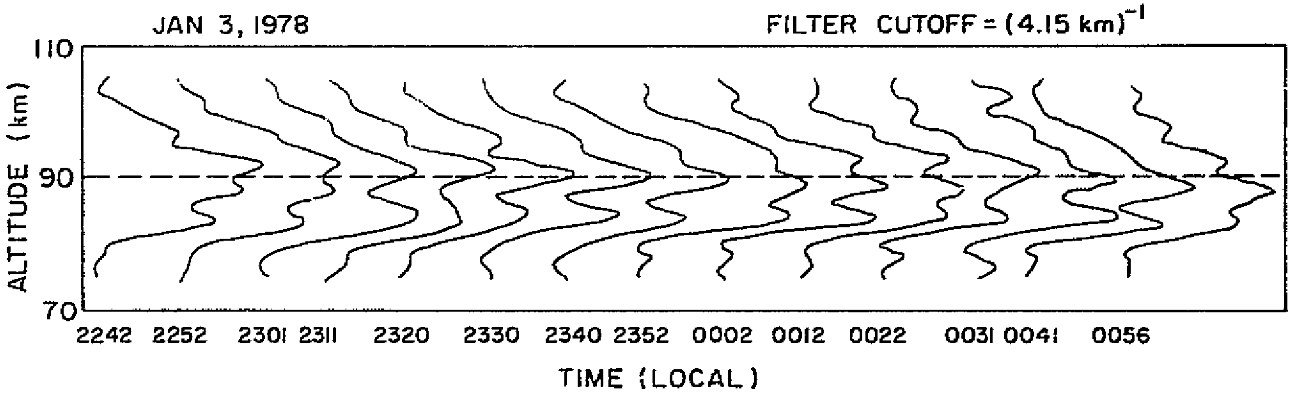


Figure 6.13 The evolution of the sodium lidar data obtained during the evening of January 3, 1978. The receiver gate was 1 μs. Each set represents the integrated return from 250 shots.

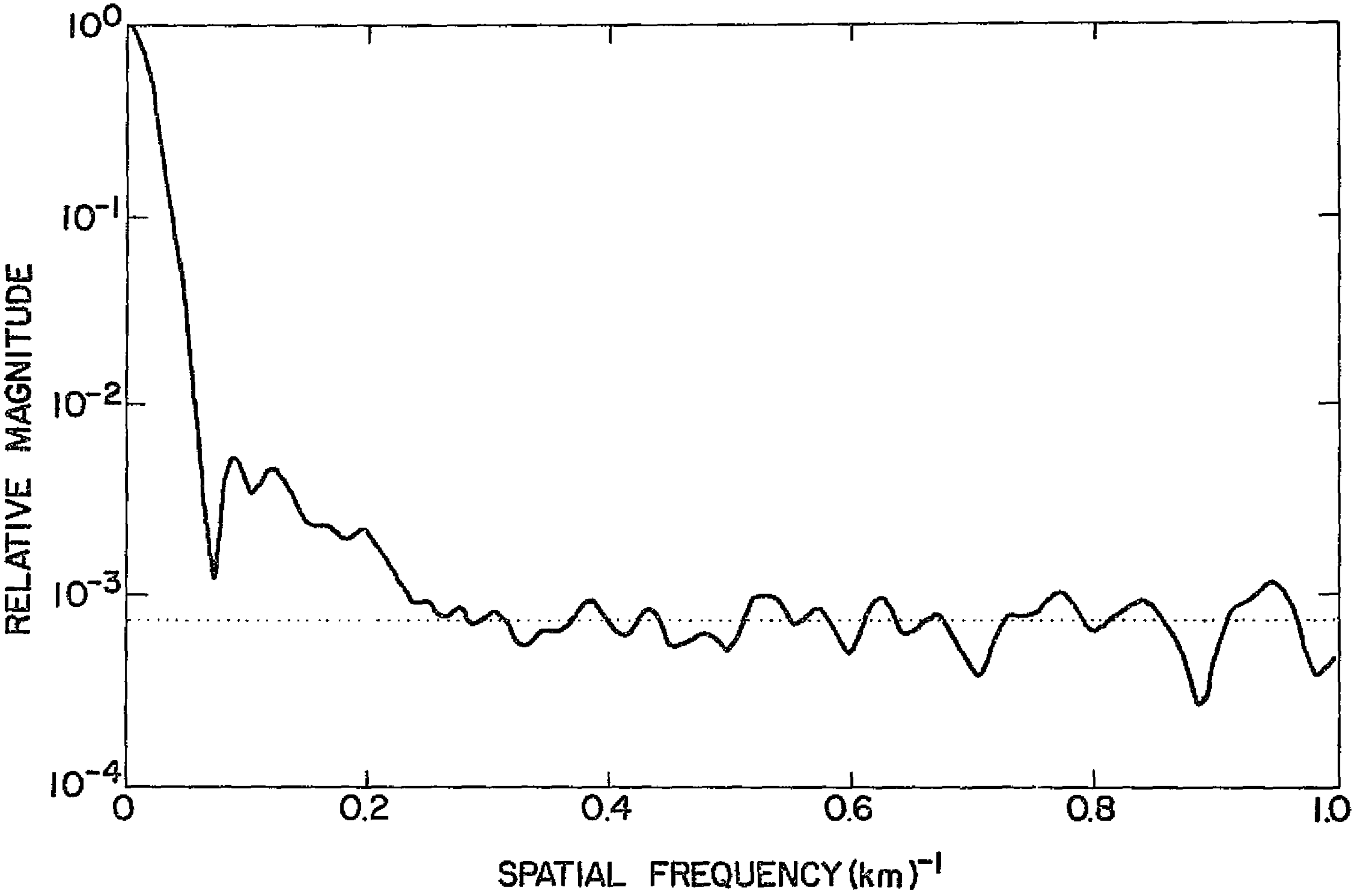


Figure 6.14 Averaged spectrum from the first half of January 3, 1978 data. Time period covers 2230-0100 hrs.

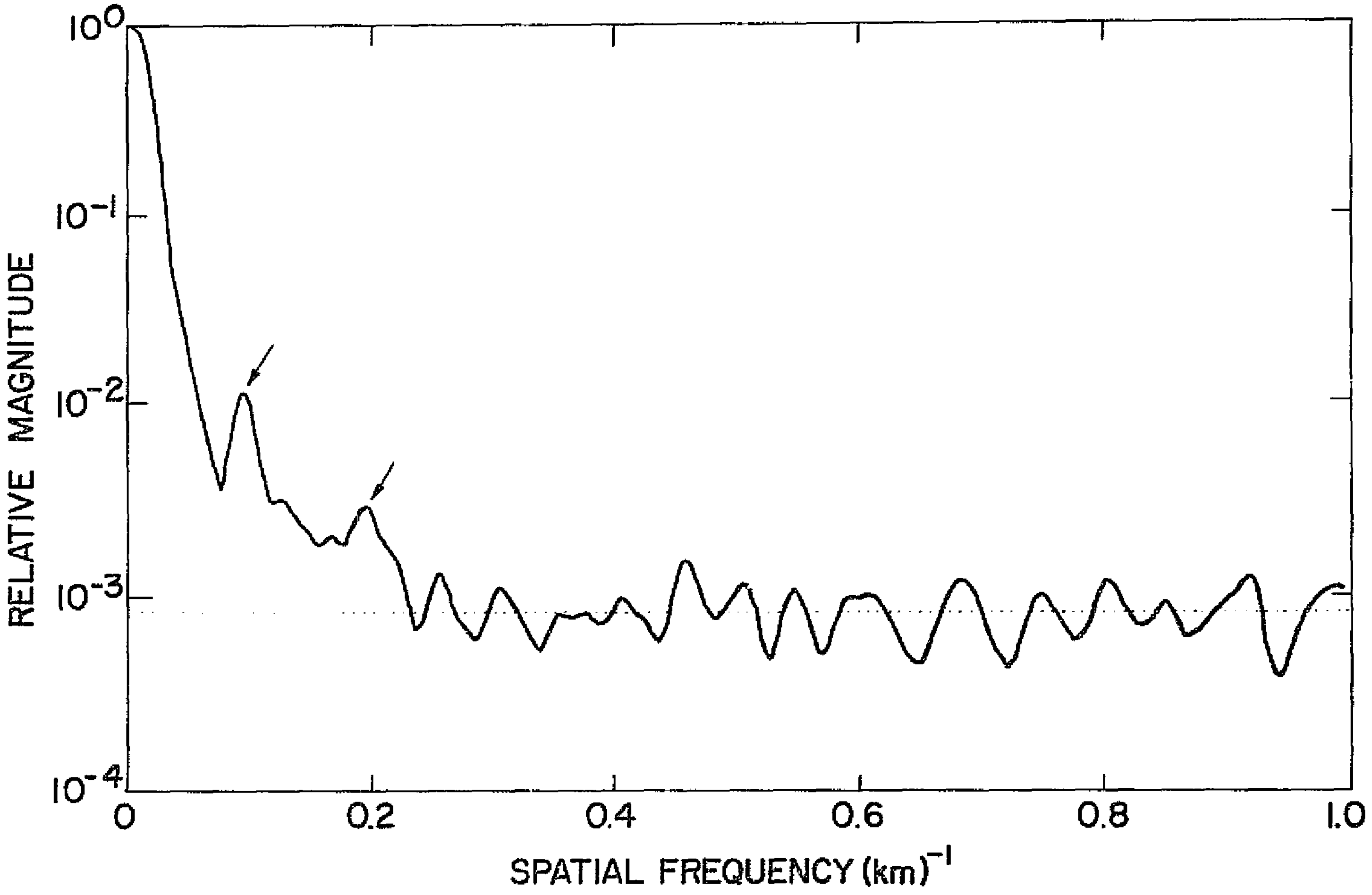


Figure 6.15 Averaged spectrum from the second half of January 3, 1978 data. Arrows indicate the growth of a spectral components which were not so pronounced in Figure 6.14. Time period covers 0012-0323 hrs.

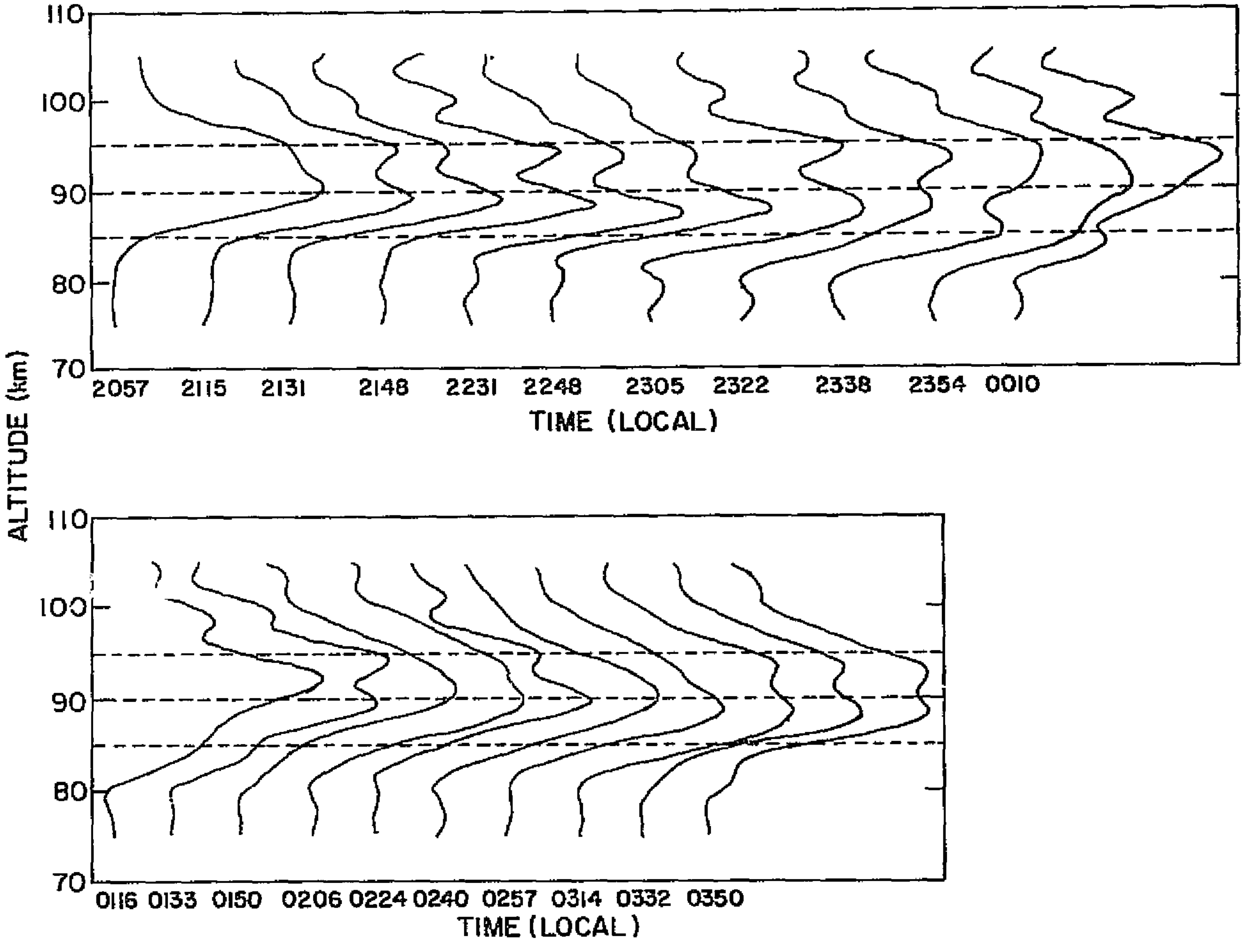


Figure 6.16 Time evolution of the sodium lidar data obtained during the evening of March 16-17, 1978. The receiver gate was 3 μs. The data have been corrected for the $1/Z^2$ dependence. Each set represents the integrated return from 400 shots and starting times of data collection are indicated.

broadening of the layer has also been observed by *Megie and Blamont* [1977] and the change in scale height with time is in good agreement with their observations.

6.4 *Purpose and Preliminary Tests for Joint Meteor Radar and Laser Radar Measurements*

The Aeronomy Laboratory Field Station meteor radar facility can be used to infer the neutral wind velocity in the 90 km region by measuring the radial Doppler shift resulting from specular reflection of the 41 MHz transmitter pulse from the ionized trails of meteors. Central to the meteor-ion theory described Chapter 7 is the wind shear mechanism. It becomes important, therefore, to determine if the wave motion detected in the sodium layer corresponds to the motion of the ambient neutral atmosphere or results from layers of ions caused by the wind shear mechanism. Since the mechanism is sensitive to the direction of the wind, and this direction can be sensed by the meteor radar experiment, it should be possible to lend experimental evidence to the corkscrew mechanism as a source of sodium. If the sodium is affected predominately by the ambient neutral atmospheric motion, then sodium may be an effective tracer for atmospheric motion at nighttime since the layer is under dynamical control at this time. The verification of this hypothesis could be made with the meteor radar facility.

The joint experiment is not a trivial procedure because the laser and lidar data collection system reside in the meteor radar transmitter room. On the evening of October 19-20, 1977 we attempted a run in which the laser radar and meteor radar systems were operated alternately. This method was chosen to avoid possible meteor radar transmitter interference problems in the lidar receiver and to guarantee the laser operator's safety since the laser is in the vicinity of some high voltage points in the meteor radar transmitter. Lidar runs of 20 minutes duration were alternated with 40 minute meteor radar runs; and during this 40 minute period, the ambient temperature increased sufficiently to require laser tuning adjustments prior to each lidar run. In the summer it is expected that the ambient temperature during meteor radar runs could reach as high as 32°C.

Unfortunately the meteor radar data were lost as a result of a system malfunction. However the attempted run was useful because it was discovered that temperature control of the laser tuning optics would be required for joint runs. In addition it was observed that if the sodium cell monitor was moved to a remote location, the operator might safely monitor the laser emission. If tuning became necessary, the meteor radar transmitter could then be turned off to perform the adjustment. This arrangement would greatly extend the operating time of both systems.

On January 10, 1978 such a simultaneous run was attempted, but we found that the lidar receiver system was severely affected by the operation of the meteor radar unit; the interference could be eliminated if the data collection system was removed from the transmitter room. In addition the meteor radar transmitter prevented the laser from pulsing properly. Much time was spent shielding properly the pulsing unit but the effort was not successful. Hence it was concluded a simultaneous run between the two systems would require that the laser be moved to a room shielded from the fields of the meteor radar transmitter.

6.5 *The Pre-Sunrise Sodium Layer*

As mentioned in Chapter 1, the concentration of sodium is believed to be proportional to electron density and atomic oxygen. In particular the steep gradient of the bottomside of the sodium layer may coincide with the electron-density ledge which is known to exist in the 80-85 km region. A simple test of this theory would be to monitor simultaneously the bottomside of the sodium layer and the electron-density ledge. An appropriate experiment for performing the latter measurement would involve the ionosonde. Unfortunately the Aeronomy Laboratory Field Station ionosonde was not available for this study, and plans are now underway to reactivate the unit. In the future, incoherent scatter radar measurements of the electron-density profile will be conducted thus permitting detailed study of the electron-density ledge.

Since a suitable experiment was not available for monitoring the electron density at night, it was decided to probe the *D* and *E* regions with a partial reflection experiment operating at 2.66 MHz. The partial reflection technique, though, is a daytime experiment, while the lidar operates only at nighttime. As a compromise it was decided to attempt a near simultaneous run at dawn. Since the *D* region is not formed at this time, partial reflections would not occur but perhaps main reflections from ionized layers in the lower *E* region would be present. Furthermore the pre-sunrise measurements would allow an opportunity to observe any significant changes in the sodium layer, possibly due to rapid chemical changes in the sunlit atmosphere.

Gough [1975] made the interesting suggestion that photoionization of sodium atoms can be a significant source of ionization in the upper *D* region during twilight. If this is true, then it follows that the sodium layer may be responsible for the initial buildup of ionization in the 80-90 km region during the sunrise transition period.

The pre-sunrise sodium layer was monitored first on January 21, 1978 (Figure 6.17). The data indicate a significant notch in the profiles at about 85 km. On February 22, 1978 (Figure 6.18) the pre-sunrise layer was again monitored and this notch became prominent about 45 minutes before ground sunrise. At 23 minutes before ground sunrise the returns from the

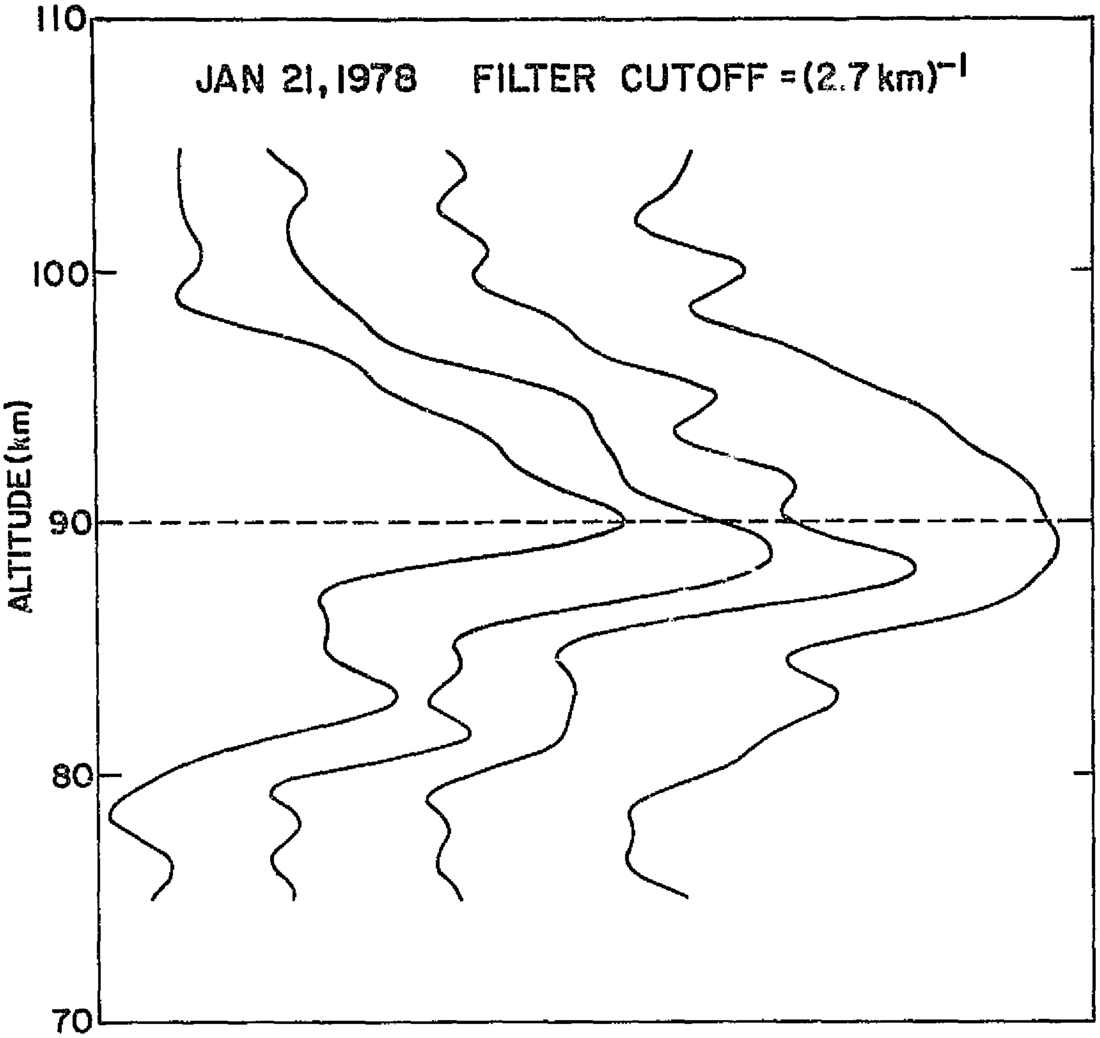


Figure 6.17 Sunrise sodium layer. Time of profiles left to right are 0537-0545, 0549-0559, 0602-0610, and 0612-0622, respectively.

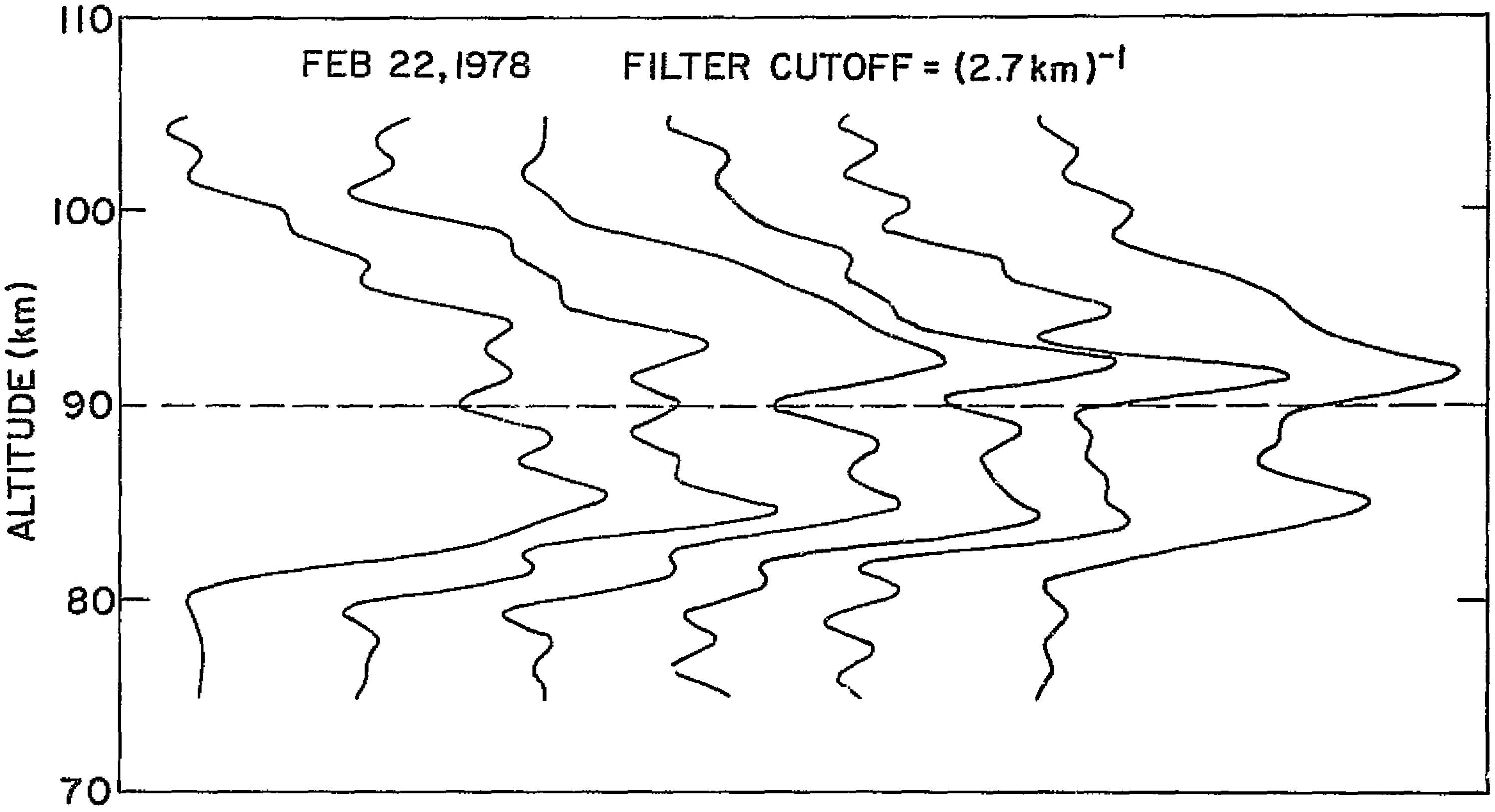


Figure 6.18 Sunrise sodium layer. Time of profiles left to right are 0425-0436, 0439-0510, 0454-0507, 0510-0527, 0523-0534, and 0537-0549 respectively. Ground sunrise is 0636 for this day.

2.66 MHz radar system indicated a strong echo from 87 km and a weaker echo from 93 km (Figure 6.19). These returns are related to the gradient in electron density and coincide with the negative gradient in the sodium returns below the notch and above the peak of the sodium layer in the last profile of Figure 6.18.

If one assumes a screening height of 40 km and neglects the atmospheric refraction of sunlight, then the geometry for this sunrise condition is shown in Figure 6.20. Taking the radius of the earth to be 6.37×10^3 km, one finds that for solar zenith angles of 97.5° and 96.4°, the altitudes of 95 km and 80 km respectively are first illuminated. Since the earth rotates at 4 min deg^{-1}, these heights are illuminated 30 minutes and 26 minutes respectively before ground sunrise. Sodium profiles on both occasions could not be taken after 45 minutes before sunrise.

If the layer did not change from the time of the last profile of Figure 6.18 to the time that the photograph in Figure 6.19 was taken ($\Delta t \sim 25$ min), then it is possible that ionization of the sodium is contributing to the increase in the electron density in the upper *D* region during sunrise. In this case sodium would be a significant source of ionization.

6.6 *Conclusion*

The experimental results of this chapter indicate the highly variable nature of the sodium layer. Wavelike structures have been observed in the layer by the use of frequency domain filters which can be used to generate continuous density profiles from the histogram data. These waves have wavelengths of 3-15 km and may travel downward with a phase velocity of 50-100 cm s^{-1}. On one occasion there was evidence for the presence of a standing wave in the layer. During two successive evenings the bottomside of the layer was observed to oscillate with a period of about 2 1/2 hours.

ORIGINAL PA[illegible]
OF POOR QUALITY

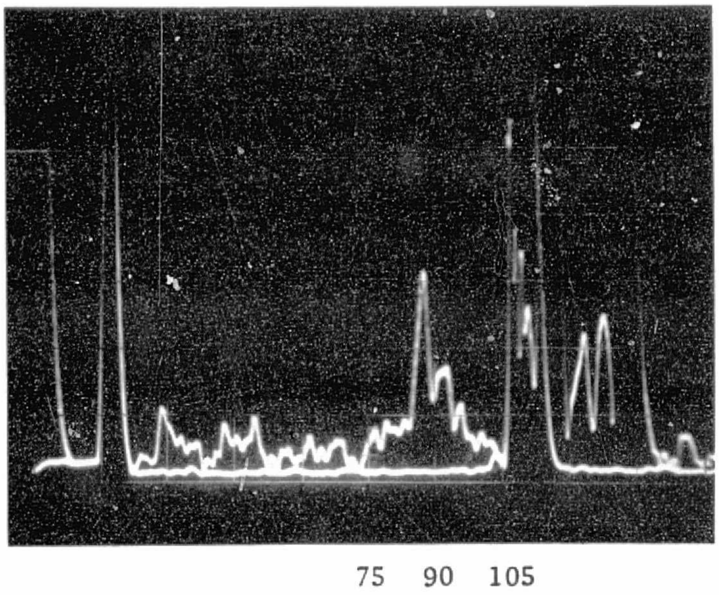


Figure 6.19 A-scope trace of returns from the 2.66 MHz transmitter taken on February 22, 1978 at 0613 hours. The main reflection from 105 km is due to the *E* layer. The reflection from 87 km was intermittent approximately 20 minutes prior to the taking of this photograph. Ground sunrise is at 0636 hours for this day.

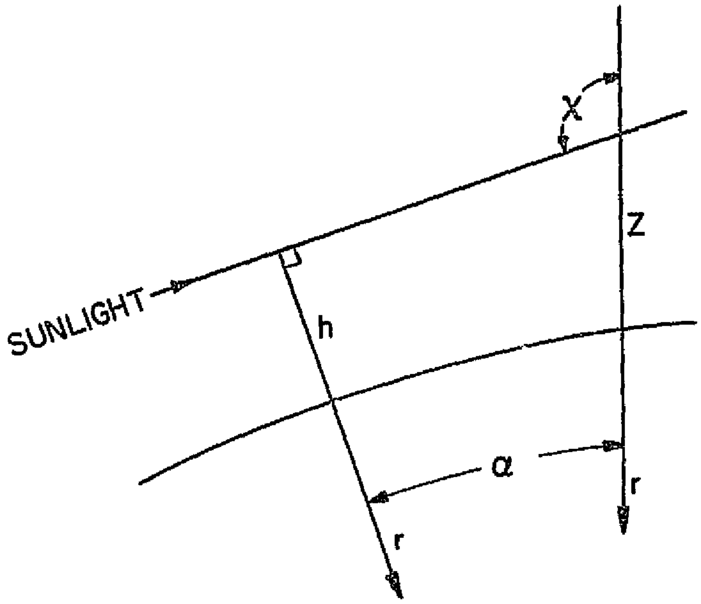


Figure 6.20 Geometry of sunrise under simplest conditions. The screening height is indicated by h. The altitude of the point of observation is z and the radius of the earth is r. The solar depression angle is $\alpha = \chi - 90°$ where χ is the solar zenith angle.

Many observations of the layer are also in agreement with those of *Megie and Blamont* [1977]. The layer is observed to broaden through the night and the topside scale height has been observed on one occasion to change on the average of 400 m hr^{-1}. In addition it appears that as the altitude of the peak of the layer lowers, the topside scale height increases.

7. A METEOR ABLATION-CLUSTER ION SODIUM THEORY

7.1 *Introduction*

The experimental data presented in the previous chapter indicate the highly variable nature of the neutral sodium layer. From rocket mass spectrometer measurements and incoherent scatter radar measurements, it is known that Na^+ resides in approximately the same altitude region. Furthermore, charged metallic species are dynamically controlled by an interaction of the neutral wind system and the geomagnetic field. This control evidences itself by the formation of metallic layers and their subsequent downward motion. It is therefore possible that variations in Na^+ are responsible for at least some of the perturbations in the Na layer.

However the theory demands that the return of Na^+ to Na be rapid and previous chemical schemes for sodium do not predict this (Appendix II). But an important part of the sodium chemistry has been largely neglected. Cluster ion reactions have been shown to occur in the laboratory and evidence exists for their presence in the upper atmosphere. In this work it is shown that the cluster ion scheme may provide a fast return of Na^+ to Na and thus remove a major obstacle in the theory which relates the dynamics of Na^+ to those of Na.

Section 7.2 will cover the meteor ablation source of sodium and the downward sweeping of ions. Diffusion time scales are discussed in Section 7.3. A cluster ion scheme is developed in Section 7.4. In Section 7.5 a chemical equilibrium model is presented and particular attention is given to the seasonal variation of sodium and the lack of diurnal variation. The experimental results are reviewed in Section 7.6 in the light of the meteor ablation-cluster ion theory. Finally the chapter is concluded with a brief summary in Section 7.7.

7.2 *The Source of Sodium and the Downward Sweeping of Sodium Ionization*

7.2.1 *Theories for the source of sodium.* The source of metallic elements in the upper atmosphere is a subject of much controversy. Theories have proposed either a terrestrial origin, such as a particulate matter from evaporated sea spray [*Allen,* 1970], or an extraterrestrial source. In the latter case another contested debate concerns the form of this source. *Donahue and Meier* [1967] suggest that it results from a layer of dust containing aerosols with radii less than 10^{-6} cm having a residence time of several years. *Hunten and Wallace* [1967] suggested that this layer of dust could also act as a sink for sodium. Usually any hypothesis about a common source for the atmospheric metallic species is interpreted in the light of the numerous studies of the sodium layer. *Fiocco and Visconti* [1973] discuss the seasonal variation of sodium in the light of the dust theory and present a numerical model of the sodium layer which involves both photochemistry and transport [*Fiocco and Visconti,* 1974]. They also indicate that the dust layer source may be either extraterrestrial or terrestrial in nature.

Gadsden [1967, 1968, 1969, 1970] has outlined a theory which employs meteor ablation as a source of sodium. His model includes a downward sweeping of ionization but depends largely on the shape of the meteoric deposition profile to predict the general features of the sodium layer. A sink for sodium is assumed below the ablation region in a portion of the atmosphere where water vapor exists. The total influx of meteoric material required to sustain the sodium layer is calculated and is found to be in agreement with similar values derived from different methods. Although it is thought that the deposition of sodium from meteor ablation is in the atomic form, the truth is not certain. The consequences of significant Na^+ deposition resulting from meteoric ablation are discussed in Section 7.6. Also an apparent weakness of Gadsden's method is its inability to explain the seasonal variation of sodium. The impact of a cluster ion scheme on this process is discussed in Section 7.5.

Megie et al. [1978] have shed some new light on the source of sodium by their simultaneous measurements of sodium and potassium using the lidar technique. They find that the abundance ratio (Na/K) varies seasonally, being a factor of 4 to 5 higher during the winter; and they argue persuasively for a dual source of sodium. A common meteoric source is taken to be constant over the year; and a terrestrial source resulting from a high latitude, winter vertical transport of salt particles, supplements the Na concentration. This conclusion is based on the assumption that the chemical and ionic properties of sodium and potassium are very similar. However, the cluster ion scheme developed for this work may make this assumption invalid and this point is discussed further in Section 7.6.

The most convincing arguments for the meteoric source are the clear correlations of metallic ion enhancement during meteor showers [*Narcisi*, 1968; *Zbinden et al.*, 1975] and of neutral sodium [*Hake et al.*, 1972; *Megie and Blamont*, 1977]. In addition the relative atmospheric abundances of metallic ions are roughly consistent with corresponding relative abundances in meteoric material [*Brown*, 1973; *Herrmann et al.*, 1977]. For these reasons and the sweeping mechanism described earlier, we will adopt meteor ablation as the source of sodium.

7.2.2 *Ablation of meteors.* The process of meteor ablation as it applies to the sodium problem has been well presented [*Gadsden*, 1968, 1969]. The rate of deposition per unit height of meteoric material is a function of initial radius and velocity of the meteor. While atoms may be released by the process higher than 250 km, the main contribution is in the 80-120 km region. In general meteors having velocities greater than 40 km s^{-1} and radii of a few microns ablate above 100 km [*Gadsden*, 1967].

The number of meteors, $n(r)$, entering the earth's atmosphere per unit time per area is related to the radius, r by

$$n(r) \propto r^{-4} dr \tag{7.1}$$

where the exponent is subject to some uncertainty. Since the ablation rate is approximately proportional to r^3, it is obvious that small meteors are just as likely to contribute to the deposition as large ones. A typical entry velocity for many meteors is near 20 km s^{-1}. The number of atoms per unit height interval ablated from a meteor having a starting radius of 100 μm is plotted in Figure 7.1. *Gadsden* [1968] shows that this deposition is only weakly dependent on the velocity and initial direction of the meteor but that the height at which ablation occurs is strongly dependent upon the initial velocity. In a later model calculation, *Gadsden* [1970] shows that meteor ablation can explain the small scale height on the topside of the layer if the gradient of the deposition profile is chosen correctly. Although Gadsden included the downward sweeping of ions in his calculation, he did not consider the layering of ions which can occur. The 95 km region is an area where sodium ions are dumped [*Rowe*, 1973; *Smith and Mechtly*, 1972], hence the wind shear mechanism may also play a role in the control of the topside scale height, perhaps perturbing it as layers of ions move downward. Furthermore since several parameters can be adjusted in Gadsden's calculation to achieve a reasonable sodium layer profile, it is not clear if the deposition profiles have been developed to agree solely with sodium. The downward sweeping of ions then provides the opportunity to collect Na^+ in the 95 km region while keeping the process largely independent of the ablation profile.

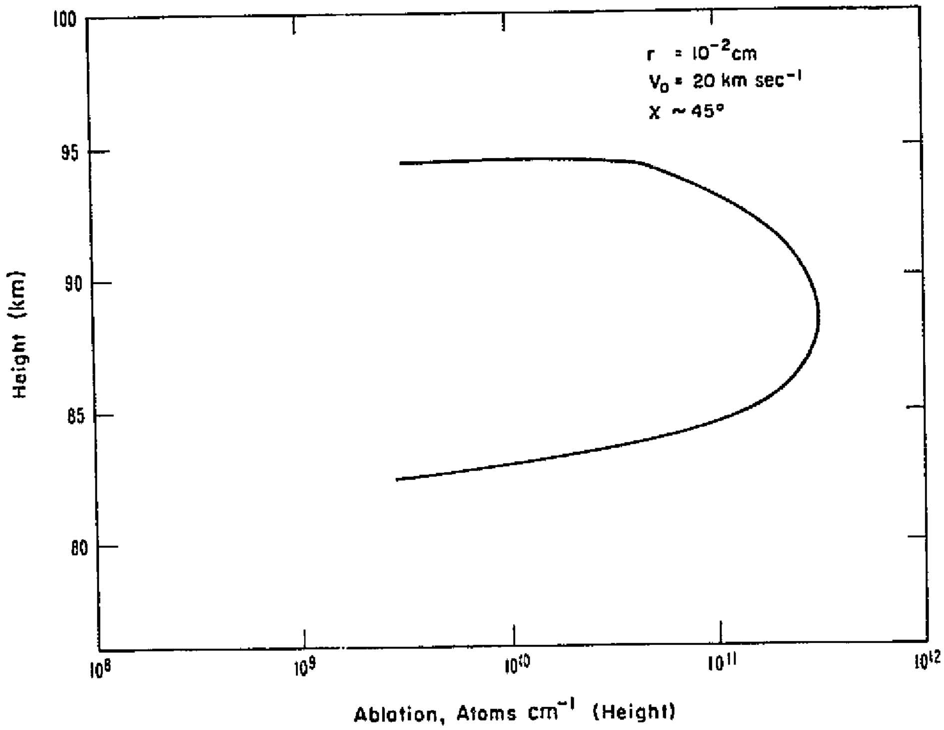


Figure 7.1 Ablation rate: atoms boiled off per unit height interval [*Gadsden*, 1968].

ORIGINAL PAGE IS
OF POOR QUALITY

7.2.3 *Na^+ lifetime.* If sodium enters the atmosphere above 100 km, the eddy transport of the neutral species is greater than 5 days. However, the lifetime of Na against photoionization,

$$\mathrm{Na} + h\nu \rightarrow \mathrm{Na}^+ + e^- \qquad I = 1.7 \times 10^{-5}\ \mathrm{s}^{-1} \tag{7.2}$$

and daytime charge exchange

$$\mathrm{Na} + \mathrm{O_2}^+ \rightarrow \mathrm{Na}^+ + \mathrm{O_2} \qquad k = 6 \times 10^{-10}\ \mathrm{cm}^3\ \mathrm{s}^{-1} \tag{7.3}$$

$$\mathrm{Na} + \mathrm{NO}^+ \rightarrow \mathrm{Na}^+ + \mathrm{NO} \qquad k = 8 \times 10^{-11}\ \mathrm{cm}^3\ \mathrm{s}^{-1} \tag{7.4}$$

is 16 hours and 10 hours respectively where it is assumed that $[\mathrm{NO}^+] = [\mathrm{O_2}^+] = 4 \times 10^4\ \mathrm{cm}^{-3}$. Hence it is clear that sodium is rapidly ionized during the daytime. Direct recombination of Na^+ with electrons has a rate constant $k \approx 10^{-12}\ \mathrm{cm}^3\ \mathrm{s}^{-1}$ and is obviously too slow of a deionization process. Furthermore other reactions which convert Na^+ to a shorter lived ion are slow because of the rarity of the atmosphere at this altitude. The only effective loss of Na^+ at midlatitudes is thus the downward transport of ions by the corkscrew mechanism.

7.2.4 *Vertical movement of long lived ions: the corkscrew mechanism.* *Dungey* [1956] first described an interaction of the wind system and the geomagnetic field which resulted in the formation of ionized layers in the ionosphere. Since that time, the wind shear mechanism has been developed in quite some detail [e.g. *Whitehead,* 1967; *Axford,* 1967] and has received experimental support [*Rowe,* 1973, 1974].

In addition to the formation of the layers, there is a net downward motion of these layers. This process has been termed the "corkscrew" mechanism because the effect is analogous to the pulling of a cork from a bottle by a corkscrew. This drawing down of ionization from the ionosphere applies to long lived ions, such as Na^+, and the process may dump these ions into the lower *E* and *D* regions. The theory has been developed in *Axford* [1961, 1963, 1967].

The predominant winds in the *E* region are those caused by the solar tides. In addition, perturbations in this large scale wind are caused by gravity waves which originate below the *E* region. The latter have been suggested as a source of sporadic-*E* layers [*Chimonas*, 1971]. In both cases the downward directed phase velocity is thought to be responsible for the downward motion of the ionization. If we neglect recombination and photoionization, we find that the corkscrew mechanism behaves differently in the *F* region than at lower altitudes. In both cases vertical movement of ions occurs, but different wind shears are involved.

At the highest altitudes the ions are constrained to move along the geomagnetic lines of force because the collision frequency of ions with neutrals is much less than their gyrofrequency. Hence the component of the wind parallel to the magnetic field drives the ions to form layers near the nulls of the wind profile. For midlatitudes it is the (geomagnetic) N-S winds which are responsible for this convergence.

In the lower *E* region, calculations indicate that the vertical motion of the ions is dependent upon the (geomagnetic) E-W wind system [*Axford and Cunnold*, 1966]. Since the collision frequency to gyrofrequency ratio for ions is large, positive ions move vertically because of the vertically induced electric fields which result from the E-W winds and the N-S component of the geomagnetic field. The ratio is still small for electrons, and therefore they continue to move along the geomagnetic lines of force. Since a vertical component to the magnetic field exists at midlatitudes, charge neutrality can be maintained.

7.2.5 *Equation of motion.* The motion of an ion is to a great extent affected by collisions with neutrals and interactions with electric and magnetic fields. The equation of motion for an ion is given by

$$m_i \frac{d\vec{v}_i}{dt} = q(\vec{E} + \vec{v}_i \times \vec{B}) + \nu m_i(\vec{u} - \vec{v}_i) - \frac{1}{n_i}\nabla p_i + m_i\vec{g} \tag{7.5}$$

where

m_i, q, $\vec{v}_i$ = mass, charge and velocity of the ion
$\vec{E}$, $\vec{B}$ = electric and magnetic field
ν = ion-neutral collision frequency
n_i, p_i = number density and partial pressure of the charged particle
$\vec{g}$ = gravitational acceleration.

It is apparent that the collision frequency term is a simplified expression. In this case $m\nu(\vec{u} - \vec{v}_i)$ is taken to be the force per unit volume exerted on the ions by collisions with the neutral gas. Collisions between ions may be neglected because of the large neutral to ion concentration ratio.

A number of simplifications and approximations may be made. The magnetic field may be taken as being equal to the geomagnetic field and is written as

$$\vec{B} = B_0\vec{1} \tag{7.6}$$

where B_0 is the field magnitude and $\vec{1}$ is a unit vector in the direction of the geomagnetic field. This is reasonable since at midlatitudes, ionospheric currents produce only minor perturbations in the magnetic declination [*Cloutier and Sandel*, 1972].

If equation (7.5) is divided by qB_0, then it becomes

$$\frac{1}{\omega_i}\frac{d\vec{v}_i}{dt} = \left(\frac{\vec{E}}{B_0} + \vec{v}_i \times \vec{\Gamma}\right) + \frac{1}{R_i}(\vec{u} - \vec{v}_i) - \frac{1}{q_i B_0 n_i}\nabla p_i + \frac{\vec{g}}{\omega_i} \qquad (7.7)$$

where $\omega_i = qB/m_i$ is the ion gyrofrequency and R_i is the ratio of the ion gyrofrequency to collision frequency. Since the gyrofrequency for Na^+ is large ($\omega_i = 230\ s^{-1}$) the acceleration and gravity terms may be neglected [*MacLeod et al.*, 1975].

By comparing the pressure term with $\vec{v}_i/R_i$ it is apparent that the former term may be neglected [*Dungey*, 1956]. This term becomes important at higher altitudes as temperature grows; or if the shape of the particular layer were a concern, then the pressure term would be important.

Finally by assuming that the neutral wind is a function of time and altitude, the vertical ion drift velocity may be obtained from the steady state solution of equation (7.5)

$$q(\vec{E} + \vec{v}_i \times B_0\vec{\Gamma}) + m_i\nu(\vec{u} - \vec{v}_i) = 0. \qquad (7.8)$$

To analyze this equation, consider the right handed coordinate system of Figure 7.2. The vertical coordinate is taken to be $\hat{z}$ with $\hat{x}$ and $\hat{y}$ pointing magnetically east and north respectively. Then $\vec{B} = B_y\hat{y} + B_z\hat{z}$ where $|\vec{B}| = B_0$. The neutral wind is assumed to have only horizontal components and hence $\vec{u} = u(z,t)\hat{x} + v(z,t)\hat{y}$. Equation (7.8) may now be solved for the vertical component, w of the ion wind $\vec{v}_i$. By writing a similar expression for electrons and equating the vertical velocity of ions to that of electrons so that charge neutrality is maintained, *Chimonas* [1973] obtained the expression for w at midlatitudes

$$w \approx \frac{(B_y/B_0)R_i}{1 + R_i^2}\left[u(z,t) - R_i v(z,t)\ B_z/B_0\right] \qquad (7.9)$$

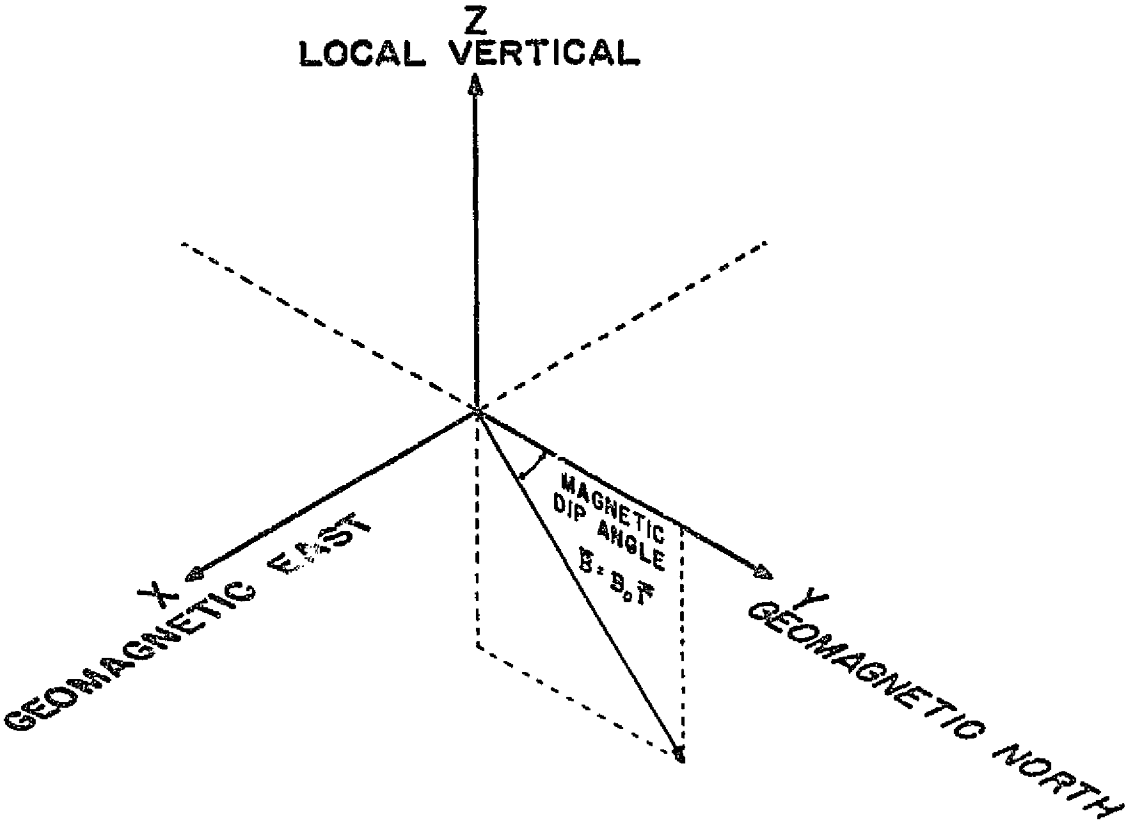


Figure 7.2 Orientation of the coordinate system employed [*MacLeod*, 1966].

where $R_i = q_i B_0 / m_i \nu_i$. The quantity R_i increases with altitude as shown in Figure 7.3.

Generally $u(z,t)$ and $v(z,t)$ have the form of tidal winds. Therefore, u (and analogously v) may be expressed as

$$u(z,t) = U_0 \exp[(z - 120/2H] \sin k(z - 120 - v_p t) \qquad (7.10)$$

where H is the local scale height and v_p is the downward directed phase velocity. Typical values for these constants for the midlatitude E region are

$U_0 = 50 \text{ m s}^{-1}$

$H = 10 \text{ km}$

$k = 2\pi/\lambda = .2 \text{ km}^{-1}$

$v_p = -1 \text{ m s}^{-1}$

$z - 120$ = distance in km from 120 km

$B_y/B_0 = .45$

[*Chimonas*, 1973; *Axford*, 1963].

To determine a numerical value for R_i, values for B_0 and ν_i are needed. For midlatitudes, $B_0 = 5.5 \times 10^{-5} T$ Gauss [*Dudziak et al.*, 1963]. Unfortunately the definition of the collision frequency, ν, may take several forms [*Stubbe*, 1968]. One must keep in mind that each definition is a convenient quantity having dimension s^{-1}. From the work of *Banks* [1966] the approximation for ν_i is given as

$$\nu_i = 6.94 \left(\frac{\eta q_i^2}{\mu_{in}} \right)^{\frac{1}{2}} \quad [X] \qquad (7.11)$$

where

η = polarizability of the neutral particle

q_i = ionic charge

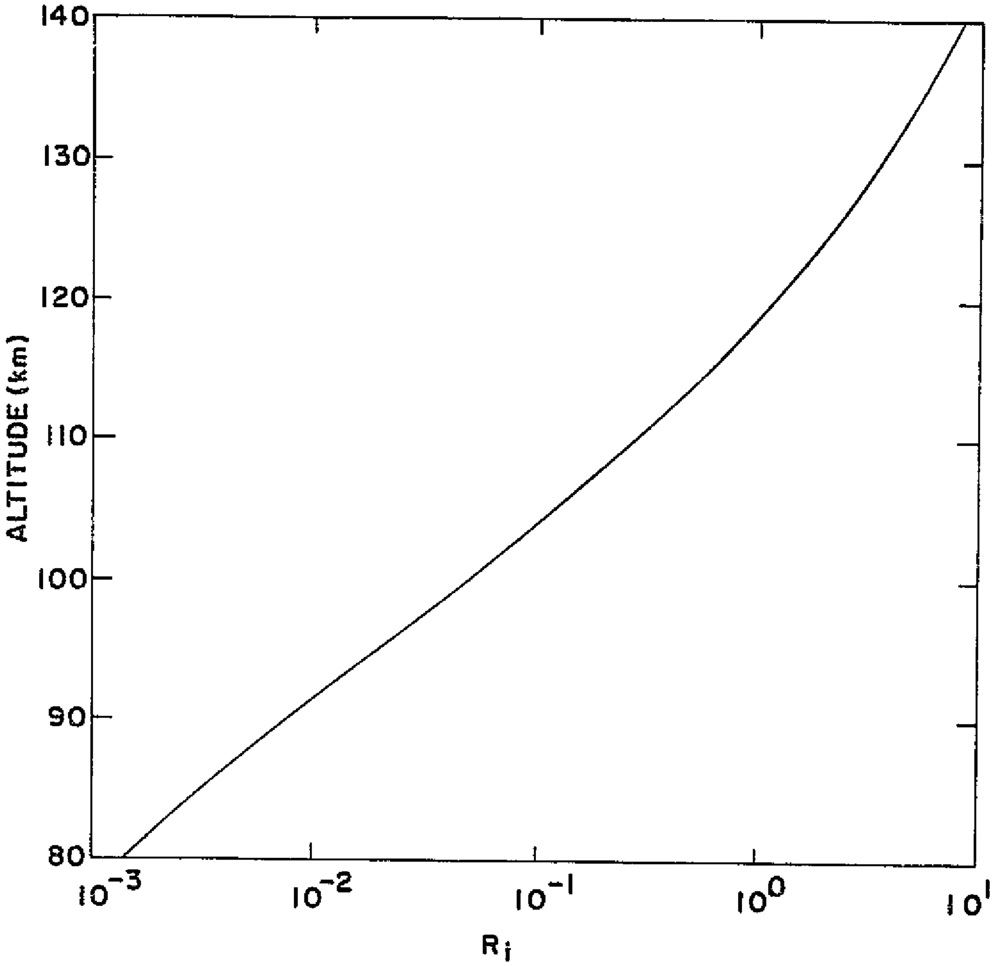


Figure 7.3 Variation of $R_i = q_i B_0 / m_i \nu$ with altitude.

μ_{in} = $m_i m_n/(m_i + m_n)$ = reduced mass

m_i, m_n = ion and neutral particle mass

[X] = neutral number density.

Using the definition of *Dalgarno* [1961] $\nu_i = kT/m_i D_i$ and the expression for the diffusion coefficient D_i given by *Chapman and Cowling* [1970], the collision frequency becomes

$$\nu_i = 6.94 \frac{\mu_{in}}{m_i} \left(\frac{\eta q_i^2}{\mu_{in}}\right)^{\frac{1}{2}} [X] \tag{7.12}$$

Finally *Stubbe* [1968] defines a "real" collision as one in which two particles have direct contact. Under these conditions

$$\nu_R = 2\pi \left(\frac{\eta q_i^2}{\mu_{in}}\right)^{\frac{1}{2}} [X] \tag{7.13}$$

For our purposes we will choose, somewhat arbitrarily, the value of ν_i given by equation (7.12). If the neutral atmosphere is taken to consist solely of O_2 and N_2, then

$$R_i = \frac{1}{1.89 \times 10^{-12}[O_2] + 1.9 \times 10^{-12}[N_2]} \tag{7.14}$$

where the number densities have units of cm^{-3}. Implicit in this calculation is the use of the values for the polarizability of O_2 and N_2 given in *Banks* [1966], i.e. $\eta_{N_2} = 1.76 \times 10^{-24}$ cm^3 and $\eta_{O_2} = 1.59 \times 10^{-24}$ cm^3. The variation of R_i with altitude is plotted in Figure 7.3.

The ion velocity may now be easily calculated as a function of time and altitude using equation (7.9) and (7.14). The profile for the ion drift velocity for $t = 0$ is shown in Figure 7.4 and compares favorably with the experimental results of *Geller et al.* [1975].

From Figure 7.4 it is apparent that above 100 km large vertical ion drift velocities will cause the formation of ionized layers. This wind profile then moves downward, drawing the ionization along. Below 100 km the mechanism breaks down and the ions are "dumped" in this region. More appropriately the ions are no longer able to continue traveling with the phase velocity of the wave but rather move downward with a velocity less than v_p. However a net downward movement still exists as each passing wave lowers the ions left behind from the previous wave.

7.3 *Eddy and Molecular Diffusion*

7.3.1 *Expressions*. Since Na and Na^+ are both distributed in narrow layers, there is always a tendency for these constituents to try to reduce the gradients of their boundaries. Molecular diffusion describes the transport of gas when the average molecular velocity of the minor constituent differs from that of the background gases; the process tends to separate the constituents according to their mass. Eddy diffusion accounts for differences in composition and describes the motion of molecules due to turbulence; this tends to mix the constituents and to restore the minor constituent's distribution height to the atmospheric scale height. Before chemical equilibrium can be assumed, the time constants for diffusion must be compared to the chemical time constants.

It is a reasonable approximation to assume that sodium diffuses through a stationary N_2 atmosphere. The flux of this $i^{\underline{th}}$ constituent may be expressed as [*Colegrove et al.*, 1965]

$$\phi_i = -D\left[\frac{dn_i}{dz} + \left(\frac{1}{T}\frac{dT}{dz} + \frac{1}{H_i}\right)n_i\right] \tag{7.15}$$

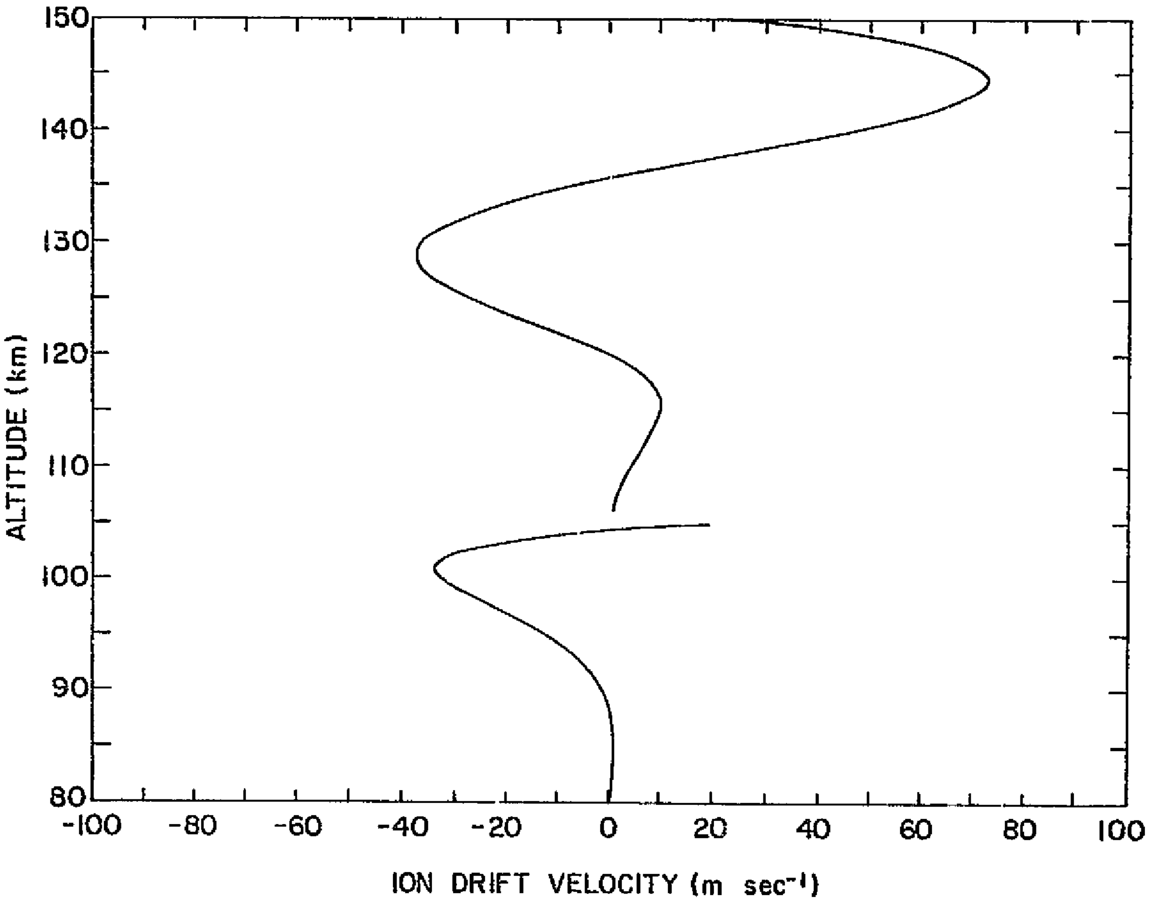


Figure 7.4 Theoretical values for Na^+ drift velocity. The neutral winds are assumed to be zero at 120 km. Drift velocities below 105 km are multiplied by 100.

where

D = molecular diffusion coefficient

H_i = $kT/m_i g$ = scale height of the $i^{\underline{\text{th}}}$ constituent

m_i, n_i = mass and concentration of the minor constituent

g = gravitational acceleration

k = Boltzmann's constant

T = absolute temperature

z = altitude.

In an analogous manner the flux due to eddy diffusion may be written as

$$\phi_i = -K\left[\frac{dn_i}{dz} + \left(\frac{1}{T}\frac{dT}{dz} + \frac{1}{H}\right) n_i\right] \tag{7.16}$$

where

K = eddy diffusion coefficient

H = scale height of the background atmosphere.

7.3.2 *Diffusion coefficient values.* If we treat the molecules as perfect spheres, then the molecular diffusion coefficient may be calculated by using gas kinetic theory [*Chapman and Cowling*, 1970].

$$D = \frac{3}{8[X]\sigma_{12}^2}\left[\frac{kT(m_1 + m_2)}{2\pi m_1 m_2}\right]^{\frac{1}{2}} \tag{7.17}$$

where

σ_{12} = $(\sigma_1 + \sigma_2)/2$ and σ_1, σ_2 are the molecular diameters of the minor and major gases respectively

k = Boltzmann's constant

T = absolute temperature

m_1, m_2 = masses of the two types of molecules

$[X]$ = molecular density of the main gas, taken to be N_2.

Unfortunately the values for many of the molecular diameters are not known. A method for estimating the diameters of simple constituents, which involves knowledge of the molecular bond, has been described for a simple sodium chemistry [*Mitra*, 1973]. However it is known that Na and Na^+ have collision diameters of 314 pm and 190 pm respectively. The altitude dependence of D_{Na} and D_{Na^+} is plotted in Figure 7.5. Molecules or ions containing sodium will have even lower molecular diffusion coefficients than those of Na and Na^+ respectively.

Values for the eddy diffusion coefficient are even more difficult to estimate. The problem is that all scales of motion are being parameterized by this coefficient. This includes not only local vertical mixing, but also larger structures such as meridional motions, large scale eddy fluxes, and tidal motions. The coefficient also accounts for horizontal as well as vertical motion. Obviously there is no satisfactory way to parameterize these motions into a single coefficient and this problem is the main disadvantage of the eddy diffusion coefficient. In fact values are often picked to give the best results for a particular measurement. In the 90 km region, values vary by about two orders of magnitude depending on the measurement technique. Analysis of chemical trails lead to K values slightly greater than 10^7 cm^2 s^{-1} [*Zimmerman and Trowbridge*, 1973]. Studies of inert gases such as helium or argon lead to much lower values between 10^5 and 10^6 cm^2 s^{-1}. For this work it will be assumed that $K = 3 \times 10^6$ cm^2 s^{-1} for the sodium region which is the same order of magnitude as that deduced by *Hunten* [1975] and derived by *Megie and Blamont* [1977] through their analysis of the Na layer dynamics.

7.3.3 *Comparison of time scales.* The turbopause is thought to occur slightly above the sodium layer. Above the turbopause, molecular diffusion dominates while below the turbopause eddy diffusion is most important.

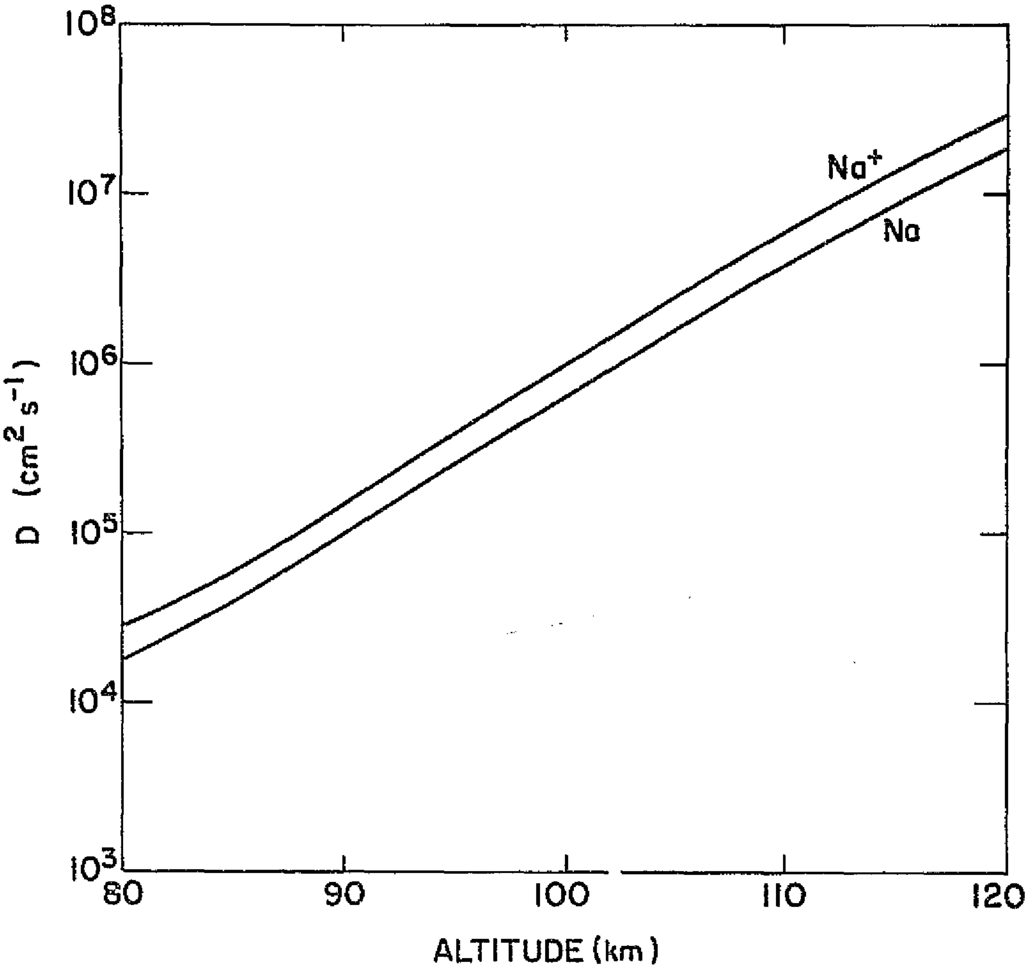


Figure 7.5 Molecular diffusion coefficients for Na and Na$^+$ assuming a *CIRA* [1972] atmosphere.

Referring to Figure 7.5 it can now be seen that most likely eddy diffusion will be the more important form of diffusion in the region of the sodium layer. If a minor constituent is distributed with a concentration $n = n_0 \exp(-z/H)$, then a diffusion time scale useful for comparing to chemical time constants is given by H^2/K. For H = 3 km which is appropriate for the topside scale height of the sodium layer, the eddy diffusion time scale has a value of about 8 hours. It will be shown that most of the cluster ion reactions at 95 km still have a reasonably fast time constant, on the order of 1 hour. The wind shear mechanism described in Section 7.2 will sweep ions downward with an ion velocity of about 30 cm s^1 [*Gadsden*, 1970]. The time scale for this process between 110 km and 90 km will be 20×10^5 cm/30 cm $s^1 \approx 20$ hours. Since the topside gradient lifetime is estimated to be 2½ times smaller than the time constant for the downward motion of Na^+, and Na^+ can be converted to Na rapidly, the topside scale height should vary as layers of ions are brought down. The topside may show a diurnal variation. More importantly, since cluster ion reactions generally are three body reactions, their time constants will decrease with decreasing altitude. The lifetime of Na against photoionization and charge exchange is less than ½ of a day, hence chemical equilibrium is an appropriate assumption for the daytime. However at night the lifetime of Na increases substantially and dynamics should play a more important role.

7.4 *The Sodium Cluster Ion Chemistry*

7.4.1 *Introduction*. The neutral and metal oxide ion chemistry is reviewed in Appendix II. A chemical equilibrium calculation, which includes 36 reactions, is performed and the calculation indicates that Na^+ should dominate at all altitudes. This result does not agree with experimental evidence and implies that the model is not a realistic chemical scheme because the return of Na^+ to Na is too slow.

In this section a new sodium chemistry based on cluster ion reactions is developed. The reactions are temperature sensitive and depend greatly on the atmospheric density since many of them are 3 body reactions. Both of these properties are investigated in Section 7.5. The scheme provides the fast chemical return of Na^+ to Na necessary to relate the dynamics of Na^+ to those of Na.

7.4.2 *Sodium cluster ion reactions*

7.4.2.1 *Temperature dependence.* Several measurements of sodium cluster ion reactions have been made and the results of these measurements are shown in Table 7.1. Three pieces of information are lacking to make the table experimentally complete: first, k_f is not known for the $Na^+ \cdot N_2$ cluster reaction; second, the relative efficiencies of third bodies is not given; and third, temperature relations for the rate coefficients are not tabulated. Referring to Figure 7.6 we see that the rates of Na^+ cluster reactions should be roughly equal to those of NO^+. Since we know $k_f = 2 \times 10^{31}$ cm^6 s^1 at room temperature for the analogous reaction

$$NO^+ + N_2 + N_2 \rightarrow NO^+ \cdot N_2 + N_2 \qquad (7.18)$$

[*Johnsen et al.*, 1975] we will take this to be the value for k_f in the $Na^+ \cdot N_2$ reaction. *Heimerl* [1976] has tabulated the ratios of rate coefficients relative to N_2 for third bodies X = He, Ar, O_2, and NO. In general these ratios (k_{N_2}/k_X) are between .9 and 1.5 with the exception of He which exhibits rates almost 4 times slower than N_2 [*Heimerl*, 1976]. Reactions involving H_2O and CO_2 are thought to proceed with the same rate as if N_2 were the third body.

Table 7.1

Experimental data available for cluster ion reactions

Cluster Reaction	$-\Delta H$ (kcal mol^{-1})*	Rate**	Source
$Na^+ + N_2 + N_2 \rightleftarrows Na^+ \cdot N_2 + N_2$	9.5	$K_c = 1(+1,-0.5) \times 10^{-18}$ cm^3 @310°K	[*Niles et al.*, 1972]
		$k_f = 2 \times 10^{-31}$ $k_r = 2 \times 10^{-13}$ @310°K	see text
$Na^+ + CO_2 + CO_2 \rightleftarrows Na^+ \cdot CO_2 + CO_2$	12.5	$k_f = 2 \times 10^{-29\pm0.5}$ $k_r = 1 \times 10^{-14\pm0.5}$ @310°K	[*Keller and Beyer*, 1971]
$Na^+ + O_2 + O_2 \rightleftarrows Na^+ \cdot O_2 + O_2$	5.7	$K_c = 7(+7,-4) \times 10^{-20}$ cm^3 @310°K	[*Niles et al.*, 1972]
		$k_f = 5 \times 10^{-32(+1.5,-0.5)}$ @310°K $k_r = 8 \times 10^{-13(+1.5,-0.5)}$	[*Keller and Beyer*, 1971]
$Na^+ + H_2O + H_2O \rightleftarrows Na^+ \cdot H_2O + H_2O$	24.6	$k_f = 1(+.5,-.3) \times 10^{-28}$ @310°K $k_r = 0$	[*Johnson et al.*, 1971]

*Calculated by *Spears* [1972]

**Forward rate constant has units cm^6 s^{-1}. Reverse rate constant has units cm^3 s^{-1}.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

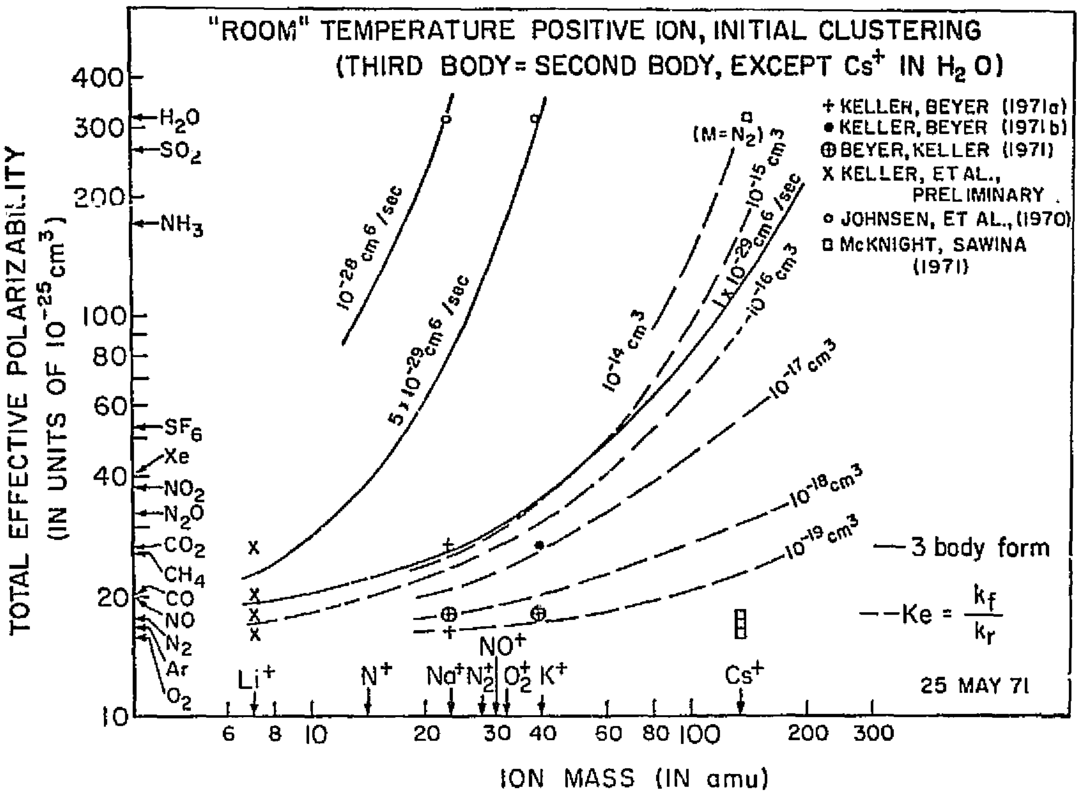


Figure 7.6 Map of minimum values for rate constants of the initial clustering reactions at about 300 K. The symbols denote measured values by the workers indicated [after *Niles et al.*, 1972].

The temperature relations for the rate coefficients may be estimated in the following manner. As mentioned in Appendix III the forward rate coefficient is proportional to $(T/{}_R)^b$ where b is between -1 and -2. We will therefore take $b = -1.5$ for all the cluster reactions of Na. It is more difficult to estimate the temperature variation of the reverse reaction. The activation term involves a change in enthalpy and therefore the exponential factor in equation (III.7) may be written as

$$\exp(-e/T) = \exp(+\Delta H/RT) \qquad (7.19)$$

Bond energies for $Na^+ \cdot O_2$, $Na^+ \cdot N_2$, $Na^+ \cdot CO_2$, and $Na^+ \cdot H_2O$ have been calculated by *Spears* [1972] and we may equate these values to the enthalpy change. Hence the temperature dependence of the reactions may now be established.

As an example consider the reaction which produces $Na^+ \cdot N_2$ (see Table 7.1). The forward rate coefficient will be taken as

$$k_f = 2 \times 10^{-31} \left(\frac{310}{T}\right)^{1.5} \text{cm}^6 \text{ s}^{-1} \qquad (7.20)$$

The change in enthalpy for this reaction is $-\Delta H = 9.5$ kcal/mole [*Spears*, 1972]. Applying equations (III.7) and (7.19) the reverse rate coefficient may be written as

$$k_r \propto \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-4750}{T}\right) \qquad (7.21)$$

where the constant of proportionality is to be determined. The extra factor of T results from the change in units of the equilibrium constant, as discussed in Appendix III. From Table 7.1, $K_c = 10^{-18}$ cm^3 at 310°K. Therefore the constant may be determined for 310° and the final expression for k_r becomes

$$k_r = 9 \times 10^{-7} \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-4750}{T}\right) \text{cm}^3 \text{ s}^{-1} \qquad (7.22)$$

One may perform several checks on the above calculation. For example one sees from equations (III.6), (III.7), (III.9), (III.10), and (III.14) that the constant factors in the rate coefficients are related to the change in entropy by the relation

$$\frac{\Delta S}{R} = \ln \frac{a}{cR(310^\circ\mathrm{K})} \tag{7.23}$$

Spears [1972] has calculated $-T\Delta S = 7.2$ kcal mol^{-1}. Taking $R = 1.99$ cal $(^\circ\mathrm{K})^{-1}$ mol^{-1} and $T = 310^\circ\mathrm{K}$,

$$\frac{\Delta S}{T} = -11.7 \tag{7.24}$$

For the right side of (7.23) we must take $R = 82$ cm^3 atm$(^\circ\mathrm{K})^{-1}$ mol^{-1} since a/c has units of cm^3. Using Avogadro's number of 6.02×10^{23} mol^{-1},

$$\ln \frac{a}{CR(310)} = \ln \frac{(2 \times 10^{-31})(6.02 \times 10^{23})}{(9 \times 10^{-7})(82)(310)} = -12.2 \tag{7.25}$$

This is reasonable agreement; the errors result from the simultaneous use of the theoretical calculations of *Spears* [1972] and the experimental results of *Niles et al.* [1972].

The other rate coefficients are obtained in a similar manner and are summarized in Table 7.2.

7.4.2.2 *Recombination.* Recombination coefficients for the Na cluster ions have not been measured. In general they are in the range of 10^{-6} to 10^{-5} cm^3 s^{-1} [*Leu et al.*, 1973], and it is thought that at temperatures below 300°K, the recombination rate coefficient displays a $T^{-0.5}$ dependence; the dependence is predicted by theory and established in the laboratory [*Bardsley*, 1970]. Only at higher temperatures (~1000°K) will the dependence take on a $T^{-1.5}$ law. Estimates for the recombination rate coefficient of $NO^+ \cdot H_2O$ have taken on values from 2×10^{-6} to 3×10^{-6} cm^3 s^{-1} at 300°K [*Arnold and*

Table 7.2

Assumed rate coefficients for cluster ion reactions

Reaction reference number*	Reaction	Rate Coefficients**
R_1	$Na^+ + N_2 + X \rightleftarrows Na^+ \cdot N_2 + X$	$k_1 = 2 \times 10^{-31} \left(\frac{310}{T}\right)^{1.5}$ cm^6 s^{-1}
		$k_{-1} = 3 \times 10^{-7} \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-4750}{T}\right)$ cm^3 s^{-1}
R_2	$Na^+ + CO_2 + X \rightleftarrows Na^+ \cdot CO_2 + X$	$k_2 = 2 \times 10^{-29} \left(\frac{310}{T}\right)^{1.5}$ cm^6 s^{-1}
		$k_{-2} = 6 \times 10^{-6} \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-6250}{T}\right)$ cm^3 s^{-1}
	$Na^+ + O_2 + X \rightleftarrows Na^+ \cdot O_2 + X$	$k_f = 5 \times 10^{-32} \left(\frac{310}{T}\right)^{1.5}$ cm^6 s^{-1}
		$k_r = 8 \times 10^{-9} \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-2850}{T}\right)$ cm^3 s^{-1}
	$Na^+ + H_2O + X \rightleftarrows Na^+ \cdot H_2O + X$	$k_f = 1 \times 10^{-28} \left(\frac{300}{T}\right)^{1.5}$ cm^6 s^{-1}
		$k_r = 0$

*Reactions lacking a reference number are not used in the final scheme.

**Positive subscripts denote forward reactions. Negative subscripts denote reverse reactions.

Krankowsky, 1977; *Reid*, 1977; *Thomas*, 1976]. Although *Keller and Beyer* [1971] suggested a rate coefficient of 5 × 10^{-6} $(T/300)^{-1}$ cm^3 s^{-1} for recombination of $Na^+ \cdot CO_2$ and $Na^+ \cdot O_2$ with electrons a more conservative view suggests that this be lowered. Therefore the recombination rate coefficient for all clusters will be taken to be $\alpha = 1 \times 10^{-6} (T/300)^{-0.5}$ cm^3 s^{-1}.

Recombination with negative ions can also take place and these rate coefficients have values on the order of 10^{-7} cm^3 s^{-1} and temperature dependences of T^{-1}. However since the model considers altitudes above 80 km, where negative ion concentrations are thought to be small, dissociative recombination of the Na cluster ions with negative ions will not be considered.

It is interesting to note that the reaction $Na^+ \cdot H_2O + e^- \rightarrow Na + H_2O$ is probably sufficiently exothermic to excite Na into its ^2P state [*Swider*, 1970; *Sipler and Biondi*, 1978]. Therefore the reaction scheme may provide an additional source of sodium nightglow and in fact one may expect a correlation between the downward motion of Na^+ layers and sodium emissions. In this theory the downward motion of Na^+ layers may also produce an apparent downward motion of Na as Na^+ is converted to Na. Hence a correlation between downward moving Na layers and sodium emissions is predicted for altitudes above 90 km where the corkscrew mechanism is most evident.

7.4.2.3 *Switching*. In previous studies switching reactions for the Na cluster ion scheme have not been included and as a result the cluster ion scheme was thought to be untenable [*Brown*, 1973; *Ferguson*, 1972]. A switching reaction is a two body reaction in which the weakly bound molecule is replaced by a more tightly bound molecule. On the basis of laboratory experience the switching rates have values on the order of 10^{-9} cm^3 s^{-1} and do not exhibit a great temperature dependence [*Niles et al.*, 1972]. Switching with CO_2 is more effective than with H_2O because of the higher concentration of CO_2 in the atmosphere. Ultimately the scheme should produce $Na^+ \cdot H_2O$ which is the most stable cluster ion of the four considered ($Na^+ \cdot O_2$, $Na^+ \cdot N_2$, $Na^+ \cdot CO_2$, and $Na^+ \cdot H_2O$).

7.4.2.4 *Photodestruction of cluster ions.* Photodissociation of Na^+ clusters represents another area where the experimental evidence is sparse; this process could be another source of Na^+. It has been found that $O_2^+ \cdot (H_2O)_{1,2}$, $O_2^+ \cdot O_2$, and $O_2^+ \cdot CO_2$ have large photodissociation cross sections in the 676.4 to 454.5 nm range; although, $H^+ \cdot (H_2O)_{n = 1 - 4}$ were found to have small or zero cross sections in this range [*Beyer and Vanderhoff*, 1976]. The cluster ion $NO^+ \cdot H_2O$ was added to the list of ions with a sizeable cross section by *Vanderhoff and Niles* [1976] and on the basis of this information *Peterson et al.* [1977] suggest sunlight photodestruction rates for $O_2^+ \cdot H_2O$ and $NO^+ \cdot H_2O$ of 0.6 and 0.06 s^{-1} respectively. However Na^+ is more tightly bound than NO^+ in the cluster ion. For example the binding energy of $NO^+ \cdot N_2$ is about .2 eV which corresponds to an enthalpy change of 4.9 kcal/mol [*Turner and Coway*, 1975]. However the calculated enthalpy change for $Na^+ \cdot N_2$ is 9.5 kcal/mol [*Spears*, 1972]. On this basis it is likely that photodestruction of the Na^+ clusters is negligible, although it must be remembered that laboratory measurements have not been made far into the ultraviolet and hence the uncertainly remains.

It is probable that the photodestruction rate of $Na^+ \cdot O_2$, $Na^+ \cdot N_2$, and $Na^+ \cdot CO_2$ will be much greater than that of $Na^+ \cdot H_2O$ since the latter is a more stable cluster. In fact $Na^+ \cdot H_2O$ may not photodissociate at all. Since it is likely that the collisional breakup rate of the weaker bound ions is comparable to or greater than the photodestruction rate, it will be assumed that photodestruction for all of the cluster ion is not an important process.

C-3

7.4.3 *The sodium chemical sink.* A likely candidate for the sodium sink mentioned in Section 7.2 is the production of higher order cluster ions from $Na^+ \cdot H_2O$. The cluster ions are known to form a ledge between 80 and 85 km and therefore conveniently approximate a perfect sink. A test of this hypothesis though, must include an estimate of the magnitude of the sink and a comparison of this value to the magnitude of the source from meteor ablation.

If we assume that the cluster ion scheme repeats itself for higher order clusters as shown in Figure 7.7, then the controlling reaction will be

$$Na^+ \cdot H_2O + N_2 + X \rightarrow Na^+ \cdot H_2O(N_2) + X \qquad \text{(R11)} \qquad (7.26)$$

where the rate must be determined. *Dzidic and Kebarle* [1970] obtained the enthalpy and Gibb's free energy for the reactions

$$Na^+ + H_2O + H_2O \rightarrow Na^+ \cdot H_2O + H_2O \qquad (7.27)$$

and

$$Na^+ \cdot H_2O + H_2O + H_2O \rightarrow Na^+ \cdot (H_2O)_2 + H_2O \qquad (7.28)$$

From the ratios of the equilibrium constants and activation energy factors of the reverse rate coefficients for reactions (7.27) and (7.28), it is clear that the higher order clustering forward rate coefficient is slower. Assuming this also holds for the clustering of N_2 to $Na^+ \cdot H_2O$, k_{11} may be conservatively set as 10^{-1} times the forward rate coefficient of the reaction

$$Na^+ + N_2 + N_2 \rightarrow Na^+ \cdot N_2 + N_2 \qquad \text{(R1)} \qquad (7.29)$$

The value for k_{11} may be derived eventually by aeronomic techniques, and this point is discussed in Section 7.5.1.

Recombination of these higher order cluster will be neglected since $[e^-]$ is small in the 80-85 km region. The breakup of $Na^+ \cdot H_2O(N_2)$ should not affect the estimate severely since the switching reaction R13 is undoubtedly

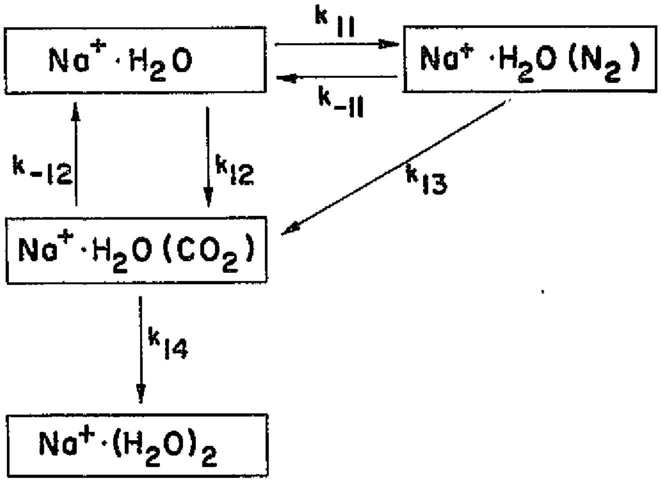


Figure 7.7 Suggested chemical map for a sodium sink in the 80-85 km region. $Na^+ \cdot (H_2O)_2$ is then lost to higher order water clusters and possibly aerosols.

faster. Hence the forward rate of R11 should give a reasonable estimate to the rate of the loss process.

Zbinden et al. [1975] measured a $Na^+ \cdot H_2O$ layer centered at 80 km with a width of 6 km. The concentration at 80 km was about 10 cm^{-3}. Since their measurement was taken during the Geminids meteor shower, values of the ions are certainly enhanced. Since the maximum Na^+ concentration was about 5 times the peak value normally quoted (100 cm^{-3}), it is reasonable to assume that a nominal value for $[Na^+ \cdot H_2O]$ is 2 cm^{-3}. This value would also account for the sporadic observations of $Na^+ \cdot H_2O$. Using the values of $[N_2]$ and [X] at 80 km from Table II.2, the column loss rate of Na^+ is about $k_{11}[Na^+ \cdot H_2O][N_2]$ $[X] \cdot 6 \times 10^5$ cm $= 4.8 \times 10^3$ cm^{-2} s^{-1}. If we take the total influx of meteoric material to be 10^9 gm yr^{-1} (earth surface)$^{-1}$ [*Zbinden et al.*, 1975], then this corresponds to a total influx of 1.5×10^3 cm^{-2} s^{-1}, in good agreement with the column loss. In the steady state the velocity for the downward motion of Na^+ is given by the ratio of the influx of Na^+ to the Na^+ density, or 1.5×10^3 cm^{-2} $s^{-1}/1 \times 10^2$ cm^{-3} = 15 cm s^{-1} which compares with *Gadsden's* [1970] suggestion of 30 cm s^{-1}. The timescale for eddy diffusion in the 80-90 km region is given by $H^2/K = (6 \times 10^5 \text{ cm})^2/3 \times 10^6 \text{ cm}^2 \text{ s}^{-1} = 1.2 \times 10^5$ s where H is taken to be the scale height of the atmosphere. Hence the velocity needed to move an ion from 90 km to 80 km in 1.2×10^5 s is 10 cm s^{-1}.

7.4.4 *The life of a sodium atom.* The picture that emerges is the following. Ions are layered regularly above 95 km by the semidiurnal tides. These relatively high concentration packets of ions are then brought downward with a velocity of about 1 m s^{-1} [*Chimonas*, 1973] and have the characteristic that the average flux travels with velocity of 15-30 cm s^{-1}. The layers may perturb the Na topside gradient in the process. Once the ions are dumped into the 90-95 km region, they diffuse downward to the sink at 80-85 km by the process of eddy diffusion. This sink is most likely total. That is, Na^+ is rapidly converted to higher order clusters, which may attach themselves to aerosols, and are then removed from the sodium chemistry system by a process of sedimentation. The sodium layer takes a small role in this process, acting somewhat like a reservoir. It provides Na^+ by photoionization and charge exchange processes and receives Na from the dissociative recombination of $Na^+ \cdot H_2O$ with electrons but at all times remains in chemical equilibrium with NaO.

7.4.5 *Adopted reaction scheme.* The results of Sections 7.4.1-7.4.4 may now be combined to form the adopted reaction scheme. As discussed in Appendix II the neutral chemistry will be approximated by the interplay of Na and NaO. From the results of Section II.2 the concept of sodium ion oxides playing an important role in the sodium chemistry will be rejected. Rather a cluster ion scheme will be used on the basis of discussions in Sections 7.5.1-7.4.4. The clustering of O_2 and H_2O with Na^+ may be neglected since the $Na^+ \cdot N_2$ route is much more efficient in producing water cluster ions. Photodestruction of $Na^+ \cdot N_2$ and $Na^+ \cdot CO_2$ will not be considered since the collisional breakup of these clusters is probably of greater importance. The cluster $Na^+ \cdot H_2O$ will be assumed to be too stable to photodestruct. Loss of Na in the form of $Na^+ \cdot H_2O$ will be taken as having a rate coefficient equal to $k_{11} = k_1/10$. The details of the sink need not be specified for this calculation and are therefore not included. The final reaction scheme is shown in Figure 7.8 and the list of equations and rates are given in Table 7.3.

7.4.6 *Relating experimental Na and Na^+ profiles as a test of the cluster ion reaction scheme.* Although several measurements of neutral sodium by both ground-based and rocket-borne photometry techniques have been made over the past few decades, near simultaneous measurements of Na and Na^+ have

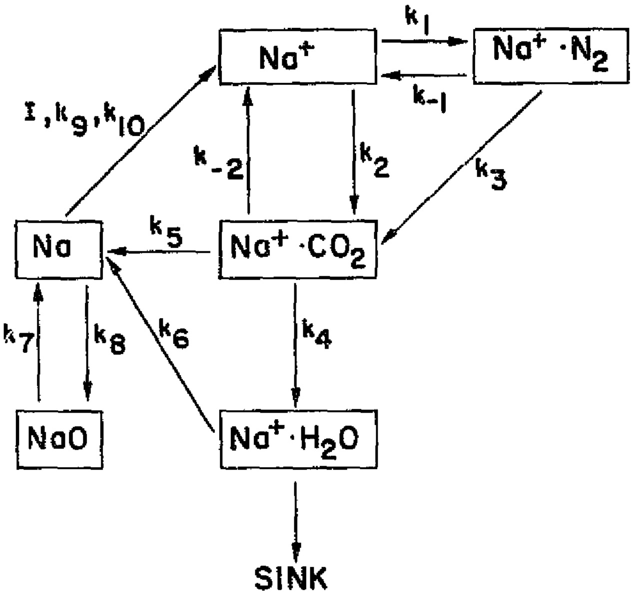


Figure 7.8 The adopted chemical scheme for the meteor ablation-cluster ion theory of sodium. A possible chemical structure for the sink is shown in Figure 7.7. Specific reactions are listed in Table 7.3.

Table 7.3

Reactions and rate coefficients for the meteor ablation-cluster ion theory of sodium

Reaction Reference Number	Reaction	Rate Coefficient
	Clustering	
R1	$Na^+ + N_2 + X \rightleftarrows Na^+{\cdot}N_2 + X$	$k_1 = 2 \times 10^{-31} \left(\frac{310}{T}\right)^{1.5}$ cm^6 s^{-1} $k_{-1} = 9 \times 10^{-7} \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-4750}{T}\right)$ cm^3 s^{-1}
R2	$Na^+ + CO_2 + X \rightleftarrows Na^+{\cdot}CO_2 + X$	$k_2 = 2 \times 10^{-29} \left(\frac{310}{T}\right)^{1.5}$ cm^6 s^{-1} $k_{-2} = 6 \times 10^{-6} \left(\frac{310}{T}\right)^{2.5} \exp\left(\frac{-6250}{T}\right)$ cm^3 s^{-1}
	Switching	
R3	$Na^+{\cdot}N_2 + CO_2 \rightarrow Na^+{\cdot}CO_2 + N_2$	$k_3 = 1 \times 10^{-9}$ cm^3 s^{-1}
R4	$Na^+{\cdot}CO_2 + H_2O \rightarrow Na^+{\cdot}H_2O + CO_2$	$k_4 = 1 \times 10^{-9}$ cm^3 s^{-1}
	Recombination	
R5	$Na^+{\cdot}CO_2 + e^- \rightarrow Na + CO_2$	$k_5 = 1 \times 10^{-6} \left(\frac{300}{T}\right)^{0.5}$ cm^3 s^{-1}
R6	$Na^+{\cdot}H_2O + e^- \rightarrow Na + {}_2O$	$k_6 = 1 \times 10^{-6} \left(\frac{300}{T}\right)^{0.5}$ cm^3 s^{-1}
	Neutral Reactions	
R7	$NaO + O \rightarrow Na + O_2$	$k_7 = 3.9 \times 10^{-11}$ cm^3 s^{-1}
R8	$Na + O_3 \rightarrow NaO + O_2$	$k_8 = 6.5 \times 10^{-12}$ cm^3 s^{-1}
	Ionization and Charge Transfer	
I	$Na + h\nu \rightarrow Na^+ + e^-$	$I = 1.7 \times 10^{-5}$ s^{-1}
R9	$Na + O_2^+ \rightarrow Na^+ + O_2$	$k_9 = 6 \times 10^{-10}$ cm^3 s^{-1}
R10	$Na + NO^+ \rightarrow Na^+ + NO$	$k_{10} = 8 \times 10^{-11}$ cm^3 s^{-1}

ORIGINAL PAGE IS
OF POOR QUALITY

been performed only once. *Goldberg et al.* [1977] observed over Esrange, Sweden two peaks in the Na^+ profile occurring at 90 and 95 km; the Na profile also exhibited a double peaked structure, with the peaks occurring at the same altitudes. The $[Na]/[Na^+]$ ratio was near 2 at these peaks. While these data indicate that $[Na] \propto [Na^+]$ much data are lacking in the report, particularly on the behavior of the cluster ions.

While simultaneous measurements which give us enough information have not been performed, we are fortunate in having two independent measurements which nearly satisfy the criterion. During the Geminids meteor shower on the night of December 13-14, 1971, *Hake et al.* [1972] measured the concentration profile of atomic sodium over Menlo Park, California [37°27'N, 122°10'W). The following day, December 14, 1971, at 12:11 UT *Zbinden et al.* [1975] launched a rocket-borne mass spectrometer at Sardinia (39°36'N, 9°26'E) and obtained Na^+ and $Na^+ \cdot H_2O$ profiles in the 80-110 km region. Selected data of *Zbinden et al.* [1975] are shown in Figure 7.9. In a later publication, *Hidalgo* [1977] derived a temperature profile based on these measurements by employing the temperature dependence of a $NO^+ \cdot N_2$ reaction scheme. The derived temperature, $T \approx 225°K$ between 80 and 86 km, is a necessary input into the cluster ion scheme.

Major constituent concentrations are taken from *CIRA* [1972]. Although this is a reference atmosphere, problems should not occur since N_2 is not believed to be highly variable. The CO_2 and H_2O concentrations were based on constant mixing ratios of 300 ppm and 1 ppm respectively. Values for NO^+ and O_2^+ were taken from *Zbinden et al.* [1975]; but since they did not simultaneously measure $[e^-]$, the electron density profile for quiet sun conditions is taken from *Mechtly et al.* [1972].

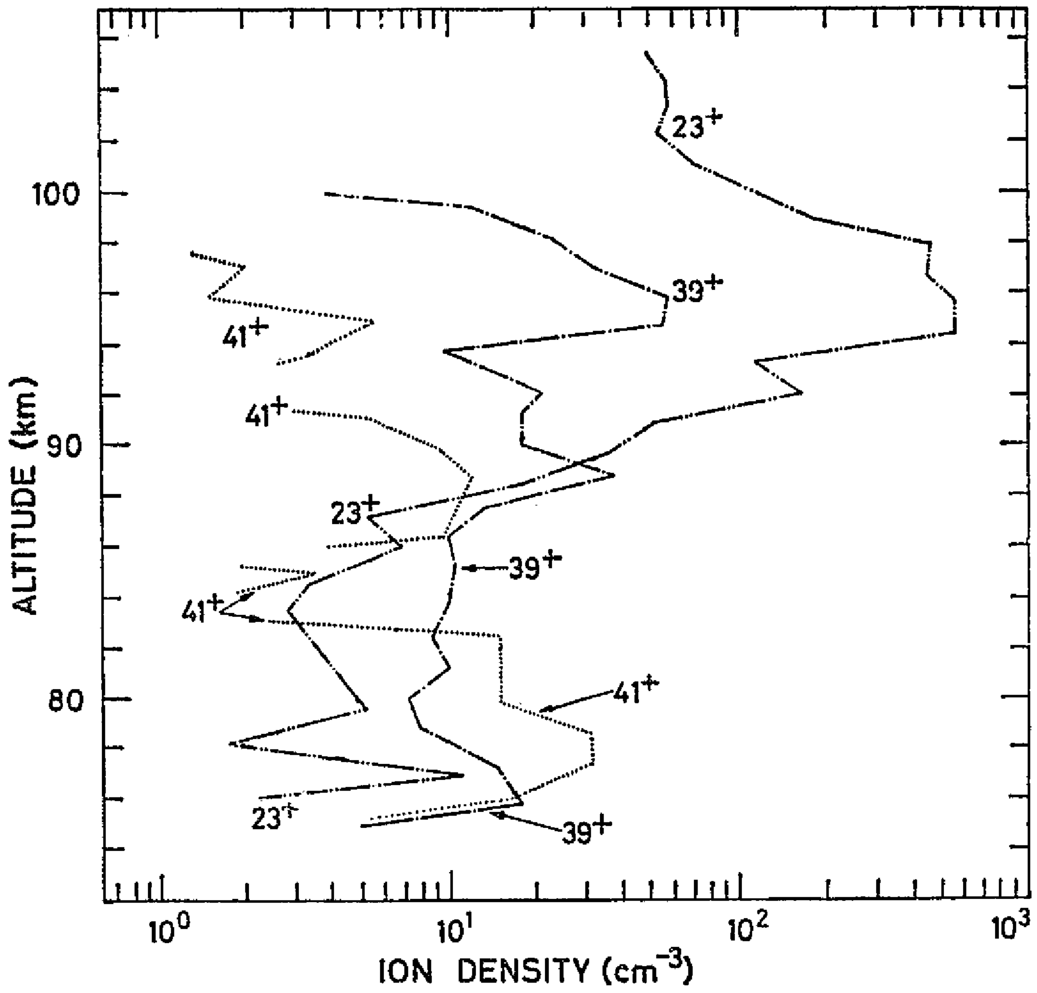


Figure 7.9 Upleg *D*-region profiles of 23^+, 39^+, and 41^+ obtained by *Zbinden et al.* [1975]. Ion 23^+ is identified as Na^+. Ion 39^+ is identified by then to be $^{39}K^+$ above 88 km, and possibly NaO^+ below 88 km. It is suggested that mass 41^+ is $Na^+ \cdot H_2O$ near 80 km and is probably $^{41}K^+$ above 90 km.

Chemical equilibrium may be assumed for these calculations since the cluster ion reactions have small time constants in comparison to the time scales associated with eddy and molecular diffusion. Using the Na^+ profile of *Zbinden et al.* [1975], Na, $Na^+ \cdot N_2$, $Na^+ \cdot CO_2$, and $Na^+ \cdot H_2O$ values are calculated every kilometer. The results of these calculations are shown in Figure 7.10.

The agreement is very good. First the peak concentration of sodium is predicted properly, and the peak occurs at the proper altitude. It is not expected that the shapes of the two sodium profiles should agree, since dynamical effects can undoubtedly perturb the layers. However since lidar measurements indicate no diurnal variation in the sodium layer [*Gibson and Sandford*, 1972], the profile of *Hake et al.* [1972] should be a good estimate to the general features of the sodium layer that existed during the *Zbinden et al.* launch.

Furthermore the cluster ion scheme accurately predicts the $Na^+ \cdot H_2O$ profile near 80 km obtained in the mass spectrometer measurements. Since significant amounts of $Na^+ \cdot H_2O$ are also predicted near 90 km, it may be improper to identify this ion as $^{41}K^+$ as *Zbinden et al.* have done. The cluster ion $Na^+ \cdot CO_2$ is calculated to remain less than 0.3 cm^{-3} except in the altitude range 89-98 km where values are close to 1 cm^{-3}. Only at the peak of the Na^+ layer, is $Na^+ \cdot CO_2$ estimated to be 2-3 cm^{-3}. It is therefore unlikely that *Zbinden et al.* [1975] detected $Na^+ \cdot CO_2$ and this is reflected in their data. The cluster ion $Na^+ \cdot N_2$ is calculated to remain below 10^{-2} cm^{-3} at all altitudes and obviously was not detected. Finally on the basis of Figure II.4 which predicts that $[NaO^+] \simeq 10^{-4}[Na^+]$ at 85 km, it is very unlikely that mass 39^+ was NaO^+ but rather $^{39}K^+$.

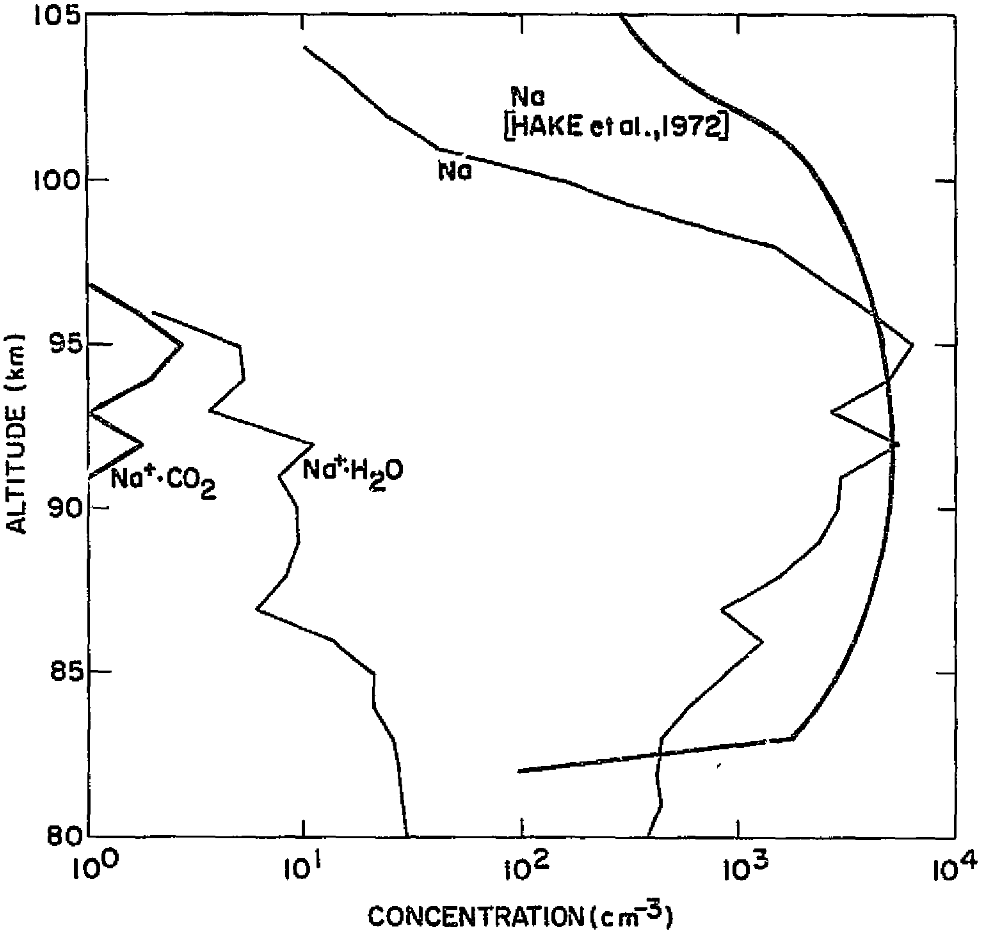


Figure 7.10. Profiles obtained from the cluster ion scheme using the Na^+ profile of *Zbinden et al.* [1975] as input. Comparison of the calculated Na profile to the experimental profile measured by *Hake et al.* [1972] shows good agreement.

In summary the simple chemical equilibrium calculation for the cluster ion scheme accurately predicts the concentrations of both neutral and ionic forms of sodium. The final test of the cluster ion scheme will concern itself with the ability of the scheme to explain the more common features of the sodium layer.

7.5 *Model Calculations*

7.5.1 *The effect of variations in second and third body concentrations.* Assuming chemical equilibrium, the expression for [Na] becomes

$$[\mathrm{Na}] = \frac{k_7[\mathrm{O}][\mathrm{NaO}] + \{k_5[\mathrm{Na}^+\cdot\mathrm{CO}_2] + k_6[\mathrm{Na}^+\cdot\mathrm{H}_2\mathrm{O}]\}[e^-]}{k_8[\mathrm{O}_3] + I + k_9[\mathrm{O}_2{}^+] + k_{10}[\mathrm{NO}^+]} \tag{7.30}$$

It is apparent that [Na] is proportional to the atomic oxygen and electron-density concentrations. Thus the sharp decrease on the bottomside of the layer is related to the electron density and atomic oxygen ledges which exist in the same region.

An examination of the daytime time constants indicates that the neutral Na system comes into equilibrium slightly faster than the neutral-ion system. Yet the latter is still of a shorter time scale than the process of eddy diffusion. Thus the chemical scheme behaves as though sodium acted as a reservoir. It is in constant equilibrium with NaO, and so the supply of sodium by the recombination of the cluster ions must be balanced by the loss of sodium by photoionization and charge exchange.

All of the reactions depend on assumed profiles; but since the important cluster ion processes involve three body reactions, and the ambient neutral atmosphere is not expected to vary greatly with time, the chemical scheme should be applicable to all seasons.

However the effect of variations in $[CO_2]$, $[H_2O]$, and $[e^-]$ should be investigated. The case for CO_2 will be considered first. Assuming chemical equilibrium and using the notation of Table 7.3.

$$[Na^+ \cdot N_2] = \frac{k_1 [Na^+][N_2][X]}{k_{-1}[X] + k_3[CO_2]} \tag{7.31}$$

The switching reaction is so fast, that $k_3[CO_2] >> k_{-1}[X]$. Hence $[Na^+ \cdot N_2]$ varies inversely as $[CO_2]$. The concentration of $Na^+ \cdot CO_2$ may be written as

$$[Na^+ \cdot CO_2] = \frac{k_2[Na^+][CO_2][X] + k_3[Na^+ \cdot N_2][CO_2]}{k_{-2}[X] + k_5[e^-] + k_4[H_2O]} \tag{7.32}$$

Since $[Na^+ \cdot CO_2]$ is more strongly supplied by the switching of CO_2 with $Na^+ \cdot N_2$ than it is by the direct clustering of Na^+ with CO_2,

$$[Na^+ \cdot CO_2] \propto k_3[Na^+ \cdot N_2][CO_2] \tag{7.33}$$

which implies with the aid of equation (7.31), that $[Na^+ \cdot CO_2]$ is independent of CO_2 concentrations on the order of 300 ppm.

The case for H_2O is a bit different. Mixing ratios on the order of 1 ppm are commonly assumed in cluster ion calculations [*Reid*, 1977; *Thomas*, 1975]. Denoting the rate coefficient for the clustering of Na^+ with H_2O as k, one obtains

$$[Na^+ \cdot H_2O] = \frac{k_4[Na^+ \cdot CO_2][H_2O] + k[Na^+][H_2O][X]}{k_6[e^-] + k_{11}[N_2][X]} \tag{7.34}$$

It is assumed in equation (7.34) that the switching of H_2O with $Na^+ \cdot N_2$ produces less $Na^+ \cdot H_2O$ than the indirect route involving the switching of CO_2 with $Na^+ \cdot N_2$. This is true since $[CO_2] >> [H_2O]$. The main question concerns the numerator of equation (7.34) and whether

$$\frac{k_4[\mathrm{Na^+ \cdot CO_2}]}{k[\mathrm{Na^+}][\mathrm{X}]} >> 1 \qquad (7.35)$$

Referring to Table 7.1, it is apparent that $k_4 \sim 10^5\ k[\mathrm{X}]$ where a temperature of 200°K is assumed and $[\mathrm{X}] = 7 \times 10^{-13}\ \mathrm{cm^{-3}}$ at 90 km. Since all calculations show that $[\mathrm{Na^+ \cdot CO_2}] \sim 10^{-2}\ [\mathrm{Na^+}]$, equation (7.35) holds and the direct clustering of H_2O with Na^+ should not be important. Below the electron-density ledge (~85 km), $k_5[e^-] << k_4[\mathrm{H_2O}]$ and so

$$[\mathrm{Na^+ \cdot CO_2}] \propto \frac{1}{[\mathrm{H_2O}]} \qquad z \lesssim 85\ \mathrm{km} \qquad (7.36)$$

Hence the production of $[\mathrm{Na^+ \cdot H_2O}]$ should be independent of $[\mathrm{H_2O}]$. Above 95 km, $k_5[e^-]$ becomes comparable to $k_4[\mathrm{H_2O}]$. However, calculations indicate that $[\mathrm{Na^+ \cdot H_2O}]$ still remains relatively independent of $[\mathrm{H_2O}]$ for water mixing ratios of up to 10 ppm.

The scheme will be affected slightly by changes in $[e^-]$. In expression (7.32), the denominator is controlled by the relative magnitudes of $k_5[e^-]$ and $k_4[\mathrm{H_2O}]$. The altitude at which $[\mathrm{Na^+ \cdot CO_2}]$ is affected by the electron density is determined by the condition

$$\frac{[e^-]}{[\mathrm{H_2O}]} = \frac{k_4}{k_5} \sim 10^{-3} \qquad (7.37)$$

Also from (7.34) $[e^-]$ becomes important when

$$\frac{[e^-]}{[\mathrm{N_2}][\mathrm{X}]} = \frac{k_{11}}{k_6} \sim 3 \times 10^{-25} \qquad (7.38)$$

The altitudes corresponding to (7.37) and (7.38) are near 95 km and 85 km respectively. Above these altitudes $[e^-]$ variation becomes less important for producing Na because increases in $[e^-]$ create decreases in $\mathrm{Na^+ \cdot CO_2}$ and

$Na^+ \cdot H_2O$. However below 85 km, the production of Na is increasingly dependent on $[e^-]$ since $[Na^+ \cdot CO_2]$ and $[Na^+ \cdot H_2O]$ are controlled more by switching and three body reactions respectively. Hence one would expect the bottomside of the Na layer to be moderately affected by the electron-density profile, particularly near the ledge.

In fact a knowledge of the value for $[e^-]$ near the ledge may be an important parameter for determining the value of k_{11}. Since the *Zbinden et al.* launch was performed under anomolous conditions, $[e^-]$ may have been enhanced in the 80-85 km region by as much as a factor of 3. A value of $k_{11} = k_1/2$ is then found to be more appropriate than the value used in the previous calculations ($k_{11} = k_{1/10}$). This higher value would also yield a more rapid bottomside falloff in [Na] under normal conditions.

7.5.2 *Seasonal variation.* It is well known that [Na] increases during the winter and that the peak of the layer may lower slightly (Figure 1.1). The cluster ion reactions are temperature sensitive, but unfortunately in the wrong directions. That is, as temperature increases, fewer cluster ions are produced which reduces the rate of dissociative recombination. Hence the temperature dependence implies that [Na] should decrease in the winter. If it is assumed that Na^+ has a Gaussian profile given by

$$[Na^+] = \begin{cases} 100 \exp\left[-((z - 90)/3)^2\right] \text{ cm}^{-3} & \text{for } [Na^+] > 1 \text{ cm}^{-3} \\ 1 \text{ cm}^{-3} & \text{otherwise} \end{cases} \tag{7.39}$$

then this increase in summer is evident in Figure 7.11. Notice that the Gaussian profile is not an accurate description, since large quantities of $Na^+ \cdot H_2O$ are predicted. A realistic Na^+ profile would fall off much more rapidly on the bottomside. Yet the technique still has merit.

The assumption used in deriving Figure 7.11 is that the Na^+ layer does not move. Three body reactions are so important to this scheme, that the

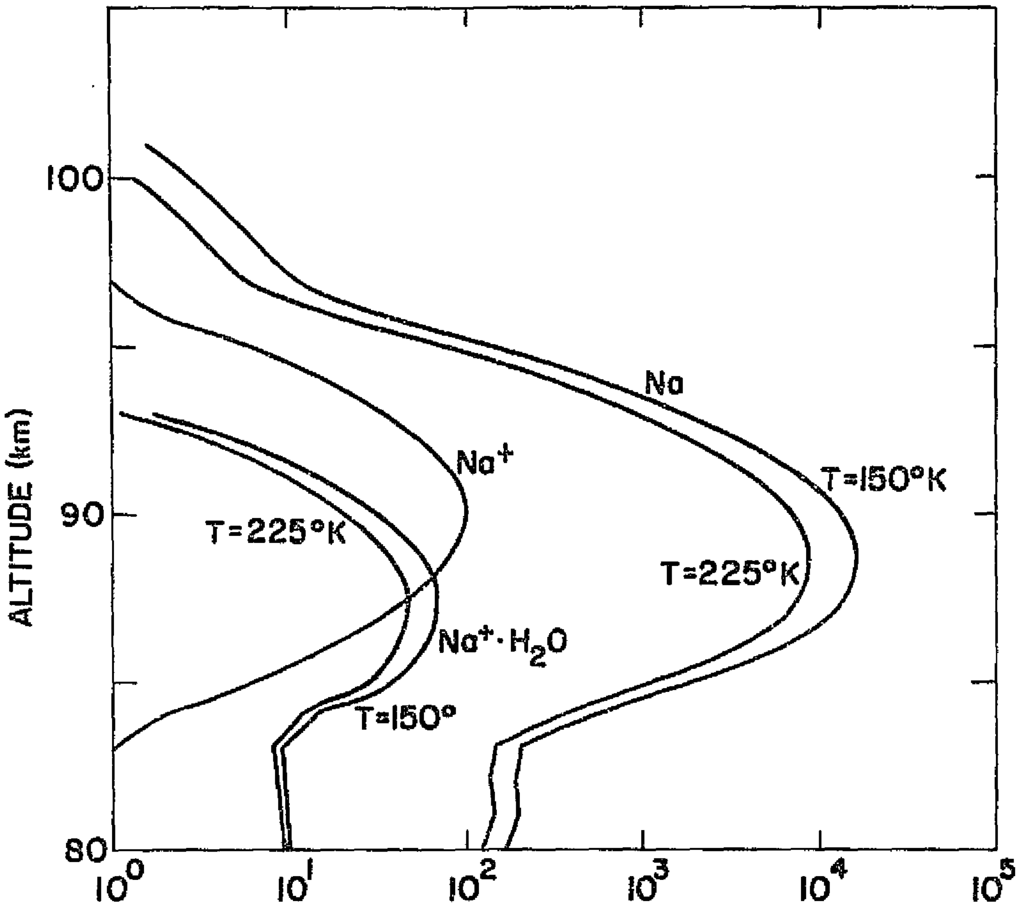


Figure 7.11 Temperature effects on the values of [Na] and [$Na^+ \cdot H_2O$]. A Gaussian profile is assumed for [Na^+]. Temperatures of 150° and 225° correspond to summer and winter mesospheric conditions.

equilibrium value of [Na]/[Na^+] will be highly altitude dependent. Let us assume that Na^+ takes on the same Gaussian shape but that it is centered about 95 km in the summertime (T = 150°K) and compare this to a Na^+ layer centered about 90 km in the winter (T = 225°K). The two resulting profiles are shown in Figure 7.12 and indicate a lowering of the peak of the layer by about 5 km and a column increase in sodium by a factor of 3. The factor is more than that observed by *Megie and Blamont* [1977] and hence the 5 km separation may be reduced to 3-4 km which is in good agreement with the profiles shown in Figure 1.1. The principle then is, if the Na^+ profile does not change its amplitude or shape with season, the peak of the Na^+ layer determines the peak of the Na layer; and the altitude at which the peak occurs strongly affects the equilibrium value of [Na]. It is interesting to note that the [Na]/[Na^+] ratio varies from 30 to 90 for this example (summer to winter) and hence this fact may explain the wide range of estimates for this ratio [*Narcisi*, 1968].

Finally a seasonal variation in the ambient background density would produce the same effect as that produced by the movement of the Na^+ layer. From *CIRA* [1972] the atmospheric concentration at 90 km is about 2 times that at 95 km. Hence an increase in number density of a factor of 2-3 during winter is required. Model calculations show this is plausible (Tables II.1 and II.2). For this assumption seasonal changes in the peak of the Na layer would not be so pronounced unless dynamical effects became more important during winter.

7.5.3 *Diurnal variation.* Motions play an increasing role at nighttime. Since photoionization of Na is removed from the scheme, and charge exchange is severely reduced because of the decrease in [O_2^+] and [NO^+], a path no

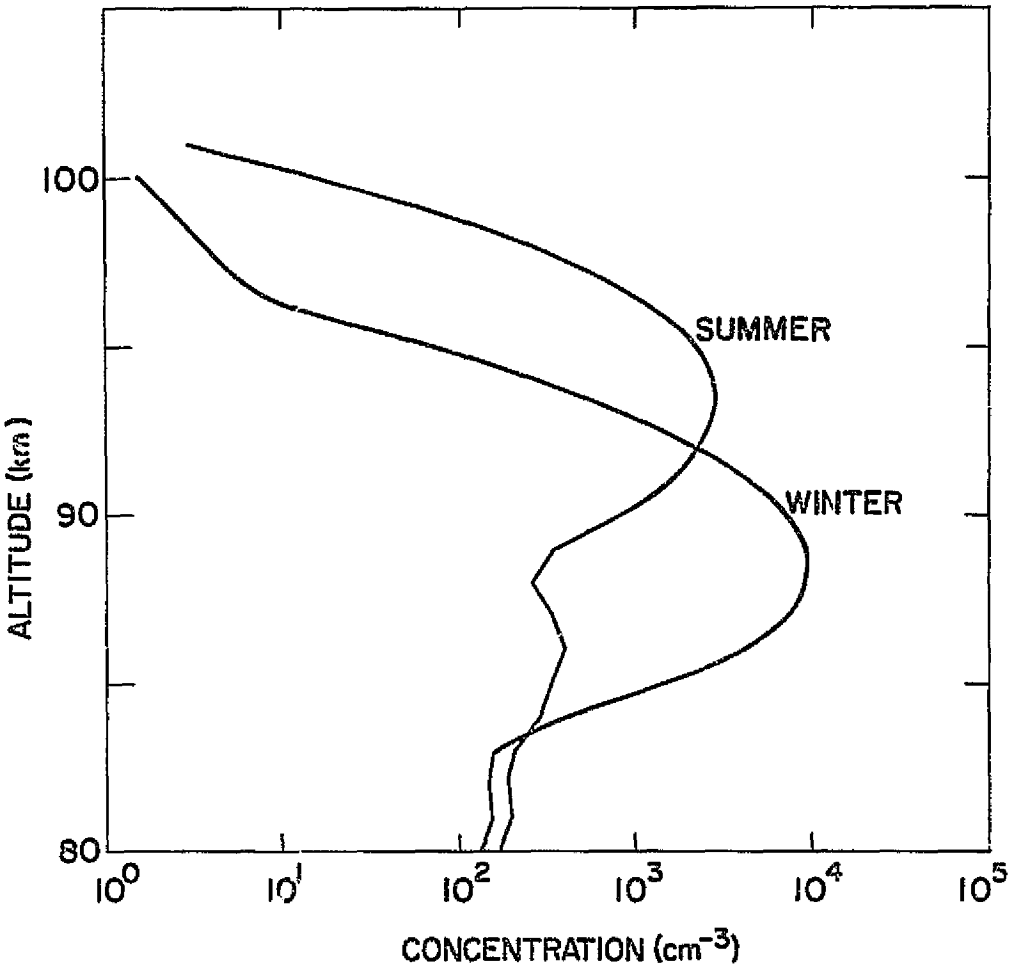


Figure 7.12 The effect of lowering the Na^+ layer from 95 km in the summer to 90 km in the winter yields an increase in the total column content of Na. A similar effect would be caused by an increase in the ambient neutral density during winter, although the altitude difference would be much reduced.

longer exists between Na and Na^+. Rather Na^+ should be depleted by the cluster ion reactions and either be converted to Na or be removed by the sink in the 80-85 km region. However the $[Na]/[Na^+]$ ratio is high enough that the column abundance of Na^+ should cause a negligible increase in the column abundance of Na. A diurnal variation should not be observed. The layer at this time may be an effective tracer for atmospheric motions and could be used to monitor wave motion having spatial scales less than the width of the sodium layer (10-20 km). The lack of diurnal variation is supported by lidar measurements [*Gibson and Sandford*, 1972] as well as by spectroscopic measurements of sodium dayglow [*Burnett et al.* 1972, 1975].

7.6 *The Interpretation of Some Experimental Results using the Meteor Ablation-Cluster Ion Theory*

7.6.1 *Variation of the topside scale height.* Although the cluster ion scheme can explain a seasonal variation in the sodium column content and a lack of diurnal variation in the layer, one of the questions to be explored by this work has been the validity of the corkscrew mechanism. A major concept is that the corkscrew mechanism is responsible for the continual supply of Na^+ to the sodium layer. The experimental observations of enhancements in the sodium topside which move downward with a velocity of about 50-100 cm s^{-1} lend support to the presence of the corkscrew mechanism. The chemistry developed in this chapter shows that chemical time constants are still rapid at 95 km for the conversion of Na^+ to Na and hence movement of the Na^+ layer will be mirrored in the Na layer. Results from other groups studying the sodium problem also lend support to the corkscrew mechanism.

A characteristic feature that *Megie and Blamont* [1977] observed is the variation of the topside gradient with time. At night the layer is under dynamical control since the ion chemistry gradually disappears as Na^+ is depleted. Under these conditions one would expect the layer to be broadened under the influence of eddy diffusion. In 90% of the cases, Megie and Blamont found that a proper choice for the eddy diffusion coefficient could explain this broadening of the layer and in fact they obtained average values for K using this technique.

The puzzling fact is that in 10% of the data, the topside scale height decreases with time particularly at the beginning of the night, and they postulate that this results from a source of sodium near 91 km. It may well be that this source is the dumping of ions into the region by the corkscrew mechanism. Near sunset, the 90-120 km region still has a supply of long-lived Na^+ ions remaining from the daytime ionization of Na. It is expected that these ions remain layered and are swept out of the region during the night. Evidence of electron-density gradients have been seen in the 90 km region during the nighttime monitoring of the layer (Section 6.5) and may be indicative of metallic layers of ions. The maximum amount of ionization available naturally occurs near the beginning of the night, however higher altitude Na^+ layers may enter the Na layer well into the evening according to this theory.

The time scale for the conversion of Na^+ to Na is limited by the time constant for the formation of $Na^+ \cdot N_2$. At 92 km this time constant is roughly 20 min. A layer of Na^+ capable of maintaining a concentration of 50 cm^{-3} having a thickness of 3 km will then provide a source having an integrated strength of 1.2×10^4 cm^{-2} s^{-1}. Interesting enough this compares favorably with a lower limit value on the source of 2×10^4 cm^{-2} s^{-1} assuming $K = 2 \times 10^6$ cm^2 s^{-1} [*Megie and Blamont*, 1977]. During enhancements in the meteoric input, it is likely that the integrated source could be greater.

Hanson and Donaldson [1967] explained the influence of the photoionization rate on the topside scale height. Their method does not assume a source for sodium. If a constant source in the same altitude region is postulated then the effect works against photoionization and tends to increase the scale height. Since layers of ions driven by the tidal oscillations are moving through the sodium layer, diurnal variations in the topside layer should occur.

7.6.2 *The [Na]/[K] ratio.* The cluster ion scheme may have an impact on the seasonal variation of the [Na]/[K] ratio as observed by *Megie et al.* [1978], K^+ is not as reactive as Na^+ and hence should not show as strong of a temperature dependence in the cluster ion-chemistry scheme. Furthermore ion separation is believed to be produced by the wind shear mechanism at higher altitudes (~ 105 km) [*Chimonas*, 1969]. The effect is presumably lost near 90 km but may still occur. Since the seasonal variation is tied to the temperature effects on reactions and the position of the Na^+ layer (or alternatively on the neutral atmospheric concentration), these temperature and layer effects need to be considered for the case of K before it is assumed that chemistry cannot explain the seasonal changes in the [Na]/[K] ratio.

7.6.3 *Removal of sodium during the day.* Finally *Megie and Blamont* [1977] notice that after meteoric activity, the enhancement in the sodium layers is removed sometimes within 12 hrs. It should be emphasized that the sink for sodium varies directly as $[Na^+]$ and is therefore continuously adjustable. Since $[Na^+ \cdot H_2O]$ varies directly as $[Na^+]$, an enhancement in $[Na^+]$ produces enhancements in $[Na^+ \cdot H_2O]$ which causes a greater loss of sodium.

7.6.4 *Na^+ meteoric input.* In this theory the source is assumed to be Na^+ resulting from the photoionization of ablated material containing sodium. This is based on the assumption that meteoric material is ablated in neutral form. However, *Lebedinets and Shushkova* [1970] calculated that near 110 km, meteor ionization is important. The effect of change transfer or photoionization in the ablation process may also increase the input of Na^+. If these facts are true then a source for Na^+ would result at night, indicating the corkscrew mechanism should supply the Na layer at night. These ideas are not inconsistent with observations or theory.

7.7 *Summary*

In this chapter a simple ion chemistry for sodium has been developed under the assumption of daytime chemical equilibrium. The chemistry includes switching and recombination of cluster ions but does not consider their photodestruction. The main cluster ion is found to be $Na^+ \cdot H_2O$. The chemistry is then used in conjunction with a meteor ablation source to explain the dynamical features of the sodium layer.

A seasonal variation may be produced by the model and lack of a diurnal variation is evident. Current problems in the sodium study are discussed in the light of this theory.

8. SUMMARY AND CONCLUSIONS

8.1 *Summary*

A laser radar system has been built and operated to monitor the nighttime sodium layer over Urbana. The signal-to-noise ratio has been increased sufficiently to obtain accurate relative concentration profiles of the sodium layer in 10 minutes. Thus the time history of the sodium layer may be monitored throughout the night.

In general the layer is observed to be highly variable during the winter time. Wavelike structures have been found which appear to have downward phase velocities of from 50 cm s^{-1} to 100 cm s^{-1} and vertical wavelengths from 3 to 15 km. The altitude of the peak of the layer is often found to be lower in the winter than in the spring and has appeared as low as 83 km. The bottomside usually falls off more rapidly than the topside.

Simultaneously with the collection of data, a theory for the sodium layer was developed. The main problem areas about the aeronomy of sodium are the following:

1. *Origin.* Suggested sources have been meteor ablation or thin aerosol layers, either of extraterrestrial or terrestrial origin.
2. *Boundary gradients.* The sodium layer falls off quite rapidly on both the topside and bottomside.
3. *Seasonal variation.* There is an enhancement in winter at midlatitudes and a possible lowering of the peak of the layer at this time.
4. *Diurnal variation.* Lidar measurements and spectroscopic dayglow measurements indicate little diurnal variation.
5. *Nocturnal variation.* The sodium layer seems to show little systematic variation during the night.
6. *Sporadic effects.* The layer shows considerable short-term variability.

8.2 *Conclusions*

To explain all the abovementioned properties a meteor ablation-cluster ion theory was developed. This theory postulates that the source of sodium is meteor ablation in the 90-120 km region. The deposited sodium is then rapidly ionized in the daytime by photoionization or charge exchange with O_2^+ and NO^+. It is shown that the lifetime of Na^+ is long enough for Na^+ to be layered by an interaction of the tidal winds with the geomagnetic field lines (corkscrew mechanism). The layers collect near the nodes of the wind profile and are then transported downward with the characteristic phase velocity of the wind profile (~30-100 cm s^{-1}). Below 100 km the ions no longer can maintain the same downward velocity of the wind profile as the neutral-ion collisional frequency increases. They gradually are deposited in a region between 90 and 95 km.

A cluster ion chemistry was developed to explain a rapid conversion of Na^+ into Na. This scheme had several advantages over the earlier metal oxide ion chemistry. The major sodium constituents then become Na and Na^+ at higher altitudes with NaO and $Na^+ \cdot H_2O$ playing an increasing role near the bottom of the layer. A sink is provided in the form of higher order cluster ions which possibly attach themselves to aerosols and thereby undergo sedimentation to the lower atmosphere.

The following explanations are offered to the problem areas:

1. The origin of sodium is taken to be meteor ablation based on the correlation of sodium column increases with meteoric activity. Also metallic ion abundances have relative ratios close to what is believed to exist for the neutral metals in meteoric material.

2. The topside gradient is controlled by photoionization, eddy diffusion, and the dumping of ions by the corkscrew mechanism. The bottomside of the layer is chemically controlled and is related to the electron density and atomic oxygen ledges known to exist in the region. The bottomside falloff of Na also depends on the atmospheric conditions which favor the formation of higher order cluster ions below 85 km.

3. The seasonal variation in column content is a result of the sensitivity of the cluster ion reactions to atmospheric density. It may be related to a repositioning of the Na^+ layer with season or an actual change in the atmospheric density with season.

4. The lack of diurnal variation follows from the large $[Na]/[Na^+]$ ratio. Na^+ is gradually depleted through the night.

5. The layer shows no systematic variation during the night because the loss and production mechanism of the sodium column content are turned off.

6. The short-term variability is undoubtedly related to the turbulence known to exist within the region. The influence of gravity waves and horizontal transport are undoubtedly important.

8.3 *Suggestions for Future Work*

8.3.1 *Experimental aspects.* Several correlative measurements are suggested by this work and a brief discussion of each will follow. In addition, some obvious improvements in the lidar system are proposed. It has become apparent that lidar measurements taken alone are insufficient to solve the aeronomy of the sodium layer, and the following topics present the core experimental effort which should be undertaken.

1. *Simultaneous measurements of electron density.* The theory predicts that the bottomside of the sodium layer is tied to the electron-density ledge.

Lacking a method for measuring electron density at night, we probed the atmosphere with a 2.66 MHz signal while simultaneously monitoring the sodium layer. Reflections of the radio signal indicated electron density gradients in the same region as the sodium layer. This then suggests that measurements of the electron density be made with an ionosonde or incoherent-scatter radar, simultaneously with the lidar experiments, for precise monitoring of the ionization.

2. *Simultaneous measurements of neutral winds.* A meteor radar system exists at the Aeronomy Laboratory Field Station and can monitor winds in the sodium region. It would be important to find any correlation between the waves identified by meteor radar and waves observed by laser radar. Furthermore the meteor radar system could monitor the east-west winds which most strongly affect the vertical ion transport below 100 km. By this joint experiment it may be possible to determine if the wave motion in the sodium layer is related to wave motion of the ambient atmosphere or rather to the corkscrew mechanism. If the former case were true, sodium could be used as a tracer to observe atmospheric wave motion.

A feasibility study conducted at the Field Station revealed interference between the systems. Also the meteor radar unit was observed to increase the ambient temperature of the room housing the laser to the point at which thermal stability of the tuning optics became a problem. If the systems are operated alternately throughout the night to avoid the interference, the data may suffer because each system depends on the integration of returns for improved signal-to-noise. Yet the experiment can be done with existing equipment and should be attempted in the near future to see if the above objections can be removed.

3. *Mobile lidar applications*. There are other experiments which require lidar mobility. Simultaneous lidar monitoring with a rocketborne positive ion mass spectrometer measurement could be accomplished over Wallops Island, Virginia. This would help to determine more exactly the relation between Na and Na^+, and thus the cluster ion scheme could be refined. Furthermore a mobile lidar could be taken to the equator where the sodium layer is predicted to be rather broad. Neither of the measurements have ever been done. In any event work should progress toward reducing the volume of the lidar system in preparation for the day when a mobile lidar is feasible. Measurements from an airplane might also be considered. The weather over Urbana many times of the year is not conducive to frequent operation; an airplane would give us the opportunity to fly above the clouds and avoid much of the attenuation due to molecular and aerosol scatter.

4. *Steerable lidar*. Although this work assumes that the movement of sodium is in the downward direction, horizontal transport may be present. A steerable lidar system could measure the effect of horizontal transport; however, it should be noted that the experiment has been performed over Winkfield (51.4°N, 0.7°W) [*Thomas et al.*, 1977]. Hence the experiment should be undertaken only if substantial improvements over success of Thomas et al. can be guaranteed. It should be noted that safety regulations and common sense dictate that the lidar should be turned off when planes are in the area if the zenith angle of the beam is great enough. A radar unit may then have to accompany the steerable lidar for detection of local air traffic.

5. *Measurement of potassium*. Measurements of other constituents, especially potassium may be conducted; but the measurements involve much lower signal levels. This study indicates that more work needs to be done in the sodium area, and a measurement of K may only dilute our ability to study the

Na layer. *Megie et al.* [1978] have conducted simultaneous K and Na measurements; and although their statistics could be improved it will be hard to better their results in the near future.

6. *Daytime measurements.* Daytime lidar measurements of Na have not been repeated. These measurements made a strong impact on the sodium theory because they showed no diurnal variation in the layer [*Gibson and Sandford,* 1972]. The measurements should be repeated. Careful monitoring of the twilight sodium layer could also be accomplished and would provide information on any rapid structural changes in the layer during this time. A smaller telescope with a narrow band filter would be required and increased laser output desirable.

7. *Laser improvements.* A general upgrading of the lidar is required. The signal-to-noise can best be increased by enlarging the aperature of the receiving telescope; this plan is being implemented. In addition laser improvement is necessary. Wavelength emission stability and spectral linewidth monitoring improvements are the two areas of prime concern. At this time we cannot compare returns from one night to the next with any accuracy because of the uncertainty in the spectral linewidth measurement; hence our approach has been to observe relative changes in the layer during a night. Laser stability can be accomplished best by a two etalon tuning in which the tuning elements are temperature controlled. A Lyot filter has been used with success by some groups as a replacement for the crude tuning etalon and should be considered. A high resolution spectrometer is sorely needed to measure the spectral width of the laser line emission. If the linewidth cannot be monitored accurately, then absolute measurements of sodium cannot be made. A high quality spectrometer is a necessity for excellent data. Ideally linewidth measurements should be recorded with the data. Some thought should

also be given to the portability of such a spectrometer. Finally it has often been proposed that an amplifier be added to the current oscillator. Before this is implemented, a transverse flow dye laser should be considered. These lasers can provide high average power as well as high peak powers [*Jethwa*, 1978] and may be commercially available in the near future. The spot geometry is not circular, but this should cause no problem. Also any second generation laser must reduce the duration of the pulse to less than 500 ns for better range resolution.

8. *Measurement of chemical rate coefficients.* Finally one area of experimental research is directed toward the scientific community in general. There is an urgent need for the laboratory chemists to measure the rate coefficients for many of the sodium reactions at mesospheric temperatures. It is crucial to determine if the important neutral sodium reactions proceed by the electron jump mechanism. Also the estimates for the rate coefficients of the cluster ions are accurate to an order of magnitude at 300°K. The extrapolation of these rates to lower temperatures therefore has even a greater chance for inaccuracy.

8.3.2 *Theoretical investigations.* The foundational work for the meteor ablation-cluster ion theory has been performed and several areas have now been opened for future research.

1. *The effect of higher order cluster ions and the electron density ledge near 85 km.* The cluster ion chemistry should be expanded to model more accurately the sink. In particular the effect of higher order clustering reactions on the $Na^+ \cdot H_2O$ topside needs to be investigated. It would be interesting to determine if the Na^+ clustering reactions can be tied to the NO^+ clustering reactions.

This work may lead to a further understanding of the electron-density ledge near 85 km. This region is particularly interesting because of strong dynamical changes in the region and temperature gradients.

2. *Seasonal variation of* [*Na*]/[*K*]. The [Na]/[K] ratio variation with season needs to be explored. Is this variation caused by changes in the source for Na or by chemical differences between the elements? The cluster ion reaction scheme may invalidate the assumption that K behaves in the same manner that Na does.

3. *Chemistry of metallic species*. A general study of metallic species in general would be a significant step forward. Although *Swider* [1969, 1970] provides an excellent review, the work needs to be updated in the light of cluster ion reactions. It may be possible that some metals follow an oxide-ion chemistry while others obey the cluster ion reactions.

4. *Atmospheric simulations*. Finally atmospheric simulations which emphasize dynamical effects on the sodium layer need to be explored. It is important to be able to eventually model the sodium layer to see if experimental observations can be duplicated. The chemistry for this model though may have to be fairly complete. Sodium probably is tied to other clusters in the region and a complete neutral chemistry may also be required. Atmospheric motions would include the wind shear mechanism, gravity waves, horizontal transport, and changes in eddy diffusion.

APPENDIX I

THE INITIAL LIDAR RECEIVER

To obtain the first sodium returns over Urbana as quickly as possible, the first lidar receiver used the PDP-15 computer facility at the Aeronomy Laboratory Field Station. This computer is used with several ground-based experiments for data collection and its inclusion into the lidar system was accomplished by minimizing hardware whenever software could accomplish the same results.

Since we required a high data rate (100 kHz) interface with direct memory access, the custom-built analog-to-digital (A/D) converter interface was chosen as the point of entry for our data. Existing data collection software was used by substituting 8 bits of lidar data for the output buffer of the A/D converter. Softward modification provided suitable masking of the unused interface bits; and when the BCD counter was used, the software converted the BCD to binary. The block diagram of this system is shown in Figure I.1. The flow of information is essentially the same as that described in Chapter 5. Schematic diagrams of important circuits are indicated in Figures I.2 - I.4.

Because of the versatility of the PDP-15 and the ample storage capabilities, the computer enabled us to define problem areas and investigate data collection routines. A serious problem for this system resulted from combining the asynchronous BCD upcounter (University of Illinois, Elementary Particle Physics Research, 10 MC Scaler S102) with the synchronous computer system; this arrangement resulted in the occasional sampling of the upcounter while it was shifting thus yielding errors. When adjacent bins were subtracted to obtain the counts per bin, these sampling errors created errors in two adjacent bins. The ripple through BCD counter also provided ample

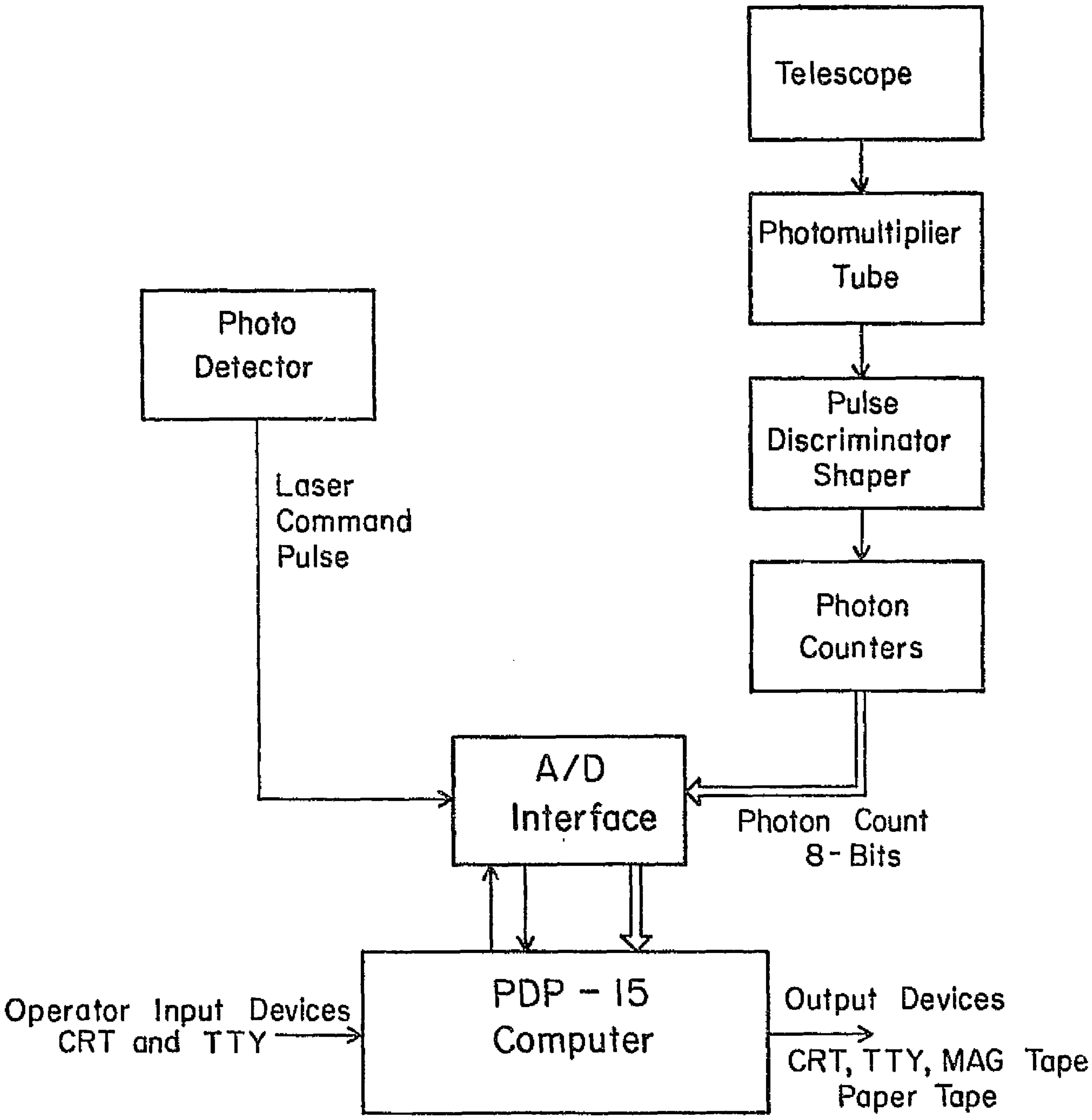
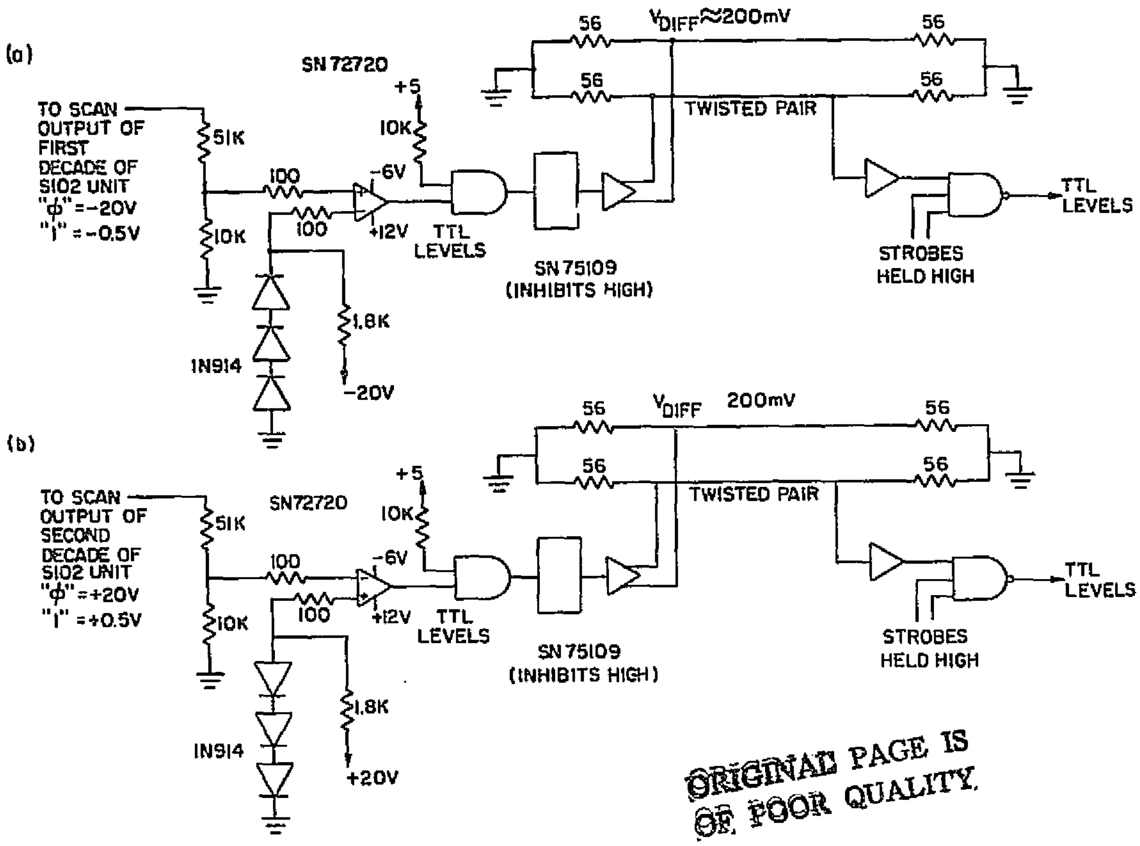


Figure I.1 Initial lidar receiver.



ORIGINAL PAGE IS OF POOR QUALITY

Figure I.2 Line driver and line receiver circuitry for the receiving system which employed the BCD upcounter (Scaler S102). The actual circuit consists of 4 copies each of circuits (a) and (b).

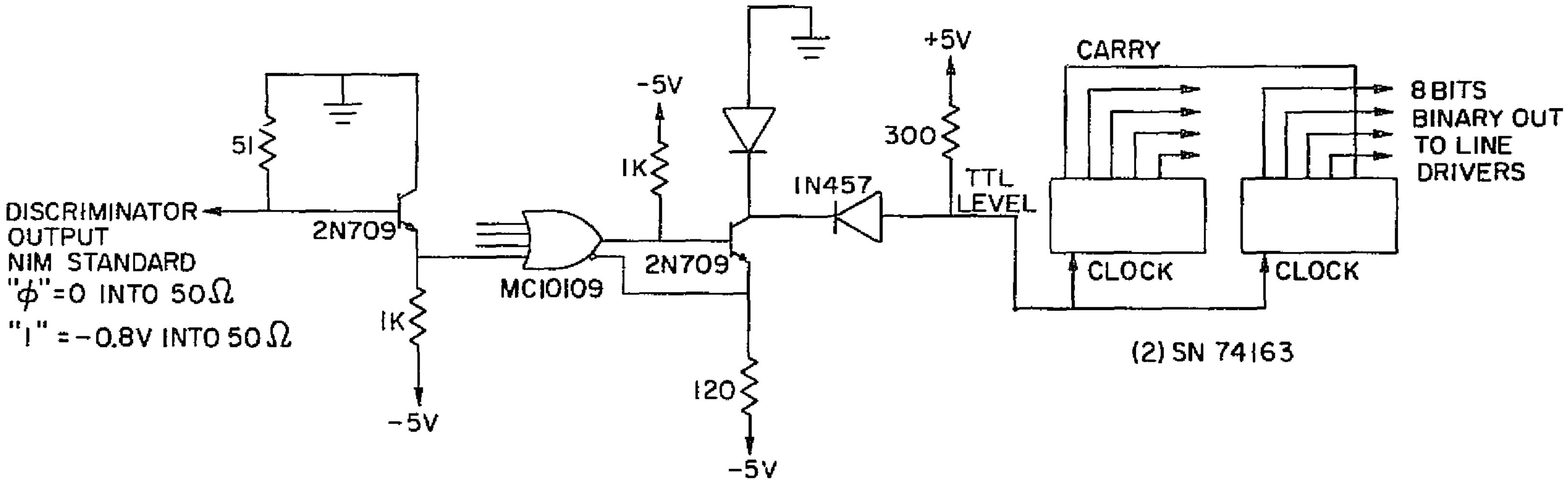


Figure I.3 Replacement circuit for the BCD upcounter.

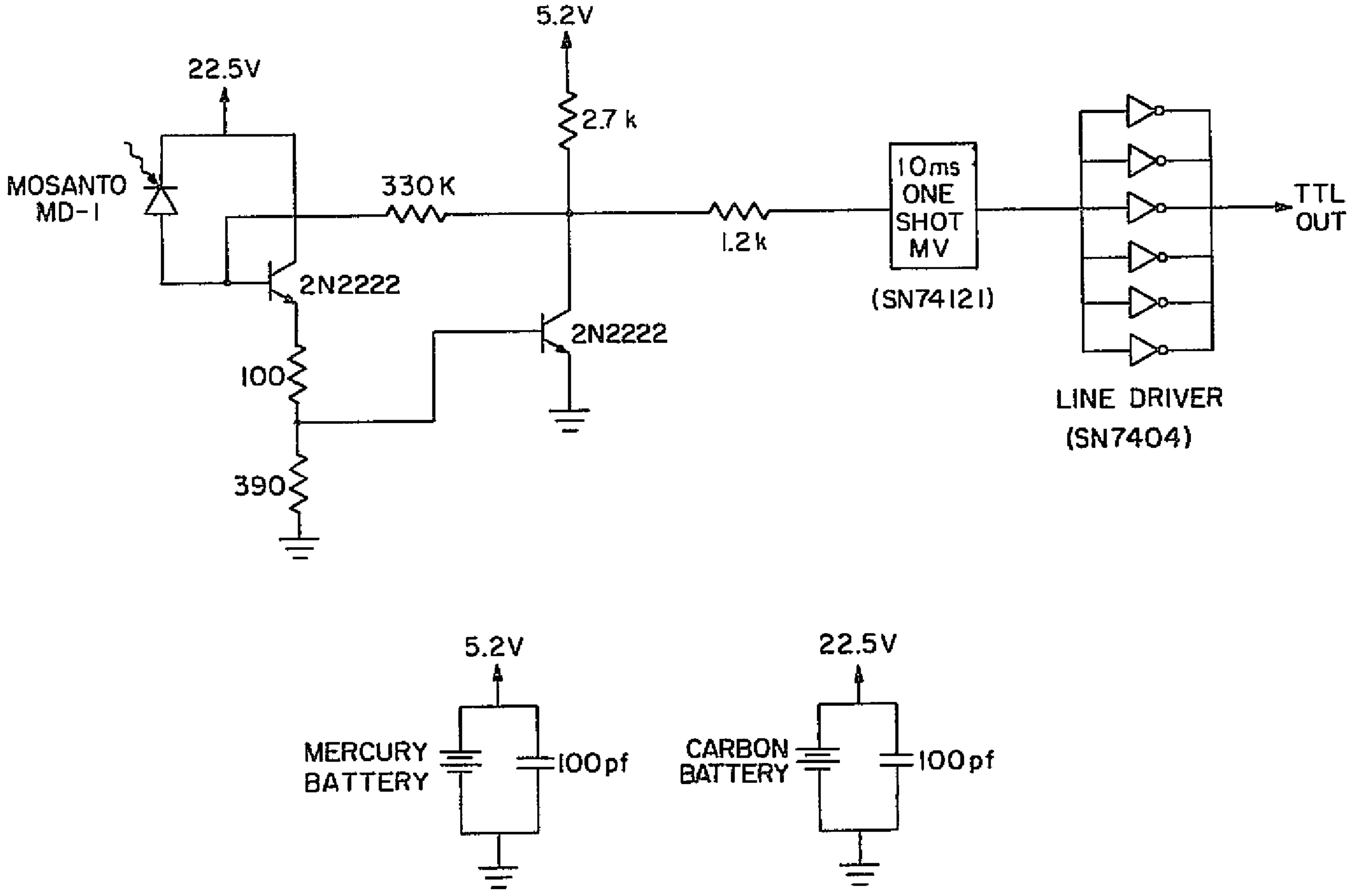


Figure I.4 Laser Command Pulse Generator.

opportunity for sampling errors to occur since it remained in a shifting state for an extended period of time. This counter was later replaced by two cascaded 4-bit 32-MHz counters which required only 25 ns for all outputs to shift (Figure I.3). These counters reduced the sampling errors substantially but did not solve the problem. The remaining errors were removed by software, but data were lost in the process.

The eventual need for a mobile system, coupled with limited manpower, suggested that a new system revolving around the HP 2114A computer should be built. It was judged more efficient to incorporate the improvements into the new system rather than to continue to modify the old system. Furthermore, the range bin size for the PDP-15 system was limited to 1.5 km which is standard for all experiments using that computer. Thus the desire for improved range resolution was also a determining factor in the construction of a new receiver system. This new system is now in use and is described in Chapter 5.

APPENDIX II

SODIUM NEUTRAL AND METAL OXIDE ION CHEMISTRY

II.1 *Neutral Sodium Chemistry*

The neutral sodium chemistry has been reviewed by many authors [*Brown*, 1973; *Kvifte*, 1973; *Hunten*, 1967]. In general theories have suggested that sodium reacts with various forms of oxygen, nitrogen, and hydrogen. Although most of the reaction rates have not been measured for the sodium chemistry, estimates may be made on the basis of laboratory experience involving reactions which are closely similar to those of sodium.

A map of the neutral species chemistry is shown in Figure (II.1). The map is useful for indicating the multitude of interactions thought to exist; although no study to date has incorporated all of these reactions. In fact most of the reactions are of secondary importance, given the current estimates for the rate constants. This is clearly indicated if we look at the time constants for these reactions. The assumed profiles for the second and third bodies of the reactions are taken from *Shimazaki and Laird* [1972] and are shown in Tables II.1 and II.2. The specific reactions and adopted rate constants are listed in Tables II.3 and II.4; these tables also include the calculated time constant. Photoionization of NaO and NaO_2 is neglected because of the lack of knowledge about these reactions. The fast bimolecular reaction of *Srivastava and Shukla* [1970] which produces the excited species NaO^* from Na is also neglected since it is likely that NaO^* will dissociate, returning Na rather than producing NaO.

It is apparent from these calculations that the neutral chemistry is dominated by the interchange of sodium with sodium monoxide via reactions involving ozone and atomic oxygen. This conclusion is in agreement with the numerical calculations of *Megie* [1977] and for the adopted reaction scheme

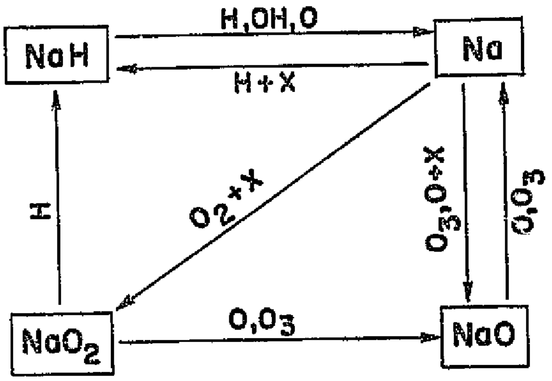


Figure II.1 Map for the neutral sodium chemical reactions.

Table II.1 Adopted profiles for major constituents during summer [*Shimazaki and Laird*, 1972]. To indicate diurnal changes, the top entry, if present, represents the daytime value (12 hrs., $\chi = 45°$) while the lower entry is the nighttime value (0 hrs., $\chi = 135°$). Omissions imply very low concentrations. The entry 4.1(10) denotes a concentration of 4.1×10^{10} cm^{-3}.

	Concentration (cm^{-3})					
Constituent	Altitude (km)					
	70	80	90	100	110	120
O(^3P)	4.1(10)	6.0(10)	6.0(11)	2.8(11)	1.0(11)	6.0(10)
	1.0(7)	1.7(10)	6.0(11)	2.8(11)	1.0(11)	6.0(10)
O_2	4.6(14)	1.0(14)	1.0(13)	1.0(12)	1.0(11)	1.9(10)
O_3	4.6(9)	3.6(8)	1.0(8)	1.0(5)		
	1.3(10)	3.6(8)	7.7(8)	7.7(6)	1.0(5)	
OH	1.3(6)	7.7(5)	3.6(3)			
	7.7(5)	2.8(6)	3.6(4)			
H	1.5(7)	1.7(8)	4.6(7)	1.0(7)	3.6(6)	2.8(6)
		1.3(8)	4.6(7)	1.0(7)	3.6(6)	2.8(6)
N_2	2.2(15)	4.6(14)	6.8(13)	7.7(12)	1.0(12)	2.8(11)
$X = N_2 + O_2$	2.7(15)	5.6(14)	7.8(13)	8.7(12)	1.1(12)	3.0(11)

Table II.2 Adopted profiles for major constituents during winter [after *Shimazaki and Laird,* 1972]. The entry format is the same as in Table II.1.

	Concentration (cm^{-3})					
Constituent	Altitude (km)					
	70	80	90	100	110	120
O(^3P)	1.5(10) 1.0(5)	7.7(9) 1.7(9)	4.6(10) 4.6(10)	3.6(10) 3.6(10)	1.1(10) 1.1(10)	8.8(9) 8.8(9)
O_2	3.2(14)	6.8(13)	1.0(13)	2.2(12)	3.6(11)	1.0(11)
O_3	6.0(8) 1.0(10)	2.2(7) 7.7(7)	6.0(6) 1.0(8)	1.0(5) 3.6(6)		
OH	2.2(5) 1.7(5)	1.7(5) 1.0(6)	1.7(3) 2.2(4)			
H	4.6(6)	6.0(7) 4.6(7)	2.5(7) 2.5(7)	1.0(7) 1.0(7)	2.8(6) 2.8(6)	1.7(6) 1.7(6)
N_2	1.7(15)	3.2(14)	6.8(13)	1.0(13)	2.8(12)	6.8(11)
$X + N_2 + O_2$	2.0(15)	3.9(14)	7.8(13)	1.2(13)	3.2(12)	7.8(11)

Table II.3 Summer time constants for neutral sodium reactions. The entry format is the same as in Table II.1.

			70	80	90	100	110	120
Loss of Na								
$Na + O_3 \rightarrow NaO + O_2$	6.5(-12)	[M]	3.3(1) 1.2(1)	4.3(2) 4.3(2)	1.5(3) 2.0(2)	1.5(6) 2.0(4)	 1.5(6)	
$Na + O + X \rightarrow NaO + X$	7.0(-33)	[K]	1.3(6) 5.3(9)	4.3(6) 1.5(7)	3.0(6) 3.0(6)	5.9(7) 5.9(7)	1.3(9) 1.3(9)	7.9(9) 7.9(9)
$Na + O_2 + X \rightarrow NaO_2 + X$	1.0(-32)	[B]	8.4(1)	1.8(3)	1.3(5)	1.2(7)	9.1(8)	1.8(10)
$Na + H + X \rightarrow NaH + X$	4.0(-32)	[K]	6.2(8)	2.6(8) 3.4(8)	7.0(9) 7.0(9)	2.9(11) 2.9(11)	6.3(12) 6.3(12)	3.0(13) 3.0(13)
Production of Na								
$NaO + O \rightarrow Na + O_2$	4.0(-11)	[K]	6.1(-1) 2.5(3)	4.2(-1) 4.2(-1)	4.2(-2) 4.2(-2)	8.9(-2) 8.9(-2)	2.5(-1) 2.5(-1)	4.2(-1) 4.2(-1)
$NaO + O_3 \rightarrow Na + 2O_2$	2.0(-12)	[M]	1.1(2) 3.9(1)	1.4(3) 1.4(3)	5.0(3) 6.5(2)	5.0(6) 6.5(4)	 5.0(6)	
$NaH + H \rightarrow Na + H_2$	4.0(-17)	[M]	1.7(9)	1.5(8) 1.9(8)	5.4(8) 5.4(8)	2.5(9) 2.5(9)	6.9(9) 6.9(9)	8.9(9) 8.9(9)
$NaH + OH \rightarrow Na + H_2O$	2.0(-15)	[K]	3.9(8) 6.5(8)	6.5(8) 1.8(8)	1.4(11) 1.4(10)			
$NaH + O \rightarrow Na + OH$	2.0(-15)	[M]	1.2(4) 5.0(7)	8.3(3) 2.9(4)	8.3(2) 8.3(2)	1.8(3) 1.8(3)	5.0(3) 5.0(3)	8.3(3) 8.3(3)
Other								
$NaO_2 + O \rightarrow NaO + O_2$	3.0(-11)	[M]	8.0(-1) 3.3(3)	5.6(-1) 2.0(0)	5.6(-2) 5.6(-2)	1.2(-1) 1.2(-1)	3.3(-1) 3.3(-1)	5.6(-1) 5.6(-1)
$NaO_2 + O_3 \rightarrow NaO + 2O_2$	1.0(-14)	[B]	2.2(4) 7.7(3)	2.8(5) 2.8(5)	1.0(6) 1.3(5)	1.0(9) 1.3(7)	 1.0(9)	
$NaO_2 + H \rightarrow NaH + O_2$	3.0(-12)	[K]	2.2(4)	2.0(3) 2.6(3)	7.3(3) 7.3(3)	3.3(4) 3.3(4)	9.3(4) 9.3(4)	1.2(5) 1.2(5)

*Two and three body reactions have units $cm^3\ s^{-1}$ and $cm^6\ s^{-1}$ respectively.
[M] = [*Megie and Blamont*, 1977], [K] = [*Kuifte*, 1973], [B] = [*Brown*, 1973].

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

Table II.4

Winter time constants for neutral sodium reactions. The entry format is the same as in Table II.1.

Reaction	Rate Constant*	Time Constants (s) with altitude (km)					
Loss of Na		70	80	90	100	110	120
$Na + O_3 \rightarrow NaO + O_2$	6.5(-12) [M]	2.6(2) 1.5(1)	7.0(3) 2.0(3)	2.6(4) 1.5(3)	1.5(6) 4.3(4)		
$Na + O + X \rightarrow NaO + X$	7.0(-33) [K]	4.8(6) 7.1(11)	4.8(7) 2.2(8)	4.0(7) 4.0(7)	3.3(8) 3.3(8)	4.1(9) 4.1(9)	2.0(10) 2.0(10)
$Na + O_2 + X \rightarrow NaO_2 + X$	1.0(-32) [B]	1.6(2)	3.8(3)	1.3(5)	3.8(6)	8.7(7)	1.3(9)
$Na + H + X \rightarrow NaH + X$	4.0(-32) [K]	2.7(9)	1.1(9) 1.4(9)	1.3(10) 1.3(10)	2.1(11) 2.1(11)	2.8(12) 2.8(12)	1.9(13) 1.9(13)
Production of Na							
$NaO + O \rightarrow Na + O_2$	4.0(-11) [K]	1.7(0) 2.5(5)	3.3(0) 1.5(1)	5.4(-1) 5.4(-1)	6.9(-1) 6.9(-1)	2.3(0) 2.3(0)	2.8(0) 2.8(0)
$NaO + O_3 \rightarrow Na + 2O_2$	2.0(-12) [M]	8.3(2) 5.0(1)	2.3(4) 6.5(3)	8.3(4) 5.0(3)	5.0(6) 1.4(5)		
$NaH + H \rightarrow Na + H_2$	4.0(-17) [M]	5.4(9)	4.2(8) 5.4(8)	1.0(9) 1.0(9)	2.5(9) 2.5(9)	8.9(9) 8.9(9)	1.5(10) 1.5(10)
$NaH + OH \rightarrow Na + H_2O$	2.0(-15) [K]	2.3(9) 2.9(9)	2.9(9) 5.0(8)	2.9(11) 2.3(10)			
$NaH + O \rightarrow Na + OH$	2.0(-15) [M]	3.3(4) 5.0(9)	6.5(4) 2.9(5)	1.1(4) 1.1(4)	1.4(4) 1.4(4)	4.6(4) 4.6(4)	5.7(4) 5.7(4)
Other							
$NaO_2 + O \rightarrow NaO + O_2$	3.0(-11) [M]	2.2(0) 3.3(5)	4.3(0) 2.0(0)	7.3(-1) 7.3(-1)	9.3(-1) 9.3(-1)	3.0(0) 3.0(0)	3.8(0) 3.8(0)
$NaO_2 + O_3 \rightarrow NaO + 2O_2$	1.0(-14) [B]	1.7(5) 1.0(4)	4.6(6) 1.3(6)	1.7(7) 1.0(6)	1.0(9) 2.8(7)		
$NaO_2 + H \rightarrow NaH + O_2$	3.0(-12) [K]	7.3(4)	5.6(3) 7.3(3)	1.3(4) 1.3(4)	3.3(4) 3.3(4)	1.2(5) 1.2(5)	2.0(5) 2.0(5)

*Two and three body reactions have units $cm^3\ s^{-1}$ and $cm^6\ s^{-1}$ respectively.
[M] = [*Megie and Blamont*, 1977], [K] = [*Kvifte*, 1973], [B] = [*Brown*, 1973].

described in Section 7.4.5 only these two reactions will be used to describe the neutral chemistry.

Some question exists as to whether these two reactions proceed by the electron jump mechanism, in which case the rate constants are about two orders of magnitude larger [*Kolb and Elgin,* 1976]. The ablation model will still be useful in this case, if much of the deposited Na is converted to Na^+ before it reaches the layer. This idea is pursued in Section 7.6. In the absence of experimental data, we will adopt those values obtained from kinetic theory analogies with hydrogen reactions.

II.2 *Metal Oxide Ion Reactions*

II.2.1 *Introduction.* Until recently the ionic reactions of the sodium chemical scheme have concentrated on the relation between Na^+, NaO^+, and NaO_2^+. However most of the reaction rates have never been measured and some of the reactions may be speculative. For this reason it is important to evaluate the ideas behind this ion chemistry; the foundation for this work was presented by *Swider* [1969, 1970].

II.2.2 *Ionization processes.* The primary source of Na^+ is photoionization of Na during the daytime,

$$Na + h\nu \rightarrow Na^+ + e^- \qquad I = 1.7 \times 10^{-5}\ s^{-1} \tag{II.1}$$

This value of the photoionization rate is quite typical for most studies of sodium which consider the ion chemistry. *Swider* [1969] obtained the value from the integrated product of the sodium absorption cross section and the unattenuated solar flux. The value is an upper limit, since 100% ionization efficiency is assumed. The calculation is of course sensitive to the resolution in the cross section and flux data and also the size of the integration step. Photoionization of Na by HLyα is negligible since unity optical depth of the solar ionizing flux for Na occurs near 40 km.

Charge exchange with NO^+ and O_2^+ is of secondary importance in the daytime but dominates at night over photoionization,

$$Na + O_2^+ \rightarrow Na^+ + O_2 \qquad k = 6 \times 10^{-10} \text{ cm}^3 \text{ s}^{-1} \tag{II.2}$$

$$Na + NO^+ \rightarrow Na^+ + NO \qquad k = 8 \times 10^{-11} \text{ cm}^3 \text{ s}^{-1} \tag{II.3}$$

Rates of these reactions are taken from *Farragher et al.* [1969]. Since the electron affinity for O_2^+ and NO^+ is about 12 ev and 9 ev respectively, while the ionization potential of Na is about 5 ev, these reactions are exothermic. The excess energy released by the reaction may appear as vibrational excitation in the diatomic product molecule or may even dissociate the molecule; it is possible that the excess energy may instead excite the sodium ion.

Sodium oxide ions may be reduced by atomic oxygen by the process

$$NaO^+ + O \rightarrow Na^+ + O_2 \qquad k \approx 10^{-10} \text{ cm}^3 \text{ s}^{-1} \tag{II.4}$$

The analogous reactions involving MgO^+ and SiO^+ have been observed in the laboratory [*Ferguson*, 1972] and are found to be quite fast [*Ferguson and Fehsenfeld*, 1968]. Therefore it is likely that process (II.4) is similarly fast. If this is so, then processes involving NaO^+ formation must be considered as potential sources for Na^+.

Since the ionization potential for NaO is thought to be about 7.6 ev [*Swider*, 1969] it is unlikely that photoionization of NaO is a major source of NaO^+. However, the charge exchange processes

$$NaO + O_2^+ \rightarrow NaO^+ + O_2 \tag{II.5}$$

$$NaO + NO^+ \rightarrow NaO^+ + NO \tag{II.6}$$

are exothermic. Yet as discussed in Section II.1, the $[O]/[O_3]$ ratio keeps [NaO] small above 85 km. Since NO^+ and O_2^+ begin to increase above 85 km, if we assume the same rate for processes (II.5) and (II.6) as we did for (II.2) and (II.3) respectively, it is unlikely that charge exchange processes with NaO are a significant source of Na^+.

Reaction (II.2) may have the branch

$$Na + O_2^+ \rightarrow NaO^+ + O \tag{II.7}$$

with the rate constant of reaction (II.2) being about 5 times that of (II.7) [*Ferguson and Fehsenfeld*, 1968]. However since NaO^+ is returned quickly to Na^+, reaction (II.7) is probably not that significant. The analogous reaction involving NO^+ is probably endothermic.

Sources of Na^+ resulting from NaO_2 will not be considered because information is lacking on the bond energy and ionization potential for NaO_2. However if NaO is not thought to be an important source of Na^+, as discussed above, NaO_2 is most likely of lesser significance since $[NaO_2]$ is thought to be much less than [NaO] [*Brown*, 1973].

Photodissociation of the sodium oxide ions may also be a source of Na^+,

$$NaO^+ + h\nu \rightarrow Na^+ + O \tag{II.8}$$

$$NaO_2^+ + h\nu \rightarrow Na^+ + O_2 \tag{II.9}$$

but experimental evidence does not exist for these reactions; these reactions may be invoked to create a diurnal variation in Na^+ [*Blamont and Donahue*, 1964].

II.2.3 *Loss processes of* Na^+. Sodium ions may be deionized by the reactions

$$Na^+ + e^- \rightarrow Na + h\nu \tag{II.10}$$

$$Na^+ + A^- \rightarrow Na + A. \tag{II.11}$$

Direct recombination with electrons is so unlikely that it can be ignored; a typical rate constant is given as $k = 3 \times 10^{-12}$ cm^3 s^{-1} [*Bates and Dalgarno*, 1962]. Mutual neutralization is much faster with a rate constant $k \approx 10^{-7}$ cm^3 s^{-1} [*Brown*, 1973], but concentrations of negative ions are not well known. Certainly above 85 km, A^- should fall off rapidly; $[A^-]$ during the day may be as high as 10^3 cm^{-3} [*Brown*, 1973]. This reaction was also considered to explain diurnal variation in ionization [*Hunten*, 1967].

Indirect processes involving various forms of oxygen have also been suggested as being important in the deionization of Na^+. Two three body reactions have been proposed which form sodium oxide ions,

$$Na^+ + O + X \rightarrow NaO^+ + X \qquad \text{(II.12)}$$

$$Na^+ + O_2 + X \rightarrow NaO_2^+ + X \qquad \text{(II.13)}$$

These reactions are fairly slow at 90 km having time constants of about 10 hours; rate constants are given for reactions (II.12) and (II.13) as $k = 10^{-30}$ cm^6 s^{-1} and $k = 5 \times 10^{-32}$ cm^6 s^{-1} respectively by *Brown* [1973] based on the experimental evidence of *Ferguson and Fehsenfeld* [1968]. It should be emphasized that NaO_2^+ may be unstable and therefore would break up easily in collisions with other atmospheric molecules; furthermore the reaction

$$NaO_2^+ + O \rightarrow NaO^+ + O_2 \qquad \text{(II.14)}$$

is possible exothermic since the NaO^+-O bond is weaker than the O-O bond [*Swider*, 1969]. Brown estimates the rate of this reaction as $k \approx 10^{-10}$ cm^3 s^{-1} which would imply a substantial proportion of NaO_2^+ is converted to NaO^+. *Anderson and Barth* [1971] have discussed the possibility of an analogous reaction involving Mg and suggest that its rate is $\approx 10^{-10}$ cm^3 s^{-1}. Since NaO^+ returns to Na^+ rapidly, as mentioned above, these loss processes for Na^+ may not be that significant.

Two body reactions have also been proposed as deionizing Na^+

$$Na^+ + O_3 \rightarrow NaO^+ + O_2 \qquad (II.15)$$

$$Na^+ + O \rightarrow NaO^+ + h\nu \qquad (II.16)$$

$$Na^+ + O_2 \rightarrow NaO_2^+ + h\nu \qquad (II.17)$$

The reaction involving ozone is possibly endothermic. Similar reactions involving Mg^+, Ca^+, Fe^+ were all observed to be fast but the reactions involving Na^+ and K^+ produced no products, implying a rate constant $k < 10^{-11}$ [*Ferguson and Fehsenfeld*, 1968]. Unfortunately the thermochemistry does not illuminate the problem; for NaO^+, the bond energy is $D_0(Na^+\text{-}O) = 1.25 \pm 0.7$ ev [*Hildenbrand and Murad*, 1970]. The reaction is exothermic if $D_0(NaO^+) > D_0(O_2\text{-}O) = 1$ ev and hence the thermochemistry is ambiguous. In any event the return of NaO^+ to Na^+ is so fast that the deionization of Na^+ by O_3 is unimportant.

Swider [1968] suggests that (II.16) and (II.17) are important near 100 km with three body reactions dominating below 90 km. An appropriate value for the rate constant is 10^{-17} $cm^3\ s^{-1}$. Although this reaction may be of value in the neutralization of sporadic *E* layers it is debatable whether the reaction can be appropriately applied to the main Na^+ layer where high atomic oxygen concentrations will dissociate NaO^+.

II.2.4 *Dissociative recombination of metal oxide ions.* Dissociative recombination by the processes

$$NaO^+ + e^- \rightarrow Na + O \qquad (II.18)$$

$$NaO_2^+ + e^- \rightarrow \begin{cases} Na + O_2 & (II.19a) \\ NaO + O & (II.19b) \end{cases}$$

are believed to be fast having rate constants on the order of 10^{-7} to 10^{-6} cm^3 s^{-1}. Based on the discussion in Section 7.4.1, the product NaO of reaction (II.19b) is rapidly returned to Na at the altitudes of interest.

The principal weakness of this ion scheme centers on the lack of knowledge for rates of reactions and the existance of certain reactions. Furthermore the rapid return of metal oxide ions to Na^+ via reaction (II.4) places a severe constraint on the theory. The severity of the problem is indicated below.

II.2.5 *The general steady-state solution based on reactions discussed in Sections II.1 and II.2.* Figure II.2 schematically displays the chemical processes involving sodium compounds which might be *a priori* of importance in the 80-120 km region. One is struck immediately by the symmetry of the scheme. In addition, an examination of time constants at 90 km reveals that the rapid return of NaO to Na and NaO^+ to Na^+ produces essentially two closed loops, one neutral and ion ionic, with the dominant daytime connection between these loops being photoionization of atomic sodium and dissociative recombination of the ionic species with electrons.

The rates of these reactions are given in Table II.5. Since reactions involving negative ions were assumed to be less important and cluster ions were not initially included, these ions were eliminated from Brown's scheme. Assumed daytime profiles are shown in Figure II.3. Other constituent values were taken from *CIRA* [1965] to parallel Brown's work.

Calculations were performed by setting the complete time expression for each constituent to zero. Molecular and eddy diffusion were not taken into consideration. The resulting set of simultaneous equations were solved for relative concentrations by setting $[Na_x] = 1$ and then all constituents were scaled appropriately so that $[Na] = 1000$ cm^{-3} at 90 km.

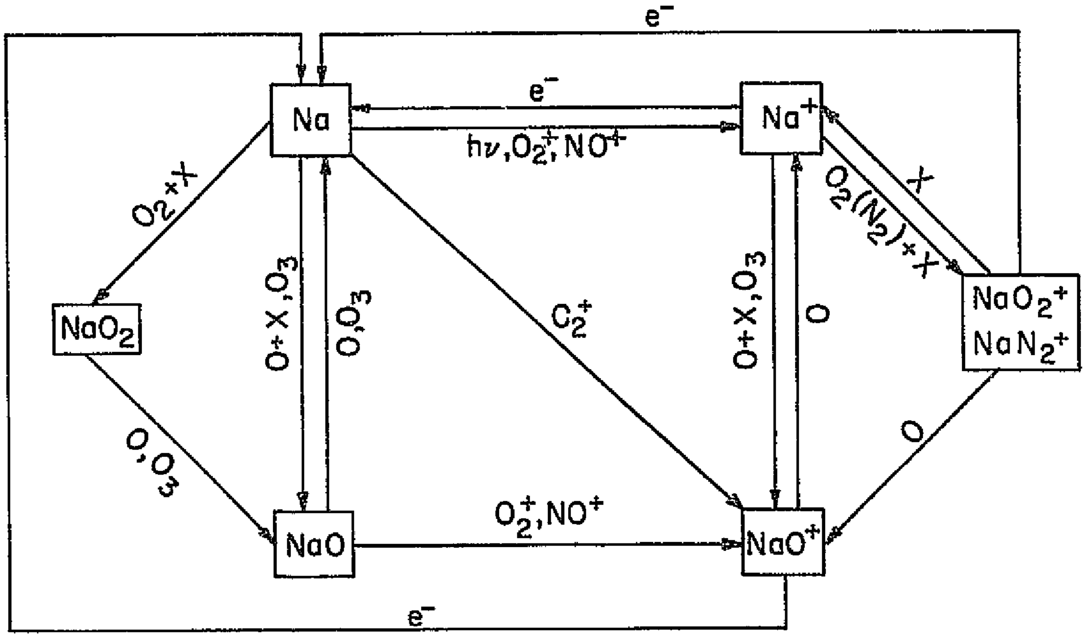


Figure II.2 Complete sodium reaction scheme after *Brown* [1973].

ORIGINAL PAGE IS
OF POOR QUALITY

Table II.5

Rate constants for sodium reactions.

Reaction	Rate Constant
1. $Na + O_2^+ \to Na^+ + O_2$	6×10^{-10} cm^3 s^{-1}
2. $Na + NO^+ \to Na^+ + NO$	8×10^{-11}
3. $Na + P^+ \to Na^+ + P$	2.4×10^{-9}
4. $Na + h\nu \to Na^+ + e^-$	$J = 1.7 \times 10^{-5}$ sec^{-1}
5. $Na + O_2 + X \to NaO_2 + X$	1×10^{-32} cm^6 s^{-1}
7. $Na + O_3 \to NaO + O_2$	1×10^{-13}
9. $Na + O + X \to NaO + X$	1×10^{-33}
10. $Na + O_2^+ \to NaO^+ + O$	1×10^{-10}
11. $Na^+ + e^- \to Na$	1×10^{-12}
12. $Na^+ + A^- \to Na + A$	1×10^{-7}
13. $Na^+ + O_2 + X \to NaO_2^+ + X$	5×10^{-32}
14. $Na^+ + N_2 + X \to NaN_2^+ + X$	1×10^{-31}
15. $Na^+ + O + X \to NaO^+ + X$	1×10^{-30}
16. $Na^+ + O_3 \to NaO^+ + O_2$	1×10^{-12}
17. $NaO_2^+ + X \to Na^+ + O_2 + X$	1×10^{-12}
18. $NaN_2^+ + X \to Na^+ + N_2 + X$	1×10^{-13}
19. $NaO_2^+ + O \to NaO^+ + O_2$	1×10^{-10}
20. $NaN_2^+ + O \to NaO^+ + N_2$	1×10^{-10}
21. $NaN_2^+ + e^- \to Na + N_2$	2×10^{-6}
22. $NaO_2^+ + e^- \to Na + O_2$	2×10^{-6}
23. $NaO^+ + O \to Na^+ + O_2$	1×10^{-10}
24. $NaO^+ + e^- \to Na + O$	4×10^{-7}
26. $NaO^+ + A^- \to NaO + A$	1×10^{-7}
27. $NaO_2 + O \to NaO + O_2$	1×10^{-11}
29. $NaO + O \to Na + O_2$	1×10^{-11}
30. $NaO + O_2^+ \to NaO^+ + O_2$	5×10^{-10}
31. $NaO + NO^+ \to NaO^+ + NO$	5×10^{-10}
33. $NaO_2 + O_3 \to NaO + 2O_2$	1×10^{-14}
36. $NaO + O_3 \to Na + 2O_2$	1×10^{-16}

Note: Branching ratio between reactions 1 and 10 has been taken as 20%. Reactions and rate constants are taken from *Brown* [1973].

ORIGINAL PAGE IS
OF POOR QUALITY

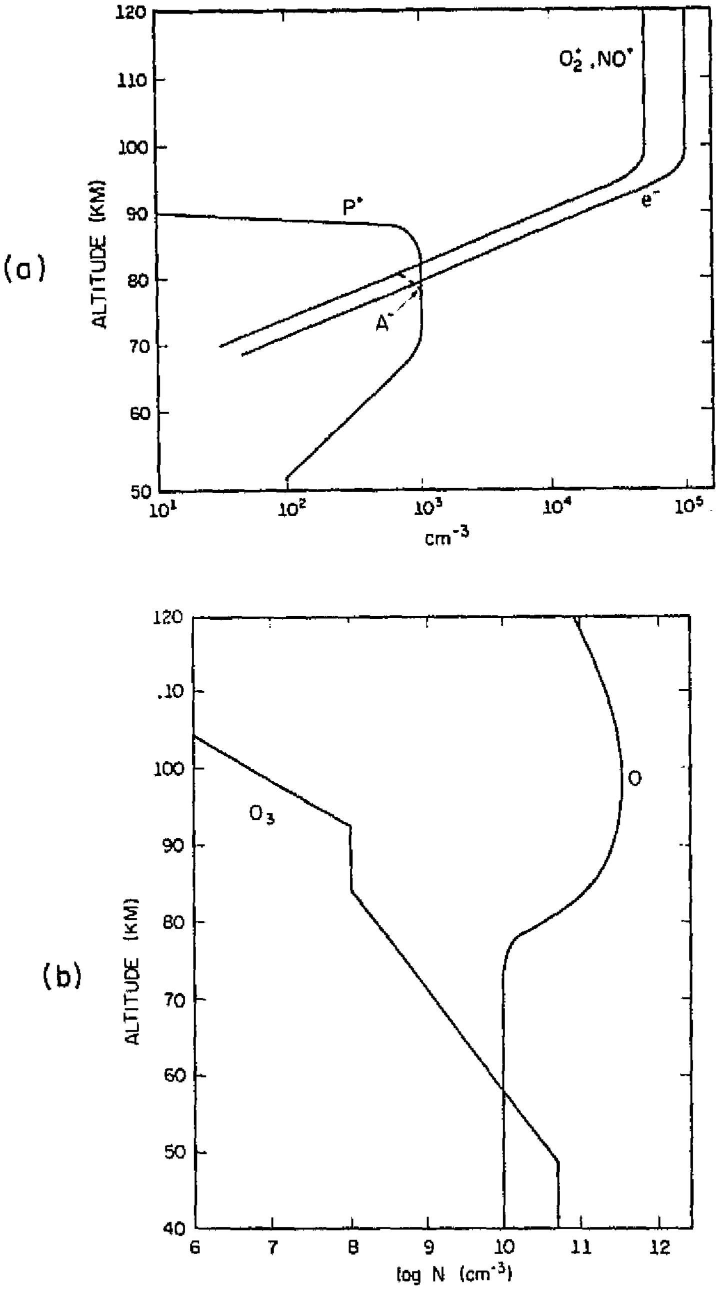


Figure II.3 (a) Assumed daytime profiles of ionic species employed in calculations; (b) Assumed daytime profiles for O and O_3 (adapted from *Brown* [1973]).

Results of these calculations are shown in Figure II.4. In interpreting Figure II.4 it should be noted that profiles for these constituents are not being plotted. Rather the curves at each altitude give the predicted relative proportions of the respective constituents. It is obvious though that Na^+ should dominate at all altitudes, and this is not observed experimentally. Furthermore it was found that inclusion of the negative ions and cluster ions aided the falloff of $[Na^+]/[Na_x]$ below 90 km. While NaO^+, ${NaO_2}^+$, and ${NaN_2}^+$ were small in number density, the reactions returning these constituents to Na^+ were sufficiently fast that they had to be included in the scheme. Finally it should be noted that the restriction $[Na_x] = 1$ creates the problem that a small change in Na^+ can make a great change in the other minor constituents. It was obvious that the very completeness of Brown's scheme with all of its uncertainties made it quite difficult to obtain agreement with experimental results.

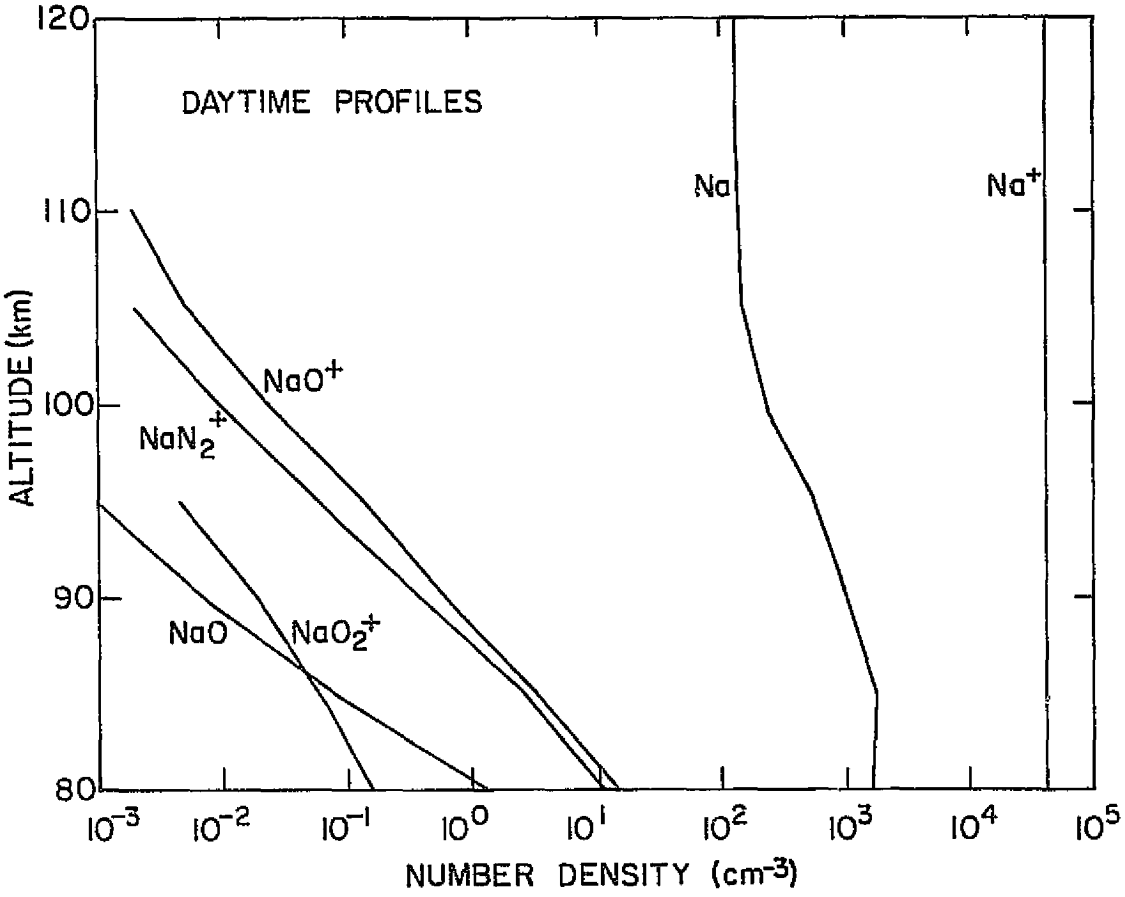


Figure II.4 Daytime profile obtained with scheme of *Brown* [1973].

APPENDIX III

CLUSTER ION REACTIONS

III.1 *Experimental Studies*

Over the past ten years mass spectrometer measurements of ion-molecule reactions have been extended to higher pressures and to a wider range of temperatures. As a result it has been shown that third order kinetics play an important role in ion-molecule reactions. In many instances the clustering reactions of the form

$$M^+ + A + X \rightarrow M^+{\cdot}A + X \qquad \text{(III.1)}$$

where M^+ is an ion, and A and X neutral species, have been studied for the case of alkali ions. These reactions have the advantage of employing a less complex ion which can be copiously produced. Minimum values for rate constants or equilibrium constants for atmospheric molecules clustering to ions can be obtained by studying the clustering of alkali ions. As a result laboratory chemists can understand the clustering process more fully; and ionospheric chemists, interested in the metallic species, benefit directly. Therefore unlike the metal oxide ion chemistry discussed in Section II.2, several clustering reactions have been measured for Na or may be easily extended to the Na case. If clustering appears to be important for the sodium chemistry, then much of the chemistry no longer needs to rely on speculation.

III.2 *The Clustering Mechanism*

The mechanism of clustering is often considered as an energy transfer process [*Good*, 1975]. Reaction (III.1) is believed to proceed in a two step process, temporarily forming an intermediate complex which may either break up or be stabilized by the third body of the reaction. This process is indicated by the following two reactions

$$M^+ + A \underset{k_d}{\overset{k_c}{\rightleftharpoons}} M^+\cdot A^* \qquad \text{(III.2)}$$

$$M^+\cdot A^* + X \underset{k_a}{\overset{k_s}{\rightleftharpoons}} M^+\cdot A + X \qquad \text{(III.3)}$$

where

k_c = combination rate constant

k_d = decomposition rate constant

k_s = stabilization rate constant

k_a = activation rate constant.

Usually $k_d \gg k_s[X]$ and assuming equilibrium, the overall forward and reverse rates of this process may be written in terms of the above rate constants

$$k_f = k_c k_s / k_d \qquad \text{(III.4)}$$

$$k_r = k_a \qquad \text{(III.5)}$$

What is interesting about cluster ion reactions is that they exhibit an apparent negative Arrhenius energy temperature dependence. Although these reactions are more complicated than common three body reactions, we may apply the general form for a reaction rate coefficient. The forward rate coefficient can be written as [*Good,* 1971]

$$k_f = a(T/T_R)^b \qquad \text{(III.6)}$$

and the reverse rate may be written as

$$k_r = c(T/T_R)^d \exp(-e/T) \qquad \text{(III.7)}$$

where a, b, c, d, and e are constants and T_R is a reference temperature, usually 300°K. *Heimerl* [1976] develops a method for estimating the errors incurred in the determination of these constants based on thermodynamical data. Because the necessary information for alkali clusters is not available, the error analysis fails. However, the basic temperature dependence may still be determined. Equations (III.6) and (III.7) indicate that the temperature dependence of the forward rate constant is the same as that of the pre-exponential factor of the reverse rate coefficient. The exponent, b, is negative as mentioned above and for D_0(M-A) >> nRT, b is related to the number of vibrational degrees of freedom in the cluster; that is $|b|$ increases as the number of degrees of freedom increase [*Durden et al.*, 1969]. Although b has been shown to be large in absolute value for $NO^+ \cdot N_2$ formation [*Arnold and Krankowsky*, 1977], it is thought that b is between -1 and -2 for reactions involving Na [*Keller and Beyer*, 1971]. As a rule of thumb k_f may increase by a factor of 2 from 300°K, where the reaction rates are measured, to 200°K, appropriate for mesospheric temperature.

Often when information is lacking specifically about k_f and k_r, the equilibrium constant

$$K = k_f/k_r \tag{III.8}$$

is measured. This quantity is independent of the third body and strongly temperature dependent. If k_f and k_r are pressure units then K will be denoted by K_p; if they are in concentration units, then K will be denoted by K_c.

III.3 *Determining the Equilibrium Constant for Cluster Ion Reactions at Mesospheric Temperatures*

The change in the Gibb's free energy, ΔG, for a reaction has long been recognized as a criterion for the equilibrium and spontaneity of the reaction. By intuition it is thought that heat of reaction, ΔH, or alternatively the change in enthalpy, is responsible for the spontaneity of the reaction; that is, if a reaction is exothermic, then it will occur. For an isothermal reaction the relation

$$\Delta G = \Delta H - T\Delta S \qquad \text{(III.9)}$$

applies where

T = absolute temperature

ΔS = change in entropy.

Hence it can be seen that both concepts apply for $T\Delta S$ small. For cluster reactions, $T\Delta S$ is relatively small, and hence either ΔG or ΔH will predict the spontaneity of the reaction.

The equilibrium constant may be related to either the change in the Gibb's free energy or the change in enthalpy. A common expression for the variation of K_p with temperature is given by

$$-\Delta G = RT\ln K_p \qquad \text{(III.10)}$$

However ΔG may be a function of temperature, and hence equation (III.10) is not useful for extrapolating K_p from room temperature, where the reactions are measured, to mesospheric temperatures. However since ΔG is independent of pressure, equation (III.10) predicts that K_p is likewise independent of temperature.

The change in enthalpy may be related to the equilibrium constant by the van't Hoff relation,

$$\ln \frac{K_p(T_2)}{K_p(T_1)} = \frac{-\Delta H}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \qquad \text{(III.11)}$$

where R = universal gas constant.

There is evidence that ΔH is relatively constant with temperature for most cluster ion reactions over a 100°K range of temperature [*Heimerl*, 1976], and this assumption is certainly true for hydrates of Na^+ as shown in Figure (III.1). Hence equation (III.2) may be used to determine the value of K_p at mesospheric temperature. Furthermore it is important to note that $-\Delta H > 0$ for cluster ion reactions, and so K_p will increase as the temperature of the

ORIGINAL PAGE IS
OF POOR QUALITY

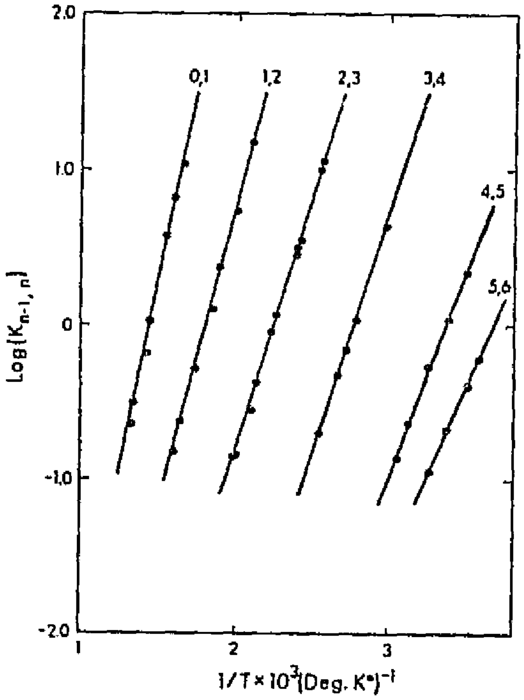


Figure III.1 Van't Hoff plots for equilibrium constants $K_{n-1,n}$ of reactions $Na^+ \cdot (H_2O)_{n-1} + H_2O + H_2O$ $Na^+ \cdot (H_2O)_n + H_2O$ [*Dzidic and Kebarle*, 1970].

reactants is lowered.

One final point must be made about the equilibrium constant. Normally K_p is expressed in standard-state units of atm^{-1}. However for aeronomy the equilibrium constant, K_c, has units cm^3. It is important to recognize the difference between K_p and K_c. By definition K_p is given for the reaction

$$A + B \rightarrow C + D \tag{III.12}$$

as

$$K_p = \frac{P_C^c \ P_D^d}{P_A^a \ P_B^b} \ \text{(at equilibrium)} \tag{III.13}$$

where $P_i^{\,j}$ is the partial pressure of the ith gas and j represents the number of times the ith gas appears in the stoichiometric equation. Assuming an ideal gas, $P_i = n_i RT/V = c_i RT$, where c_i is the concentration of the ith gas, we can rewrite (III.13) as

$$K_p = \frac{C_C^c \ C_D^d}{C_A^a \ C_B^b} \ (RT)^{c+d-a-b} = K_c (RT)^{\Delta\nu} \tag{III.14}$$

where $\Delta\nu$ is the number of moles of products minus the number of moles of reactants. For cluster ion reactions, $\Delta\nu = -1$. Hence when the equilibrium constant is given in units of cm^3, k_f/k_r should have an extra factor of T; that is if k_f varies as T^{-x} then k_r should vary as T^{-x-1} [*Reid*, 1977]. The point is a minor one because the factor of T may be assumed into the activation term of the reverse rate coefficient.