PREFACE

There are several forms that a book about the ionosphere might take, and it is perhaps appropriate to mention what this book does not do. It does not deal with the mathematics of radio propagation, nor with the use of ionospheric data for communications and making predictions; nor does it discuss experimental techniques in any detail. Its purpose is to discuss the physical processes at work in the ionosphere, and to relate them to the observed facts of ionospheric behavior, without too much detail. We think this approach leads to a fairly straightforward description, even though it might seem contrary to the usual scientific ordering of observation followed by theory. Our approach does carry a danger that we might select the observational description to match the theoretical framework, but we do not think we have glossed over the many difficulties that arise in fitting theory to observation.

It would be easy to overload our account with observational facts, for they exist in profusion. Most ionospheric data have come from vertical soundings although a considerable number of ionograms has been obtained from satellite-borne sounders above the ionosphere. Data on important parameters, particularly the composition, temperature, and motions of the atmosphere as a whole, which are essential to complete the physical picture, are becoming available mainly through rocket and satellite experiments. Unresolved questions include magnetic storms and the accompanying ionospheric phenomena, and the relationships between the ionosphere and other parts of the atmosphere. We shall refer to these topics at several places in our book.

This book had its origins at Stanford University when, in 1963, we started to write a fairly simple account of the ionosphere. Our account was intended to be useful to physicists, especially graduate students, possessing no great knowledge of the upper atmosphere. It was issued as a Stanford Electronics

Early in 1966 we decided to revise this report completely, updating it and expanding it as seemed appropriate to cover various new fields. This book is the result. It is intended for the same audience as was the original report; we hope it will also be useful to specialists in fields related to ionospheric physics who are not themselves experts on the ionosphere.

We have included six tables to summarize numerical and other detailed information which we consider useful. Here, as elsewhere in the book, we encounter the problem of units. Although we might wish to use only Système Internationale m.k.s. units, we realize that most papers on ionospheric physics use c.g.s. units, especially for such basic quantities as electron concentration and production and loss rates. We therefore use S.I. units as much as seems expedient, while making concessions to common usage in areas where c.g.s. units are most firmly entrenched. In diagrams reproduced from other publications we retain the original units.

The book "Bibliography of the Ionosphere" by L. A. Manning (Stanford Univ. Press, Stanford, California, 1962) lists about 4000 papers published up to 1960. Since then thousands of further papers have appeared. From the vast literature in this field we have selected a few hundred papers which illustrate the material we describe. We do not intend our bibliography to be complete in any area, nor to be relevant to questions of priority. Occasionally several papers on a particular topic appear together in one issue of a journal (sometimes being the record of a conference), and in such cases we find it convenient to give collective references, instead of citing every paper individually.

May, 1969

HENRY RISHBETH
OWEN K. GARRIOTT
ACKNOWLEDGMENTS

A number of acknowledgments are called for. The early stages of the work were carried out under grant NsG30-60 of the U.S. National Aeronautics and Space Administration; later the work became an official project of the Radio and Space Research Station at Slough, within the U.K. Science Research Council. Permission for this publication has been granted by the Director of the Radio and Space Research Station, and by the National Aeronautics and Space Administration.

Of the personal acknowledgments, we must first thank Mr. J. A. Ratcliffe, formerly of the Cavendish Laboratory, Cambridge, and of the Radio and Space Research Station, for his encouragement of this project, and for reading our draft manuscript. The manuscript was also read by Dr. A. V. da Rosa of Stanford University and Professor J. A. Gledhill of Rhodes University, South Africa. Some of the theoretical derivations given in this book have been adapted from lectures at the Cavendish Laboratory, Cambridge, by Dr. K. Weekes. We are also grateful to several colleagues at the Radio and Space Research Station for their reading of portions of the text. The comments and criticism we have received from all these people have enabled us to make a great many improvements in our original draft.

We would like to thank all those who have given us permission to reproduce diagrams. Lastly, we must express appreciation of all the excellent work of the typing, drawing office, and library staff, principally at R.S.R.S. but also at Stanford, which has made possible the production of this book.

vii
THE NEUTRAL ATMOSPHERE

1.1 Atmospheric Nomenclature

The scientific study of the upper atmosphere has become known as aeronomy. As generally happens in the study of a complex natural system, nomenclature has been developed to describe the different parts of the atmosphere (Chapman, 1950). The description may be based on chemical composition or on temperature or on the dominant physical processes. For each of these alternatives there exists a series of words terminating in "-sphere," each using some property to characterize the atmosphere within a certain range of altitude. The upper boundary of any "-sphere" may be denoted by a similar word ending in "-pause"; thus, the "tropopause" is the upper boundary of the troposphere. Sometimes these levels can be defined to within a few kilometers, but in other cases, the "-pause" nomenclature is inappropriate because the divisions can only be specified within tens or hundreds of kilometers. Figure 1 contains all the "-sphere" terms that occur in this book. The figure also shows typical daytime profiles (i.e., distributions with respect to altitude) of two quantities of particular interest to us: the neutral gas temperature and the electron concentration (otherwise known as electron density).

The lowest atmospheric layer is the troposphere, in which the composition is uniform and the temperature generally decreases upward. The emission and absorption of infrared radiation by molecules such as water vapor, carbon dioxide, and ozone provide efficient transfer of heat between different levels in this region. Radiative transfer leads to a decrease of temperature with height, but convection prevents the "lapse rate," or negative temperature gradient, from exceeding a limiting value, about 10°K km⁻¹. Generally, the lapse rate is smaller than this, and under certain conditions, particularly
at night, "inversions" may be set up in which the temperature gradient near the ground is positive.

Until the turn of the century, it was supposed that the temperature continues to decrease upward and that the atmosphere terminates at about 50 km, there merging into cold interplanetary space. However, experiments with balloon-borne thermometers, like those of L. P. Teisserenc de Bort in 1898, revealed a nearly isothermal region at a temperature of about 220°K, beginning at about 11 km altitude in mid-latitudes. This region has since been named the stratosphere, and its base is known as the tropopause.

The existence of a temperature inversion, or positive temperature gradient, above the stratosphere was suggested by observations of sound propagation over distances of 100 km or more, which seemed to result from the refraction of sound waves in the upper atmosphere. We now know that such an inversion results from the presence of a trace of ozone in the atmosphere. Although it is a very minor constituent, contributing only a few millionths of the total ground-level pressure, the presence of ozone is very important. It absorbs all solar ultraviolet radiation of wavelength less than 2900 Å, and partially absorbs wavelengths between 2900 and 3600 Å. The atmosphere is heated
by this absorption and the earth's surface is shielded from the otherwise lethal radiation. This region of elevated temperature is known as the mesosphere. It does not seem to have a well-defined lower boundary (stratopause), but its upper boundary (mesopause) lies at 80 to 85 km, and is the coldest level in the entire atmosphere (about 1800K). Heat flows toward this level by conduction from above and is removed by radiation in the infrared and visible airglow, and by downward eddy transport into the mesosphere (Sec. 1.3). The existence of the temperature minimum arises from the lack of any strong heating mechanism at this height. Ozone cannot exist in appreciable quantities at this level or higher, being rapidly destroyed by photochemical reactions.

The shorter ultraviolet radiations are absorbed at greater heights, in the thermosphere, and are responsible for the high temperatures existing there. Most of the heat liberated in the thermosphere is removed by downward conduction, so that the temperature increases upward. Finally, the heat conductivity becomes so good that the region of the upper thermosphere is maintained in a nearly isothermal condition at a relatively high temperature (1000-2000K). In the exosphere, collisions between molecules are so infrequent that neutral particles move in ballistic orbits subject only to gravity, whereas the ionized particles are constrained by the magnetic field. This region will be discussed in detail later in this chapter (see Sec. 1.5).

The foregoing description is closely linked to the vertical temperature profile, shown in Fig. 2. Parallel to this description, a classification in terms of composition can be made. The term ozonosphere may be principally associated with the mesosphere, but it is not too well defined because appreciable concentrations of ozone exist in the stratosphere. The vertical distribution of ozone extends, roughly, between 10 and 80 km with a peak of concentration at about 25 km (London, 1967).

The ionosphere may be defined as the part of the earth's upper atmosphere where ions and electrons are present in quantities sufficient to affect the propagation of radio waves. It extends down to perhaps 50 km and thus overlaps the ozonosphere; the symbols D, E, F1, and F2 are used to distinguish its various parts. It has no well-defined upper boundary, but merges into (or may be extended to include) the heliosphere, where neutral and ionized helium are important constituents, and the protonosphere, which is composed principally of ionized hydrogen. The latter "-spheres" are still poorly defined in extent.

The remaining terms used in Fig. 1 refer to the physical regimes at different levels in the atmosphere. Up to about 100 km, the atmosphere is well
mixed by turbulence. The relative abundances of major constituents may be assumed independent of height, but the abundances of chemically active constituents, such as ozone, are subject to variation, as are constituents such as water vapor and contaminants in the troposphere. In the lower thermosphere, the composition can be modified by photochemical reactions, such as dissociation of molecular gases.

The level at which turbulence ceases may be called the turbopause; it is rather sharply defined and lies at about 100 km. At greater heights, the lack of turbulence enables a condition of diffusive separation to be established, in which the vertical distribution of each neutral gas depends on its molecular weight. The distribution of chemically active gases, however, may be influenced to some extent by photochemical reactions as well as by diffusion. This is especially true of the ionization which attains a diffusely controlled distribution only at heights well above the F2 peak, at 400 km or so.
Finally, there is the magnetosphere, the region in which the earth's magnetic field controls the dynamics of the atmosphere. It is difficult to define a lower limit, since the movement of ionization is geomagnetically controlled at all heights above about 150 km (or even less); but the magnetosphere certainly includes the whole atmosphere above the level at which ionized constituents become predominant over neutral constituents, probably at about 1500 km. Because of this magnetic control, the earth's atmosphere may be said to terminate at the magnetopause, the boundary of the geomagnetic field which lies at about ten earth radii on the day side of the earth and at a greater distance on the night side.

In this brief introduction, we have not attempted to justify our generalized statements, or to provide full literature citations. In the remainder of this chapter, we will consider the structure of the upper atmosphere in some detail.

1.2 The Hydrostatic Equations of Atmospheric Structure

1.21 Basic Equations

The distribution with height \( h \) of a neutral atmospheric gas may be assumed to be subject to the perfect gas law

\[
p = nkT
\]

and the hydrostatic or barometric equation

\[
- \frac{dp}{dh} = nmg = \rho g
\]

The symbols \( p, \rho, \) and \( n \) are used to denote pressure, density, and concentration (sometimes called number density), respectively. If it is necessary to specify any particular gas, the chemical symbol is appended in square brackets, e.g., \( n[N_2] \). \( T \) denotes absolute temperature and \( g \) the acceleration due to gravity. Molecular mass is denoted by \( M \) (in atomic units) and \( m \) (in conventional units) so that \( \rho = nm \). The ratio \( M/m \) is denoted by \( \mathcal{N} \), which is \( 6 \times 10^{23} \) molecules gm-mole\(^{-1}\) (Avogadro's number), and takes the value \( 6 \times 10^{26} \) molecules kg-mole\(^{-1}\) in m.k.s. units. If \( R \) stands for the gas constant and \( k \) for the Boltzmann constant, then

\[
\frac{R}{k} = \frac{M}{m} = \mathcal{N}
\]

By combining the above equations, we find that

\[
- \frac{1}{p} \frac{dp}{dh} = \frac{Mg}{RT} = \frac{mg}{kT} = \frac{1}{H}
\]
Equation (103) provides two alternative definitions of the scale height $H$ (more precisely, the pressure scale height). For each individual gas, these definitions are consistent at any level where diffusive separation exists and where the distribution is determined solely by the balance of the partial pressure gradient against gravity. They are also valid for a gas in a completely mixed atmosphere, in which processes such as turbulence are strong enough to cause each gas to conform to the scale height of the mixture as a whole. The masses $M$ and $m$ then refer not to the individual gas, but to the "composite" gas, air.

It is often convenient to express altitude in terms of a dimensionless parameter $z$, known as reduced height. The unit of $z$ is the scale height $H$. Since $H$ may vary with height, we may define an increment of $z$ by the equation

$$dz = dh/H$$

Taking $z = 0$ at some convenient reference height $h_0$, we then have

$$z = \int_{h_0}^{h} {dh \over H} = {h - h_0 \over H}$$

in which the second equality is only valid if $H$ is independent of height. We shall use the suffix 0 to indicate any quantity evaluated at the height $h_0$. On integrating Eq. (103) with respect to height, we find that

$$\ln(p/p_0) = \int_{h_0}^{h} {dh \over H} \equiv z$$

From this equation and the gas law (100) we may obtain a general expression for the variation of pressure and concentration with $z$:

$$p \over p_0 = nT \over n_0 T_0 = e^{-z}$$

Equation (107) applies not only to the pressure and gas concentration of the atmosphere as a whole but also to the partial pressure and concentration of any constituent that conforms to the hydrostatic equation, provided the proper scale height is used in calculating $z$.

So far our equations have been quite general. We now consider the case of a single gas in hydrostatic equilibrium, and neglect the variation of gravity with height. Then $T$ is the only variable in the formula for the scale height, $H = kT/mg$, so that $dH/H = dT/T$. From the gas law (100), by taking logarithms and differentiating with respect to height, we obtain the useful
1.2 THE HYDROSTATIC EQUATIONS

relations

\[ - \frac{dz}{z} = - \frac{dh}{H} = \frac{dp}{p} = \frac{dn}{n} + \frac{dH}{H} \]  

(108)

We can also find the total number of particles in a column of unit cross section above the height \( h_0 \), by using the gas law \( p = nkT \) and our definitions of \( H \) and \( z \). This number is simply

\[
\int_{h_0}^{\infty} n \, dh = \int_{0}^{\infty} nH \, dz = \int_{0}^{\infty} (nkT/mg) \, dz \\
= (p_0/mg) \int_{0}^{\infty} e^{-z} \, dz = n_0H_0
\]  

(109)

Hence, if the whole atmosphere above the height \( h_0 \) were compressed to a uniform pressure \( p_0 \), its vertical extent would be the scale height \( H_0 \), wherever \( h_0 \) may be chosen. This result does not require \( T \) to be independent of height.

1.22 GEOPOTENTIAL HEIGHT

For accurate calculations of atmospheric parameters, the variation of \( g \) with height must be taken into account. Neglecting a small contribution due to centrifugal force (which is inappreciable at the heights with which we are concerned), the variation of \( g \) is given by the inverse square law

\[
g(h) = \frac{GM_E}{r^2} = g(0) \frac{R_E^2}{(R_E + h)^2}
\]  

(110)

where \( G \) is the gravitational constant, \( M_E \) and \( R_E \) are the earth's mass and radius, \( r = R_E + h \) is radial geocentric distance, and \( g(0) \) is the sea-level value of \( g \) (which varies somewhat with latitude).

Sometimes a coordinate known as "geopotential height" \( (h*) \) is defined to take account of the altitude variation of \( g \). The simplest way of defining this quantity is to equate the work done in raising unit mass to height \( h \) to the work done in moving it through height \( h* \) in a constant gravity field \( g(0) \). Then

\[
g(0)h* = \int_{0}^{h} g(h) \, dh
\]  

(111)

so that

\[
h* = \int_{0}^{h} \frac{R_E^2 \, dh}{(R_E + h)^2} = \frac{R_Eh}{R_E + h}
\]  

(112)

Thus, \( h = h* \) near the earth's surface, where \( h \ll R_E \). The quantity \( h* \) can be used to replace \( h \) in the equations of Sec. 1.21.
We can include the variation of $g$ in the equations for hydrostatic equilibrium. Here, we shall only consider equations for a single gas in an isothermal atmosphere. From Eq. (110) we can set $g = GMe/r^2$ in (103); then, since $d/dh = d/dr$, we can integrate with respect to $r$. Taking $r = r_0$ as any convenient reference level, we have

$$
\frac{p}{p_0} = \exp \left[ - \int_{r_0}^{r} \frac{GM_em}{kTr^2} \right] dr = \exp \left[ \frac{GM_em}{kT} \left( \frac{1}{r} - \frac{1}{r_0} \right) \right] \quad (113)
$$

### 1.23 EXPANSION AND CONTRACTION

Let us now examine the changes in the vertical pressure distribution which occur when a column of air is heated. We assume that horizontal motions of air, if they occur, cause no net gain or loss of air in the column; and that any vertical acceleration is very much smaller than $g$, so that hydrostatic equilibrium can be assumed to hold.

For the moment, we assume the base of the column to be at ground level, $h = 0$. We now consider a small “cell” of air, of vertical thickness $\delta h$. During the heating, the weight of air above any cell does not change; therefore, the pressure within the cell remains constant and so does its reduced height $\bar{z}$. If we assume that the column of air is initially at a uniform temperature $T$, and is then heated to a uniform temperature $\theta T$, the gas concentration within each cell is reduced from $n$ to $n/\theta$ and the cell thickness expands to $\theta \delta h$. Since the scale height has increased from $H$ to $\theta H$, the real altitude $h$, measured from the base of the column ($z = 0$), is increased in the same ratio, that is, from $zH$ to $z\theta H$.

During the course of the heating, any cell $z$ moves vertically with speed $w = dh/dt$. Since $h = zH$, this speed is $z dH/dt$, and as $H \propto T$ we find that

$$
w = \frac{h}{T} \frac{dT}{dt} = h \frac{d\theta}{dt} \quad (114)
$$

provided that $T$—and hence, $dT/dt$—remains independent of height at all times (Ratchiffe and Weekes, 1960, p. 390). The speed $w$, therefore, increases with height throughout the heated column.

We now compare the distributions before and after the heating. If $n_0$ is the initial concentration at $z=0$, then the concentration at a height $h$ is $n_0 e^{-h/H}$, or $n_0 e^{-z}$. After the heating, the vertical distribution is given by

$$
n(h) = (n_0/\theta) \exp (-h/\theta H) \quad (115)
$$

$H$ being the original scale height. For small values of $h$, this equation shows
that the concentration \( n \) at a fixed height is decreased by the heating. But at greater heights \( n \) is increased, because the reduction of the exponent in the equation more than compensates for the factor \( 1/\theta \). At some intermediate level, therefore, \( n \) is unchanged, and for small changes of temperature (\( \theta \approx 1 \)), this may be shown to be just one scale height above \( z = 0 \). This is known in meteorology as the isopycnic level. Obviously, the total integrated concentration throughout the atmosphere, \( \int n \, dh \), is unaffected by the heating in the absence of horizontal flow.

![Graph showing hydrostatic distributions](image)

**Fig. 3.** Hydrostatic distributions of two gases (identified by subscripts 1, 2) in an isothermal atmosphere, for two different temperatures \( T', T'' \); see Eq. (115). The total quantities of the gases are the same at both temperatures. The concentrations are plotted on a natural logarithmic scale, with an arbitrary zero. The height scale is the same for all graphs in terms of real height \( h \), but is labeled in terms of reduced height \( z_1 \) for the lighter gas 1; the left-hand scale (\( z_1' \)) for temperature \( T' \), and the right-hand scale (\( z_1'' \)) for temperature \( T'' \). Full lines refer to gas 1, broken lines to gas 2, and points \( I_1 \) and \( I_2 \) are the respective isopycnic levels. The points at which \( n_1 = n_2 \) are \( E' \) for temperature \( T' \), and \( E'' \) for temperature \( T'' \). Numerical values used in the graph: mass ratio of gases, \( M_2/M_1 = 2 \); temperature ratio \( T''/T' = \theta = 1.5 \); concentration ratio at base level, \( (n_2/m_1)_0 = e^2 = 7.4 \).

These changes are illustrated by the full lines in Fig. 3, which represent the hydrostatic distributions of a fixed quantity of gas of species 1 at two different temperatures, \( n_1' (h) \) at temperature \( T' \) and \( n_1'' (h) \) at temperature \( T'' \). The scale of \( h \) is the same for both graphs, but is labeled in terms of reduced height \( z_1 \), which takes different values \( z_1' \) and \( z_1'' \) at the two temperatures. \( I_1 \) is the isopycnic level, which would be found at \( z_1 = 1 \) for an infinitesimal
change of temperature; but for the finite temperature change assumed for drawing the diagram \( T'' = 1.5T' \) this level is not exactly isopycnic.

The broken lines in Fig. 3 refer to another gas, species 2, which is distributed according to its own scale height. It is assumed that the ratio of molar masses \( M_2/M_1 = 2 \). Hence, the reduced heights for gas 2 are related to those for gas 1 by \( z_2' = 2z_1' \) for temperature \( T' \), and \( z_2'' = 2z_1'' \) for \( T'' \). \( I_2 \) is the isopycnic level for gas 2, and \( E' \) and \( E'' \) are the levels at which the concentrations of gases 1 and 2 are equal at temperatures \( T' \) and \( T'' \), respectively. We note that \( E' \) lies at the same level in terms of \( z_1' \) (actually, \( z_1' = 2 \) for the numbers used in the diagram), as does \( E'' \) in terms of \( z_1'' \). More generally, for any chosen value of \( z_1' \) the ratio \( n_1'/n_2' \) (represented in Fig. 3 by the horizontal separation of points on the full line \( n_1' \) and the broken line \( n_2' \)) equals the ratio \( n_1''/n_2'' \) at the level where \( z''_1 \) takes the same value. This follows from the geometry of the figure. Since reduced height is uniquely related to pressure, this result implies that at any fixed pressure level the relative concentrations of the gases are unaffected by the temperature change; no forces exist to cause one gas to diffuse relative to the others (Garriott and Rishbeth, 1963).

Although the above discussion relates to an isothermal air column in contact with the ground, it can be generalized to apply to the diurnal temperature variations in the thermosphere. There are three important differences that complicate the analysis but do not affect the principle under discussion.

First, the part of the atmosphere subject to the heating does not extend to the ground, so we have to shift the zero of \( z \) in our analysis to an appropriate level at the base of the thermosphere. Second, the temperature in the thermosphere is not uniform, but varies with height. The analysis is then carried through with \( z \) defined by an integral, as in Eq. (105); the speed of expansion is given by a rather complex integral in place of Eq. (114) (Harris and Priester, 1962a). Third, the atmosphere at this height contains several gases; provided that these do not interact chemically, the equations can be solved for each constituent separately, and the considerations deduced from Fig. 3 will apply. Existing calculations generally do not take account of possible composition changes at the lower boundary, nor of horizontal transport of air, which might have appreciable effects on thermospheric structure.

1.24 Model Atmospheres

A “model atmosphere” consists of tables or graphs showing distributions of pressure, temperature, and gas concentrations, based on experimental
data and computed according to the principles we have described. It may present data for a variety of conditions; e.g., different times of day, different levels of solar activity, different latitudes, and so on. At altitudes where the mean molecular weight is independent of height, knowledge of the vertical distribution of one of these basic parameters—plus boundary assumptions—enables the others to be computed from Eqs. (100) and (101). This is not true in the thermosphere above 100 km, with which we are primarily concerned here, because the molecular weight varies with height. Between about 120 and 200 km lies a region in which the neutral atmosphere is very difficult to observe; above 200 km the best determined parameter is density, found from satellite drag data. But even if the complete density distribution \( \rho(h) \) is available, further information is needed in order to compute the other quantities. Most existing calculations start from an assumed chemical composition at a lower boundary height, variously taken as 100 to 150 km. If diffusive separation is assumed to exist above this height, the variation of mean molecular weight with reduced height can be found at once by applying an equation of the type \( p/p_0 = e^{-z} \) to the partial pressure of each gas. However, conversion of reduced height \( z \) to real height \( h \) involves knowledge of the temperature profile \( T(h) \). This may be found by trial-and-error fitting of the model to the observed density profile (e.g., Yonezawa, 1960) or by solution of the heat-conduction equations of Sec. 1.32 (Nicolet, 1961). In the detailed models of Harris and Priester (1962a,b), the time-dependent distribution \( T(h, t) \) is computed for different heat inputs corresponding to different levels of solar activity. The same approach is used in the standard models (CIRA, 1965), from which some representative data have been plotted in Fig. 4.

We see from the above discussion that both the temperature distribution and the chemical composition must be known before accurate model atmospheres can be computed. These topics are discussed in the next two sections.

### 1.3 The Heat Balance in the Thermosphere

In the previous section, it was shown that the temperature distribution plays a large part in determining the structure of the atmosphere. Accordingly, we now discuss the factors which make up the heat balance in the thermosphere, the part of the atmosphere which concerns us most.

We start by listing the processes contributing to the heat balance:

*Production* (rate \( Q \) per unit volume):

(a) Absorption of solar ultraviolet and X-radiation, leading to photo-
I. THE NEUTRAL ATMOSPHERE

Fig. 4. Cospar International Reference Atmosphere (1965). Vertical distribution of density and temperature for high solar activity (10 cm solar flux $S = 250$) at noon (1) and midnight (2); and for low solar activity ($S = 75$) at noon (3) and midnight (4). These models assume the pressure, composition, and temperature at 120 km to be invariant.

ionization, photodissociation, and consequent chemical reactions which liberate heat.

(b) Absorption of energetic charged particles entering the atmosphere.

(c) Dissipation of tidal motions, gravity waves and hydromagnetic waves, by molecular viscosity and turbulence.

(d) Joule heating by ionospheric electric currents.

Loss (rate $L$ per unit volume): Radiation at wavelengths for which the thermosphere is transparent. This includes the visible airglow emissions of oxygen and nitrogen (Sec. 3.7); the numerous spectral bands of the hydroxyl radical; and the infrared line of atomic oxygen at 63$\mu$. The form of $L$ has been given by Bates (1951).

Transport (rate $\nabla \cdot \Phi$ per unit volume, $\Phi$ being the heat flux):

(a) Molecular conduction, giving rise to a heat flux which depends on the temperature gradient, $\Phi \propto \nabla T$.

(b) Eddy transport, important below 100 km. This process is capable of
transferring heat from the thermosphere to the mesosphere, and derives its energy from wind motions.

(c) Transport by large-scale winds. In the high thermosphere, the diurnal heating and cooling is believed to drive large-scale winds that carry heat horizontally.

(d) Chemical transport, due to motion of ionized or dissociated gas, which may eventually release its energy at a distance from where it was originally absorbed.

The thermal balance is very complicated. We must here note that the present discussion concerns only the temperature of the neutral gas which, of course, contains most of the thermal energy in the thermosphere. In recent years, it has become established that, at least above 150 km, the electron temperature may considerably exceed the temperature of the ions and of the neutral gas, especially by day (Hanson, 1963). The subject of electron and ion temperatures, which properly belongs to the ionosphere, is dealt with in Sec. 6.5. Great complexity arises if it is included in calculations of the thermospheric heat balance, therefore it is omitted from our present first-order discussion.

1.31 Heating Processes in the Thermosphere

In the lower thermosphere—say between the mesopause and turbopause, or roughly 80–105 km—many processes must be considered. The region acts as a heat sink, not only for heat produced locally but also for energy absorbed higher in the thermosphere. Although some heat is lost by radiation in the infrared and visible ranges, eddy transport is probably the principal means of removing heat from the thermosphere, being more rapid than molecular conduction at levels below the turbopause (Johnson and Wilkins, 1965; F. S. Johnson, 1967). We note that, in any case, molecular conduction could not carry heat through the temperature minimum at the mesopause; also that the temperature gradient is too small in the upper mesosphere, and of the wrong sign in the thermosphere, to support natural convection.

According to the eddy transport theory, the bulk of the heat absorbed in the thermosphere is transported down to the mesosphere, and deposited at heights around 50 km. This input plays only a trivial part in the heat balance of the mesosphere, which largely depends on absorption by ozone of solar ultraviolet radiations not significantly absorbed in the thermosphere. Calculations by Murgatroyd and Goody (1958) show that radiative equilibrium exists in most of the atmosphere between 30 and 90 km; there is, however,
a discrepancy in the polar mesopause region which is observed to be warmer in winter than in summer. Kellogg (1961) and Young and Epstein (1962) have suggested that this region is heated in winter by the recombination of atomic oxygen transported from greater heights by a slow downward motion of the air. It is envisaged that atomic oxygen is produced primarily in low latitudes where solar irradiation is strongest, and carried polewards by a general atmospheric circulation (G. A. M. King, 1964; F. S. Johnson, 1964). This provides an example of chemical transport.

The heat sources in the lower thermosphere include solar ultraviolet radiation in the range absorbed by molecular oxygen, between 1026 and about 1750 Å, which totals about 30 mW m⁻² (30 erg cm⁻² s⁻¹) (Detwiler et al., 1961). This radiation is absorbed mainly below 95 km, and much of its energy goes into photodissociation. Apart from losses by radiation, most of this energy eventually appears as heat; it may be redistributed by chemical transport due to the advection of atomic oxygen, mentioned above. Solar ultraviolet radiation shorter than 1026 Å, the ionization limit of molecular oxygen, can produce ionization; it amounts to about 3 mW m⁻² at sunspot minimum (Allen, 1965; Hinteregger, 1965), and is mostly absorbed between 95 and 250 km.

Another source of heat is provided by dynamic processes, such as the dissipation of gravity waves. Hines (1965a) has assessed the heat input as about 0.1 W kg⁻¹ at 95 km. To compare this estimate with the input of solar energy, quoted above as 30 mW m⁻², we note that the total columnar mass of air above 95 km is 7 g m⁻² (U.S. Standard Atmosphere). Most of this mass is contained within one scale height of 95 km, so we can make a rough estimate of 0.7 mW m⁻² for the gravity wave input. Hines considers that tidal energy may provide a comparable input. Fortunately, these estimates do not depend on the mechanism by which the energy is dissipated, which might be turbulence or viscosity, or joule heating by the electric currents produced by the tides (Sec. 1.7).

In addition to the electric currents flowing in the undisturbed ionosphere, large currents flow in the auroral zones at the time of magnetic disturbance. Their joule dissipation may be an important source of heat, related to observed temperature increases at high latitudes (Cole, 1962). Particle precipitation may also cause heating, but accurate estimates of it do not seem to be available.

At greater heights, well above the turbopause, the situation is more simple. Radiative loss is negligible and, apart from the still uncertain part played by large-scale horizontal winds, the heat balance of the neutral gas may be
discussed in terms of a simple model. This assumes heat to be produced by the absorption of solar ionizing radiation and lost by downward thermal conduction. The lower thermosphere is then regarded as an unvarying sink of heat, and the details of the thermal balance there are ignored. From now on, we adopt this model as a first-order description of the upper thermosphere.

1.32 Equations of Production and Conduction of Heat

We now consider the equations applicable to the production-conduction model of the upper thermosphere. The equations for $Q$, the rate of heating by solar ionizing radiation, are very similar to the equations for the rate of production of ionization. Since we will discuss these equations in detail in Sec. 3.2, we give only a simplified account here. Consider monochromatic radiation incident at a zenith angle $\chi$ on a plane atmosphere, containing a single gas whose concentration is $n(h)$ and scale height $H(h)$. Let $\sigma$ be the absorption cross section and $\epsilon$ the amount of heat liberated per unit of radiation absorbed. Then, the intensity of the radiation $I(h)$ diminishes downwards from its value $I_{\infty}$ outside the atmosphere. The attenuation depends on the gas concentration along the path of the radiation; in an element $ds$ of the path, it is given by

$$- \frac{dI}{I} = \sigma n ds$$  \hspace{1cm} (116)

This equation defines an increment $d\tau$ of optical depth $\tau$. Since for a plane atmosphere $ds = -dh \sec \chi$, Eq. (116) can be expressed in terms of height. It can then be integrated by using the result of Eq. (109) for the total content of a unit column of gas above any height $h$, giving

$$\tau(h, \chi) = \ln \frac{I_{\infty}}{I(h)} = \int_{h}^{\infty} \sigma n dh \sec \chi = \sigma n(h) H(h) \sec \chi$$  \hspace{1cm} (117)

The rate of production of heat is

$$Q(h, \chi) = I(h) \epsilon \sigma n(h) = I_{\infty} e^{-\tau} \epsilon \sigma n(h)$$  \hspace{1cm} (118)

It is easily shown that $Q(h, \chi)$ peaks near the level where $\tau = 1$, which occurs at heights in the range 100–180 km for the radiations which heat the thermosphere. To make the calculations more realistic, as is done in more detail in Sec. 3.3, we can:

(i) Take account of the earth's curvature; for constant $H$ this is accomplished by replacing $\sec \chi$ by the grazing-incidence integral (Chapman, 1931b) when $\chi$ is near $90^\circ$. 
(ii) For an atmosphere containing several gases, with individual parameters \( n_i, H_i, \sigma_i, \epsilon_i \), replace \( \sigma n H \) by \( \sum \sigma_i \rho_i H_i \) in (117) and \( \epsilon \sigma n \) by \( \sum \epsilon_i \sigma_i \rho_i \) in (118).

(iii) Include the dependence of \( \sigma_i \) and \( \epsilon_i \) on ultraviolet wavelength \( \lambda \); then \( \tau \) becomes a function of \( \lambda \), and the calculation of \( Q \) requires an integration in which \( dQ/d\lambda \) replaces \( Q \) and \( dI_\infty/d\lambda \) replaces \( I_\infty \) in (118). In practice, the integration is approximated by a summation over a number of finite bands of wavelength.

Even when these points are included, the resulting equation neglects two further complications connected with photoionization heating. One is that ejected photoelectrons may travel some distance before giving up their kinetic energy, so that some of the energy may appear in remote locations. Below 250 km, however, where most of the ionizing radiation is absorbed, this effect is slight. The other is that part of the energy is stored in the ionization, and does not appear as heat until the ionization recombines. This implies a time lag which is, however, only of the order of minutes between 100 and 200 km.

The other process to be considered in the thermosphere is molecular conduction. If only vertical variations are considered, the heat flux \( \Phi \) due to conduction may be written

\[
\Phi = -AT^{1/2} \frac{\partial T}{\partial h}
\]

The product \( AT^{1/2} \) represents the thermal conductivity, \( A \) being a constant given by kinetic theory. According to Nicolet (1961), the value of \( A \) (\( \text{J m}^{-1} \text{s}^{-1} \text{C}^{-3/2} \)) is 0.021 for atomic hydrogen, 0.0036 for atomic oxygen, and 0.0018 for \( N_2 \) or \( O_2 \).

1.33 Solving the Heat Balance Equation

To compute the temperature distribution in the thermosphere, a continuity equation for the heat balance is constructed which expresses conservation of energy. The derivation must take into account changes of gravitational and internal energy arising from vertical motion and thermal expansion of the gas, as well as heat sources and heat flow. One convenient approach (Nicolet, 1961) is to consider the rate of gain of heat \( G_v \) within a unit volume fixed in space, namely

\[
G_v = Q - L - \nabla \cdot \Phi = \rho c_v \frac{\partial T}{\partial t}
\]

where \( c_v \) is the specific heat per unit mass at constant volume. Unless a steady
state exists, the density within the volume may change, and account must be taken of this in computing the production function $Q$ from Eq. (118) and other functions. Moreover, the equation is inexact in that it does not fully include the effects of gravity and thermal expansion on the temperature of the gas.

Another approach, used by Harris and Priester (1962a) and Cummack (1962), is to follow a “cell” of gas, as it moves with velocity $v$, and compute the rate of change of temperature within the cell, $dT/dt$. This makes use of the principles described in our previous discussion of thermal expansion, in Sec. 1.2. If hydrostatic equilibrium exists, as was assumed, then the cell remains at constant pressure, and because of the work done in expansion, the appropriate specific heat is now $c_p$. The rate of gain of heat per unit volume is

$$G_p = \rho c_p \frac{dT}{dt} = \rho c_p \left( \frac{\partial T}{\partial t} + v \cdot \nabla T \right)$$  \hspace{1cm} (121)$$

In evaluating the contributions to $G_p$, such as the production rate, we must take account of the changes of density within the cell, according to the perfect gas law $p = nkT = \text{const}$. This is simpler, however, than the corresponding calculation for the terms in $G_v$. In particular, the optical depth $\tau$ depends on pressure only, and so remains constant within a cell as long as the solar zenith angle does not change.

In many calculations, only vertical motion is considered, and the term $v \cdot \nabla T$ in Eq. (121) reduces to $w(\partial T/\partial h)$. In practice, the simple formula given in Sec. 1.23, Eq. (114), is inadequate because the heating rate $\partial T/\partial t$ varies with height throughout the thermosphere, so $w$ must be expressed as an integral (Harris and Priester, 1962a, 1965).

The conditions at great heights in the thermosphere are largely governed by conduction, since in Eq. (120) the terms $Q$ and $L$ decrease upward as the atmospheric density decreases. If, as is generally supposed, the flux $\Phi$ tends to zero at great heights, then $\partial T/\partial h \to 0$ by Eq. (119) so that the upper thermosphere is isothermal. To justify the assumption that $\Phi \to 0$ at great heights, we note that the outermost atmosphere is fully ionized, and its thermal conductivity is large, parallel to the magnetic field but extremely small across it. Hence, the flow of heat across the magnetopause, which effectively marks the termination of the atmosphere, is usually negligible (but it could be appreciable when hot plasma from the sun actually penetrates into the geomagnetic field, as may happen during magnetic storms). We shall use the term “limiting (thermospheric) temperature” for the tem-
I. THE NEUTRAL ATMOSPHERE

perature in the isothermal region, although other terms such as “thermo-pause temperature” and “exospheric temperature” have been used.

Because of the importance of thermal conductivity in the upper thermosphere, the limiting temperature is closely controlled by the variations of temperature at lower heights. As illustrated by the graphs of Fig. 4, the daily temperature change causes a large relative change of density above 300 km—described as the “diurnal bulge”—which is detected by analyzing the atmospheric drag on artificial satellites. From these observations, good estimates of the limiting thermospheric temperature can be made (Jacchia, 1963). The data indicate that the limiting temperature is greatest at about 1400 hours local time, whereas the computed models give a later maximum, at about 1700 hours. This discrepancy led Harris and Priester to postulate that a substantial fraction of the heating is due to a “corpuscular” source, which peaks at about 0900 hours local time. Although the existence of this source has not been established, the discrepancy has yet to be resolved, so the question is still regarded as open. Though the difficulty might partly be due to imperfect knowledge of the many parameters involved, it is quite likely that the diurnal variation is influenced by further processes, such as gravity-wave heating (Hines, 1963) or horizontal heat transport by winds or molecular conduction (Volland, 1966, 1967). Actually, Harris and Priester (1965) find that no great difference is made to their results if the height of the lower boundary, or the temperature or composition there, is made to undergo a diurnal variation.

To delineate the “diurnal bulge,” empirical functions of latitude and local time are used, in which the numerical constants are chosen to give the best agreement with satellite drag data (Jacchia, 1965; Jacchia and Slowey, 1967). The “bulge” is centered in tropical regions. For any given conditions the limiting temperature may be obtained from the formulas given by Jacchia (1965). Apart from the diurnal and solar cycle variations, illustrated in our Fig. 4, there are annual and semiannual variations of order 100°K. In addition, the limiting temperature increases with magnetic disturbance, by about 1°K per unit of the parameter $Ap$. We must note that all these formulas are purely empirical; they are constructed to describe the observations, not to explain them. For the most part, quantitative physical explanations are still lacking.

The heat input to the thermosphere above 120 km may be estimated by assuming equilibrium between production and conduction. For this case, Eq. (120) reduces to $Q = d\Phi/dh$. Then Eq. (119) may be used to substitute for $\Phi$, and the equation integrated to yield a relation between the tempera-
ture gradient at any height $h$ and the columnar heat input above that height. Using our assumption that $\Phi \to 0$ at great heights, as discussed previously, we have

$$\int_{h}^{\infty} Q \, dh = -\Phi(h) = AT^{1/2} \frac{dT}{dh}$$  \tag{122}$$

This equation shows that $dT/dh > 0$ at all heights, so there can be no temperature inversion ($dT/dh < 0$) under equilibrium conditions. Also, $dT/dh$ increases with decreasing height, until a "heat sink" is encountered where the loss $L > Q$.

On solving Eq. (122) for the temperature profile $T(h)$, and comparing the results with the data, the input $\int Q \, dh$ above 120 km is found to be of order 1 mW m$^{-2}$ (erg cm$^{-2}$ s$^{-1}$), qualitatively consistent with the measured flux of solar radiation (Hunt and Van Zandt, 1961; Allen, 1965). The solar-cycle variation may amount to a factor of about five.

1.4 Dissociation and Diffusive Separation

Above the mesopause the composition of the neutral atmosphere varies with height, because of the dissociation of molecular oxygen and the occurrence of diffusive separation above the turbopause.

The dissociation of oxygen has been studied in detail by Nicolet and Mange (1954). They point out that even though $O_2$ molecules are dissociated by solar radiation, notably Lyman $\alpha$ (1216 Å) and the Schumann-Runge continuum (below about 1750 Å), the lifetime of an $O_2$ molecule above 100 km is some days. However, the $O_2$ concentration in the thermosphere is replenished by diffusion in a time shorter than this, so the shape of the vertical distribution of $O_2$ should more closely correspond to the hydrostatic equation (101) than to a condition of photochemical equilibrium (see also Colegrove et al., 1965).

The $N_2$ molecule, unlike $O_2$, has a very small cross section for photodissociation, and most of the nitrogen in the upper atmosphere is thought to be in molecular form. Such atomic nitrogen as does exist is produced by other photochemical processes, such as those involved in the production of ionization.

The question of diffusive separation of atmospheric gases has been of interest ever since the composite nature of air was recognized in the eighteenth century. It was once thought that diffusive separation would take effect just above the ground, but it is now clear that major constituents are well mixed.
up to at least 100 km, and that diffusive separation begins near this altitude.

From observations with rocket-borne mass spectrometers, Meadows and Townsend (1958) found molecular nitrogen and argon to be diffusively separated at heights above 115 km. The situation concerning the gases $N_2$, $O_2$, and $O$ remained obscure, partly because the chemical activity of atomic oxygen creates difficulties in mass-spectrometer sampling. The loss of atomic oxygen by chemical reactions within the apparatus and contamination by gases from the vehicle have been major problems, and there is considerable divergence between the results obtained in the several rocket experiments made in the height range 100–200 km.

![Fig. 5. Ratios of the concentrations of the major neutral constituents of the upper atmosphere as functions of altitude [after Nier et al. (1964)].](image-url)

The results obtained by Nier et al. (1964)—illustrated in Fig. 5—and by Schaefer and Brown (1964) show that the $O/O_2$ concentration ratio increases rapidly upwards from about 100 km, indicating the existence of diffusive separation at this altitude. The available results have been compared by Pokhunkov (1966). Explorer XVII satellite data for the range 250–800 km (Reber and Nicolet, 1965) show that atomic oxygen is the dominant neutral constituent from roughly 300 to 600 km, with helium dominating at greater heights.

The neutral composition can also be determined from observations of the solar spectrum at a series of heights, such as those of Hinteregger and
Watanabe (1962). It is necessary to know the ultraviolet absorption cross sections, as a function of wavelength, for each gas. This method, essentially a trial-and-error approach, has been applied by Norton et al. (1963).

At present, the concentration ratio of the major thermospheric constituents, O and N₂, seems to be poorly known. According to Fig. 5, the level at which \( n[O] = n[N₂] \) is near 200 km, but both the experimental data and the computed models differ widely among themselves on this point. For instance, the CIRA (1965) models place this level at 190-300 km, depending on temperature, although Norton et al. (1963) place it at 130 km. On the basis of the discussion of thermal expansion in Sec. 1.23, one would expect the level to depend on temperature. Unfortunately, the accuracy of the present data is hardly sufficient for any detailed discussion of diurnal, seasonal, or geographical variations of composition.

![Fig. 6. IQSY daytime ionospheric and atmospheric composition, based on mass spectrometer measurements. Ion and neutral distributions below 250 km are from two daytime rocket measurements above White Sands, New Mexico (32°N, 106°W). The helium distribution is from a nighttime measurement. Distributions above 250 km are from the Elektron II satellite results of Istomin (1966) and Explorer XVII results of Reber and Nicolet (1965) [C. Y. Johnson, U.S. Naval Research Laboratory, Washington; reprinted from "Ion and neutron composition of the ionosphere," by C. Y. Johnson, in "Annals of the IQSY," Volume 5, 1969, by permission of the M.I.T. Press, Cambridge, Massachusetts. Copyright 1969 by M.I.T.]

Figure 6 may be taken to illustrate average daytime conditions for sunspot minimum. It includes data on ionic constituents (to be discussed in Sec. 3.6), and incidentally demonstrates the great preponderance of neutral gases over ionic constituents in the upper atmosphere.

The question of diffusive separation is linked to that of turbulence, which is believed to be absent at heights above about 100 km. Blamont and de Jager
(1961) have observed the rate of expansion of sodium vapor trails released from rockets. Below 102 km the trails are rapidly distorted by turbulence, but above this level they expand more uniformly at a rate consistent with molecular diffusion. From the results of similar experiments, Rosenberg (1963) places the turbopause at 110 to 115 km.

The theory of mixing and diffusive separation is incomplete. A number of different approaches have been compared by Mange (1957). Most of these define a criterion for estimating the time taken for a minor constituent to attain a diffusive equilibrium, starting from an initial nonequilibrium (or mixed) condition. This time is, in general, of order \((\text{scale height})^2/(\text{diffusion coefficient})\). If it is less than some time-constant characteristic of the process responsible for mixing, then mixing will not be maintained. Naturally, the problem is more complicated if photochemical processes are operative as well as diffusion.

Hines (1963) has discussed the problem of mixing of major constituents, with special reference to turbulence. If mixing were suddenly to cease, the gases would separate with a relative velocity which is found by balancing the resistance to motion, due to collisions, against the "buoyancy" forces which tend to separate the gases. The power dissipation of the collisional forces can be estimated; up to 100 km, it is found to be less than the power available from turbulence, so that it appears that turbulence can just maintain mixing of the major constituents up to 100 km. The precise level of the turbopause presumably also depends on the strength of the mechanism causing turbulence, which may be gravity waves. Further discussion of turbulence, and of the pertinence of parameters such as Reynolds and Richardson numbers, is given by Hines in the review cited above.

Unless some other mechanism for opposing separation is effective at greater heights, diffusive separation of atmospheric gases must be established not far above the turbopause. If the level of the turbopause should change, the atmospheric composition at higher levels would be affected. As mentioned before, the experimental data do, indeed, suggest that diffusive equilibrium exists for the neutral gases above about 110 km, with the exception of hydrogen which we discuss in Sec. 1.5.

1.5 The Exosphere

The exosphere is generally defined as the region in which the mean free path exceeds the scale height. For the U.S. Standard Atmosphere, this definition places the base of the exosphere at about 600 km altitude. A molecule
moving upwards at the base of the exosphere is unlikely to make any collisions above this level, and thus moves in a ballistic orbit under the influence of gravity. This orbit is elliptical if the upward velocity of the molecule is less than the velocity of escape $v_E$ from the earth's gravitational field, which is about 11.4 km s$^{-1}$. If a molecule acquires a velocity exceeding 11.4 km s$^{-1}$ as the result of collisions, it enters a hyperbolic orbit and escapes from the atmosphere unless it suffers further collisions. Some molecules possess elliptical "satellite" orbits which do not intercept the base of the exosphere, so that their chances of colliding with other molecules are small.

In spite of the rarity of collisions, the barometric formula is applicable to a gas possessing a Maxwellian velocity distribution (F. S. Johnson, 1966). Assuming the inverse square law variation of gravity, Eq. (110), the barometric distribution is given by Eq. (113) of Sec. 1.22. For isothermal conditions, the gas pressure varies radially as $p \propto \exp(e/c/r)$, where $c$ is a constant, implying that $p \to$ constant as $r \to \infty$. This difficulty can be avoided by an approach based on Liouville's equation, which retains the concept of a Maxwellian velocity distribution (Chamberlain, 1963).

The barometric formula is applicable to the heavy gases in the earth's exosphere, but not to light gases which are able to escape. If we define an "escape temperature" $T_E$ by the simple equation $\frac{1}{2}mv_E^2 = \frac{3}{2}kT_E$, we find $T_E$ to be 5200 K for hydrogen atoms, 21,000 K for helium atoms, and 84,000 K for oxygen atoms. At sunspot maximum, the temperature of the exosphere exceeds 2000 K and there is a steady loss of the fastest hydrogen atoms. Consequently, the velocity distribution for hydrogen atoms is not Maxwellian, and the barometric formula becomes inaccurate at altitudes comparable to one earth radius (Chamberlain, 1963). Helium is also lost, but much more slowly. This gas is liberated at a steady rate by radioactive decay processes in the earth's crust, at a rate which can be estimated from the geological data. Since the rate of escape depends on the temperature of the upper atmosphere, it was once thought that this temperature could be estimated from the observed helium abundance. However, it seems that this problem is too complicated to enable any straightforward conclusions to be drawn (Chamberlain, 1963). The heavier gases hardly escape at all.

Even in the thermosphere, the vertical distributions of light gases may depart from the barometric formula. The departure is partly due to the process of thermal diffusion, which takes place in the presence of a temperature gradient (Chapman and Cowling, 1952); but it is also connected with the flow of light gases up to the exosphere, from which they escape (Bates and Patterson, 1961; Kockarts and Nicolet, 1962, 1963). In the case of atomic
helium, the departure from hydrostatic equilibrium is slight; Kockarts and
Nicolet find that the lifetime of a He\textsuperscript{4} atom in the thermosphere is some
years. On the other hand, at some heights, the distribution of atomic hy-
drogen departs very considerably from hydrostatic equilibrium. Hydrogen
is generated by the dissociation of water vapor and methane at the base of
the thermosphere, and diffuses upwards to the exosphere where it is even-
tually lost, the average lifetime of a hydrogen atom in the atmosphere being
a few days. Near the level where it is generated, the atomic hydrogen con-
forms to a "mixed" distribution, with a scale height approximately equal to
that of the entire neutral atmosphere. At greater heights, however, molecular
diffusion is so rapid that only a slight departure from diffusive equilibrium
is sufficient to maintain the upward flux, and the hydrogen scale height
differs little from the value appropriate to hydrostatic equilibrium.

Because the rate of escape from the exosphere is very temperature-depen-
dent, the hydrogen content of the atmosphere should be greater by night
than by day, though the difference is reduced by the lateral flow of hydrogen
around the earth (Hanson and Patterson, 1963; Donahue and McAfee, 1964;
Patterson, 1966). The content must also vary with the solar cycle, as illus-
trated by the computations of Joseph (1967). The existence of these diurnal
and solar cycle variations is borne out by the observational data, that have
been summarized by Donahue (1966a).

Although there are comparatively few experimental observations which
are related to the hydrogen content of the exosphere, they have aroused
considerable interest. Purcell and Tousey (1960) first observed the solar
Lyman $\alpha$ line from a rocket above 130 km. They found it to be about 1 Å
wide, with a narrow absorption core attributable to a column of hydrogen
atoms, of order $3 \times 10^{16}$ m\textsuperscript{-2} at a temperature of order 1000°K. This tem-
perature is more likely to be associated with the terrestrial atmosphere than
the interplanetary gas, and moreover, the absence of Doppler shift suggests
that the hydrogen is stationary with respect to the earth (Johnson and Fish,
1960). Diffuse Lyman $\alpha$ radiation has also been detected in the night sky by
rocket experiments (Kupperian et al., 1959), and it is believed to be solar
radiation scattered by hydrogen in the outer atmosphere. The amount of
hydrogen required to account for the observed Lyman $\alpha$ intensity is several
times greater than the daytime content quoted above. Brandt (1961) and
Donahue and Thomas (1963) have suggested that the nighttime Lyman $\alpha$
radiation is scattered by a hydrogen cloud situated at a distance of many
earth radii, which might be supplied from the solar wind or by terrestrial
escape.
1.6 Atmospheric Dynamics

In this section we shall consider some aspects of upper atmosphere motion which are relevant to the ionosphere. For our purposes, we do not require a very complete discussion of atmospheric dynamics, nor do we develop the mathematical analysis in any detail. Fuller discussions exist elsewhere; see for example the textbook by Craig (1965), the monograph on atmospheric oscillations by Wilkes (1949), and the essay by Hines (1963).

In discussing winds, we make use of the terms "eastward," "northward," etc. to describe a direction toward which motion takes place, as is the convention in ionospheric physics (and oceanography), in contrast to the meteorological description of a wind by the direction from which it blows.

Various types of atmospheric motion are summarized as follows:

**Prevailing winds** of global scale in the mesosphere and lower thermosphere, driven by long-lived—though seasonally varying—pressure inequalities, as described in Sec. 1.62.

**Thermospheric winds** above about 120 km, driven by the pressure inequalities due to daily temperature variations. Because of their basic daily periodicity, these winds could be regarded as tidal in nature. We do not treat them as such, however, because we can discuss them (Sec. 1.63) without the complicated tidal theory used for atmospheric tides below 100 km.

**Planetary waves** in the lower atmosphere, with periods of days. Charney and Drazin (1961) find that these waves are almost entirely trapped in the lower atmosphere, as a result of mesospheric winds. However, at some seasons a small fraction of their energy may penetrate the mesosphere and contribute to observed motions near 100 km (Hines, 1963). These waves are not considered here in any further detail.

**Tidal oscillations**, global in scale, with periods related to the solar and lunar days. The theory and observation of atmospheric tides are discussed in Sec. 1.7, though we note that they can also be treated as a special case of atmospheric gravity waves.

**Internal gravity waves**, with periods of minutes or hours, vertical wavelengths of a few kilometers and horizontal wavelengths of up to thousands of kilometers. These are discussed in Sec. 1.64.

**Turbulence**, which serves as a "sink" of energy; larger-scale motions are degraded into small-scale turbulent motions and ultimately lost as heat. As
noted in Sec. 1.4, turbulence ceases abruptly near the 100 km level. Just below the turbopause, at meteor heights, Hines (1963) quotes typical parameters for the smaller-scale motions: size 20 m, velocity \( \frac{1}{2} \) m s\(^{-1}\), time scale 40 s; and a total turbulent energy density of order \( 10^{-2} \) W kg\(^{-1}\). We do not propose to discuss turbulence further in this book.

1.61 Equations of Motion of the Neutral Air

The neutral air velocity \( U \) is assumed to satisfy a fluid equation of motion. Making some approximations, we write

\[
dU/dt + 2\Omega \times U = g - (1/\rho) \nabla p - \nabla \Psi + (\mu/\rho) \nabla^2 U - v_{ni}(U - V_i) \quad (123)
\]

in which, as usual, \( p \) = pressure, \( \rho \) = density; the operator \( dU/dt \) represents "differentiation following the motion" of the air, so that

\[
dU/dt \equiv \partial U/\partial t + (U \cdot \nabla) U \quad (124)
\]

The final term in (124) represents transport of momentum, and renders the equation nonlinear, though it is often negligible in practice. The other terms in Eq. (123) are as follows: \( 2\Omega \times U \) is the Coriolis acceleration, \( \Omega \) being the earth's angular velocity. \( g \) is the gravitational acceleration, with which the small centripetal acceleration \( \Omega \times (\Omega \times r) \) may be combined (\( r \) is the geocentric distance). \( \Psi \) is a scalar potential due to tide-raising forces. \( \mu \) is the coefficient of molecular viscosity: the simple form of the viscous term in Eq. (123) assumes the air to be incompressible and \( \mu \) to be a constant (Sutton, 1953), though actually \( \mu \) depends somewhat on temperature and molecular mass. \( v_{ni}(U - V_i) \) represents what is termed "ion-drag" due to collisions between the neutral gas and the ionospheric plasma, \( V_i \) being the drift velocity of the ions. This term is important above 150 km, where the ionic motion is strongly controlled by the geomagnetic field, as described in Sec. 4.24. The parameter \( v_{ni} \) is related to the frequency of collisions with ions experienced by any neutral particle, as will be discussed in Sec. 4.12.

In a static atmosphere \( U = 0 \) and Eq. (123) reduces to the hydrostatic equation \( \rho g = \nabla p \) (see Eq. (101), Sec. 1.21). This equation holds extremely well, even when the atmosphere is in motion, because the vertical accelerations are generally very much smaller than \( g \). We may recall from Eq. (106) that, upon integrating the hydrostatic equation with respect to height, we obtain the equation (in which \( p_0 \) is the pressure at ground level, \( h = 0 \))

\[
p/p_0 = \exp \left[ - \int_0^h dh/H \right] \quad (125)
\]
The large-scale winds in the lower atmosphere generally conform to the "geostrophic approximation," in which the horizontal pressure gradient balances the Coriolis force in Eq. (123). If \( x \) and \( y \) denote local Cartesian coordinates at any point, conveniently taken in the southward and eastward directions, then the wind speeds \( U_x, U_y \) are given by

\[
-2\Omega U_x \sin \varphi = \rho^{-1} \frac{\partial p}{\partial y}, \quad 2\Omega U_y \sin \varphi = \rho^{-1} \frac{\partial p}{\partial x}
\]

(126)

for any nonzero latitude \( \varphi \).

In the atmosphere the temperature (and hence, the scale height \( H \)) may vary horizontally as well as vertically, in which case the pressure distribution \( p(h) \) is a function of \( x \) and \( y \). The horizontal pressure gradients result in winds, which conform to the "thermal wind equation" obtainable from Eqs. (125) and (126) in the following way. First, we differentiate Eq. (125) with respect to \( x \) and \( y \) to obtain expressions for the horizontal derivatives \( \partial (\ln p)/\partial x, \partial (\ln p)/\partial y \), which are substituted in (126). By using the perfect gas law in the form \( p = \rho g H \) (and assuming that variations in \( H \) are due only to variations in \( T \)), we obtain equations which can be differentiated with respect to height, yielding

\[
-\frac{\partial T}{\partial y} = A \frac{\partial (U_x/T)}{\partial h}, \quad \frac{\partial T}{\partial x} = A \frac{\partial (U_y/T)}{\partial h}
\]

(127)

where the constant \( A = 2\Omega T^2 g^{-1} \sin \varphi \). These equations (with proper boundary conditions) are useful in relating wind shears to horizontal temperature gradients.

1.62 LARGE-SCALE WINDS IN THE MESOSPHERE

The wind pattern and temperature distribution in the mesosphere are fairly well known, and are presented in contour diagrams by Murgatroyd (1957) and Craig (1965). The winds are essentially geostrophic, as described by Eq. (126). By using the thermal wind equations (127), Murgatroyd (1957) has shown that the zonal winds are reasonably consistent with the latitudinal temperature distribution, though discrepancies exist in moderately high latitude regions in winter. There are, in fact, marked seasonal changes in both winds and temperature distributions, as detected by the rocket experiments of Stroud et al. (1960) at Churchill (59°N). Strong eastward winds (100 m s\(^{-1}\) or more) in winter change to more moderate westward winds in summer, at 60–80 km altitude.

At 65 km the summer and winter temperatures are roughly equal, and higher up the seasonal variation reverses, the mesopause region around
80 km being about 40°K warmer in winter than in summer. This anomaly was noted in Sec. 1.31 in connection with the heat-balance problem, together with its suggested explanation in terms of the transport of atomic oxygen. The anomalous temperature distribution does seem to be consistent with the sense of the vertical wind gradients, for most of the zonal wind data at meteor heights (Hines, 1963). At 90 km the winds are generally eastward, of order 10 m s$^{-1}$ at Jodrell Bank (53°N) and 40 m s$^{-1}$ at Adelaide (35°S), but they are westward for periods in the spring. There are also meridional components of prevailing wind, leading to net latitudinal flow of air which must presumably be balanced by return flows at other heights. At the equinoxes, there appears to be a transition (lasting about a month) when the atmosphere changes between the patterns characteristic of the solstices.

1.63 LARGE-SCALE WINDS IN THE THERMOSPHERE

In the thermosphere above about 120 km, data on large-scale winds become sparse. The only methods yet available for measuring neutral air motions depend on the introduction of neutral contaminants, by means of rockets or artillery, as described in Sec. 1.82. Although results of several experiments have been published, there are not yet sufficient data to determine the prevailing winds in the thermosphere.

The daily temperature variations described in Sec. 1.33 lead to a daytime expansion of the atmosphere—termed the "diurnal bulge"—which is centered in low latitudes at around 1400 hours local time. The horizontal pressure gradients around this "diurnal bulge" provide the driving force for the thermospheric winds. In a sense, the motion constitutes a thermally driven tide; and, in principle, a tidal potential function $\Psi$ could be formulated for insertion into the basic equation (123). However, a simpler approach is to compute horizontal pressure gradients from thermospheric data and insert them in the equation, without using tidal theory. There is no need in this treatment to resolve the motion into diurnal, semidiurnal, or other harmonic components. The calculations of Lindzen (1967b), Kohl and King (1967), and Geisler (1966, 1967b) are of this type. Volland (1966) has developed a more complex computation for solving a two-dimensional version of the equation of motion (123), the thermal conduction equation (120), and the continuity equation (or conservation of mass equation) for the air, using the heat input function $Q(h, t)$ as basic data. This is an extension of the Harris and Priester approach (Sec. 1.33) to include horizontal motion—as well as vertical motion—of the air.

We assume that the equation of motion (123) applies to the winds. The
1.6 ATMOSPHERIC DYNAMICS

The equations governing the horizontal speeds $U_x, U_y$ are as follows (usually only the vertical gradients are important in the viscosity terms):

$$\frac{dU_x}{dt} - 2\Omega U_y \sin \varphi = \frac{\mu \partial^2 U_x}{\rho \partial h^2} - \frac{1}{\rho} \frac{\partial p}{\partial x} - v_{ni} (U_x - V_{ix})$$ (128)

$$\frac{dU_y}{dt} + 2\Omega U_x \sin \varphi = \frac{\mu \partial^2 U_y}{\rho \partial h^2} - \frac{1}{\rho} \frac{\partial p}{\partial y} - v_{ni} (U_y - V_{iy})$$ (129)

These equations become linear in $U_x$ and $U_y$ if the operator $d/dt$ is replaced by $\partial/\partial t$, the nonlinear term in Eq. (124) being neglected. We can deduce the order of magnitude of the wind speeds by considering the size of the different terms; thus, King and Kohl (1965) show that wind speeds of at least 30 m s$^{-1}$ occur by day at 300 km. Without ion-drag, the speeds would be an order of magnitude greater.

In solving Eqs. (128) and (129), it generally suffices to assume that the ion motion is parallel to the geomagnetic field, and is entirely caused by the winds. For the simplest case in which the magnetic field points northward (in the negative x-direction) and downward at the dip angle $I$, we have $V_{ix} = U_x \cos^2 I, V_{iy} = 0$.

As a lower boundary condition, it is generally assumed that the air is stationary, and all parameters are constant at some height in the lower thermosphere, say 120 km. This assumption isolates the motions in the thermosphere from any motions at lower heights, and though arbitrary, it may be quite reasonable for the purpose of calculating winds at greater heights. It cannot, of course, give an accurate description of conditions near the 120 km level itself; but one or two scale heights above this level, the solutions for $U_x$ and $U_y$ are almost independent of the assumed boundary conditions (Lindzen, 1967b).

The viscosity terms in Eqs. (128) and (129) express mathematically the tendency for shears of velocity to be removed by viscosity. The ratio $\mu/\rho$, known as the kinematic viscosity, increases exponentially upwards, since $\mu$ is relatively height-independent. The derivatives $\partial^2 U_x/\partial h^2, \partial^2 U_y/\partial h^2$ must therefore become small at great heights unless there is a corresponding upward increase in the acceleration provided by other forces. In fact, there is not; the pressure gradient terms (as computed from atmospheric models) do increase upwards, but only approximately linearly with height (Geisler, 1966), whereas ion-drag decreases upwards above the F2 peak at about 300 km. We can learn more about the behavior of $U_x$ and $U_y$ at great heights by considering the upper boundary for the solutions of Eqs. (128) and (129). This boundary is taken as the base of the exosphere, which may
be regarded as a free surface beyond which the fluid equations fail. No external force capable of maintaining a velocity shear exists at the boundary, so we expect $\partial U_x/\partial h$ and $\partial U_y/\partial h$ to become small at great heights as well as the second derivatives. Though the fluid equations of motion are inaccurate near the exospheric boundary, this should not affect the velocities at lower levels.

Apart from the effects of viscosity, the thermospheric wind speeds are controlled by Coriolis force and ion-drag. Without ion-drag, the wind is geostrophic and could be described by the thermal wind equation (127). Ion-drag results in a wind component across the isobars, comparable to the geostrophic component along the isobars when $\Omega$ and $v_{ni}$ become comparable. Geisler (1966) finds the two components to be equal for an ion concentration $2 \times 10^{11}$ m$^{-3}$, which is generally exceeded throughout the daytime ionosphere between 200 and 400 km altitude even at sunspot minimum, and considerably so at sunspot maximum. The smoothing-out effect of viscosity compels the winds at other levels to conform to the pattern imposed by ion-drag at 200–400 km, more or less. Thus, we can say that the thermospheric winds are dominated by ion-drag and blow across the isobars outwards from the center of the diurnal bulge. They blow from the daylit hemisphere to the night side of the earth, both across the polar regions and across the sunrise-sunset line in other latitudes (Fig. 7). At night and in high latitudes, where

![Fig. 7. Computed wind vectors at 300 km, shown against a background of isobars at this level ($H$ = high, $L$ = low pressure). The bar (at bottom left) represents the length of a wind vector of 200 m s$^{-1}$. Symmetry about the geographic equator is assumed [Geisler (1967b)].](image-url)
the ion concentration is small, a substantial geostrophic component exists. Geisler (1966, 1967b) computes typical speeds of order 100 m s\(^{-1}\), which imply that a given "cell" of air may travel some thousands of kilometers in the course of a day.

To account for certain perturbations in the orbits of artificial satellites, it has been suggested that the atmosphere above 200 km rotates about 30% faster than the earth, so that a prevailing eastward wind exists (King-Hele and Allen, 1966). No detailed explanation of this phenomenon has been established.

1.64 Gravity Waves

All the experimental data on upper atmosphere motions show great complexity, which seems most naturally interpreted in terms of the presence of waves. Winds deduced from drifts of meteor trails or of artificial clouds show structure of a few kilometers in vertical scale. The noctilucent clouds which sometimes appear at 80–90 km altitude, especially in auroral latitudes, display a wavelike structure. In addition, the ionization in the lower ionosphere appears to be stratified to some extent (Ellyett and Watts, 1959; Dieminger, 1959).

Some of the motions are due to turbulence, but this disappears above 100 km, as has been noted already. Partly they are due to tidal motions that represent a special case of wave motion. Hines (1960, 1965b) discussed the general problem of wave motions in the upper atmosphere, and in particular the class of "internal gravity waves." The energy of these waves is generally derived from larger-scale motions. They can be produced in the upper atmosphere by the breakdown of tidal motions, which attain such large amplitudes at these heights that nonlinear processes occur and cause dissipation of energy. Alternatively, wind systems in the lower atmosphere can generate gravity waves, though the penetration of the waves to the upper atmosphere then depends on the transmission and reflecting properties of the mesosphere.

The equations which are applied to the problem of gravity-wave propagation are the equation of motion in the form

\[
dU/dt = g - \rho^{-1} \nabla p
\]  

and the continuity equation, expressing continuity of air mass,

\[
\partial \rho/\partial t + \nabla \cdot (\rho\mathbf{U}) = 0
\]

as well as the adiabatic equation of state \( p \propto \rho^\gamma \) (where \( \gamma \) is the ratio of
specific heats) and the perfect gas law \( p = \rho kT/m \). A useful parameter is the speed of sound \( C \), given by

\[
C^2 = \gamma kT/m = \gamma gH
\]  

We shall assume here that \( g, C, \) and \( H \) are independent of height.

We wish to discuss oscillatory solutions of these equations, representing wave motions. As written, the equations contain various approximations. In neglecting Coriolis force, we restrict attention to periods much shorter than one day; and it is convenient also to assume the wavelengths to be much less than one earth radius, so that Cartesian coordinates may be used. There are no dissipative forces included in these equations. We also assume that the oscillations produce only small perturbations of pressure and density \( \Delta p \ll p, \Delta \rho \ll \rho \) and that second-order terms in \( U \) may be neglected. In this case the operator \( \partial/\partial t \) replaces \( d/dt \) in Eq. (130), and the nonlinear terms disappear from the equations.

We now specify that the pressure and density fluctuations and the horizontal and vertical velocities \( U_x, U_z \) conform to a wave equation of the type

\[
\frac{\Delta p}{p} \sim \frac{\Delta \rho}{\rho} \sim U_x \sim U_z \sim \exp i(\omega t - K \cdot r)
\]  

in which \( \omega \) is the angular frequency, \( r \) is a position coordinate and \( K \) the wavenumber vector, whose components are in general complex. It is then possible to derive a “dispersion equation” which is quadratic in \( \omega^2 \). In the absence of gravity, this may be reduced to the very simple equation for sound waves, \( \pm \omega = C|K| \). With gravity included, however, the equation has complex coefficients, and there are no solutions with \( K \) purely real. We therefore look for solutions in which the horizontal component of \( K \) is a real number \( k_x \), since these represent waves which travel horizontally without attenuation and may therefore be propagated over long distances. It is then found either that the vertical component of \( K \) is purely imaginary or that it takes the form \( (k_z + i/2H) \), with \( k_z \) real.

The first alternative represents a wave that contains no vertical phase propagation, and is too simple to represent the complicated vertical structure of actual atmospheric motions. This is termed a “surface wave.” The other alternative represents an “internal” wave, the amplitude of which increases upward with height \( h \) as \( \exp(h/2H) \). Since the air density \( \rho \sim \exp(-h/H) \) in the assumed isothermal atmosphere, the product \( \frac{1}{2} \rho |U|^2 \), which is the kinetic energy density, is independent of height; this is consistent with our neglect of dissipative processes.
The analysis may then be continued to obtain expressions for the "polarization factors," which play the part of constants of proportionality in Eq. (133). These are complex numbers, giving the relative phases and the magnitudes of $\Delta p/p$, $\Delta g/\rho$, $U_x$, and $U_z$ for a wave of unit amplitude, in terms of $\omega$, $k_x$, and $k_z$, and the constants $C$, $g$, and $\gamma$.

A study of the dispersion equation reveals that two classes of waves exist for which $\omega$ can be real. One class, known as "acoustic waves," possesses periods $2\pi/\omega$ less than a limiting value $T_a$; the other, known as "gravity waves," possesses periods greater than a limiting value $T_g$. These limits are given by

$$T_a = 4\pi C/g \gamma \approx 4.4 \text{ [min]}, \quad T_g = 2\pi C/g (\gamma - 1)^{1/2} \approx 4.9 \text{ [min]} \quad (134)$$

the numerical values being for typical mesospheric conditions.

An equation can be derived to show that, even within the gravity-wave range, the wavenumber $(k_x^2 + k_z^2)^{1/2}$ may become imaginary if either the scale height $H$ is sufficiently small or (in an extended theory, in which $H$ can vary with height) the vertical gradient $dH/dh$ is sufficiently positive. This implies a reflection of the waves, which may lead to their being trapped within certain regions of the atmosphere. A similar equation applies to tidal oscillations [Sec. 1.72, Eq. (137)].

The dissipation of gravity waves has been discussed by Hines (1960) and Pitteway and Hines (1963). Atmospheric molecular viscosity and thermal conductivity provide important dissipative mechanisms. Generally speaking, viscosity tends to remove the smaller-scale waves. For any given height there exist limiting values of $k_x$ and $k_z$ which represent the largest wavenumbers (smallest wavelengths) that can be propagated without prohibitive attenuation (i.e., of order one neper per wavelength). Waves for which either $k_x$ or $k_z$ exceeds these limits are too heavily damped to be observable. The limits increase with increasing height, indicating that progressively larger waves are destroyed by viscosity. Hines (1964c) has shown that the minimum scale sizes observed in the wind structure above 100 km (Kochanski, 1964; Zimmerman, 1964) are broadly consistent with the calculated limits.

Above 100 km, the waves tend to set the ambient positive ions in motion, and the resulting ion-drag is an important mechanism of loss. Since the ion motion is accompanied by the flow of electric currents, it may also be regarded as joule loss. Since the charged particle motions are affected by the geomagnetic field, directional effects may appear in the attenuation of gravity waves.

At greater heights, the "traveling ionospheric disturbances" (TID) ob-
served by Munro (1950, 1958) have been identified as gravity waves by Hines (1960, 1963). These disturbances have been observed to travel several thousand kilometers, their speeds being some hundreds of kilometers per hour. On the whole, they seem to originate at high latitudes and travel towards the equator. Because they travel so far, it seems that their energy must be trapped in some kind of horizontal "duct." If the "ducting" results from large-scale winds, it may be possible to explain variations in the dominant directions of motion noted by Munro. The energy is largely trapped at heights around 100 km, but some energy leaks upward to the heights at which the disturbances are observed (200–300 km). We shall discuss the ionospheric observations in Sec. 6.44.

For much fuller accounts of gravity wave propagation and related topics, see Hines (1960) and Tolstoy (1963).

1.7 Atmospheric Tides

We begin this section with some general remarks on tides, which are followed by a summary of pertinent experimental observations in the upper atmosphere (Sec. 1.71) and by an outline of tidal theory (Sec. 1.72), in which we do not attempt any detailed mathematical development. Wilkes (1949) has devoted a monograph to the topic "Oscillations of the Earth's Atmosphere," which not only deals with the theory and observations in detail but also traces the historical development of the subject. The theoretical study of tides dates from the work of Laplace in the eighteenth century.

For our purposes, the basic facts may be stated as follows:

1. The sun and moon produce tidal forces in the atmosphere, the periods being related to the solar day (24 hr) and lunar day (24.8 hr).
2. These forces set up tidal waves in the atmosphere, which result in (primarily horizontal) air motions. The atmosphere responds differently to forces of different periods.
3. The motion of the air across the geomagnetic field induces electromotive forces, which drive currents at levels in the ionosphere where the air is electrically conducting, thus causing the periodic solar and lunar magnetic variations.
4. The system of electric fields, currents, and charges produced by this process has important influences on the ionosphere itself.

Of the above, (3) and (4) come within the province of the "atmospheric dynamo theory," and are dealt with in Sec. 7.4. Here we confine attention
to the dynamical aspects (1) and (2). The earliest observational fact to be recorded was the 12-hr variation of barometric pressure in the tropics, probably known in the seventeenth century. It amounts to about 1 mb in amplitude and is visible on ordinary barograms, though in mid-latitudes it is concealed by the irregular meteorological variations. Some typical amplitudes of ground-level pressure oscillations are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Tropics (mb)</th>
<th>Mid-latitudes (mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar diurnal</td>
<td>24</td>
<td>0.8</td>
</tr>
<tr>
<td>Solar semidiurnal</td>
<td>12</td>
<td>1.3</td>
</tr>
<tr>
<td>Lunar semidiurnal</td>
<td>12.4</td>
<td>0.08</td>
</tr>
</tbody>
</table>

For gravitational tides, the principal period is semidiurnal (though small components of longer period might be significant), and the lunar driving force exceeds the solar force. This is consistent with the observed domination of ocean tides by the moon, and we might expect the same situation to apply in the atmosphere. However, there is also a driving force due to the periodic input of solar thermal energy. This has a fundamental 24-hr period, but its harmonics (12-, 8-, 6-hr, etc.) may be appreciable. Thus, it is not obvious why the solar 12-hr period should predominate in the oscillations.

To overcome this difficulty, Lord Kelvin suggested in 1882 that the atmosphere might have a natural period of oscillation close to 12 hr, which would amplify the solar semidiurnal tide, but would be too selective to respond to the lunar 12.4-hr driving force. This coincidence is possible, but it might seem too good to be true since resonant phenomena are rare in nature. There are other difficulties with the resonance theory, and it has now been abandoned in its original form. Modern theory indicates that the atmosphere responds differently to driving forces of different periods, and explains the observed phenomena accordingly; but this selectivity is much weaker than was envisaged by the resonance theory, which we mention for its intrinsic interest.

1.71 Observations of Upper Atmosphere Tides

At ground level, barometric observations provide the principal data on atmospheric tides. These are summarized by means of various empirical functions of latitude, longitude, and local time, such as G. C. Simpson's formula (Wilkes, 1949). The pressure oscillations give rise to air motions...
which are principally horizontal. The tidal wind speeds at the ground are only of order \( 0.05 \text{ m s}^{-1} \), however, and are very difficult to detect in the presence of much larger meteorological winds. Owing to the phenomenon of "tidal amplification," which results from the upward decrease of air density (Sec. 1.72), the winds are much larger at 100 km altitude, being of order \( 30 \text{ m s}^{-1} \). At these heights the winds are observed by the drifts of meteor trails and of artificially introduced contaminants; the tidal pressure oscillations are not observed directly.

In the height range 80-110 km, meteors produce ionized trails which drift with the neutral air and can be tracked by radar. This technique has the advantage that it can be used by day or by night. Observation of artificial clouds is not yet possible during daylight hours, and unless the cloud is self-luminous the observations can only be made at twilight. Noctilucent clouds provide further data, but they are seldom observed except at high latitudes. Another technique is the measurement of the drifts of ionospheric irregularities (Sec. 6.4); such drifts are only likely to be related to neutral air winds at heights below 110 km, and are usually only measured by day at these heights, owing to the lack of ionization at night.

In view of their importance, it is rather surprising that long sequences of wind measurements by the meteor-trail method have been carried out at two stations only. These are Jodrell Bank, England, 53°N latitude, from 1953 to 1958 (Greenhow and Neufeld, 1961), and Adelaide, South Australia, 35°S latitude, from 1952 to 1955 (Elford, 1959). The prevailing winds were mentioned in Sec. 1.62, but at both stations the periodic components generally exceed the prevailing winds. At Adelaide, the diurnal wind, typically 30 m s\(^{-1}\), slightly exceeds the semidiurnal wind on most occasions. Although the details of the winds, such as their variations with height and with season, are very complicated, the semidiurnal winds are found to be broadly in agreement with tidal theory. At Jodrell Bank the diurnal component is small and variable, typically only 5-10 m s\(^{-1}\), and here the semidiurnal winds are dominant, being typically 20 m s\(^{-1}\). An interesting feature of the semidiurnal winds, which is present only weakly in the surface pressure data, is a marked phase change during the autumn months, September to November. Although the difference between Jodrell Band and Adelaide might be due to differences between the Northern and Southern hemispheres, it is generally considered to be due to a latitudinal variation.

The semidiurnal winds deduced from radio observations of drifts in the ionospheric E region, at about 110 km, seem roughly consistent with the meteor results. At Cambridge, England (52°N), the vertical phase gradient
of the winds agrees with that found at lower heights, by the Jodrell Bank meteor experiments (Jones, 1958), and a similar phase shift in autumn is found.

Further evidence for the existence of tidal motions in the upper atmosphere has been derived from the experiments with chemiluminescent trails by Rosenberg and Edwards (1964) and Murphy et al. (1966) (Sec. 1.82).

1.72 An Outline of Tidal Theory

The mathematical theory of tides is devoted to finding functions to describe possible modes of oscillation of the atmosphere, with periods related to the solar or lunar day. In this section, we outline the theory but do not derive the equations in detail. Full treatments have been given by Wilkes (1949) and Craig (1965). The theory has much in common with the theory of gravity waves.

The tidal functions have to satisfy certain basic equations, similar to those used for gravity waves, but including Coriolis force. First, the equation of motion for the air velocity $U$; omitting ion-drag and viscosity which are unimportant in the region of present interest, below 100 km, we have from Eq. (123):

$$\frac{dU}{dt} + 2\Omega \times U = g - \rho^{-1} \nabla p - \nabla \Psi$$

(135)

where $\Psi$ is the tidal potential function. The motion also satisfies the continuity equation (131) and the adiabatic gas equation. If the tide is thermally driven by some source of heat (such as absorbed solar radiation), then another equation relating $\frac{dT}{dt}$ to the heat input $Q$ is obtained from the First Law of Thermodynamics (see Kato, 1966).

To solve these equations, the functions $p$, $\rho$ and $T$ are expressed as the sum of an undisturbed term (representing the stationary atmosphere), and a perturbation term (representing the tide) which is assumed to be so small that Eq. (135) can be made linear, with $\partial / \partial t$ replacing $d / dt$. The undisturbed atmosphere is assumed to be spherically symmetric and in hydrostatic equilibrium, so that $\rho^{-1} \nabla p$ cancels the gravity term.

With all these assumptions, it happens that the equations can be reduced to a single differential equation which is soluble by the method of separation of variables. The velocity divergence $\chi = \text{div} U$ is generally used as the dependent variable; it is written as a function of geocentric distance $r$ and colatitude $\theta$ (or latitude $\phi$), and must be periodic in longitude $\lambda$ and time $t$. The solar oscillation $\chi_s$ is written as a superposition of solutions

$$\chi_s = \sum_m \sum_n R_{mn}(r) \Theta_{mn}(\theta) \exp \left[ im (2\pi t/T_s + \lambda) \right]$$

(136)
where $T_S$ is the length of the solar day and the index $m = 1, 2, 3, \ldots$ for diurnal, semi-diurnal, ter-diurnal (etc.), oscillations. These oscillations travel westward, keeping pace with the subsolar point. Standing oscillations (with no term in $\lambda$) are possible, but are of less interest to us. A similar equation exists for the lunar function $\chi_L$, with $T_L$ replacing $T_S$. The expression "$(m, n)$ mode" is used to refer to an oscillation characterized by the indices $m$ and $n$. The functions $\Theta_{mn}$ are known as Hough functions, and can generally be written in terms of associated Legendre functions. The theory has been extended to include oscillations with negative values of $n$ (Kato, 1966; Lindzen, 1967a). Always $|n| > m$, and the function $\Theta_{mn}$ has $|n| - m$ nodes between the north and south pole (Hines, 1967), though special considerations apply if $m = 1$.

When the method of separation of variables is applied to the differential equation for $\chi$, an eigenvalue equation results. The eigenvalues have the dimension of length and are termed the "equivalent depths," denoted by $h_{mn}$ for the $(m, n)$ mode of oscillation; their values depend on the parameters of the atmosphere, particularly the vertical temperature profile. Typically, the equivalent depths are a few kilometers for the lowest semi-diurnal modes; they are negative if $n$ is negative. For each mode $(m, n)$ there exists a wave-number $k_{mn}$ describing the vertical propagation of the wave, given by

$$k^2_{mn} = (1/H h_{mn}) [(\gamma - 1)/\gamma + dH/dh] - 1/4H^2 \quad (137)$$

Then at any level, the radial variation approximates locally to the form $R_{mn} \sim \exp(ik_{mn}r)$, though $k_{mn}$ varies with $r$. If $k_{mn}$ is real, the wave is said to be "internal" and its energy can propagate vertically. If $k_{mn}$ is imaginary, the wave is "evanescent," and the energy does not travel far in the vertical direction from the level at which it is introduced. All the waves of negative $n$ are evanescent, even though they may contain appreciable energy. For waves of positive $n$, $k_{mn}$ may become imaginary at particular heights, which implies a reflection of the wave energy. We see that this can happen if $dH/dh$ becomes sufficiently negative. For the semi-diurnal $(2, 2)$ wave this occurs at the temperature decline in the upper mesosphere, but the reflection is partial because some energy leaks through to provide the semi-diurnal tide in the lower thermosphere; other semi-diurnal modes also contribute, such as $(2, 4)$. The reflections lead to the setting up of standing waves below the mesopause.

Apart from the existence of these reflections, the energy density in tidal oscillations (given by $\frac{1}{2} \rho |U|^2$) is independent of height. Thus $|U| \propto \rho^{-1/2}$ and the amplitude of the oscillations increases upwards, as in the case of gravity waves. This is known as "tidal amplification." The fractional oscillations of
pressure and density ($\Delta p/p$ and $\Delta \rho/\rho$) are of order $10^{-3}$ at the ground; but because $\rho$ decreases by a factor of $10^6$ between 0 and 100 km, they would approach unity at 100 km, were it not for the limitations set by partial reflection and other processes.

The linear theory outlined here is only valid for small oscillations, for which $\Delta p/p \ll 1$. When the "tidal amplification" becomes so great that this condition no longer holds, the wave "breaks" and its energy is degraded into smaller scale motion. Under these conditions, the gravity waves described in Sec. 1.64 are generated. Whether or not this nonlinearity becomes important, the energy of the tidal oscillations is dissipated at heights above 100 km by viscosity and ion-drag.

![Figure 8](image-url) Amplitude (full line) and phase angle (broken line) of the semidiurnal pressure variation, as a function of height [after Weekes and Wilkes (1947)]. The dotted curve shows an upward growth of amplitude proportional to $\exp(z/2)$, where $z =$ reduced height. The departure of the full curve from the dotted curve near 30 km represents a node in the standing wave pattern, produced by internal reflections from the ground and the mesosphere; there is a phase reversal at this level.

Figure 8 shows the amplitude and phase of $\Delta p/p$ for the solar semidiurnal (2, 2) oscillation, as a function of height (Weekes and Wilkes, 1947). There is a node at about 30 km, and the small gradient of $\Delta p/p$ between 60 and 80 km is due to the partial reflection of energy. The phase gradient above 80 km is connected with the leakage of energy from the upper boundary. The available observational data, however, do not show the node at 30 km.

If Eq. (135) is solved with no driving force ($\Psi = 0$), the solutions represent free oscillations of the atmosphere. These have periods of a few hours, but
none is close to 12 hr, as required by the now-abandoned "resonance" theory, that we mentioned earlier.

The theory becomes more complicated if the excitation is assumed to be thermal, rather than gravitational, as is, indeed, likely for the solar tide (Green, 1965). Craig (1965) and Kato (1966) give theoretical equations for thermal tides. Solar heating at the earth's surface is not an effective tide-raising force, but absorption of solar radiation by water vapor in the troposphere and by ozone at higher levels is very effective (Small and Butler, 1961). In detailed calculations, Butler and Small (1963) find the atmospheric response to ozone heating to be predominantly semidiurnal. Of course, a strong diurnal component exists in the heating function; but it turns out that the vertical wavelength $2\pi/k_1$, associated with the diurnal $(1, n)$ mode is too small, as compared to the vertical extent of the ozone layer in which the input of energy occurs, to permit efficient excitation (Hines, 1967). The diurnal tide observed at meteor heights is now thought to be primarily due to the $(1, -1)$ oscillation (Hines, 1967; Lindzen, 1967a), the form of which is consistent with the observation that the diurnal tide is stronger in lower latitudes.

1.8 Experimental Techniques

In this chapter we have frequently referred to experimental data on the structure and motions of the neutral upper atmosphere. We now wish to consider some of the experimental techniques in a little more detail, but without describing the results at any length.

1.81 Determination of Atmospheric Constitution

Below about 30 km, it is possible to measure pressure, temperature, and density of the neutral atmosphere with relatively inexpensive balloons. A searchlight probing technique has permitted ground-based measurements as high as 67 km (Elterman, 1954). In this method a powerful searchlight beam, which may be modulated to discriminate against the steady background radiation from the night sky, is directed upward. Another remote photosensitive detector then scans the searchlight beam to determine the intensity of the scattered light as a function of altitude. When only Rayleigh scattering is observed, as appears to be the case above about 10 km, the received intensity is proportional to the concentration of air molecules. Furthermore, when the equation of hydrostatic equilibrium is used, the temperature and pressure profiles may be deduced from the measured concentration data. On one occasion, Elterman obtained temperature profiles from rocket, sound
propagation and balloon measurements nearly simultaneously with the searchlight experiment and all were in satisfactory agreement.

The intense pulses of light obtainable from lasers make possible the application of radar techniques to the study of upper atmosphere densities. The first results were obtained by Fiocco and Smullin (1963) and Fiocco (1965), in Massachusetts. They reported the existence of dust layers between 80 and 120 km, and suggested correlation with ionospheric sporadic E layers and meteoric activity. Observations in Jamaica, West Indies (Clemesha et al., 1966), and in England (Bain and Sandford, 1966) gave results consistent with Rayleigh scattering, between 40 and 60 km approximately. At greater heights, between 70 and 80 km, the returned signal somewhat exceeded the expected Rayleigh scattering, though the dust layers mentioned above were not found.

Another fruitful method for study of the atmosphere below about 90 km is the measurement of sound delay times from grenade explosions at high altitudes. As a rocket climbs to 100 km or more, the grenades are ejected and exploded every few kilometers. The delay times are recorded at a network of ground stations, and from the records it is possible to deduce not only the temperature profile, accurate to within ±3°K, but also the horizontal winds in the mesosphere (Stroud et al., 1960). A simplified presentation of the theory and a number of other references are given in this paper, which describes the strong seasonal variation in the Arctic winds (see Sec. 1.62).

Most of the other methods require rockets to place the instruments within the environment it is intended to measure. For example, in an experiment to measure density, a collapsed sphere is launched to a high altitude and then ejected and inflated after the rocket power has been exhausted and the atmospheric drag has reached a tolerable magnitude (Bartman et al., 1956). Owing to the large area/mass ratio, the drag force is great enough to be measurable with a sensitive accelerometer even in the very tenuous upper atmosphere. The drag force may be written as

\[ D = \frac{1}{2} \rho v^2 AC_D \]  (138)

where \( \rho \) is the air density, \( v \) is the sphere velocity, \( A \) is the cross-sectional area and \( C_D \) is a dimensionless drag coefficient. Although the value of \( C_D \) is a rather complicated function of how the air molecules are reflected from the surface of the sphere, Faucher et al. (1963) believe that it can be calculated to within a few percent.

In a rocket flight reaching an altitude of 178 km, Faucher et al. (1967) detected measurable atmospheric drag below 130 km and computed density
values below this down to 88 km. The results generally conform to within
20% to the density profile of the U.S. Standard Atmosphere, 1962. They
provide valuable information about a region of the neutral atmosphere which
has proved difficult to investigate experimentally, and whose variations are
still little known.

LaGow et al. (1959) have compared the density and pressure profiles ob-
tained in the Arctic and in mid-latitudes from a number of rocket flights.
A variety of pressure gauges were used in the measurements and from these,
both the ambient pressure and the gas density were calculated up to an
altitude of 210 km. Their results show rather large differences in scale height
between White Sands, New Mexico and Fort Churchill, Canada, which are
not confirmed by other data.

At altitudes above 200 km, different methods must be employed for the
measurement of atmospheric density. A "ribbon microphone," in which the
incident flux of neutral molecules is modulated by a rotating shutter, has
been described by Sharp et al. (1962). The resulting ribbon displacements of
only a few Angstrom units produce an alternating voltage which is amplified
with sufficient accuracy to obtain the density (proportional to the ram
pressure oscillations) at an altitude of 550 km. Their results are consistent
with values deduced from satellite drag observations and have the out-
standing advantage of being nearly instantaneous measurements.

Below the turbopause, a measurement of gas density gives the concentra-
tion of each major constituent, since the composition is nearly the same as
at the ground. However, above the turbopause the composition varies, owing
to photodissociation and diffusive separation, so independent measurements
of each species are necessary. While density values can be obtained by the
measurement of ram pressure in a fast moving rocket, it is much more
difficult to obtain the concentration of each neutral species.

Nier et al. (1964) have employed a magnetic mass spectrometer in which
the entering neutral particles are ionized by a 90 eV electron beam. After the
 neutrals have been ionized they are accelerated, sorted by mass and collected.
They have shown results in the altitude range of 100 to 210 km. In addition
to the principal constituents $N_2$, $O_2$, and $O$, many other mass peaks are
found which are due to various impurities also present in the spectrometer.
Water vapor and $CO_2$ are responsible for many of these, but Nier et al. do
not believe the impurities affect appreciably the accuracy of the measure-
ments of the principal constituents. We have already shown the ratios of the
main constituents versus altitude in Fig. 5.

The first satellite launched primarily for aeronomy measurements was
Explorer 17 in 1963 (Spencer, 1965). The satellite contained four independent pressure gauges (of two different types) from which the local atmospheric density was calculated; two identical mass spectrometers for the measurement of the concentration of each principal constituent and two independent Langmuir probes, from which the electron temperature and ion concentration were deduced. The pressure gauges and mass spectrometers were cleaned and sealed in a laboratory vacuum and only exposed to the ambient atmosphere after the satellite was in orbit.

From the ram pressure measurement, the density of the atmosphere is calculated, since the thermal velocities of the particles (about 1 km s\(^{-1}\)) are much smaller than the satellite velocity (8 km s\(^{-1}\)) (Newton et al., 1965). At 360 km, the density varies diurnally by about a factor of five, with a maximum in the afternoon between 1200 and 1500 hours local time and the minimum near 0300 hours. It was found that the densities are very sensitive to even minor geomagnetic activity; a change from 0+ to 2+ in the magnetic Kp index (Sec. 7.33) is accompanied by an apparent 70°K temperature increase in the atmosphere. Other fluctuations do not seem to correlate with any index of solar or magnetic activity.

From the mass spectrometers, altitude profiles of N\(_2\), O, and He concentrations for both nighttime and daytime conditions have been obtained (Reber and Nicolet, 1965). Their results have been considered by C. Y. Johnson, along with other sources, in the preparation of our Fig. 6.

### 1.82 Release of Chemicals in the Upper Atmosphere

Following a suggestion by Bates (1950), a number of experiments have been performed in which sodium and other contaminants have been released from rockets at high altitudes. When the vapor is released all along the rocket trajectory, a trail develops which becomes highly distorted owing to wind shears and spreads out because of molecular diffusion. The sodium trail observed at twilight is believed to be due to resonant scattering of the 5890 Å sunlight, and it is rapidly extinguished following sunset at the appropriate altitude (Bedinger et al., 1958). However, even at night a brief trail is observed with the energy for the photoemission apparently derived from chemical reactions with atmospheric constituents. They provide important data on the existence of turbulence (Sec. 1.4). Other chemicals have been used for such experiments, notably trimethyl aluminium (TMA) which reacts with atomic oxygen in the atmosphere to produce a luminous trail (Rosenberg et al., 1963).

Alternatively, chemical contaminants have been carried aloft in gun-
launched projectiles. The gun launchings at Barbados, West Indies (Murphy et al., 1966) provide a relatively inexpensive measurement of neutral winds above the mesopause. Many projectiles have been fired from a 420-mm gun, reaching a height of 140 km; trimethyl aluminium (TMA) is usually released along the trajectory. Phototriangulation from several ground stations permits the trail to be tracked and horizontal winds to be calculated between 90 and 140 km. Since the vertical wind component is very small by comparison, and it is difficult to find any clearly identifiable points along the trails, only horizontal values are computed. It is possible to obtain several different wind profiles on a single night. We shall discuss the results in Sec. 6.3 in connection with the "wind shear" theory of ionospheric sporadic E. In later work (Murphy et al., 1967), both horizontal and vertical motions of an artificial electron-ion cloud were measured.

Nuclear explosions provide a quite different source of information on the upper atmosphere, particularly on the propagation of atmospheric waves. The conditions they produce in the atmosphere and ionosphere may be very abnormal, however. Some collected papers on this topic are contained in the New Zealand Journal of Geology and Geophysics and the Journal of Geophysical Research.

An important innovation in the release of chemicals from rockets has been made by Haerendel et al. (1967). They released a few kilograms (or less) of barium which lead rapidly to the formation of a few grams of Ba$^+$ ions. In addition to the ion cloud, the presence of a small amount of strontium provides a visible cloud of neutral atoms. A sequence of twilight photographs at several ground locations allows the motion of each cloud to be determined by triangulation. As might be expected, the neutral cloud is roughly spherical, while the ion cloud is elongated along the magnetic field direction. Near 200 km, the motions of both the clouds parallel to the earth's magnetic field are roughly the same, but the neutral wind perpendicular to the field is typically several times larger than the ion drift velocity (about 80 and 40 m s$^{-1}$, respectively). These experiments are useful not only for wind measurements but also may provide one of the first methods suitable for calculating values of the small electrostatic field believed to exist in the ionosphere (Sec. 7.4). Their initial results give values of 1–3 mV m$^{-1}$, which seem reasonable, although the theory is greatly complicated by the finite extent of the ion clouds.

1 5 (6) (1962).
2 68 (9) (1963).
1.83 Satellite Drag Analysis

Since 1957, it has been possible to measure atmospheric density by means of optical and radio tracking of satellites, from which can be found the rate of decrease of the orbital period. Very important advances have been made in our knowledge of the upper atmosphere from these observations. Although the gas density at 300 km is about $10^{-10}$ of that at the ground, appreciable drag is experienced by satellites at this altitude, and useful results have been obtained between 200 and 1100 km (Jacchia, 1963). The method of analysis is particularly simple for circular orbits, as will be shown below, but such an orbit is very difficult to achieve in practice, and the gas density is not spherically symmetric in any case. The accurate equations applicable to elliptical orbits have been derived by Sterne (1958) and others, and King-Hele (1963) has given a relatively simple derivation of the most important terms.

We shall only give some equations relating to a circular orbit of radius $r$. Let $G =$ gravitational constant, $M_E =$ mass of earth, $m_s =$ mass of satellite, and $v =$ satellite velocity. We may equate the centripetal acceleration $m_s v^2 / r$ to the gravitational force, obtaining

$$GM_E = v^2 r$$

(139)

The kinetic energy is given by

$$KE = \frac{1}{2} m_s v^2 = GM_E m_s / 2r$$

(140)

From first principles we know that the potential energy is

$$PE = -GM_E m_s / r$$

(141)

We should note that the magnitude of the PE is just twice that of the KE at any radius, and that their derivatives with respect to $r$ are also in the ratio of 2:1.

When atmospheric drag is included, the rate at which satellite energy is lost is $Dv$, where $D$ is the drag force given in Eq. (138). This may be equated to the rate of change of total energy:

$$- Dv = d / dt (KE + PE) = d / dr (KE + PE) \frac{dr}{dt}$$

(142)

Solving for the rate of change of altitude, we have

$$- \frac{dr}{dt} = \frac{\frac{1}{2} \rho v^3 AC_D}{\frac{1}{2} m_s GM_E / r^2} = \frac{\rho v AC_D r}{m_s}$$

(143)

Provided $\rho$ is sufficiently small that $r$ changes only slightly in one orbital
I. THE NEUTRAL ATMOSPHERE

period $T$, the satellite gradually spirals deeper into the atmosphere, gaining kinetic energy while losing potential energy twice as fast, in what may be termed a "quasi-circular" orbit.

We observe the paradoxical situation in which the drag force acts in the direction opposite to the velocity vector, yet the velocity of the satellite continually increases. This may be understood as follows: the impact of a molecule at one point in the orbit does reduce the KE and thus the satellite velocity by a very small amount. This puts the satellite in a slightly elliptical orbit with the point of impact at apogee. When the satellite reaches perigee half an orbit later, the velocity will have increased to a value greater than prior to impact at the expense of a reduction in PE. For the quasi-circular orbit, the variations of KE, PE and $r$ are, of course, continuous.

The measurements of atmospheric density rely upon the observation of the average rate of change of satellite period. Since the period $T = 2\pi r/v$, we find from Eq. (139) that

$$T^2 = \frac{(4\pi^2/GM_E)}{} r^3$$  \hspace{1cm} (144)

$$2T \frac{dT}{dt} = \frac{(4\pi^2/GM_E)}{} 3r^2 \frac{dr}{dt}$$  \hspace{1cm} (145)

Using Eqs. (139) and (143),

$$\frac{dT}{dt} = - (3\pi AC_D r/m) \rho$$  \hspace{1cm} (146)

The value of $C_D$ is usually considered to be about 2.2 for satellites (Cook, 1966). Equation (146) can then be used to calculate the density at the satellite altitude. The analogous equations for elliptical orbits will not be derived, but it develops that the quantity best determined is the value of $\rho H^{1/2}$ at the perigee of the satellite. Given an estimated value of $H$, this information leads to a value of $\rho$ at a level $\frac{1}{2}H$ above the perigee. Another method of determining scale height, based on observations of changing orbital characteristics, has been described by May (1963). We mentioned in Sec. 1.63 that satellite orbital data give information about mean zonal winds, in addition to the density variations we have discussed elsewhere.

A drawback of satellite orbital observations is their limited time resolution, amounting to hours or even days. By mounting a sensitive accelerometer within a satellite, as in the Italian "San Marco" project, the time resolution of the data can be greatly improved (Broglio, 1967).
2.1 Introduction

2.1.1 Some Historical Notes

The first suggestion of a conducting layer in the earth's atmosphere was associated with geomagnetism. The small daily variations of the magnetic field had been known since the eighteenth century, and as described by Kaiser (1962), speculations were made by C. F. Gauss in 1839 as to their origin in atmospheric electric currents. Lord Kelvin in 1860 speculated on the existence of a conducting layer, in connection with the phenomena of atmospheric electricity (Chalmers, 1962).

If at some level in the atmosphere the air is sufficiently conducting, then electric fields and currents are generated when winds drive the conducting layer across the earth's magnetic field. This is the basis of the "dynamo theory" of the magnetic daily variations, put forward in Balfour Stewart's article in the "Encyclopaedia Britannica" and subsequently developed by A. Schuster and others. The physical nature of the electrical conductors could not be established until J. J. Thomson's discovery of the electron and the subsequent growth of atomic physics, around 1900.

The "conducting layer" theory soon became topical in quite another way. In 1901, G. Marconi succeeded in transmitting radio signals across the Atlantic. This successful experiment implied that radio waves were deflected around the earth's surface, to a much greater extent than could be attributed to diffraction. In 1902, A. E. Kennelly and O. Heaviside suggested that free electric charges in the upper atmosphere could reflect radio waves, and their names became generally associated with the conducting layer, even though

1 9th Ed., 1882.
Heaviside's work was never published in detail (see Ratcliffe, 1967). J. E. Taylor in 1903 and J. A. Fleming in 1906 were among the first to suggest solar ultraviolet radiation as a source of the electrical charges, with its implication of solar control of radio propagation. This was not universally accepted and the "academic myth" of a conducting layer provoked controversy for another twenty years. Final experimental proof of the existence of reflecting layers came from the "frequency change" experiments of Appleton and Barnett (1925a, b), who demonstrated the existence of downcoming waves by an interference technique; and the "pulse sounding" experiments of Breit and Tuve (1925, 1926). Both series of experiments led to estimates of the height of the reflecting layers, and might be said to mark the beginning of ionospheric physics as an experimental subject in its own right. These developments were quickly followed by theoretical work, and the papers by Hulburt (1928) and Chapman (1931a) could be regarded as starting points of modern theory.

The name "ionosphere" was proposed by R. A. Watson-Watt, in a letter to the United Kingdom Radio Research Board dated 8 November 1926. Although it did not enter the literature till later (Watson-Watt, 1929), it has now virtually superseded the older term "Kennelly-Heaviside layer." In order to describe stratifications, the ionosphere is conventionally divided into regions (D, E, F), using a notation originated by E. V. Appleton, as described by Silberstein (1959) and Ratcliffe (1967). The boundary between the D and E regions is often taken as 90 km altitude, and that between the E and F regions at 150 km. However, for our physical description of the ionosphere, we prefer to avoid setting precise boundaries. Within any region, distinct "layers" or "ledges" of ionization may be observed (such as F1, F2); and there may exist a C layer at 50–60 km below the principal D-region ionization.

Interesting descriptions of early ionospheric history, including many literature citations, are to be found in articles by Kenrick and Pickard (1930) and Green (1946). A historical review of geomagnetism is contained in Chapter XXVI of "Geomagnetism" (Chapman and Bartels, 1940).

2.12 Some Results from the Magnetoionic Theory

The magnetoionic theory, which historically is particularly associated with E. V. Appleton, deals with the propagation of electromagnetic waves in ionized gases. We shall need some of its formulas in our discussion of experimental techniques for studying the ionosphere. We state here a few results, derivations of which are given in various books, such as Ratcliffe (1959), Budden (1961), and Davies (1965).
Let $N$ denote the concentration of free electrons, $m$ and $-e$ the electron mass and charge, $c$ the free space velocity of light, and $\epsilon_0$ the electric permittivity of free space. $B$ is the flux density of the earth's magnetic field; $f$ stands for radio wave frequency; and $\omega$ for angular frequency. We shall altogether neglect the influence of positive and negative ions on wave propagation. Two particular frequencies are defined by the equations:

Plasma:  \[(2\pi f_N^2 = \omega_N^2 = Ne^2/m\epsilon_0) \quad (200)\]

Gyro:  \[2\pi f_H = \omega_H = Be/m \quad (201)\]

The three customary magnetoionic parameters are defined as:

\[X = \omega_N^2/\omega^2, \quad Y = \omega_H/\omega, \quad Z = v/\omega \quad (202)\]

where $v$ is the collision frequency of electrons with heavy particles. If $\theta$ is the angle between the direction of the wave normal and the magnetic field, then we define $Y_L = Y \cos \theta$, $Y_T = Y \sin \theta$.

In the presence of the magnetic field, the ionosphere is a doubly refracting medium, and two modes of propagation exist for which the names "ordinary" and "extraordinary" are taken from crystal optics. Wherever the sign $\pm$ appears in this section, the $+$ sign refers to the "ordinary" wave and the $-$ sign to the "extraordinary" wave. We shall only need to evaluate the refractive index $\mu$ for the case $Z \ll 1$ (negligible collision frequency). From the fundamental Appleton–Hartree equation, it is found that

\[\mu^2 = 1 - \frac{X(1 - X)}{(1 - X) - \frac{1}{2}Y_T^2 \pm \left[\frac{1}{4}Y_T^4 + (1 - X)^2Y_L^2\right]^{1/2}} \quad (203)\]

A wave vertically incident on the ionosphere (assumed horizontally stratified) is reflected at a level where $\mu^2 = 0$. This occurs for the "ordinary" wave where $X = 1$, just as though there were no magnetic field. For the "extraordinary" wave, it is found (after some algebraic manipulation) that reflection occurs at the level where $X = 1 - Y$ if $Y < 1$ ($f < f_H$), and where $X = 1 + Y$ if $Y > 1$ ($f > f_H$).

If the wave normal lies in the direction of the magnetic field ($\theta = 0$), then $Y_L = Y$ and $Y_T = 0$ and Eq. (203) is considerably simplified. Even if $\theta$ is not zero, a good approximation to $\mu$ may be obtained by taking $Y_T = 0$; this is known as the "quasi-longitudinal" approximation, and it generally holds for high-frequency waves in the ionosphere provided $X$ is not near 1 and $\theta$ is not near $90^\circ$. (The actual condition for its validity is $\frac{1}{4}Y_T^4 \ll$
The refractive index is then given by

$$\mu^2 = 1 - X/(1 \pm Y_L)$$

(204)

The two modes are (nearly) circularly polarized in opposite senses, and a plane polarized wave traversing the ionosphere can be regarded as the sum of "ordinary" and "extraordinary" components. Because the two components have different phase velocities, the plane of polarization continually rotates along the path of the wave. This phenomenon resembles the "Faraday effect" of optics, in which the plane of polarization of light is rotated as it travels through a transparent medium along the direction of an imposed magnetic field; it is therefore known as "Faraday rotation."

We shall only require the high-frequency case in which $X \ll 1, Y_L \ll 1$. Along an element $ds$ of their path, the "ordinary" and "extraordinary" waves undergo phase changes of $d\phi_{o,x} = 2\pi \mu_{o,x} (ds/\lambda)$, where $\lambda$ is the free space wavelength. The plane of polarization of the resultant plane wave rotates through an angle $d\Omega = \frac{1}{2} (d\phi_o - d\phi_x)$. From Eqs. (204) and (202) we then have the Faraday rotation formula

$$\frac{d\Omega}{ds} = \frac{\pi}{\lambda} X Y_L = \frac{NB e^3}{8\pi^2 m^2 c \varepsilon_0 f^2} \cos \theta$$

(205)

in which we have converted $\omega$ to $f$ and used $c = f\lambda$.

Another formula we shall need is that for the absorption coefficient of high-frequency waves under "nondeviative" conditions, when $\mu = 1$, but collisions are important so that $Z$ need not be small. The refractive index is then a complex quantity $(\mu - i\kappa c/\omega)$; using the quasi-longitudinal approximation, the absorption coefficient (per unit length of path), $\kappa$, is given by

$$\kappa = \frac{\omega}{2\mu c Z^2 + (1 \pm Y_L)^2} = \frac{e^2 N}{2mc\varepsilon_0} \frac{\nu}{\mu \nu^2 + (\omega \pm \omega_L)^2}$$

(206)

where $\omega_L = \omega \mu \cos \theta$. We shall introduce numerical values into these formulas as they are needed, later in this chapter.

2.2 Vertical Incidence Sounding

After forty years, the pulse sounding technique of Breit and Tuve (1925) is still a basic tool of ionospheric research. A sounder is a type of radar which is capable of obtaining echoes from the ionosphere over a wide range of operating frequencies. The model C-4 ionosonde developed by the National Bureau of Standards has been employed at several stations for a
2.2 VERTICAL INCIDENCE SOUNDING

number of years and is representative of the many designs in use. In a typical mode of operation, the sounder is swept from 1–25 MHz in 15 s, using a pulse repetition frequency of about 60 s⁻¹ and a peak power of up to 20 kW. Newer types of ionosonde have been developed which provide better resolution and greater versatility in operation.

A newer development is the swept-frequency c.w. “chirp” sounder (Fenwick and Barry, 1966). This gives high quality ionograms with far smaller transmitter powers than are required in pulse sounding, but it has mainly been used for oblique-incidence studies that do not concern us here.

In a conventional recording system, the echo received from the ionosphere is used to modulate the intensity of a spot of light on an electronic time base. Distance along the time base represents the “time of flight” of the radio pulse, which, if divided by the free-space velocity, gives the equivalent path length, this being twice the “virtual height” $h'$ of the reflection point in the ionosphere. In the recorder, a photographic film is moved at right angles to the time base as the frequency is varied, so that the spot of light traces a graph of virtual height $h'$ against radio frequency. Range and frequency calibration markers are usually inserted automatically. These recordings are commonly known as “ionograms” or $h'(f)$ curves.

The pioneer ionospheric observatory is situated at Slough in England, where an ionosonde has been operated regularly since 1932. Other stations were opened in the thirties at Washington, D.C., Watheroo in Western Australia, and at Huancayo on the magnetic equator in Peru. Since World War II, many other stations have operated, and during the International Geophysical Year, 1957–1958, the total number was about 150. Normally, each station makes a sounding every hour (sometimes more frequently), and the total number of ionograms in existence amounts to some tens of millions. Two specimen ionograms, obtained at Lindau, Germany, are shown in Fig. 9.

2.21 CRITICAL FREQUENCIES AND $h'(f)$ CURVES

To understand the form of $h'(f)$ curves, we need to consider the propagation equations derived from the magnetoionic theory (Ratcliffe, 1959). From the Appleton–Hartree equation, expressions can be obtained for the phase refractive index $\mu$ and the group refractive index $\mu'$. Equation (203) gives $\mu$ for the case of negligible collision frequency (an adequate approximation for our present discussion); if the magnetic field is negligible then $\mu' = 1/\mu$, but otherwise, the theory gives a more complicated formula for $\mu'$. We stated in Sec. 2.12 that the “ordinary” wave is reflected at the level
II. IONOSPHERIC MEASUREMENTS
where $X = 1$, at which the wave frequency is equal to the local value of plasma frequency $f_N$. Inserting numerical values in Eq. (200), we find that, for $N$ in electrons per cubic meter,

$$f_N = (80.6 \, N)^{1/2} = 9 \, N^{1/2} \quad \text{[Hz]} \quad (207)$$

The “extraordinary” wave is reflected at the level where $X = 1 - Y$ (provided $Y < 1$; otherwise, at $X = 1 + Y$). At the peak of an ionospheric layer, where $N$ is greatest, we may apply the conditions $X = 1$ and $X = 1 - Y$ to find the critical frequencies $f_o$ and $f_x$, the minimum radio frequencies that can normally penetrate the layer in the two modes at vertical incidence. The separation of the “ordinary” and “extraordinary” critical frequencies can then be found from the definitions of $X$ and $Y$, Eq. (202); we have

$$f_o = \left[ f_x^2 - f_x f_H \right]^{1/2} = f_x - \frac{1}{2} f_H \quad (208)$$

The approximation is good if the critical frequencies are much greater than $f_H$, as is generally the case for the F layer (a typical mid-latitude value of $f_H$ is 1.4 MHz). Under certain circumstances, a third trace is seen on ionograms, notably at high latitudes. The critical frequency $f_z$ of this so-called z-trace is approximately $(f_o - \frac{1}{2} f_H)$. A discussion of this phenomenon is given by Ratcliffe (1959).

Figure 10 is an idealized sketch of daytime “ordinary” and “extraordinary” $h' f$ curves. The virtual height is large near the “critical frequencies” (or “penetration frequencies”) of the layers, which are shown as $f_z E$ and $f_z E$ for the E layer and $f_z F2$ and $f_z F2$ for the F layer. Sometimes (though not in this example), the F layer is divided and “cusps” at the F1 layer critical frequencies, $f_z F1$ and $f_z F1$, are then observed between the E layer and F2 layer.

---

**Fig. 9.** Daytime ionograms from Lindau/Harz, Germany (52°N, 10°E; local time 15°E meridian) showing virtual height $h'$ (km) against frequency $f$ (Mc/s = MHz). On the ionogram of 20 September 1961, the ordinary critical frequencies are $f_o E = 3.1$ MHz, $f_o F1 = 4.5$ MHz, $f_o F2 = 6.0$ MHz, approximately. The corresponding extraordinary critical frequencies are 0.7 MHz greater; e.g., $f_x F2 = 6.7$ MHz. There is an Es (sporadic E) layer with $f_x E = 3.4$ MHz and $f_x E = 4.1$ MHz. Owing to “blanketing” by Es (and to absorption), the F1 layer ordinary and extraordinary echoes are not seen at frequencies less than 3.4 MHz and 4.3 MHz, respectively. On the ionogram of 24 August 1960, $f_o E = 2.9$ MHz and $f_x F2 = 8.3$ MHz, approximately; there is no distinct F1 layer. An Es layer exists, which “blankets” the F-layer ordinary echo between 2.9 and 3.3 MHz, and the extraordinary echo between 3.6 and 4.0 MHz approximately. Second-hop echoes from most layers are seen on both ionograms. Some other features are due to interference [W. Becker, Max-Planck-Institut für Aeronomie, Lindau].
critical frequencies. Waves of frequencies exceeding $f_x F_2$ cannot be reflected at vertical incidence but penetrate the ionosphere completely. At times additional stratifications are observed; in fact, ionograms reveal a wide variety of phenomena, some of which are touched upon in subsequent sections. Also shown as the dotted curve in Fig. 10, is a plasma frequency profile that could have produced the virtual height curves (though, as we shall see in Sec. 2.22, this profile is not uniquely determined).

The ordinary critical frequencies such as $f_o F_2$ (the subscript "o" is sometimes omitted) are the most widely studied parameters of the ionosphere. The formula $f_o = (80.6 N_m)^{1/2}$, derived from Eq. (207), gives the peak electron concentration $N_m$ within each layer. Apart from $N_m F_2$, however, the values $N_m$ may not represent actual peaks (maxima) of the electron distribution since it is found that an inflexion of the $N(h)$ profile is sufficient to produce a maximum of virtual height which gives a "cusp" on an ionogram.

Actually $h'$ becomes infinite where $dN/dh = 0$; this condition occurs at the F2 peak but it does not necessarily occur in the E and F1 layers where there may be only a small "ledge" in the electron concentration profile (sketched for $f_o E$ in Fig. 10). We conclude that the virtual heights observed near the critical frequencies have little physical meaning. The real heights at these points, namely $h_m E$, $h_m F_1$, and $h_m F_2$, are important quantities, however, and we shall discuss how they are computed in Sec. 2.22.

Fig. 10. An idealized daytime ionogram, showing virtual height (km) versus frequency (Mc/s or MHz). Only E and F2 layers are present. The full and broken curves show the ordinary and extraordinary wave echoes, respectively. The dotted curve represents a possible profile of plasma frequency versus real height, which could produce the virtual height curves.
2.2 VERTICAL INCIDENCE SOUNDING

Besides their use for measuring peak electron concentrations, ionograms can also be used for estimating the thickness and total electron content of the various layers. A very full discussion of ionograms and their interpretation has been given by Piggott and Rawer (1961).

2.2.2 OBTAINING ELECTRON DISTRIBUTIONS FROM IONOGRAMS

Considerable analysis is needed to obtain the vertical electron distribution, or \( N(h) \) profile, from \( h'(f) \) curves. The virtual height (or group height) \( h' \) and the real height \( h \) at which a given frequency \( f \) is reflected are connected by an equation that depends on the electron distribution below the reflection point; it is

\[
h' = \int_0^h \mu'(f, N) \, dh
\]  

(209)

Since the group refractive index \( \mu' \geq 1 \), the group height \( h' \geq h \). We wish to solve this equation to obtain \( N(h) \). This is accomplished by writing it as an integral with respect to plasma frequency instead of height (which assumes that \( N \) and \( /fN \) vary monotonically with height)

\[
h'(f) = \int_0^f \mu'(dh/df_N) \, df_N + h(0)
\]  

(210)

where \( h(0) \) is the height of the base of the ionosphere, below which it is assumed that \( f_N = 0 \). At this base the real and virtual heights are equal.

A good review of the ways of inverting this equation has been given by J. O. Thomas (1959). The availability of electronic computers has made obsolete the earlier methods in which a special form, such as a parabola, is assumed for the \( N(h) \) profile.

The present-day methods of solving the integral equation (210) fall into two main classes, "lamination" and "polynomial." The former class, represented by the "matrix method" of Budden (1955) replaces the integration in (210) by a summation over a number of thin slabs, each corresponding to a discrete interval of plasma frequency. For this purpose, some simplifying assumption is made, the commonest being that the gradient \( dh/df_N \) is constant within each slab so that the function \( h(f_N) \) is represented by a series of linear segments.

Suppose there are \( n \) slabs, and that \( h'(f_i) \) is the virtual height for the frequency \( f_i \) reflected at the top of the \( i \)th slab, at a real height \( h_i \). We can
then form a set of \((n + 1)\) equations, examples of which are

\[
\begin{align*}
   h'(0) &= h(0) \\
   h'(f_1) &= h(0) + \frac{h(f_1) - h(0)}{f_1 - 0} \int_0^{f_1} \mu'(f_1, f_N) df_N \\
   &\vdots \\
   h'(f_j) &= h(0) + \sum_{i=1}^{j} \frac{h(f_i) - h(f_{i-1})}{f_i - f_{i-1}} \int_{f_{i-1}}^{f_i} \mu'(f_j, f_N) df_N
\end{align*}
\]

(211)

Additional equations are written up to \(j = n\) (note that we take \(f_{i-1} = 0\) for \(i = 1\)). By using the abbreviation

\[
   M_{ji} = (f_i - f_{i-1})^{-1} \int_{f_{i-1}}^{f_i} \mu'(f_j, f_N) df_N 
\]

(212)

we can write the equations compactly in the form

\[
   h'(f_j) - h(0) = \sum_{i=1}^{j} [h(f_i) - h(f_{i-1})] M_{ji}
\]

(213)

or

\[
   h^{*'} = \sum_{i=1}^{j} (\Delta h_i) M_{ji}
\]

(214)

Note that \(h^{*'}\) is the virtual height at frequency \(f_j\) measured from \(h(0)\), the base of the layer; \(\Delta h_i\) is the real width of the \(i^{th}\) slab; while \(M_{ji}\) is the average value of \(\mu'\) within the \(i^{th}\) slab for waves of frequency \(f_j\). This equation can obviously be written in matrix form, \(h^{*'} = M \cdot \Delta h\), its solution being \(\Delta h = M^{-1} \cdot h^{*'}\). From the inverted equation, which gives the width of each slab, the real height is obtained as

\[
   h_j = h(0) + \sum_{i=1}^{j} \Delta h_i
\]

(215)

Given the data necessary to evaluate the matrix elements for the required location, Eqs. (214) and (215) may be conveniently solved with the aid of a computer to obtain the real heights \(h_j\) for the plasma frequencies \(f_j\).

It may be advantageous to use an unequally spaced set of frequencies \(f_i\). These can be so chosen that the summation reduces to a simple averaging of a number of values of \(h'\), which can be carried out manually (Kelso, 1952). Schmerling (1958) has shown that this method can be adapted to include the magnetic field. Some authors do not explicitly include the term \(h(0)\), or \(h'(0)\), in their equations. For example, J. O. Thomas (1959) imposes an additional
condition on the matrix elements which takes account of this term, but neglects the effect of ionization in the first lamination (i.e., below the height \( h(f_1) \) at which the first frequency \( f_1 \) is reflected).

The assumption that \( \frac{dh}{df} = \text{constant} \) within each slab may be varied. In a comparison of different methods, G. A. M. King (1957) advocates the assumption \( \frac{dh}{d(\ln f_N)} = \text{constant} \), whereas Doupnik and Schmerling (1965) use a parabolic variation within each slab.

In the "polynomial" method of Titheridge (1961), it is assumed that the function \( h(f_N) \) can be represented by a polynomial in \( f_N \). Using \( h^* \) and \( h'^* \) to denote real and virtual heights measured from the base level \( h(0) \), we write

\[
h - h(0) \equiv h^* = \sum_{j=1}^{n} \alpha_j f_N^{j+1}
\]

First a set of sampling frequencies \( f_1, f_2, \ldots, f_n \) is chosen. Then the heights \( h_i^* \) at which the plasma frequency takes the values \( f_i \) can be included in a matrix equation, \( h^* = A \cdot \alpha \), in which each element of the column matrix \( \alpha \) is the coefficient appearing in Eq. (216), and the element \( A_{ij} \) is equal to \( f_i^{j+1} \). Now the set of virtual heights at frequencies \( f_i \) can be written as \( h'^* = B \cdot \alpha \), in which the elements of \( B \) contain integrals with respect to frequency of the group refractive index \( \mu' \). Hence \( h^* = AB^{-1} \cdot h'^* \). The elements of the matrix \( AB^{-1} \) depend on the frequencies \( f_i \) and the parameters of the geomagnetic field, and can be computed for a given location and a given set of sampling frequencies. Thus, sets of values of \( h' \) and \( h \) are connected by a matrix equation, as in the lamination method. But whereas the lamination method represents the function \( h(f_N) \) by a series of linear or other specially shaped segments, the polynomial method assumes that a smooth polynomial of degree \( n+1 \) in \( f_N \) can be drawn to pass through the \( n \) sampling points \( (h_i, f_i) \), although the coefficients of this polynomial are not required to be known. For a given number of sampling points, the polynomial method gives greater accuracy than the lamination method. Titheridge (1961) also describes a "modified polynomial method" of greater complexity, which is suitable for reducing ionograms containing a number of cusps. However, polynomial methods possess certain drawbacks, such as the tendency for spurious oscillations to appear in the \( N(h) \) profiles.

There are several difficulties to be faced in the reduction of \( h'(f) \) curves to \( N(h) \) profiles. One of these is termed "valley ambiguity." All the methods of analysis that we have described give unambiguous results only if electron concentration increases monotonically with height. If the \( N(h) \) profile is re-
entrant so that the same value of $N$ occurs at more than one height, then the reduction procedures do not give a unique result. For instance, it is possible that $N$ decreases with height immediately above the E layer to give a "valley" between the E and F layers. Indeed, this was generally thought to be the case until rocket observations showed the valley to be small on most occasions. It is also possible, though less likely, that there could be a valley between the F1 and F2 layers.

Conventional methods of reduction also become inaccurate near the peak of a layer, where $h' \to \infty$. This is particularly noticeable in calculations for the F2 layer, so that some workers prefer to compute the height at which $(N/N_m)$ takes a fixed value, such as 0.9, rather than the actual height $h_m$ of the peak (Schmerling, 1960). Quite accurate values of $h_m$ should be obtainable from the use of a model "parabolic" distribution in the vicinity of the peak (Titheridge, 1966a).

Another difficulty arises because conventional ionosondes have a lower frequency limit, usually about 0.5–1 MHz, and record no echoes below this limit. Consequently, they obtain no information about the electron distribution below the E layer in the daytime or below the F layer at night. Special sounders have been built which operate down to 50 kHz (Watts, 1957), but this equipment is not in general use. The lack of data about these low frequencies can lead to errors of 20 km or more in the computed real heights corresponding to higher plasma frequencies.

Titheridge (1959a) has shown that if data from the extraordinary ray virtual height are included in the analysis, it is possible to correct for the low-level ionization not included in the ionograms, and also to make some estimate of the ionization contained in the valleys. Although the method does not give a detailed distribution of electrons in the valley, Titheridge was able to show that, on some occasions, only a very small valley existed between the E layer and F layer at night. Paul and Wright (1963) have shown some results obtained with a very refined method of $N(h)$ analysis. Given ionograms of sufficiently high quality, it should be possible to compute the real height at which a given electron concentration occurs to within about 1 km, whereas with older methods of reduction in which the difficulties described above have not been fully overcome, the errors might be about 10–20 km (Thomas et al., 1958). A group of papers on the analysis of ionograms and computation of electron distributions appears in *Radio Science.*

$^2$ 2 (10) (1967).
2.23 TOPSIDE SOUNDING

The successful launching of the Canadian satellite "Alouette" in September 1962 inaugurated the era of the "topside sounder" (Warren, 1963). This satellite was placed in a nearly circular orbit at a height of just over 1000 km and carried a compact sweep-frequency ionosonde operating on radio frequencies of 0.5 to 11.5 MHz. On command from ground stations, the satellite made soundings at intervals of 20 s, corresponding to intervals of about 100 km in horizontal distance. A topside ionogram, matched to a conventional bottomside ionogram taken at almost the same place and time, is shown in Fig. 11.

The interpretation of the ionograms obtained from topside sounders, and their reduction to $N(h)$ profiles, present some interesting problems. The

![Fig. 11. Matching topside and bottomside ionograms recorded above Port Stanley, Falkland Islands (52°S, 58°W), near 1100 hours local time on 2 January 1963. The virtual height scales indicate depth below the topside sounder satellite (Alouette I) for the upper ionogram, and height above ground for the lower ionogram. The critical frequency $f_{\circ} F2$ is at 5.8 Mc/s (MHz) [P. A. Smith, Radio and Space Research Station, Slough (Crown Copyright)].]
II. IONOSPHERIC MEASUREMENTS

analysis is complicated by the fact that neither the electron concentration nor its gradient $dN/dh$ is zero at the sounder, as is the case with a conventional ionosonde on the ground. On the other hand, there is generally less structure in the topside than in the bottomside of the ionosphere, so that there is less difficulty with valleys. Nelms (1963) has developed a reduction program which is based on the original Budden matrix method, whereas Fitzenreiter and Blumle (1964) use a lamination method which takes $dh/d(\ln N)$ = constant within each slab. The “parabolic lamination” method of Doupnik and Schmerling (1965) is also useful for topside ionograms. Thomas et al. (1963) have used a polynomial method in which each real height is expressed as a polynomial in $f_N - f_V$, $f_V$ being the plasma frequency at the satellite vehicle.

The topside records show a number of plasma phenomena. These include resonances at the plasma frequency $f_N$, the electron gyrofrequency $f_H$ and its multiples, and the electron hybrid resonance $(f_N^2 + f_H^2)^{1/2}$; and cutoffs at the minimum frequencies at which particular propagation modes are possible (Lockwood, 1963; Calvert and Goe, 1963). These resonances have also been studied by a fixed-frequency topside sounder (Calvert and Van Zandt, 1966). They provide means of measuring both the electron concentration and the earth’s magnetic field. In addition, phenomena detected by the very-low-frequency receivers aboard Alouette and other satellites have provided information on the positive ion composition at the satellite (Barrington et al., 1965; Gurnett et al., 1965).

Another topside sounder, Alouette II, was launched in November 1965 into an orbit with apogee of nearly 3000 km, which carried it into regions of low electron concentration (Nelms et al., 1966). In addition to the cutoffs and resonances seen before, it was possible to see an interference effect between the electron gyro-resonance $f_H$ and the electron hybrid resonance $(f_N^2 + f_H^2)^{1/2}$, which converge as $f_N$ becomes small. Hagg (1967) has shown that electron concentrations of 8–100 cm$^{-3}$ can be computed from such interference patterns on the topside ionograms. We do not discuss the physics of the resonance phenomena here, although it has become a very active field of research.

2.3 Ground-Based Radio Investigation of the Lower Ionosphere

2.31 Some Basic Principles

In contrast to the relative ease with which the electron distribution in the E and F regions can be measured by ionosondes, the D region below 90 km
2.3 RADIO INVESTIGATION OF THE LOWER IONOSPHERE

presents formidable experimental difficulties. This section describes the ground-based radio techniques which are used to obtain D-region parameters. These include experiments with the long waves, several kilometers in wavelength, which are reflected from the D region; and experiments with shorter waves which penetrate the D region and are reflected in the E and F regions, such as absorption measurements and the newer techniques of partial reflections and wave interaction. These methods are supplemented by rocket experiments, using both radio propagation and direct measurement techniques, which are described in Secs. 2.4 and 2.5. The physical information which emerges, concerning the quiet and disturbed D region, is discussed in Secs. 5.1 and 8.3.

All the techniques to be described depend on the fact that the D region is both a refracting and an absorbing medium for radio waves. When a radio wave travels through the ionosphere, it sets the electrons into oscillation, and these oscillations are damped by collisions between electrons and gas molecules. In the E and F regions, these collisions are sufficiently infrequent that for many purposes—such as ionosonde measurements of electron concentration—their effects may be neglected. In the D region this is not so; both the electron concentration \(N(h)\) and the collision frequency \(v(h)\) are basic parameters, and the refractive index is a complex quantity. Its imaginary part is related to the absorption coefficient \(\kappa\), which we may take from Eq. (206) of Sec. 2.12. Inserting m.k.s. values,

\[
\kappa = 5.31 \times 10^{-6} \frac{N v^2}{\mu v v^2 + (\omega \pm \omega_L)^2} \text{ [m}^{-1}] \tag{217}
\]

in which the + and − signs apply to the “ordinary” and “extraordinary” modes, respectively; \(\omega = 2\pi f\) is the angular frequency of the radio wave; and \(\mu\) is the real part of the refractive index derived from the Appleton–Hartree equation. As in Sec. 2.12, the angular frequency \(\omega_L\) is related to the gyrfrequency \(f_H\) of electrons in the earth’s magnetic field; if \(\theta\) is the angle between the direction of the field and the direction of radio propagation, then we define \(f_L\) and \(\omega_L\) by the equations

\[
f_L = f_H \cos \theta, \quad 2\pi f_L = \omega_L = (Be/m) \cos \theta \tag{218}
\]

Throughout the D region the collision frequency \(v\) decreases exponentially with height, being roughly proportional to the neutral gas pressure, whereas \(N\) increases rapidly with height. The sensitivity with which any radio technique can measure \(N\) and \(v\) depends on the relative sizes of \(\omega\), \(v\), and \(\omega_L\). In experiments using medium-frequency and high-frequency waves, it may
generally be assumed that $\mu = 1$ in the D region and that $\omega \gg \omega_L$. If $v \gg \omega$, as is the case in the lower D region, then $\kappa \propto N/v$; but on the other hand $\kappa \propto Nv$ if $v \ll \omega$, as occurs in the upper D region. For waves of a few megahertz the transition level where $v \approx \omega$ occurs at around 70 km. As a very general rule, many of the experiments described in this section yield most detailed information at heights where $v$ and $\omega$ are comparable.

The simple theory based on Eq. (217) is inaccurate because it does not take into account that the collision frequency of electrons with neutral particles depends on electron energy. The laboratory experiments of Phelps and Pack (1959) have established that the collision frequency $v_M$ of monoenergetic electrons in nitrogen gas varies linearly with temperature, quite accurately. This result is generally assumed to apply to the D region, and it is used in conjunction with the assumption that electrons possess a Maxwellian distribution of energy. By means of a statistical calculation, known as the "Sen–Wyller generalization of the Appleton–Hartree theory," it is possible to define an "effective collision frequency" $v_{\text{eff}}$, which can be used in the Appleton–Hartree equation. For a given ambient temperature $T$, $v_{\text{eff}}$ lies in the range $1.5 v_M(T)$ to $2.5 v_M(T)$, depending on the values of $\omega$ and $\omega_L$; $v_M(T)$ applies to electrons with energy $kT$ (Sen and Wyller, 1960; Kane, 1962; Thrane and Piggott, 1966).

For some calculations, particularly those concerned with the reflection of long waves, even this refinement is inadequate. Instead of an approach based on ray optics, in which the refractive index and absorption coefficient are computed from the Appleton–Hartree equations, it is necessary to solve the "full wave" propagation equations to obtain the spatial distribution of the electromagnetic wave fields; see Budden (1961). The "full wave" method is appropriate if the properties of the ionosphere (such as $N$ and $v$) change considerably within one radio wavelength.

### 2.32 Phase and Amplitude of Long Waves

The data obtained from radio waves actually reflected in the D region consist of phase and amplitude measurements of signals in the range 10–200 kHz, propagated over distances of up to about 1000 km. One of the most basic parameters, obtained from the phase of the received signal, is the "equivalent height of reflection," which depends on wave frequency $f$ and on the angle of incidence at the ionosphere, $i$. The comparison of different data is often simplified by using an approximation known as Martyn's theorem, which states that in a plane stratified ionosphere the reflection height of an obliquely incident wave is equal to that of a vertically incident wave of
frequency $f \cos i$. The equivalent height for vertical incidence is found to show considerable diurnal changes; for 16 kHz waves it is of order 70–75 km by day, and roughly 90 km by night (Bracewell et al., 1951). The equivalent height is reduced during ionospheric disturbances of various types, and provides a useful indicator of disturbed conditions, as discussed in Sec 8.3.

Further important information is obtained from the amplitude and polarization of the reflected signals. If a plane polarized wave is transmitted from the ground, the downcoming wave reflected from the ionosphere is, in general, elliptically polarized. It may be resolved into two plane polarized components, whose electric vectors are respectively parallel to and perpendicular to the electric vector of the upgoing wave. The ratio of the amplitudes of the upgoing and downcoming components, with similar polarizations, gives a "reflection coefficient"; the ratio of the amplitudes of the upgoing wave and the perpendicular downcoming component gives a "conversion coefficient." At vertical incidence the "reflection" and "conversion" coefficients are generally similar and the downcoming wave is roughly circularly polarized. At oblique incidence, two pairs of "reflection" and "conversion" coefficients are defined, one pair for an upgoing wave polarized parallel to the plane of incidence on the ionosphere, the other for an upgoing wave polarized perpendicular to this plane.

Many attempts have been made to use the phase and amplitude data to determine the height profiles of electron concentration and collision frequency. This problem is so difficult that it is approached by a "trial and error" method; model distributions are postulated, and their reflecting properties computed and compared with the data. After a series of adjustments, a best-fitting model is found (Nertney, 1953). As previously mentioned, the "ray theory" approach of the Appleton–Hartree equations is inadequate for this purpose, and it is necessary to solve the complete wave equations (Gibbons and Nertney, 1952). The "full wave" approach was developed for computer use by Barron and Budden (1960), Pitteway (1965) and Piggott et al. (1965). It is now a very powerful tool for investigating D-region electron distributions, and has been used by Deeks (1966) to obtain best fitting $N(h)$ profiles from a large amount of earlier long-wave propagation data. A further development in the calculations, which makes use of the "stationary phase" principle of optics, has been described by Bain and May (1967).

2.33 Absorption Measurements

For over thirty years, the absorption of radio waves has been used as a means of investigating the lower ionosphere. Three types of experiment are
known by the following conventional designations, and the frequency ranges which are most generally used are also shown:

- **A1**: Vertical incidence pulse absorption 1.5–6 MHz
- **A2**: Cosmic noise absorption (riometer) 5–30 MHz
- **A3**: Field strength of c.w. signals 0.01–5 MHz

The pulsed signals used in the A1 experiment are reflected from E or F layers, at which level they suffer some "deviative" absorption, which is particularly severe if the wave frequency is close to one of the E or F layer critical frequencies. This absorption, to which Eq. (217) does not apply, must be allowed for in the determination of D-region "nondeviative" absorption. The relative size of the "deviative" and "nondeviative" components may be in question at times, though in principle their different frequency dependences enable the two components to be separated, as explained below.

The absorption is often expressed in terms of the "apparent reflection coefficient" \( \rho \), defined as the ratio of the amplitude actually received to that which would have been received with no absorption present. This definition implies that the effects of spatial attenuation are removed before \( \rho \) is calculated. For a signal reflected at vertical incidence, which traverses the D region twice, the quantity known as "the absorption" is

\[
-\ln \rho = 2 \int_0^h \kappa \, dh
\]

(219)

in which the integration is taken up to the reflection height. The symbol \( L \) is often used to denote the absorption in decibels, so that \( L = -8.68 \ln \rho \). Under normal conditions in the D region, practically all the absorption occurs at heights where \( v \ll \omega \) and \( \mu = 1 \); so that for the ordinary mode, which is generally observed in practice, the absorption is

\[
L = \frac{2.33 \times 10^{-6}}{(f + f_L)^2} \int_0^h Nv \, dh = \frac{A}{(f + f_L)^2}
\]

(220)

in which \( L \) is in decibels and the numerical constant is appropriate for m.k.s. units (for c.g.s. units its value is \( 2.33 \times 10^{-2} \)), and \( f_L = \omega_L/2\pi \); see Eqs. (217) and (218). Hence, a graph of \( L^{-1/2} \) versus \( f \) should give a straight line, with a slope \( A^{-1/2} \) (where \( A \propto \int Nv \, dh \)) and intercept \( -f_L \). The experimental data generally follow a relation of this type, except where \( f \) is near one of the critical frequencies \( f_E \), \( f_F1 \) (etc.), at which deviative absorption is important (Appleton and Piggott, 1954); see Fig. 12. This provides evidence that nondeviative absorption is dominant, away from the critical frequencies.
Fig. 12. Variation of absorption with radio-wave frequency, observed at Slough, England (52°N, 1°W) near noon on 11 July 1950. The measured values are shown by open circles, and the straight line shows the theoretical frequency variation (see text). As frequency increases, reflection takes place successively in the E, Es, F1, and F2 layers (Es is sporadic E, see Sec. 6.3). The quantity $f_g$ is proportional to the electron gyrofrequency, as in Eq. (218) [after Appleton and Piggott (1954)].

The use of cosmic radio noise to measure ionospheric absorption—now known as the A2 method—was described by Mitra and Shain (1953) who observed at a frequency of 18 MHz. The cosmic radio noise flux incident on the ionosphere varies with direction in space, and thus with sidereal time. To determine the ionospheric absorption at a given local time, the noise level is compared with the level observed at the same sidereal time, with the same antenna system, at a time when no absorption is present (i.e., at night under magnetically quiet conditions); the difference gives the ionospheric absorption. An instrument called a riometer (for relative ionospheric opacity meter) is used for routine absorption measurements (Little and Leinbach, 1958). It is designed to reject interference from broadcasting stations, by scanning across a bandwidth of order 100 kHz and determining the minimum noise level observed. At frequencies of 25–30 MHz, generally used for riometer operation, the normal D and F regions typically contribute about one decibel each to the total daytime absorption (Lusignan, 1960). Periods of greater absorption are mostly due to increases of D-region electron concentration, such as occur during auroral and polar cap absorption events and sudden ionospheric disturbances (Secs. 6.1 and 8.3).
II. IONOSPHERIC MEASUREMENTS

If riometer data for several different frequencies are available, one can exploit the fact that the absorption at any one frequency is mainly determined by the electron concentration at the height where $\omega = v$ for that frequency. Parthasarathy et al. (1963) have described a method in which the electron distribution $N(h)$ is represented by a polynomial containing $n$ parameters, these being determined from data for $n$ radio frequencies. This method thus bears some resemblance to the polynomial method of determining $N(h)$ profiles from ionograms (Sec. 2.22). There are certain limitations in the method, one of which arises from the finite beamwidth of the riometer antenna systems; typically the antenna beam is some tens of kilometers wide at D-region heights, so that for accurate results the distribution of ionization must be uniform over horizontal distances of this order (unless the polar diagrams are identical at every frequency, which is unlikely). Parthasarathy et al. used the method to determine the $N(h)$ distribution at high latitudes during a polar cap absorption event (Sec. 8.3); for values of $N > 100$ cm$^{-3}$ (10$^8$ m$^{-3}$) the real heights $h$ were considered to be determined to within 3 km.

Field-strength measurements of c.w. signals from distant low-frequency and medium-frequency transmitters constitute the A3 method of measuring absorption. The transmission paths are typically a few hundred kilometers. The receiving antenna system must be capable of distinguishing between the "ground wave," received direct from the transmitter, and the "sky wave," reflected from the ionosphere. Within certain assumptions, the amplitude of the sky wave can be used to determine an equivalent value of absorption at vertical incidence, by use of Martyn's theorem (see Sec. 2.32). Great care must be taken to avoid ambiguity, whenever there is a possibility of more than one ionospheric propagation path from the transmitter to the receiver. Some results of the A3 method have been described by Lauter (1966) and Schwentek (1966).

A crude measure of absorption is given by $f_{\text{min}}$, the minimum frequency at which echoes are recorded on ionograms. Since $f_{\text{min}}$ is routinely tabulated by a number of ionospheric stations, its values are widely available.

2.34 PARTIAL REFLECTIONS

The partial reflection method, described by Gardner and Pawsey (1953), relies on weak reflections from small discontinuities of refractive index in the lower ionosphere. Short pulses at a wave frequency of 2–3 MHz are transmitted, and the amplitude of the echo measured as a function of delay time, which is related to reflection height. The "ordinary" and "extraordinary"
modes of the echo are received separately. Because of the weakness of the echoes, a low-noise site is required for this experiment.

For either mode, the received amplitude \( A \) depends on the reflection coefficient \( R \) and the total attenuation along the two-way path between the ground and the reflection height \( h \). Hence

\[
A \propto R \exp \left\{ - 2 \int_0^h \kappa \, dh \right\}
\]  

(221)

Using subscripts \( o \) and \( x \) to distinguish "ordinary" and "extraordinary" modes, the ratio of the received amplitudes is

\[
\frac{A_x}{A_o} = \frac{R_x}{R_o} \exp \left\{ - 2 \int_0^h (\kappa_x - \kappa_o) \, dh \right\}
\]

(222)

The quantities \( R \) and \( \kappa \) may be evaluated from the generalized magnetoionic theory. They depend on the electron concentration and the monoenergetic collision frequency \( v_M \) introduced in Sec. 2.31 (see Belrose and Burke, 1964). When a wave is reflected from a small discontinuity \( \Delta \mu \) of refractive index, the reflection coefficient \( R \propto \Delta \mu \). It is found that the ratio \( R_x/R_o \) is a function of \( v_M \) and wave frequency, but not of electron concentration. The same applies to the "absorption coefficient per unit electron concentration," \( K = \kappa/N \), as may be seen from Eq. (217). Hence, Eq. (222) may be written in the form

\[
\ln \left( \frac{A_x}{A_o} \right) - \ln \left( \frac{R_x}{R_o} \right) = - 2 \int_0^h (\kappa_x - \kappa_o) \, dh = - 2 \int_0^h [K_x(h) - K_o(h)] N(h) \, dh
\]

(223)

To obtain the electron profile \( N(h) \), the collision frequency profile \( v_M(h) \) must be assumed known from independent data, and used to calculate \( K_x, K_o, \) and \( R_x/R_o \). On differentiating Eq. (223), \( N(h) \) is obtained from the experimental values of \( A_x/A_o \) versus height.

Some measurements of the ratio \( A_x/A_o \) are shown in Fig. 13, together with the derived \( N(h) \) profile. At heights below about 60 km, the absorption is small and the right-hand side of Eq. (223) is approximately zero, so that \( A_x/A_o \approx R_x/R_o \). Since \( R_x/R_o \) depends on \( v_M \), Belrose and Burke (1964) were able to determine \( v_M \) from measured values of \( A_x/A_o \). According to theory the ratio of \( R_x/R_o \) increases monotonically upwards; but \( A_x/A_o \) reaches a peak and then decreases because of the rapid upward increase of differential absorption, given by the right-hand side of Eq. (223). Piggott and Thrane...
II. IONOSPHERIC MEASUREMENTS

Fig. 13. (a) Observed amplitude ratios \( |A_x/A_o| \) of partially reflected extraordinary and ordinary waves (open circles); the bars show upper and lower quartile errors of the data points. The reflection coefficient ratio \( |R_x/R_0| \) is calculated from an assumed collision frequency profile. (b) The deduced electron concentration profile. The measurements were made in August 1962, near Tromsø, Norway (69°N, 19°E) [O. Holt, Norwegian Defence Research Establishment, Kjeller, Norway].

(1966a) have pointed out that partial reflections may be caused by gradients of \( v \) as well as by gradients of \( N \), and have considered how the application of the theory to the observations is thereby affected.

2.35 WAVE INTERACTION

The phenomenon of wave interaction or cross-modulation depends on the heating of ionospheric electrons by a strong radio wave, customarily known as the “disturbing signal.” Because the collision frequency depends on electron temperature, the heating affects the absorption of any other radio wave passing through the disturbed region. Such a wave is customarily known as the “wanted signal.”

The phenomenon of wave interaction was described by Bailey and Martyn (1934), and received considerable attention in subsequent years, mostly in connection with commercial radio transmissions. A survey of the subject was given by Huxley and Ratcliffe (1949). A method based on pulsed radio signals, specifically intended as a tool for investigating electron concentration and collision frequency in the lower ionosphere, was described by Fejer (1955), and subsequently used by Fejer and Vice (1959) and Barrington and Thrane (1962). The “disturbing” and “wanted” transmitters are both pulsed, their pulse repetition rates being in the ratio 2:1 (typically 50 and 25 pulses s\(^{-1}\)). The timing of the pulses is arranged so that every alternate “wanted”
pulse, after reflection from the E layer, meets a "disturbing" pulse in the D region at a known height, which can be varied. The other "wanted" pulses do not encounter a "disturbing" pulse, so the experiment depends on measuring the difference in amplitude of the two interlaced trains of "wanted" pulses. The pulse repetition rates are sufficiently high to prevent errors due to fading.

In Fig. 14, $T_1$ and $T_2$ are the "wanted" and "disturbing" transmitters respectively, and $R$ is the receiver. At the instant shown, the downcoming "wanted" pulse is just about to meet the upgoing "disturbing" pulse at the height $h_o$. In the diagram $T_1$, $T_2$, and $R$ are shown separated for clarity, but in reality they may all be at one location, so that the pulses travel vertically. The height resolution obtained depends on the sum of the pulse lengths of the "wanted" and "disturbing" signals.

![Figure 14](image)

**Fig. 14.** Illustrating the measurement of electron concentration and collision frequency by the radio-wave interaction technique. A pulse from the "disturbing" transmitter $T_2$ interacts with alternate pulses from the "wanted" transmitter $T_1$ at the altitude $h_o$, the "wanted" pulse having been reflected from the E layer. A region of slightly enhanced collision frequency follows in the wake of the "disturbing" pulse, where the electron temperature is increased by $\Delta T$ and then recovers exponentially as indicated by the curve.

When a radio wave travels through the D region, it forces the electrons to oscillate at the wave frequency; this ordered energy is converted to random energy, or heat, by collisions with the neutral particles. If a fraction $G$ of the excess electron energy is lost at each collision, the electron temperature will return to its undisturbed value in a time $\tau \sim (Gv)^{-1}$, which is much shorter than the interval between successive pulses. Thus, the excess electron temperature $\Delta T$ will diminish behind the pulse in a distance $\Delta h \sim c\tau$. This is indicated for the "disturbing" pulse in Fig. 14; with a typical "disturbing"
transmitter power of 50 kW, the electron temperature may be increased by 20%, leading to significant increases in collision frequency.

The effect of increased collision frequency on the “wanted” signal varies with height. We recall from Sec. 2.31 that the absorption coefficient $\kappa$ is proportional to $Nv$ if $v \ll \omega$ and to $N/\nu$ if $v \gg \omega$ (here for simplicity we assume $\omega >> \omega_L$, though this condition is only moderately well satisfied in practice). For experiments at about 2 MHz, the level where $v = \omega$ lies at about 70 km. Above this level $v < \omega$, and an increase in $v$ leads to increased absorption, i.e., positive cross-modulation. Below this level $v > \omega$, and an increase of $v$ leads to decreased absorption, i.e., negative cross-modulation. At the level where $v = \omega$, there is no cross-modulation. If this level can be determined experimentally, it provides a useful determination of $v$.

Since the experiment basically depends on the variation of collision frequency with temperature, an adequate theory of this effect must be used. The analysis is carried out in terms of the “monoenergetic collision frequency” $v_M$, assuming a Maxwellian distribution of electron velocity. Allowance should also be made for the actual waveforms of the “disturbing” and “wanted” pulses.

Actual experimental results do not contain enough information to determine all the parameters. Barrington and Thrane (1962), by considering also the data obtained from contemporary partial reflection experiments, were able to place fairly strict limits on the collision frequency and the “electron cooling parameter” $G$ (found to be of order $2 \times 10^{-3}$). The measured interaction effects are extremely weak, being generally less than $10^{-3}$ of the “wanted” signal amplitude. Consequently, considerable averaging of the data is required, and the resulting electron density distributions are limited to resolutions of about 5 km in height and half-an-hour in time. The $N(h)$ and $v(h)$ profiles, measured at Kjeller, Norway (60°N), have been found to be reasonably consistent with those given by other propagation methods (Piggott and Thrane, 1966b). We shall compare the D region $N(h)$ profiles obtained by different techniques in Sec. 2.5, Fig. 18.

2.4 Propagation Experiments Using Rockets and Satellites

Soon after the World War II was concluded, rockets were launched into the lower ionosphere for research purposes. Radio propagation experiments through this region were among the first to be attempted. Both the Doppler shift and the polarization of radio waves are affected by the presence of
2.4 PROPAGATION EXPERIMENTS USING ROCKETS AND SATELLITES

ionization, and these effects were measured to determine the vertical distribution of electrons and ions (Sec. 2.41).

By 1956 it became possible to reflect VHF signals from the moon, and the measured polarization rotation (the Faraday effect) permitted calculation of the total content of electrons in the ionosphere (Sec. 2.42). Only a year later, in 1957, the first Sputniks were launched and both Faraday and Doppler effects on the transmitted signals from satellites have been widely studied since then (Sec. 2.43). A major advance in 1962 was the initiation of topside sounding with Alouette, as we have discussed in Sec. 2.23.

In 1964, the geostationary satellite Syncom 3 was positioned over the Pacific Ocean. As described by Garriott and Little (1960), many advantages accrue when the line of sight between the satellite and ground stations remains fixed. These experiments are discussed in Sec. 2.44. In 1965, precise Doppler tracking of Mariner 4, as it was occulted by Mars, permitted a measurement of another planet's ionosphere for the first time. In the next two years, 1965-1967, radio transmissions to Pioneers 6 and 7 from the earth have provided measurements of the average interplanetary electron concentration (Sec. 2.45).

2.41 ROCKET EXPERIMENTS

Propagation experiments using V-2 rockets were successful as early as 1947 (Seddon, 1953). The basic technique used in these early flights is still accepted as the standard for comparison of \( N(h) \) measurements in many rocket flights today. Seddon employed two harmonic frequencies (about 4.3 MHz and its sixth harmonic), radiated from the vehicle and received at one or more stations on the ground. The rocket flights were nearly vertical, with the ground stations beneath the trajectory, which is an essential difference between the results obtained here and those for satellite orbits.

This experiment yields the electron concentration profile \( N(h) \) up to near the peak altitude of the rocket flight. To obtain this result the phase path lengths for one (or both) of the magnetoionic modes at the two frequencies are compared, from which the local electron concentration can be determined. To see how this is done, we write the expression for the phase path (in wavelengths) at either frequency as

\[
P = \int \frac{ds}{\lambda_m} = \int \frac{\mu \, ds}{\lambda_0}
\]

in which \( \lambda_m \) is the wavelength in the ionospheric medium, \( \lambda_0 \) is the free-space
wavelength, \( \mu \) is the refractive index, and the integral extends along the ray path from the ground station vertically to the rocket.

The higher frequency (subscript 1) is selected so that \( \mu_1 \approx 1 \) to sufficient accuracy, although a correction can be included if desired. At the lower frequency (subscript 2) we take \( X \ll 1 \) and \( Y_L = 0 \) in Eq. (204), thereby obtaining

\[
\mu_2 \approx 1 - \frac{40.3 N}{f^2} \quad (225)
\]

(m.k.s. units), although again corrections can be included for electron collisions and the magnetic field. In the absence of the ionosphere, the phase path lengths would be exactly in harmonic relation, but the presence of electrons slightly decreases the phase path length at the lower frequency. To compare the relative lengths, we multiply the lower frequency by the harmonic ratio \( f_1/f_2 = n \) and take the difference in the two path lengths; thus

\[
\Delta P = P_1 - nP_2 = (40.3n^2/cf_1) \int N \, ds \quad (226)
\]

Now the Doppler shift at either frequency (or for either mode) is just the rate of change of phase path in wavelengths. Similarly, the rate of change of \( \Delta P \), which we call “differential Doppler frequency,” is the difference in their Doppler shifts, referred to the higher frequency,

\[
\Delta f = \frac{d(\Delta P)}{dt} = \frac{40.3n^2vN}{cf_1} \quad (227)
\]

The quantity \( \Delta f \) is measured, \( v \) is the known vertical velocity of the rocket, and so the electron concentration profile can be determined.

This method has been used in other countries (Gringauz, 1958, 1961). A variation of it makes use of the Faraday rotation of a plane polarized wave, which is caused by the difference in phase path between the “ordinary” and “extraordinary” components, as described in Sec. 2.12 (Bauer and Jackson, 1962).

The electron distribution in the lower ionosphere has been measured by the differential absorption technique (Seddon, 1958; Kane, 1959, 1962), using a frequency of 7.75 MHz. For a wave propagated vertically to the ground from a rocket at height \( h \), the amplitude ratio of the two magnetoionic components varies as

\[
A_x/A_o \propto \exp \left[ \int_0^h (\kappa_o - \kappa_x) \, dh \right] \quad (228)
\]
in which the absorption coefficients $k_\omega, k_x$ are given by Eq. (206). When data on $A_x/A_\omega$ are combined with the phase path measurement described above, it is possible to obtain both $N(h)$ and $v(h)$ in the D region. Bowhill and Smith (1966) have described rocket experiments in which a combination of Faraday rotation, differential absorption and a direct measuring probe are used to obtain the complete $N(h)$ profile throughout the D region; Mechtly et al. (1967) also measured $v(h)$ above 80 km.

In another technique a rocket carrying three orthogonal receiving antennas is used to measure the electric wave field produced by a ground transmitter, operating at about 200 kHz. Full wave propagation calculations (Sec. 2.32) are carried out with various assumed model ionospheres, and a best fitting $N(h)$ profile deduced (Hall and Fooks, 1965). We shall illustrate results of these experiments, compared with those of direct measurement techniques, in Sec. 2.5.

2.42 MOON ECHOES

The first measurements of the electron content of the ionosphere were made by Browne et al. (1956) and extended by Evans (1957). They reflected radio waves at 120 MHz from the moon and measured the total polarization rotation along the two-way path. Two closely spaced frequencies were used to resolve the possible ambiguity of extra half rotations. The total Faraday rotation, in radians, for a one-way passage through the ionosphere is obtained by integrating Eq. (205); in m.k.s. units it is

$$\Omega = \frac{2.36 \times 10^4}{f^2} \int_0^\infty (B \cos \theta \sec \chi) N \, dh$$  \hspace{1cm} (229)

where $B$ is the local magnetic flux density (tesla or Wb m$^{-2}$), $\theta$ is the angle between radio wave normal and magnetic field direction, and $\chi$ is the angle between wave normal and vertical. Since $B$ decreases inversely as the cube of the geocentric distance, the integral is heavily weighted near the earth. The electron concentration also decreases rapidly with altitude, so that for most purposes, the integral can be considered to provide values of electron content below about 1000 km. The term $(B \cos \theta \sec \chi)$ is usually reasonably constant below 1000 km, and if evaluated near the centroid of the electron concentration distribution (about 400 km), it can be taken outside the integral sign.

Moon echo measurements have been largely superseded by similar experiments with satellites, both at geostationary altitudes and in lower orbits. This is partially due to the major effort in labor and equipment required.
to obtain the data, whereas satellite signals are readily recorded at locations all over the globe.

2.43 Satellite Transmissions

Satellite studies rely on the same fundamental equations (224) and (229) used for rocket and moon echo experiments. Turning first to the case of Faraday rotation, we see that when the ray path becomes perpendicular to the magnetic field, \( \cos \theta \) in Eq. (229) approaches zero and, therefore, the angle \( \Omega \) also approaches zero. Although the quasi-longitudinal approximation used in (229) is not valid at precisely \( \theta = 90^\circ \), it is usable to within a few degrees of this value, and it is sufficiently accurate to assume \( \Omega \approx 0 \) at \( \theta = 90^\circ \).

For a station near the magnetic equator, the condition \( \theta = 90^\circ \) is satisfied near the time when a satellite passes east or west of the station. For stations at mid-latitudes, the surface on which \( \theta = 90^\circ \) looks rather like an inverted bowl, resting on the globe on the pole side of the observing station. This implies that \( \Omega \approx 0 \) at a time when the satellite is observed on the poleward side of the observing station. At geomagnetic latitudes above about 45° to 50°, it becomes impossible to achieve \( \theta = 90^\circ \) at ionospheric altitudes.

A linearly polarized signal radiated from a satellite to a ground station undergoes a variable polarization rotation during the course of a passage. This produces strong fading when observed with a linear antenna on the ground. In fact, the rotation rate (or fading rate) is constant, to a first order (Bowhill, 1958). Since at mid-latitudes we can usually identify one point on the fading record where \( \Omega \approx 0 \) at \( \theta = 90^\circ \), we may count fades (each is a half-rotation change in \( \Omega \)) to any other time, and compute the electron content from Eq. (229) (Garriott, 1960; Blackband, 1960). At the higher latitudes, particular care must be exercised to avoid confusion with refraction and path splitting effects. Some help is provided by noting that the rotation angle is clockwise when the direction of propagation is parallel to the magnetic field and counterclockwise when antiparallel (Garriott, 1960). A number of other authors have made extensive use of Faraday rotation observations, including Yeh and Swenson (1961), Lawrence et al. (1963), and Bhonsle et al. (1965).

The differential Doppler technique has also been widely used with radio transmissions from satellites (Ross, 1960; de Mendonça, 1962). The experimental method is basically identical to that employed with vertically ascending rockets, i.e., the Doppler shifts from two harmonic transmitters in the satellite are received and compared on the ground. However, in this case
it is the electron content below the satellite altitude and not the local electron concentration at the satellite that is measured, just as is true in the case of Faraday rotation measurements from a satellite. This can be understood in the following way.

For simplicity, let us consider only a circular orbit and a spherically stratified ionosphere. The Doppler shift imposed on a satellite signal at any time is just the time derivative of the phase path length, Eq. (224). Within the ionosphere, the index of refraction \( \mu \) varies all along the ray path. Any change in the local electron concentration at the satellite produces only an infinitesimal change in the phase path integral and therefore in the Doppler shift, and its effect can be made as small as wished by simply restricting the electron concentration change to a very thin shell at the satellite altitude. However, as the satellite orbit becomes eccentric, the local concentration does begin to contribute (usually very slightly) to the Doppler shift.

Horizontal gradients in the ionosphere and, to a smaller extent, altitude variations of the satellite also affect the observed polarization rotation rate and differential Doppler frequency. They can therefore lead to some error in the computed values of electron content. A combination of Faraday and Doppler techniques described by Burgess (1962), permits these difficulties to be overcome. Writing (229) as

\[
\Phi = K_1 \langle B \cos \theta \rangle_{av} \int N \, ds
\]  

(230)

(where the \( \text{av} \) notation denotes an average along the ray path) we may take its time derivative as

\[
\frac{\partial \Phi}{\partial t} = K_1 \left\{ \langle B \cos \theta \rangle_{av} \left( \frac{\partial}{\partial t} \right) \int N \, ds + \int N \, ds \cdot \left( \frac{\partial}{\partial t} \right) \langle B \cos \theta \rangle_{av} \right\} \tag{231}
\]

We may also obtain the time derivative of (226) to find the differential Doppler frequency for a satellite orbit,

\[
\Delta f = K_2 \frac{\partial}{\partial t} \int N \, ds \tag{232}
\]

Note that we cannot simplify this to the result of (227) because the ray path changes its location with time, unlike the vertical rocket trajectory. However, (232) can be inserted into (231), and when \( \Delta f \) and \( \partial \Phi / \partial t \) are measured, the value of electron content \( \int N \, ds = \langle \text{sec} \chi \rangle_{av} \int N \, dh \) can be calculated.
This simple derivation demonstrates the utility of the "hybrid" method. For actual computation, a somewhat modified approach given by Golton (1962) or de Mendonça and Garriott (1962) is probably more useful. A comparison of the accuracy of various methods has been made by Garriott and de Mendonça (1963).

These methods also rely on simplifying assumptions concerning the refractive index and the coincidence of ray paths for the two magnetoionic modes. Rather complicated computer programs have been devised (Lawrence and Posakony, 1961) to circumvent these difficulties and have provided insight into the magnitude of the approximations involved. In addition, Ross (1965) has developed formulas which carry these approximations to second order and provide a significant improvement in accuracy without excessive complexity.

2.44 GEOSTATIONARY SATELLITES

When satellites are placed in a geostationary orbit (circular, in the equatorial plane with a period of exactly 24 hr, which occurs at a geocentric distance of about 6.6 earth radii), propagation studies are greatly facilitated. The ray path to the observer is fixed in earth coordinates, so that both Faraday rotation and the apparent "Doppler shift" observations provide a direct measurement of time variations of electron content. The accuracy of the measurement is much better than that obtainable with lower altitude satellites. The Faraday rotation method provides absolute values accurate to a few percent at the relatively high frequency of 136 MHz and changes of electron content as small as 0.1% are detectable. Even these accuracies may be improved upon by an order of magnitude, when a suitable two-frequency harmonic transmitter is placed in orbit. An example of some of the first curves showing the diurnal variation of electron content, observed at the University of Hawaii from the transmissions of Syncom 3 is given in Fig. 15 (Garriott et al., 1965).

When the electron content is divided by $N_mF2$, found from an ionogram, the resulting "slab thickness" is a first-order measure of F-region scale height. The accuracy of the total electron content measurement permits the incident flux of solar extreme ultraviolet radiation to be calculated from the rate at which ionization builds up at sunrise. Transient effects can be clearly seen, such as those produced by solar flares. For several flares, estimates of X-ray and EUV output have been made (Garriott et al., 1967). Many of these topics will be more fully discussed in the appropriate sections of Chapters V and VI.
2.45 SPACE PROBES AND PLANETARY "FLY-BYS"

The two-frequency experiment is ideally suited to measure changes in integrated electron concentration. In the case of satellites or space probes, the observations are usually made intermittently, whenever the spacecraft is visible, and the absolute value of the integrated concentration must also be established. This is most conveniently done by modulating the two carriers at some lower frequency and comparing the phase of the demodulated signals at the receiver (Eshleman et al., 1960). Such an experiment, with the receiver in the space probe, has been carried out in Pioneers 6 and 7, now in orbit about the sun. When the space probes are far enough away from the earth, the interplanetary contribution can be separated from the near-earth electron content, revealing the average electron concentration in interplanetary space. Values averaged about 8 electrons cm\(^{-3}\) in the early analysis.
II. IONOSPHERIC MEASUREMENTS

Fig. 16. Electron number density (concentration) profile in the Martian ionosphere, deduced from observations made aboard the Mariner 4 "fly-by" [Fjeldbo et al. (1966)].

(Staff, 1966), with more recent results showing fluctuations beyond the limits of 5 and 20 cm\(^{-3}\) depending on the state of the solar wind.

A remarkable observation was made in the 1965 fly-by of Mars using Doppler tracking data (Kliore et al., 1965). After a very precise determination of the trajectory of Mariner 4, the very small refraction and phase velocity changes in the Martian ionosphere and atmosphere could be detected as the space probe was occulted by the planet. These amounted to only 10 and 30 wavelengths, respectively, although a total phase change of some \(3 \times 10^{11}\) wavelengths occurred during the transit of the spacecraft to the vicinity of Mars. The electron concentration profile shown in Fig. 16 was deduced (Fjeldbo et al., 1966). In addition to measurements of the Martian ionosphere, it was possible to estimate the surface pressure and scale height of the neutral atmosphere as the ray path approached the surface of the planet. A surface pressure of 5 mb and a scale height of 9 km were deduced. From these results, various models of the Martian atmosphere have been proposed by Fjeldbo et al. and by other workers.

2.5 Direct Measurements Using Rockets and Satellites

Instruments have been designed for the direct measurement of many variables of the atmosphere and ionosphere. When mounted in satellites
they are most useful for long-term monitoring above about 200 km. At lower altitudes, satellite orbits rapidly decay and rockets are essential. Rockets are also required when vertical profiles need to be measured, for correlation with special events such as aurora or solar eclipses and for other purposes.

The parameters most frequently sought are the vertical distribution of each of the neutral and charged constituents of the atmosphere, and the temperatures of the electrons and the ion and neutral species. Rocket flights are obviously advantageous here, although satellites have shown very interesting geographical variations in the ion species and temperatures. The methods of measuring neutral composition and mass densities have been discussed in Sec. 1.8, so we will next turn to the ionized component.

At the lower altitudes of the D region, a quadrupole mass spectrometer has been employed by Narcisi and Bailey (1965) to sort out the concentrations of each constituent. In this device, the ambient positive ions are accelerated through a small aperture into a “drift zone,” with four longitudinal rods surrounding the ion beam. When a large, r.f. voltage (as much as 310 V at 6 MHz) is applied between the rods, the ion trajectories acquire a transverse oscillatory component in addition to their longitudinal drift. An analysis shows that for given applied potentials, only ions of a particular mass-to-charge ratio have bounded trajectories; lighter or heavier ions experience growing oscillations and are collected at the sides before reaching the ion detector at the end of the “drift zone.” As the potentials are varied, different ions are collected, the collector currents being proportional to the ion concentrations.

An experiment complementary to the mass spectrometer is a “spherical electrostatic analyzer” flown by Sagalyn and Smiddy (1964). This device is basically a spherical collector, surrounded by a concentric wire mesh grid. With the outer grid at vehicle potential and with a negative voltage on the collector, the current is proportional to the total positive ion concentration (all voltages are measured with respect to the vehicle potential). However, when the outer grid is biased near space potential and a small positive voltage is placed on the sensor collector, the current is proportional to the negative ion and electron concentration. These modes are usually used, but when the voltages applied to the outer grids are swept rapidly, the operation of the device becomes that of a Langmuir probe, from which measurements of vehicle potential, charged particle energy and temperatures can be made. Results of one rocket flight in 1963, in which both experiments were performed, are shown by Narcisi and Bailey (1965) for the D and lower E regions (Sec. 3.53).
Another type of mass spectrometer devised by Bennett (1950) has been widely employed in space applications. In this device, discussed in more detail by C. Y. Johnson (1958), a beam of monoenergetic ions traverses a "stage," or set of grids, in which an r.f. voltage is applied to the center grid. If the transit time across the stage, for ions of a particular mass, coincides with the period of the r.f. voltage, then the ion velocity is selectively increased. With three or more stages, sufficient mass resolution can be achieved. After the final stage, a retarding potential is applied to one grid, so that only those ions whose velocities have been sufficiently amplified will be collected. Devices of this type were used in the Soviet satellites Elektron 1 and 2 (Istomin, 1966) and in a number of rocket flights. A summary of the results of the available ion concentration measurements, made by C. Y. Johnson, is shown in Fig. 6. The figure applies to daytime solar minimum conditions.

A dumbbell-shaped electrostatic probe has been devised by Spencer et al. (1962) for the measurement of ion concentration and electron temperature. Their instrument is launched into the ionosphere and then ejected from the rocket to avoid any local perturbations. A saw-toothed waveform is applied to one end of the dumbbell and the resulting plasma current is measured as in a Langmuir probe.

A spherical ion trap has been launched several times in a similar manner by Nagy et al. (1963). Their device is constructed of two concentric spheres. The outer sphere (essentially a grid) has many holes, allowing ions and electrons to pass through it, to be collected by the inner sphere. The potential of the inner collector is then swept with respect to the outer grid into the range of both electron and ion saturation currents. A separation of these two contributions permits the ion concentration and the ion and electron temperature to be calculated. Their flights also included a cylindrical Langmuir probe for redundant electron temperature measurements. As expected, their results have shown substantial temperature differences for the electrons and ions, with \((T_e/T_i) \approx 2\) in the daytime F region. In addition to the results of Nagy et al. (1963), Brace et al. (1963) have compared high-latitude and mid-latitude observations in both the daytime and nighttime. Their measurements will be illustrated in Sec. 6.5, where the differences of ion and electron temperature are considered in more detail.

A number of ionospheric parameters were measured by the satellite Explorer VIII as early as 1960 (Bourdeau and Donley, 1964). In addition to
temperature measurements, an early indication of the presence of He$^+$ was obtained.

In 1962, the first UK-USA international satellite Ariel I was placed in orbit with ionospheric measurements as its principal objective. Many of the preliminary results from the experiments on Ariel I are reported in the Proceedings of the Royal Society. The vertical distribution of O$^+$, He$^+$, and H$^+$ ions were investigated at 400–1200 km altitude, and diurnal and latitude variations of their relative abundances were found (Sec. 5.51). Other experiments in Ariel I measured electron concentrations and temperature and the X-ray flux between 4 and 14 Å, showing very interesting flare effects.

A resonance rectification probe devised by Aono et al. (1962) has been flown on several Japanese rockets for electron concentration measurements. Analysis has shown, however, that the frequency of the observed resonance is not exactly at the local plasma frequency but occurs somewhat below it (Crawford and Harp, 1965; Buckley, 1967). Probe geometry and other factors must be considered in computing the electron concentration from the resonant frequency. Oya and Obayashi (1967) have made $N(h)$ measurements with a "gyro-plasma" probe, which determines the "hybrid" resonance frequency $(f_N^2 + f_H^2)^{1/2}$.

A number of direct measurement experiments were flown on the same rocket by Heikkila et al. (1967) for the purpose of intercomparison. Both spherical and cylindrical electrostatic probes were flown and found to be affected by the ion sheath, vehicle potential and the presence of large r.f. fields. A capacitance probe was similarly affected, although correction factors can presumably be found empirically. The self- and mutual-admittances of a sphere and rod pair were found to exhibit sharp resonances at the "plasma," "gyro," and "upper hybrid" frequencies (Sec. 2.23). These resonances appear to be characteristic of the surrounding undisturbed medium and are not affected by the sheath or other local perturbations.

Two good comparisons of electron profiles obtained by a variety of methods are shown in Figs. 17 and 18. Figure 17 (Bauer et al., 1964) compares two F-region $N(h)$ profiles obtained from a sounding rocket (one by a direct measurement and one by a propagation experiment) with an Alouette I topside sounding (Sec. 2.23) and an incoherent scatter profile (Sec. 2.6) obtained at almost the same place and time. Figure 18 (Van Zandt and Knecht, 1964) compares typical D-region profiles obtained by rocket-borne direct measurement and propagation experiments, and by the partial
II. IONOSPHERIC MEASUREMENTS

Fig. 17. Comparison of electron or ion density (concentration) profiles in the F region and topside ionosphere, obtained nearly simultaneously by different techniques, within the region 37°-42°N, 69°-74°W. The data are identified as follows (times are 75°W meridian time on 2 July 1963): ion trap, two-frequency c.w. propagation experiment, both carried by NASA rocket No. 8.14 (0921-0926 hours); Alouette topside sounder satellite (0922 hours); Lincoln Laboratory incoherent scatter (J. V. Evans) (0930-1044 hours) [after Bauer et al. (1964)].

reflection and wave interaction techniques described in Sec. 2.3. In this case, however, the experiments were carried out at different times and places, though the data were selected to represent normal conditions at an intermediate level of solar activity.

2.6 Incoherent Scatter

Conventional ionospheric sounding depends on the reflection of radio waves by an ionized gas, as discussed in Sec. 2.2. This reflection process depends on the collective behavior of electrons, which can be described (at least approximately) in terms of the refractive indices of the Appleton-Hartree equation. A second kind of echo, much weaker than the first, arises from scattering from irregularities or sharp gradients of ionization when the frequency of the radio waves exceeds the local plasma frequency. A third kind of echo, even weaker than the others, is due to the classical Thomson scattering of waves from individual electrons. This phenomenon has become known as "incoherent scatter" or "Thomson scatter"; the power returned is directly proportional to the electron concentration.

In his initial paper on this subject, Gordon (1958) discussed the conditions
under which ionospheric electrons scatter incoherently, so that the return is of the third kind. He found that incoherent scattering occurs if the electron mean free path exceeds the radio wavelength and exceeds the scale size of the irregularities that occur statistically in the electron gas, owing to random thermal motions; and concluded that these conditions hold at meter wavelengths at heights above 100 km. Given a sufficiently powerful radar, the method could be used to measure electron concentrations throughout the F region and beyond, to a distance of one earth radius or more.

Experiments first carried out by Bowles (1961), with a 41 MHz radar, at 4–6 MW peak power and a 100 μs pulse length, yielded measurements of electron concentration from about 100 to 700 km altitude. A high-powered
scatter radar has been built at Jicamarca in Peru, close to the magnetic equator. It operates at 50 MHz and has a maximum peak power of 4 MW feeding a large array of dipoles covering a square of side 288 m. Such powerful equipment can be used for purposes other than incoherent scatter, such as investigation of ionospheric irregularities by coherent scatter, and even planetary radar. Examples of electron profiles to 7000 km altitude have been given by Bowles (1963). Another very large installation exists at Arecibo, Puerto Rico, using a reflector 300 m in diameter situated within a natural hollow. It operates at 430 MHz with 2 MW power. Other installations have been built in the auroral zone at Prince Rupert, Canada, and at mid-latitudes in the USA, UK, and France.

Most scatter systems use the pulse radar technique, the height from which scattering is measured being selected by varying the interval between transmission and reception of the pulses. The height resolution then depends on the pulse length, and is typically 20 km or more. In the French installation, however, the transmitter and receiver are separated by 300 km, and the height of observation is determined geometrically by setting the two antenna beams to intersect at the desired height. The resolution varies with height, from 3 km at 100 km to 30 km at 500 km. In this system c.w. signals (75 kW at 935 MHz) are used instead of pulses, which facilitates the spectrum measurements and enables less power to be used than is needed for a pulse system of comparable sensitivity.

Following the first observation, a number of theoretical papers on incoherent scatter were published (Fejer, 1960; Salpeter, 1960; Dougherty and Farley, 1960). These are largely concerned with the broadening of the spectrum of the returned signal due to thermal motions of the scattering particles. It is found that the breadth of the spectrum depends on the ion (rather than the electron) thermal velocity, and that the spectrum does not have a simple Gaussian form. In fact the scattering is not fully incoherent, but depends on the weak statistical fluctuations of electron concentration; in this case, both the electron and ion thermal velocities (or temperatures $T_e$ and $T_i$) are involved.

The shape of the spectrum of the scattered signal is influenced by the inequality of $T_e$ and $T_i$; by the existence of more than one kind of positive ion, such as $O^+$, He$^+$, and H$^+$ in the upper F region; and by the presence of the geomagnetic field (though the last does not seem too important in practice). In pulsed systems, the shape is also influenced by the finite spectrum of the transmitted pulses. In practice these different influences lead to some ambiguity, and it may not be possible to determine $T_e/T_i$ if more than one ion
species is present, nor to determine the ion composition unless \( T_e / T_i \) is known. These effects have been discussed in numerous papers, such as those of Hagfors (1961), Farley et al. (1961, 1967), Salpeter (1963) and Moorcroft (1964).

In addition to the spectral measurements, it is also possible to observe Faraday rotation of the scattered signal (Millman et al., 1964; Farley, 1966). From our previous discussion (Sec. 2.42), we can see that the angle of rotation depends on the quantity \( \int N B \, dh \), evaluated up to the height of scattering. From a series of rotation measurements at different heights the \( N(h) \) profile can be found. Given proper calibration, it is also possible to measure the total power of the signal scattered from height \( h \). The ratio of received to transmitted power is

\[
\frac{P_r}{P_t} \propto \frac{N \sigma_e}{h^2 \left( 1 + \frac{T_e}{T_i} \right)}
\]

where \( \sigma_e \) is the Thomson cross-section, related to the classical electron radius \( r_e \) as follows:

\[
\sigma_e = 4 \pi r_e^2 = 1.00 \times 10^{-28} \quad [\text{m}^2]
\]

\[
r_e = \frac{e^2}{4 \pi \epsilon_0 m c^2} = 2.82 \times 10^{-15} \quad [\text{m}]
\]

By this means Greenhow et al. (1963) investigated the diurnal variation of \( T_e / T_i \), using ionosonde data for \( N \). Absolute calibration of scatter systems is difficult, however, and often ionosonde determinations of \( N_m F_2 \) are used to calibrate the observed \( N(h) \) profiles.

Besides the "total power" and "spectral shape" methods of investigating ionospheric temperature, further information is obtainable from the shape of the \( N(h) \) profiles at heights well above the F2 peak (Van Zandt and Bowles, 1960). At such heights diffusive equilibrium is assumed to exist (as we shall discuss in detail in Sec. 4.33), in which case the gradient \( d(\ln N)/dh \) is related to the ratio \( M_\perp/(T_e + T_i) \), \( M_\perp \) being the mean molecular mass of the ions. Since this relation applies strictly to the electron distribution along a geomagnetic field line, it cannot be used near the magnetic equator.

Although the analysis of incoherent scatter data is very complex, the method is extremely powerful, since it makes possible the study of the diurnal variations of \( N, T_e, T_i \), and ion composition over a wide range of height. The most complete published set of data were obtained at Millstone Hill, U.S.A. (latitude 43°N) during sunspot minimum (Evans and Loewenthal, 1964; Evans, 1967c). It is possible to investigate temperature and ion composition even in the F1 layer and upper E layer (Carru et al.,...
1967; Evans, 1967b). We shall refer to the results of these experiments in several places later in the book.

The scatter technique is also useful for measuring ionospheric motions, since a detectable Doppler shift in the echo can be caused by motions of the ionospheric electrons in the line of sight, as observed in the F region by Carru et al. (1967). Using a beam inclined to the vertical, Balsley (1966) has found a day-night reversal of motion in the equatorial electrojet (Sec. 7.43), though in this case the observations are not interpreted in terms of a simple Doppler shift.

Another proposed application of high-power radar equipment may be mentioned here. This is the artificial heating of electrons in the F region. Farley (1963a) shows that using $\frac{1}{2}$ MW at 50 MHz, the thermal expansion due to the heating might produce small but probably detectable changes in the electron profile. If a lower frequency, near $f_0$F2, were employed, it should be possible to increase $T_e$ by several hundred degrees Kelvin and produce appreciable reductions of $N_m$F2 with average powers as small as 100 kW.
III

PHOTOCHEMICAL PROCESSES IN THE IONOSPHERE

3.1 The Balance of Ionization

We now turn to consider the physical processes which control the ionosphere. In this chapter and the next, we present a fairly simple theory which we can compare with the behavior of the actual ionospheric layers in Chapter V.

The processes can be divided into two broad categories: those that result in production or destruction of ionization, and those that result in movement of ionization. The terms “photochemical” and “transport” serve as convenient, though not ideal, labels for these two categories. Later, we shall see that the relative importance of these categories varies with height; photochemical processes dominate the lower ionosphere (D and E regions), and we shall find that the F2 layer lies at a level of transition between the “photochemical” and “transport” dominated regimes in the ionosphere.

The principal production process for the creation of ion-electron pairs is generally accepted to be the absorption of solar extreme ultraviolet (EUV) and X-ray radiation, at least in low and mid-latitudes. Photons with energies greater than about 12 eV can ionize one or more of the major atmospheric constituents. This process also provides the heat input which is necessary to maintain the high temperatures of the thermosphere, discussed in Sec. 1.3. The excess energy of the photon is transformed into the kinetic energy of the ion-electron pair, and the remainder of the ionization energy is transformed eventually to heat upon recombination. We discuss photoionization, and the associated theory—the Chapman equations—in Sec. 3.2 and 3.3.

At high latitudes and during magnetic storms (and perhaps at other times), ions and electrons are produced by collisions between energetic charged particles precipitated into the atmosphere and the neutral molecules. The
III. PHOTOCHEMICAL PROCESSES IN THE IONOSPHERE

importance of this "corpuscular ionization" is not yet well established and is a topic of considerable current interest. Finally, if negative ions should be formed in the lower ionosphere by an attachment process, the electrons can be released by photodetachment which provides another mechanism for electron production.

The important loss processes may be summarized as atomic ion and electron (radiative) recombination; molecular ion and electron (dissociative) recombination; and, in the lower ionosphere, the attachment of electrons to neutral molecules. We discuss these processes in Sec. 3.4.

It is usual to form an equation of "continuity" or "balance," whose terms represent the effects of the various processes which alter the electron concentration $N$. Continuity equations can also be written for either the positive and negative ions or, indeed, for any constituent whose concentration is subject to change. Within a cell of unit volume, we have:

\[ \text{[Rate of change of electron concentration]} = \text{[Gain by production]} - \text{[Loss by destruction]} - \text{[Change due to transport]} \]

If the transport processes result in a net drift velocity $V$, then the change due to transport is the divergence of the flux $NV$. Using symbols $q$ and $l$ to represent production and loss, we can write:

\[ \frac{\partial N}{\partial t} = q - l(N) - \text{div}(NV) \quad (300) \]

In the ionosphere below 200 km two simplifications can often be made. Transport processes are not very important and, if they are neglected entirely, a "photochemical" equation results, containing only the one derivative $\frac{\partial N}{\partial t}$. Furthermore, the "time constant" associated with the loss term $l(N)$ may be so short that $\frac{\partial N}{\partial t}$ is much smaller than the other terms, so that the "photochemical equilibrium" equation $q = l(N)$ is adequate. This is generally the case in the D, E, and F1 layers by day, except for rapidly varying phenomena such as eclipse effects. Transport can then be included as a small perturbation if required. This situation is referred to as a "photochemical regime." We study it with the aid of the continuity equation, for the D region (below about 90 km) in Sec. 3.5, and for the E and F1 layers (roughly 90–200 km) in Sec. 3.6. Above about 250 km, the photochemical terms $q$ and $l$ no longer dominate the continuity equation (300), so that at these heights a "transport regime" exists (although, in contrast, the distributions of neutral constituents are dominated by diffusive transport at a much lower height, namely the turbopause region around 100 km, as we
3.2 THE BASIC THEORY OF PHOTOIONIZATION

saw in Sec. 1.4). Finally, we consider the relation between ionospheric pro-
cesses and optical airglow emission in Sec. 3.7.

3.2 The Basic Theory of Photoionization

Before we consider the complex processes of electron production in the
real ionosphere, it is useful to develop the basic theory of photoionization
in the atmosphere. Quite a good description of many features of the iono-
sphere can be obtained from this theory, which was developed by Chapman
(1931a,b), though some similar equations had been given previously by
Hulburt (1928).

3.21 EQUATIONS FOR THE ABSORPTION OF RADIATION AND PRODUCTION OF
IONIZATION

The theory considers the attenuation of solar radiation as it travels down-
wards through the atmosphere, and derives a general formula for the rate
of production of ionization, \( q \), as a function of height \( h \) and the sun’s zenith
angle \( \chi \). We may always assume that no exhaustion of the neutral atmo-
sphere occurs, as the fraction of gas ionized is always extremely small. We
start with some other simplifying assumptions, but later we consider how
to relax them and so generalize the theory. They are:

(i) The radiation is monochromatic, its photon flux being \( I(h) \).

(ii) The atmosphere consists of a single absorbing gas, its concentration
being \( n(h) \).

(iii) The atmosphere is plane and horizontally stratified. We extend the
theory to include the curvature of the earth in Sec. 3.23.

Later, in Sec. 3.22, we introduce the further simplification:

(iv) The scale height \( H \) is either independent of height or varies linearly
with height (gradient \( dH/dh = \Gamma \)).

Let \( \sigma \) be the cross section for absorption of radiation in the gas, and \( \eta \) the
ionizing efficiency, the number of photoelectrons produced per photon
absorbed. For a single gas illuminated by monochromatic radiation, the
probability per unit time that a given molecule absorbs a photon is \( I \sigma \), and
the probability per unit time of producing an ion pair is therefore \( \eta I \sigma \). We
may call this quantity the “ionization rate coefficient.” So the basic equation
for the rate of production per unit volume is

\[
q = I \eta \sigma n \tag{301}
\]

The attenuation of the radiation along the path depends on the absorption
coefficient per unit length, \( \sigma n \) [some derivations replace \( \sigma n \) by \( A \rho \), \( A \) being the mass absorption coefficient and \( \rho \) the density]. For an element \( ds \) of the path of the radiation, we can define an increment of optical depth \( \tau \) by the equations

\[
- \frac{dI}{I} = d\tau = \sigma n \, ds
\]

so that the intensity varies as

\[
I = I_\infty e^{-\tau}
\]

where \( I_\infty \) is the unattenuated flux at the top of the atmosphere. From Eqs. (301) and (302) we see that \( q = -\eta \frac{dI}{ds} \), which expresses the obvious fact that the production rate is proportional to the rate of attenuation of the radiation.

By simple geometry the variation of altitude along the path of the radiation is given by \( ds = -dh \sec \chi \). Thus, in a horizontally stratified atmosphere we can rewrite (302) as

\[
- d(\ln I)/dh = d\tau/dh = -\sigma n \sec \chi
\]

For a plane earth \( \sec \chi \) does not vary along the path; furthermore \( g \) is independent of height [because then \( R_e \to \infty \) in the gravity equation (110) of Sec. 1.22]. We can then use the result obtained in (109), namely that the integrated content of a column of gas, of unit cross section, above any height \( h \) is \( n(h) \, H(h) \), the product of the local gas concentration and the local scale height. Integrating (304):

\[
\tau(h, \chi) = \int_h^\infty \sigma n \sec \chi \, dh = n(h) \, H(h) \sec \chi
\]

Combining Eqs. (301), (303), (305) we have:

\[
q(h, \chi) = I_\infty \eta n(h) \, e^{-\tau(h, \chi)}
\]

The integrated production rate per unit column, \( Q \), can be obtained easily by integrating (306) with respect to \( \tau \) from the top of the atmosphere (\( \tau = 0 \)) to the bottom (\( \tau \to \infty \)). Using (304):

\[
Q = \int_0^\infty q \, dh = \int_0^\infty q \frac{dh}{d\tau} \, d\tau = \frac{I_\infty \eta n}{\sigma n \sec \chi} \int_0^\infty e^{-\tau} \, d\tau = I_\infty \eta \cos \chi
\]

To locate the peak of \( q \), it is simplest to take logarithms in (306) and set \( d(\ln q)/dh = 0 \). Since \( I_\infty \eta \sigma \) is a constant, we find from (304) that the peak occurs where

\[
\frac{1}{n} \frac{dn}{dh} = \frac{d\tau}{dh} = -\sigma n \sec \chi
\]
This shows that \( q \) is greatest at the level where the downward increase of gas concentration \( n \) just compensates for the increasing attenuation of the radiation, as measured by \( \tau \). We can reduce (308) by using the perfect gas law (100), \( p = nkT \), and the definition (102) of \( H \). Since we are assuming that \( g \) is constant and only one gas is present, we can take \( H \propto T \), so that

\[
-\frac{1}{H} = -\frac{1}{p} \frac{dp}{dh} = -\frac{1}{n} \frac{dn}{dh} + \frac{1}{T} \frac{dT}{dh} = -\frac{1}{n} \frac{dn}{dh} + \frac{1}{H} \frac{dH}{dh}
\]  

(309)

Substituting for \( (1/n) (dn/dh) \) in (308) and multiplying by \( H \), we find the peak of \( q \) occurs at the level where

\[
1 + (dH/dh) = \sigma nH \sec \chi = \tau
\]  

(310)

The altitude of this level increases with increasing \( \chi \). From Eqs. (306) and (310) the peak value of production, \( q_1 \), can be expressed in terms of the local values of scale height \( H \) and its vertical gradient \( \Gamma \):

\[
q_1 = \frac{\eta I_\infty \cos \chi}{H(1 + \Gamma)} e^{-(1 + \Gamma)}
\]  

(311)

This equation does not require either \( H \) or \( \Gamma \) to be independent of height.

We now introduce the "reduced height" \( z = \int (dh/H) \) defined by Eq. (105) of Sec. 1.21. A convenient choice of the reference height \( h_0 \), from which \( z \) is measured, is the level of unit optical depth \( (\tau = 1) \) for overhead sun. At this level, vertically incident radiation is attenuated to a fraction \( e^{-1} \) of its original intensity. From Eq. (305) we see that (using suffix "0" to denote any quantity evaluated at height \( h_0 \))

\[
1 = \sigma n_0 H_0
\]  

(312)

Since \( H \propto T \), Eq. (107) gives us

\[
e^{-z} = \frac{p}{p_0} = \frac{nH}{n_0H_0}
\]  

(313)

Again using Eq. (305), we obtain the general formula

\[
\tau = e^{-z} \sec \chi = (p/p_0) \sec \chi
\]  

(314)

so we see that optical depth varies linearly with pressure.

From Eq. (310), we can see that the peak of production and the level of unit optical depth coincide if \( \Gamma = dH/dh = 0 \). Otherwise, the production peak is found at a distance of \( \ln(1 + \Gamma) \) scale heights below the level \( \tau = 1 \) (approximately \( \Gamma \)'s scale heights if \( \Gamma \) is small). This holds for any \( \chi \).

We can now write the production function, Eq. (306), in terms of reduced
height:

\[ q(z, \chi) = \frac{\eta I_{\infty}}{e H(z)} \cdot \exp \left[ 1 - z - e^{-z \sec \chi} \right] \]  

(315)

When thus written, with a factor of \( e \) in the denominator, the expression in square brackets has a maximum value of zero when \( z = 0, \chi = 0 \).

3.22 Chapman's Production Function and the Chapman Layer

To obtain the classical Chapman formula for the production function, we make the additional assumption that the scale height \( H \) is independent of height (Chapman, 1931a). The level of unit optical depth then coincides with the production peak, at which \( z \) takes the value

\[ z_1 = \ln \sec \chi \]  

(316)

The peak rate is

\[ q_1 = q_0 \cos \chi \]  

(317)

where

\[ q_0 = \frac{\eta I_{\infty}}{eH} \]  

(318)

is the peak rate for overhead sun. The production function can be written in the alternative forms

\[ q(z, \chi) = q_0 \exp \left[ 1 - z - e^{-z \sec \chi} \right] \]  

(319)

\[ q(z, \chi) = q_1 \exp \left[ 1 - (z - z_1) - e^{z_1 - z} \right] \]  

(320)

showing that the function retains the same shape as \( \chi \) changes, but its amplitude is scaled by a factor \( \cos \chi \) and its peak is shifted to \( z = z_1 \) (Figs. 19 and 20).

If we now assume \( H \) to vary with height, with a constant gradient \( \Gamma \), but still assume \( g \) to be constant, we have

\[ H(h) = H_0 + \Gamma (h - h_0) \]  

(321)

\[ H/H_0 = T/T_0 = e^{\Gamma z} \]  

(322)

The variation of gas concentration is then given by

\[ n/n_0 = e^{-z(1 + \Gamma)} \]  

(323)

An expression for \( q(z, \chi) \) can be found merely by combining (322) with (315). Alternatively, if we now reckon reduced height \( z' \) from the level of peak production for \( \chi = 0 \) (instead of the level of unit optical depth, used as \( z = 0 \) in our previous formulas), we find (e.g., Nicolet, 1951)

\[ q(z', \chi) = q_0 \exp \left[ (1 + \Gamma) (1 - z' - e^{-z' \sec \chi}) \right] \]  

(324)
3.2 THE BASIC THEORY OF PHOTOIONIZATION

Fig. 19. Normalized Chapman production function versus reduced height \( z \), parametric in solar zenith angle \( \chi \) [Eq. (319)].

Fig. 20. Normalized Chapman production function versus solar zenith angle \( \chi \), for several values of reduced height \( z \) [Eq. (319)]. The broken line is the envelope, \( q_1/q_0 = \cos \chi \).
The reduced height of the peak, $z_1'$, and the peak production rate, $q_1$, are

$$ z_1' = \ln \sec \chi \quad (325) $$
$$ q_1 = q_0 (\cos \chi)^{1+r} \quad (326) $$

where $q_0$ is given by Eq. (318), with $H = H_0$. As previously mentioned, the difference between the two scales of reduced height is

$$ z' - z = \ln (1 + \Gamma) \approx \Gamma \quad \text{if } |\Gamma| \ll 1 \quad (327) $$

Given these formulas for the production function $q(z)$, we can write down the "photochemical equilibrium" electron density distribution, which is the solution of the continuity equation (300) with the transport term omitted and $\partial N/\partial t = 0$. If electrons are assumed to be lost at a rate $\alpha N^2$, where $\alpha$ is the recombination coefficient, the electron density distribution corresponding to the simple production formula (319) is

$$ N(z) = \left(\frac{q_0}{\alpha}\right)^{1/2} \exp \left(\frac{1}{2} (1 - z - e^{-z} \sec \chi)\right) . \quad (328) $$

In the literature, this distribution is described as a "Chapman alpha," "alpha Chapman," or simply a "Chapman" layer.

Using Eq. (317) for the peak production rate, and Eq. (207) of Sec. 2.2 to connect the peak electron concentration $N_m$ (in $m^{-3}$) and the critical frequency $f_0$ (in Hz), we have for an idealized Chapman layer

$$ f_0 = 9 N_m^{1/2} = 9 \left(\frac{q_0}{\alpha}\right)^{1/4} (\cos \chi)^{1/4} \quad (329) $$

[If $N_m$ is in $cm^{-3}$ then the numerical factor is $9.0 \times 10^3$ in this equation]. If $\alpha$ varies with height, the exponent of $\cos \chi$ will be different, because the height of the layer varies with $\chi$ (Eq. (316)). Even with $\alpha$ constant, the existence of a scale height gradient will lead to a different zenith angle dependence. For a constant gradient, we find from Eq. (324) that

$$ f_0 \propto (\cos \chi)^{4(1+r)} \quad (330) $$

We discuss the application of these equations in Sec. 5.2.

3.23 CHAPMAN'S THEORY FOR GRAZING INCIDENCE

When $\chi$ is near $90^\circ$, as near sunrise and sunset, the "plane earth" approximation is not good enough. The factor $\sec \chi$ varies along the path of the radiation, and can no longer be taken outside the optical depth integration of Eq. (305). To overcome this difficulty, Chapman (1931b) defined a "grazing incidence function" which replaces $\sec \chi$ in the preceding analysis, and which has been tabulated by Wilkes (1954). It applies accurately only to a spherically symmetric atmosphere with $H$ independent of height.
The grazing incidence function $Ch(x, \chi)$ depends on the radius of the earth $R_E$, which enters the calculation in the parameter $x$, the ratio of geocentric distance $r$ to scale height $H$, namely

$$x = \frac{r}{H} = \frac{R_E + h}{H} \quad (331)$$

We require that the function $Ch(x, \chi)$, when used to replace sec$^2\chi$ in Eq. (305), shall give the correct value of optical depth as found by integrating (302). Hence, the function must satisfy the equations

$$\tau(h, \chi)/\sigma = \int n \, ds = n(h) \, H \, Ch(x, \chi) \quad (332)$$

A useful approximation to $Ch(x, \chi)$ may be obtained in the following manner. In Fig. 21, $h_g$ is the "grazing height" for radiation reaching $P$, at which $Ch(x, \chi)$ is to be calculated. We use suffixes $p, g$ to denote quantities evaluated at $P$ and $G$. Then $x_g = x_p \sin \chi$. It is convenient to write $ds = H \, dy$ for an element of the ray path, so that $y$ is reckoned positively along $GP$ from zero at $G$, in units of $H$. Then $y \to -\infty$ toward the sun. At $P$ the value of $y$ (positive if $\chi > 90^\circ$) is

$$y_p = -x_p \cos \chi = -x_g \cot \chi \quad (333)$$

Let $\zeta = (h - h_g)/H$ be the reduced height, measured in scale heights from the
grazing altitude $h_g$. Then the geocentric distance of any point on the path of the radiation, in scale heights, is given by

$$x + \zeta = \left( x_g^2 + y^2 \right)^{1/2}$$  \hspace{1cm} (334)

Hence $\zeta = y^2 / 2x_g$ if we assume $\zeta \ll x_g$. The local gas concentration is $n_g e^{-\zeta}$ in general, and is in particular $n_p = n_g e^{-\zeta_p}$ at $P$, where $\zeta_p = x_p - x_g$. Substituting for $n$ and $ds$ in (332), we find that

$$n_p H \left[ \text{Ch}(x, \chi) \right]_p = \int_{-\infty}^{\zeta_p} n_g \exp \left( -y^2 / 2x_g \right) H \, dy$$  \hspace{1cm} (335)

Substituting for $n_p$ and using $\zeta_p = y_p^2 / 2x_g$:

$$\left[ \text{Ch}(x, \chi) \right]_p = e^{\zeta_p} \left( \frac{1}{2} \pi x_g \right)^{1/2} \left[ 1 \pm \text{erf} \left( \frac{\zeta_p}{\sqrt{2}} \right) \right]$$  \hspace{1cm} (336)

where the $\pm$ sign refers to $\chi \gg 90^\circ$. Within the approximations made earlier, we have $x_g = x_p \sin \chi$ and

$$\zeta_p = x_p - x_g = x_p \left( 1 - \sin \chi \right) = \frac{1}{2} x_p \cos^2 \chi$$  \hspace{1cm} (337)

We can now write Eq. (336) entirely in terms of quantities evaluated at $P$. Dropping the subscript $p$, we have:

$$\text{Ch}(x, \chi) \approx \left( \frac{1}{2} \pi x \sin \chi \right)^{1/2} e^{\frac{1}{2} x \cos^2 x} \left[ 1 \pm \text{erf} \left( \frac{1}{2} x \cos^2 \chi \right)^{1/2} \right]$$  \hspace{1cm} (338)

Since tables of the error function are readily available, Eq. (338) provides a useful approximation which is good enough for most purposes. The function $\text{Ch}(x, \chi)$ has been tabulated by Wilkes (1954) and, as shown in Fig. 22, it departs significantly from $\sec \chi$ when $\chi \approx 80^\circ$. At $\chi = 90^\circ$, $\sec \chi$ becomes infinite whereas $\text{Ch}(x, 90^\circ) = \left( \frac{1}{2} \pi x \right)^{1/2}$ from Eq. (338). Significant production can occur even at $\chi = 100^\circ$ when $H$ is large.

In calculating production rates at large zenith angles, $\sec \chi$ must be replaced by $\text{Ch}(x, \chi)$ wherever it occurs. It must be noted that different values of $x$ apply to atmospheric gases with different scale heights.

3.24 The Generalized Production Function

In the actual atmosphere, the production function $q(h, \chi)$ is more complicated because our basic assumptions—(i) monochromatic radiation and (ii) a single ionizable gas—do not hold. If there are several gases, differently distributed, the optical depth $\tau$ in Eq. (305) is expressed as a summation

$$\tau(h, \chi) = \sum_i \sigma_i n_i H_i \text{Ch}(x_i, \chi)$$  \hspace{1cm} (339)

Except for grazing incidence, $\text{Ch}(x_i, \chi)$ may be replaced by $\sec \chi$ in every
3.2 THE BASIC THEORY OF PHOTOIONIZATION

Fig. 22. The function Ch(χ) for three values of r/H (where r = geocentric distance, H = atmospheric scale height) compared with sec χ.

Because the radiation is not monochromatic, we now have to take account of the variation of \( n_i \) and \( \sigma_i \) with ultraviolet wavelength \( \lambda \). Eq. (340) is then rewritten as an expression for the production \( dq(h, \chi) \) due to the photon flux \( dI_\infty \) within the wavelength range \( (\lambda, \lambda + d\lambda) \). The total production

\[
q(h, \chi) = \int (dq/d\lambda) d\lambda
\]

may be evaluated if the wavelength dependence of each parameter is known (Chapman, 1939). In practice, the integration is replaced by a summation over a number of discrete wavelength ranges within which the parameters
\( \eta \) and \( \sigma \) do not vary too much, and \( q(h, \chi) \) is then computed numerically with available data.

Allen (1965) has suggested that for atmospheric gases the ionizing efficiency can be approximated by

\[
\eta = \frac{360}{\lambda}
\]  

with \( \lambda \) in angstroms. This implies that the production rate is more closely linked to the incident energy flux \( F_\infty \) than to the incident photon flux \( I_\infty \), since by Planck's equation \( F_\infty \propto I_\infty / \lambda \). Inserting numerical values, we find that the integrated production rate is

\[
Q = \int_0^\infty q \, dh = 1.8 \times 10^{16} F_\infty \quad [\text{c.g.s.}]
\]

\[
= 1.8 \times 10^{17} F_\infty \quad [\text{m.k.s.}]
\]

This formula is equivalent to the assumption that a fixed energy of 34 eV is required for each ionization, regardless of the photon energy.

3.3 Production of the Ionospheric Layers

3.31 Solar Photoionization

Ionization is produced in the earth's atmosphere by a wide spectrum of solar X-ray and extreme ultraviolet (EUV) radiation. The radiation consists of numerous emission lines generated in the chromosphere and corona, and some continuum radiation. Its intensity is measured with the aid of high-altitude rockets (e.g., Hinteregger and Watanabe, 1962; Hall et al., 1965; Friedman, 1963) and artificial satellites (e.g., Kreplin et al., 1962; Lindsay, 1963; Bowen et al., 1964b). The currently available data have been reviewed and summarized in several articles (Allen, 1965; Hinteregger et al., 1965; Yonezawa, 1966).

To compute the rate of production of ionization, we need to know the composition of the atmosphere as well as the spectrum of the incident radiation. We need not be concerned with the details of the latter; only a few emission lines are important enough to be mentioned individually, and it is sufficient to divide the spectrum into a limited number of wavelength bands. The bands are best defined in relation to the depth of penetration of the radiation. We specify the penetration in terms of the level of unit optical depth, \( h_0 \), at which vertically incident radiation is attenuated to a fraction \( e^{-1} \) of its intensity above the atmosphere. We recall from Sec. 3.2 that this level depends on the absorption cross sections \( \sigma_i(\lambda) \) for the various atmo-
3.3 PRODUCTION OF THE IONOSPHERIC LAYERS

spheric gases; at any wavelength, it is determined by the condition (cf. Eqs. (312) and (339))

$$\sum_i \sigma_i(\lambda) n_i H_i = 1$$

(344)

The cross sections have been measured for O_2 and N_2 (e.g., Ditchburn, 1956; Cook and Metzger, 1964), but for atomic gases they are generally derived from theoretical calculations, such as those of Dalgarno et al. (1964).

3.32 PRODUCTION OF THE IONOSPHERIC LAYERS

The upper part of Fig. 23 shows the height \(h_0\) as a function of wavelength,

![Fig. 23](image)

**Fig. 23.** (a) Height of unit optical depth for vertically incident radiation, as a function of wavelength. The break at 31 Å corresponds to the \(K\) absorption limit of N_2; those at 796 Å, 911 Å, and 1027 Å to the ionization limits of N_2, O, O_2, respectively. (b) Solar flux for a moderate level of activity (sunspot number \(R \approx 60\)), for several wavelength bands between 8 and 1027 Å. The area below the broken lines indicates photon flux; the area below the full lines indicates energy flux. The photon flux contained in some important spectral lines are indicated by the lengths of the arrows [based on data of Allen (1965) and Norton et al. (1963)]. [Note: 1 erg cm\(^{-2}\) s\(^{-1}\) Å\(^{-1}\) = 1 mW m\(^{-2}\) Å\(^{-1}\).]
computed from the atmospheric data of Norton et al. (1963). The lower part of this figure shows the intensity of solar radiation within a number of bands (Allen, 1965). For each band, the area below the broken line represents the total photon flux, and the area below the full line represents the flux of energy. The positions of a few strong spectral lines are shown by arrows, whose length indicates the photon flux of the line radiation. These fluxes relate to a fairly low level of solar activity, corresponding to a mean sunspot number $R \approx 60$.

With the aid of Fig. 23, we can discuss the production of the ionospheric layers by the solar spectrum. At vertical incidence the most heavily absorbed wavelengths, for which $\sigma \sim 1 \times 10^{-17} \text{ cm}^2$, reach unit optical depth at 170 km. These are in the 500–600 Å range, and include the strong line emission He I, 584 Å. Other strong lines which reach unit optical depth above 150 km are He II, 304 Å and O V, 630 Å. Line and continuum radiations in this range are responsible for producing the F1 layer. The long wavelength limit for F-region ionization may be placed at 796 Å, the longest wavelength which can ionize $\text{N}_2$; at this point, the curve of $h_0$ versus $\lambda$ drops by some 10 km. The short wavelength limit is not so well defined, but may be taken to include the strong 170–200 Å band, which contains a group of Fe emission lines. Wavelengths below 100 Å are absorbed in the E region, and the intervening 100–170 Å band is quite weak.

In Fig. 24 the curve marked “F” shows the photoionization profile $q(h)$ for the entire spectral range 140–796 Å. In spite of the wide range of wavelengths included, the curve is quite similar to an idealized Chapman curve with a peak at 150 km.

Radiations longer than 796 Å can ionize oxygen, of which the ionization limits are at 911 Å for $\text{O}$ and 1027 Å for $\text{O}_2$. However, the wavelengths up to about 1000 Å are significantly absorbed in $\text{N}_2$, and the absorbed energy may contribute to ionospheric heating even if it does not appear as ionization. The absorption coefficient varies with wavelength in a most irregular way, but no attempt has been made to show this irregularity in the $h_0(\lambda)$ curve of Fig. 23, which is, therefore, approximate in this region. The absorption coefficient in $\text{N}_2$ for the emission line H Ly $\gamma$, 973 Å, is so large that the line is barely detectable at 200 km altitude (e.g., Friedman, 1963). In contrast, the neighboring line C HI, 977 Å has an absorption coefficient about $10^2$ times smaller, and penetrates well into the E region. The strong line H Ly $\beta$, 1026 Å can just ionize $\text{O}_2$. Together with other lines of the Lyman series, and the Lyman continuum, these radiations provide most of the “UV” component of E-region production shown in Fig. 24.
3.3 PRODUCTION OF THE IONOSPHERIC LAYERS

Fig. 24. Electron production profiles \( q(h) \) for the E and F regions, for vertically incident radiation at sunspot number \( R = 60 \). The curves refer to the following wavelength bands: \( X(E) \), 8–140 Å; \( UV(E) \) 796–1027 Å; \( E = UV(E) + X(E) \); \( F \), 140–796 Å; \( E + F \), total 8–1027 Å [after Allen (1965)]. [Note: \( 1 \, \text{cm}^{-3} \, \text{s}^{-1} = 10^{6} \, \text{m}^{-3} \, \text{s}^{-1} \).]

A quite separate contribution to E-region production is made by X-rays in the approximate range 8–140 Å. According to Fig. 24, this band somewhat exceeds the 796–1027 Å UV band in importance. However, the relative importance of the X-ray and UV contributions is not accurately known, and may well vary with the solar cycle. The photoelectrons produced by X-rays can produce secondary ionization, so the production processes in the E region are quite complex.

The total photoionization profile for the range 8–1027 Å is marked "E + F" in Fig. 24. It does not show any pronounced division between the
E and F layer productions, although Allen (1965) points out that its shape could easily be altered by varying some of the assumptions made. Of course, the electron density $N(h)$ profile depends also on the loss coefficients, so it need not resemble the $q(h)$ profile in shape. This point is relevant, not only to the division between the E and F layers but also to the upwards increase of electron density above 150 km.

Ultraviolet radiation in the range 1027–1340 Å does not ionize any major atmospheric constituent, but it can ionize the trace constituent nitric oxide, NO (Sec. 3.51). The very strong emission line H Ly $\alpha$, 1216 Å, lies in this range. Much of the Lyman $\alpha$ is expended in dissociating molecular oxygen (Sec. 1.4), but Nicolet and Aikin (1960) estimated that sufficient Lyman $\alpha$ is absorbed by NO to provide a major source of D-region ionization. Another source of D-region ionization is X-radiation of roughly 1–8 Å, which might be regarded as the “tail” of the E-region ionization. A third source is corpuscular radiation by cosmic rays, important below 70 km. Poppoff et al. (1964) consider that X-rays are the principal ionizing source for the D region at sunspot maximum, but that Lyman $\alpha$ becomes important at sunspot minimum.

Figure 25 shows the ionization versus height profiles computed by Allen (1965) for Lyman $\alpha$, and for solar X-rays under “quiet” and “disturbed” conditions. We must emphasize that the relative importance of Lyman $\alpha$ and X-rays is still not well determined, partly because of uncertainties about the NO concentration. During the “disturbed” conditions, representing a moderate solar flare (“importance 2”), the X-ray intensity below 8 Å increases by many orders of magnitude. Because the absorption cross section $\sigma$ decreases with decreasing wavelength, the additional ionization is produced at a lower level, consistent with observations of the disturbed D region (Sec. 6.1).

Photoionization is, of course, essentially a daytime process, but we shall consider in Sec. 5.21 its possible role in the nighttime E region.

3.33 Corpuscular Ionization

Ionization can be produced not only by solar photon radiation but also by energetic charged particles which enter the atmosphere most easily at high magnetic latitudes. The depth to which particles penetrate the atmosphere depends on their energy. Some calculations of penetration depths have been given by Bailey (1959), Rees (1963), and Kamiyama (1966); we show Bailey’s results in Fig. 26. The calculations of production rate are complex because the incident particles possess a spectrum of energy and direction of motion. For particles within a given energy range, the production function will broadly resemble a Chapman function.
3.3 PRODUCTION OF THE IONOSPHERIC LAYERS

Fig. 25. Production profiles $q(h)$ for the D region, for vertically incident radiation. The arrows indicate the heights of peak production for four X-ray wavelength bands. Curve $A$ represents quiet conditions at sunspot number $R = 60$, the contribution of Lyman $\alpha$ being shown. Curve $B$ represents the additional production due to a moderate solar flare (importance 2) [after Allen (1965)]. The dashed curves indicate production rates due to galactic cosmic rays, at magnetic latitude 50°, at sunspot minimum and maximum [after Webber (1962)].

We have already mentioned the cosmic ray contribution to D-region ionization. The cosmic ray particles are so energetic that the production rate only becomes appreciable at 70 km altitude, and increases downwards (Fig. 25). This production is present both by day and by night, but varies with latitude. Its solar cycle variation is opposite to that of the solar photon radiation because the flux of galactic cosmic rays reaching the earth is smaller at sunspot maximum than at minimum. This is attributed to the
Fig. 26. Approximate heights of penetration in the atmosphere for vertically incident electrons, protons and helium ions, together with data on the geomagnetic cutoff latitude (the lowest geomagnetic latitude at which a cosmic-ray particle can reach the atmosphere), this being a function of “magnetic rigidity” or momentum-to-charge ratio [Bailey (1959)].

shielding effect of the enhanced interplanetary magnetic field. Besides the cosmic ray ionization, the enhancements of D and E region ionization observed at high latitudes during magnetic disturbance are certainly due to fast electrons or protons. The mechanisms by which these particles acquire their energy are part of the general problem of high-latitude and magnetic-storm
3.4 LOSS REACTIONS

phenomena, to which we return in Chapter VIII. It has been suggested that corpuscular ionization is responsible for much of the ionization and heating in the F2 layer (Antonova and Ivanov-Kholodnii, 1961; Harris and Priester, 1962a). The only particles likely to be absorbed at these heights would be soft electrons or protons, of not more than a few hundred electron volts. Any major corpuscular contribution to the undisturbed ionosphere, outside polar regions, is problematical. Limits can be placed on the rate of production because of the airglow emission that would accompany corpuscular ionization (Sec. 3.7).

3.4 Loss Reactions

The kinds of photochemical reaction which are thought to control the ionospheric electron density are considered next. The mechanisms for loss of ionization have been discussed in two articles by Bates and Massey (1946, 1947), and by many other authors (e.g., Nicolet and Swider, 1963; Whitten and Poppoff, 1962, 1964). The pertinent reactions are listed below; in most cases, symbols X, Y, etc. may denote either an atom or a molecule. Typical magnitudes of the coefficients are given in Table I.

(a) Ion-ion recombination (coefficient $\alpha_r$):

$$X^+ + Y^- \rightarrow X + Y$$

(b) Electron-ion recombination (coefficient $\alpha_e$):

Three-body: \hspace{1cm} $X^+ + e + M \rightarrow X + M$

Radiative: \hspace{1cm} $X^+ + e \rightarrow X^* \rightarrow X + \hbar\nu$

Dissociative: \hspace{1cm} $XY^+ + e \rightarrow X^* + Y^*$

The asterisks indicate that the atoms may be left in excited states, and subsequently lose this energy by radiation or during collisions with other particles. The symbol M denotes a neutral particle which exchanges energy and momentum but does not take part in the chemical reaction. Three-body processes can occur in the lower D region, but are so rare at greater heights as to be quite unimportant. Apart from this means, recombination of electrons and atomic ions can take place only by the very slow radiative process. Only in the uppermost F region is radiative recombination likely to be the fastest loss process, and we shall see that at such heights transport processes (especially diffusion) are so completely dominant that loss coefficients are irrelevant. Elsewhere in the F and E regions, dissociative recombination (Biondi, 1964) is the most important loss process.
### TABLE I
PRODUCTION AND LOSS PROCESSES

<table>
<thead>
<tr>
<th>Process</th>
<th>Contribution to continuity equation (reactions/unit volume/unit time) and rough values of coefficients</th>
<th>D Region 50–90 km (approx.)</th>
<th>E Region 90–150 km (approx.)</th>
<th>F Region 150–600 km (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRODUCTION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solar photoionization</td>
<td>$q(h)$</td>
<td>[Ly $\alpha$ 1216 Å]</td>
<td>[Ly $\beta$ 1026 Å]</td>
<td>EUV 911–1027 Å</td>
</tr>
<tr>
<td>(principal radiations shown [1])</td>
<td>(ionizes NO)</td>
<td>O$_3$ ionized by $\lambda &lt; 1027$ Å</td>
<td>X-rays 1–10 Å</td>
<td>[He II 304 Å, He I 584 Å]</td>
</tr>
<tr>
<td>Corpuscular ionization</td>
<td>$q(h)$</td>
<td>Electrons $&gt; 30$ keV</td>
<td>Electrons 1–30 keV cause some nighttime and sporadic E ionization</td>
<td>N$_2$ ionized by $\lambda &lt; 796$ Å</td>
</tr>
<tr>
<td>(more important at high latitudes, especially auroral zone)</td>
<td></td>
<td>Protons $&gt; 1$ MeV</td>
<td></td>
<td>Electrons $\leq 1$ keV (probably small; might be significant at night)</td>
</tr>
<tr>
<td><strong>LOSS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ion-ion recombination</td>
<td>$\alpha_i N_+ N_-$</td>
<td>Important</td>
<td>Few negative ions exist</td>
<td>Very few negative ions exist</td>
</tr>
<tr>
<td></td>
<td>$\alpha_i \approx 10^{-7}$ cm$^3$ s$^{-1} = 10^{-13}$ m$^3$ s$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electron-ion recombination</td>
<td>$\alpha_e N_+ N_e$</td>
<td>Important</td>
<td>Gas densities too low</td>
<td>Gas densities too low</td>
</tr>
<tr>
<td>Three-body recombination</td>
<td>$\alpha_e - \alpha(h)$</td>
<td>Important</td>
<td>Not important</td>
<td>Not important</td>
</tr>
<tr>
<td>Radiative recombination</td>
<td>$\alpha_e \approx 10^{-12}$ cm$^3$ s$^{-1} = 10^{-18}$ m$^3$ s$^{-1}$</td>
<td>Insignificant</td>
<td>Not important</td>
<td>Principal loss mechanism</td>
</tr>
<tr>
<td>Dissociative recombination</td>
<td>$\alpha_e \approx 10^{-7}$ cm$^3$ s$^{-1} = 10^{-13}$ m$^3$ s$^{-1}$</td>
<td>Important</td>
<td>Principal loss mechanism</td>
<td>Important</td>
</tr>
<tr>
<td>Ion-atom interchange</td>
<td>$\beta(h) N_{A^+} \equiv \gamma [M] N_{A^+}$</td>
<td>Not important, because few atomic ions exist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>($N_{A^+} =$ atomic ion concentration)</td>
<td>$\gamma \approx 10^{-11}$ cm$^3$ s$^{-1} = 10^{-17}$ m$^3$ s$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 1 (continued)**

<table>
<thead>
<tr>
<th>Process</th>
<th>Contribution to continuity equation (reactions/unit volume/unit time) and rough values of coefficients</th>
<th>D Region 50–90 km (approx.)</th>
<th>E Region 90–150 km (approx.)</th>
<th>F Region 150–600 km (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attachment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiative</td>
<td></td>
<td>Three-body attachment is most important</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three-body</td>
<td></td>
<td>Can maintain some negative ions at night</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collisional detachment, etc.</strong></td>
<td></td>
<td>Radiative attachment provides a very weak source of negative ions</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collisonal detachment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Photodetachment by solar visible and long UV radiation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process</th>
<th>Contribution to continuity equation (reactions/unit volume/unit time) and rough values of coefficients</th>
<th>D Region 50–90 km (approx.)</th>
<th>E Region 90–150 km (approx.)</th>
<th>F Region 150–600 km (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attachment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiative</td>
<td></td>
<td>Three-body attachment is most important</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three-body</td>
<td></td>
<td>Can maintain some negative ions at night</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collisional detachment, etc.</strong></td>
<td></td>
<td>Radiative attachment provides a very weak source of negative ions</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collisonal detachment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Photodetachment by solar visible and long UV radiation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Radiative: $a_r \sim 10^{-15} \text{ cm}^3 \text{ s}^{-1} = 10^{-21} \text{ m}^3 \text{ s}^{-1}$
- Three-body: $a_t \sim 10^{-30} \text{ cm}^6 \text{ s}^{-1} = 10^{-42} \text{ m}^6 \text{ s}^{-1}$
- Collisional: $\delta(h) N_e \equiv \kappa [M] N_e$
- Photodetachment by solar visible and long UV radiation: $\rho \approx 1 \text{ s}^{-1}$
Since a large proportion of the ions in the E and F regions are originally atomic, dissociative recombination must be preceded by reactions involving formation of molecular ions, namely:

(c) Ion-atom interchange (rate coefficient $\gamma$):

$$X^+ + YZ \rightarrow XY^+ + Z$$

While it is generally accepted that ion-atom interchange followed by dissociative recombination is the principal loss process in the E and F regions, there is controversy as to precisely which reactions are important. Although charge-exchange reactions of the type

$$X^+ + YZ \rightarrow YZ^+ + X$$

can also lead to the formation of molecular ions, ion-atom interchange reactions are probably more rapid (Bates, 1955).

(d) Collisional detachment (coefficient $\delta$) in forward direction; three-body attachment (rate coefficient $a_t$) in reverse direction:

$$X^- + M \rightarrow X + e + M$$

The detachment reaction may be facilitated if the molecule M is in an excited metastable state (Megill and Hasted, 1965), in which case the reaction can be written

$$X^- + M^* \rightarrow X + e + M$$

(e) Associative detachment in forward direction; dissociative attachment in reverse direction:

$$X^- + Y \rightarrow XY + e$$

(f) Photodetachment (coefficient $\rho$) in forward direction; radiative attachment (rate coefficient $a_r$) in reverse direction:

$$X^- + hv \rightarrow X + e$$

The coefficient $\rho$ is the product of the flux of solar radiation responsible for the detachment, and an appropriate cross section. Except at grazing incidence, the optical depth throughout the D region is small for the visible and long ultraviolet radiation involved, so that $\rho$ varies little with height.

Table I summarizes ionospheric production and loss processes, with brief comments on their importance. Wherever a quantity is shown as height-varying, such as $q(h)$, this implies rapid (e.g., exponential) variation because
of dependence on some gas concentration. Quantities not marked in this way might be somewhat height-dependent for other reasons, such as temperature variations. The numerical values of rate coefficients are generally derived from theory or laboratory experiments, and may disagree with those deduced from ionospheric considerations. They have been taken from Branscomb (1964), Dalgarno (1961), Megill and Hasted (1965), Whitten and Poppoff (1964), and Biondi (1964).

We give expressions for the terms in the charged-particle continuity equations which result from the different processes. To economize in notation we do not introduce separate symbols for every individual process, except where we are dealing with quantities possessing different physical dimensions (such as \( a, a_\alpha, \) and \( a_\beta \)). All detachment processes, except photodetachment, are included in a single coefficient \( \delta \). The relative importance of the different contributions to \( \delta \) is difficult to assess, since it depends on the coefficients \( \kappa \) and on the concentrations of various minor constituents; many of these parameters are poorly known.

### 3.5 Chemistry of the D Region

#### 3.51 Neutral Constituents

In Sec. 3.32 we mentioned three sources of ionization for the normal D region. Two of these, solar X-rays and cosmic rays, ionize all atmospheric gases; but the third, solar Lyman \( \alpha \), ionizes only nitric oxide \( \text{NO} \), a trace constituent. We should therefore consider how nitric oxide is formed.

According to Nicolet (1965), \( \text{NO} \) can be produced and destroyed by various reactions, of which the most important at D-region heights are probably

\[
\begin{align*}
N + O + M &\rightarrow \text{NO} + M \\
N + O_2 &\rightarrow \text{NO} + O
\end{align*}
\]

(where \( M \) is a neutral particle). In this scheme the rates of production and loss of \( \text{NO} \) are both proportional to the atomic nitrogen concentration \( n[N] \). Consequently the equilibrium \( \text{NO} \) concentration is independent of \( n[N] \), but it depends on temperature because the rate coefficients (especially that of the reaction involving \( \text{O}_2 \)) vary with temperature. The \( \text{NO} \) concentrations around 70 km computed from this model are roughly \( 10^3 \) times smaller than those measured by Barth (1966) in rocket-borne spectrometric experiments. However, Sechrist (1967) considers that this discrepancy may result from the temperature dependence of the \( \text{NO} \) concentration (Sec. 5.12). It is also
possible that transport processes could influence the distributions of minor constituents such as NO.

Other minor constituents may play a part in D-region chemistry, notably molecular oxygen in the metastable $^1\Delta_g$ state and ozone; but we shall not consider their photochemistry in this book.

3.52 D-Region Balance Equations

In the lower ionosphere, transport of ionization may be neglected and only photochemical terms appear in the continuity equations for the ion concentrations $N_+$, $N_-$ and the electron concentration $N_e$. For the present, we add the subscript "e" for the sake of clarity. These equations are therefore written as

$$dN_+/dt = q - \alpha_e N_+ N_e - \alpha_i N_+ N_-$$

$$dN_e/dt = q - \alpha_e N_+ N_e - (\rho + \delta) N_-$$

$$dN_-/dt = \alpha N_e - (\rho + \delta) N_- - \alpha_i N_+ N_-$$

We further require the ionosphere to be electrically neutral, which implies

$$N_+ = N_- + N_e = (1 + \lambda) N_e$$

in which $\lambda = N_-/N_e$ is the negative ion ratio. From Eq. (349) we see that the three continuity equations are not independent because necessarily Eq. (346) is the sum of Eq. (347) and (348).

There are several simplifications that can be made in practical conditions. Except when the ionosphere is grossly perturbed (as by a solar flare or perhaps a solar eclipse), the time derivatives are generally considered to be small in comparison to the other terms, though this may not always be justified. If we take $dN_-/dt = 0$ in Eq. (348), we find that

$$\lambda = N_-/N_e = a/(\rho + \delta + \alpha_i N_+)$$

During the day, the denominator of this expression is largely dominated by the photodetachment coefficient $\rho$. Since $\rho$ is proportional to the flux of radiation capable of destroying negative ions, it is relatively independent of height because the D region is optically thin to this radiation. In contrast, the coefficients $a$ and $\delta$ decrease exponentially upward since they depend on the concentrations of the molecules which participate in the attachment and detachment processes. Hence, by day, $\lambda$ decreases upward, and is probably small above 90 km. At night, however, $\rho \approx 0$ and $\lambda$ depends on the ratio $a/\delta$. Above 100 km, ionic recombination would limit the equilibrium nighttime
3.5 CHEMISTRY OF THE D REGION

The value of $N_-$, and at greater heights where $a^{-1}$ exceeds a few hours, equilibrium conditions scarcely apply. Furthermore, day-to-night variations of $\delta$ and $a$ could exist if the dominant reactions involve a species (possibly O or O$_3$), the abundance of which varies diurnally. The uncertainties are such that $\lambda$—although a very basic ionospheric quantity—is still poorly known. Some possible $\lambda(h)$ profiles are given in Fig. 27, but opinions on them differ; Hultqvist (1963) advocates much lower negative ion densities, with $\lambda = 1$ near 48 km by day and 60 km by night.

By day, attachment and detachment are very rapid; the value $\rho \sim 1 \text{s}^{-1}$ quoted in Table I implies that the lifetime of negative ions is of order seconds. Our assumption of equilibrium is then justified, and we can substitute Eq. (349) into Eq. (346) to obtain

$$\left(1 + \lambda\right) \frac{dN_e}{dt} = q - (1 + \lambda) (\alpha_e + \lambda \alpha_i) N_e^2$$

$$= q - \alpha_e N_e^2$$

(351)

The "effective" coefficient $\alpha_E$, as defined in these equations, rapidly approaches $\alpha_e$ with increasing altitude because of the decrease of $\lambda$.

3.53 IDENTITY OF IONS IN THE D REGION

So far, we have described the chemical processes in rather general terms. In reality, discussion of the D region is hampered by the fact that the negative ions have not yet been identified. Since nitrogen does not form stable negative ions, it has generally been assumed that O$^-$ or O$_2^-$ are the most
important. These ions can be destroyed by visible light; but instead it has been suggested that \( \text{O}_3^- \) and \( \text{NO}_2^- \) ions, which require ultraviolet light for electron detachment, might be dominant (Reid, 1961, 1964). This question of electron affinity influences the interpretation of the changes in D-region ionization near sunrise and sunset, in relation to the variation of solar zenith angle \( \chi \). If ultraviolet light is the detaching agent, no effect should occur when the grazing height \( h_g \) in Fig. 21) is below about 30 km because of absorption in the ozone layer. Visible light, however, is screened only by clouds and the densest atmosphere within a few kilometers of the ground, so that detachment occurs at larger values of \( \chi \). Most of the radio data are concerned with the "polar cap absorption" (PCA) events, for which the data on sunrise changes seem clearcut; in the mid-latitude D region, rocket observations at sunrise also indicate that UV is required for detachment (Bowhill and Smith, 1966).

Because of the difficulties of explaining the diurnal changes, both in PCA events and in the normal mid-latitude D region, it is necessary to consider complicated models. Such models may involve a number of different negative ions and several of the reactions considered in Sec. 3.4. The possibilities include:

- **Associative detachment:** \( \text{O}_2^- + \text{O} \rightarrow \text{O}_3 + e \)
- **Collisional detachment**
  - by excited metastable \( \text{O}_2 \) (\( \text{O}_2^* \)): \( \text{O}_2^- + \text{O}_2^* \rightarrow 2\text{O}_2 + e \) (352)
- **Charge transfer with unidentified particle:** \( \text{O}_2^- + X \rightarrow X^- + \text{O}_2 \)

The concentrations \( n[\text{O}] \), \( n[\text{O}_2^*] \) and \( n[X] \) may all vary diurnally. In particular \( X \) might be \( \text{O}_3 \), the diurnal variation of which has been discussed by Hunt (1966). Adams and Megill (1967) claim that with reasonable values of the various parameters, the model can account for the daytime, twilight and nighttime absorption data on PCA events.

So far we have not mentioned the positive ion composition in the D region. Mass spectrometer data obtained from rocket experiments indicate that many kinds of positive ions are present (Narcisi and Bailey, 1965). Above about 83 km, the dominant ions have mass numbers 30\(^+\) and 32\(^+\), and are presumed to be \( \text{NO}^+ \) and \( \text{O}_2^+ \), as in the E-region situation described in Sec. 3.6. Metallic ions such as \( \text{Na}^+ \), \( \text{Mg}^+ \), and \( \text{Ca}^+ \) are also recorded. At lower heights, down to 64 km, ions of mass 19\(^+\) and 37\(^+\) are prominent and are identified as hydrated protons, \( \text{H}_3\text{O}^+ \) and \( \text{H}_5\text{O}_2^+ \). Narcisi (1966) gives reasons for believing that these observations are genuine, and do not result from contamination, a danger always present in such experiments. Since the
role of these complex ions in the D region has yet to be understood, we cannot attempt to relate them to our foregoing discussion. More mass-spectrometer data on ions of both signs are obviously needed. Fortunately, laboratory measurements of reaction rates are now becoming available (Fehsenfeld et al., 1967) and should help to clarify the photochemistry of the D region.

3.6 Chemistry of the E and F1 Regions

Let us now examine the situation at greater heights, in the E and F1 regions, where negative ions are virtually absent. Although negative ions are slowly produced by radiative attachment, they are destroyed rapidly by photodetachment in daylight, and quite rapidly by ionic recombination, otherwise. Under these circumstances, the only important recombination process is dissociative recombination of electrons and molecular positive ions. Since a large proportion of the neutral atmosphere is atomic, especially in the F region, atomic ions are produced by photoionization, but these do not recombine with electrons directly, except by the very slow radiative process. Instead, they undergo an ion-atom interchange reaction, and the molecular ions thus formed combine with electrons. We shall consider the electron distributions resulting from this situation in a general way, before discussing the actual reactions which occur.

3.61 Linear and Square-Law Loss Formulas

In writing the continuity equations for the concentrations of electrons, \( N \), atomic ions \( N_{A^+} \) and molecular ions \( N_{M^+} \), we assume that the ion-atom interchange reaction involves a molecular gas, the concentration of which is \( n [M] \), but that we can neglect the direct production of molecular ions by photoionization. This is not unreasonable, since \( O_2 \) is a minor constituent and \( N_2^+ \) ions are likely to be rapidly removed by dissociative recombination (for the present we neglect the possibility that \( N_2^+ \) ions undergo reactions which produce other ions). Proceeding on these assumptions, we write

\[
\frac{dN}{dt} = q - \alpha NN_{M^+}
\]

\[
\frac{dN_{A^+}}{dt} = q - \gamma n [M] N_{A^+}
\]

\[
\frac{dN_{M^+}}{dt} = \gamma n [M] N_{A^+} - \alpha NN_{M^+}
\]

Charge neutrality requires that

\[
N = N_{A^+} + N_{M^+}
\]
Let us assume equilibrium conditions \((d/dt = 0)\) and write \(\beta = \gamma n [M]\). From Eq. (355), the ratio of atomic and molecular ion concentrations is then

\[
N_{A^+}/N_{M^+} = \alpha N/\beta
\]  

We can use Eqs. (356) and (357) to eliminate the ion concentrations and thereby obtain a quadratic equation for the electron concentration \(N\) (Hirsh, 1959):

\[
\alpha\beta N^2 - \alpha q N - \beta q = 0
\]  

This equation can be rewritten as

\[
\frac{1}{q} = \frac{1}{\beta N} + \frac{1}{\alpha N^2}
\]  

There are two important limiting cases of this equation, namely

\[
q = \alpha N^2 \quad \text{if} \quad \beta \gg \alpha N, \quad \text{so that} \quad N_{M^+} \gg N_{A^+}
\]

\[
q = \beta N \quad \text{if} \quad \beta \ll \alpha N, \quad \text{so that} \quad N_{M^+} \ll N_{A^+}
\]

These equations demonstrate how the two-stage loss process proposed by Bates and Massey (1947) gives rise to a transition between a square-law loss formula \((\alpha N^2)\) and a linear formula \((\beta N)\). When the rate of electron loss is determined by the dissociative recombination reaction, the ions are mainly molecular and the \(\alpha N^2\) formula applies. The coefficient \(\alpha\) may depend on temperature, but is otherwise not height dependent, unlike \(\beta\) which varies as the molecular concentration \(n [M]\) and, therefore, decreases rapidly upward. So at greater heights, we expect the condition \(\beta \ll \alpha N\) to apply, in which case the ion-atom interchange reaction controls the rate of loss, which is linear in \(N\), and the ions are mostly atomic.

A more general expression for \(N\) is given by the positive root of Eq. (358), namely

\[
N = (q/2\beta) \left[ 1 + (1 + 4\beta^2/\alpha q)^{1/2} \right]
\]

This reduces to

\[
N = N_\alpha = (q/\alpha)^{1/2} \quad \text{if} \quad 4\beta^2 \gg \alpha q
\]

\[
N = N_\beta = q/\beta \quad \text{if} \quad 4\beta^2 \ll \alpha q
\]

The conditions \(\gg\) and \(\ll\) in Eq. (362) are essentially the same as those in (360).

### 3.62 SPLITTING OF THE F LAYER

As we shall see, rocket data on the ionic composition confirm that the transition between the domains of the \(\alpha N^2\) and \(\beta N\) loss formulas occurs in
the lower F region at about 160–200 km. This happens to coincide with the level at which the F-region production rate \( q \) is greatest, and Ratcliffe (1956a) showed that this might account for the splitting of the F layer into F1 and F2 components. To demonstrate this, we investigate the shape of the equilibrium electron distribution by evaluating Eq. (361), using the Chapman production function \( q(z, \chi) \) for an isothermal layer, Eq. (319), and a height-independent recombination coefficient \( a \). If we introduce a new quantity \( K = H_\beta / H_\beta \), the ratio of the scale heights of the ionizable gas and the linear loss coefficient, we can write

\[
\beta \propto e^{-Kz}.
\]

Since \( \beta = \gamma n [M] \), \( K \) depends on the molecular mass of the gas participating in the ion-atom interchange reaction, which may be \( \text{O}_2 \) or \( \text{N}_2 \). Since the principal ionizable gas is \( \text{O} \), we have \( K = 32/16 = 2 \) for \( \text{O}_2 \) and \( K = 28/16 = 1.75 \) for \( \text{N}_2 \) provided these gases are diffusively separated, which is found to be the case above the E region. If the coefficients \( \gamma \) were very temperature dependent, the value of \( K \) might be modified, but probably is not very different from the above mentioned values in the F region.

We found in Sec. 3.21 that the shape of the Chapman production function does not change as the solar zenith angle \( \chi \) varies, but that the level of the peak varies according to the relation \( z = \ln \sec \chi \). The form of Eq. (361) suggests that a single parameter, \( \beta^2 / aq \), determines the shape of the electron distribution \( N(z) \). Let \( G \) denote its value at the level of peak production. Then, using Eq. (362) and evaluating all quantities at the peak of \( q \), we have

\[
G = \beta^2 / aq = N_\beta^2 / N_\beta \quad \text{[in general]}
\]

\[
G = \beta_0^2 / aq_0 = N_\beta^0 / N_\beta^0 \quad \text{[for } \chi = 0 \text{]} \quad (363)
\]

Provided \( G \) and \( K \) are kept constant, changes of \( q \) and \( \beta \) can alter the magnitude of the function \( N(z) \) and displace it with respect to the \( z \)-axis, but do not affect its shape. In Fig. 28, we plot \( N(z) \) for \( G = \frac{1}{4}, 1, 4, 9 \). The graph is drawn in such a way that the dotted curve, which represents \( N_\beta = (q/\alpha)^{1/2} \) for \( \chi = 0 \), is the same in each case, but different values of \( \beta_0 \) are chosen so as to give three dashed curves \( N_\beta \), and three solid curves for \( N \). We see that if \( G = \frac{1}{4} \), the \( N(z) \) profile is almost smooth, but that a "ledge" appears for \( G = 4 \) and is more prominent for \( G = 9 \). On an \( h'(f) \) curve, such a ledge would produce a prominent "cusp," as shown by Hirsh (1959), and we may identify this with the splitting of the F layer into F1 and F2 components.

The relevance of this analysis of the F region is that, since \( G \) is evaluated at the level of peak \( q \), it varies with solar zenith angle \( \chi \). Remembering that the level of peak production occurs at \( z = \ln (\sec \chi) \), that the peak value of
Fig. 28. Electron concentration versus reduced height for the “transition region,” assuming a Chapman production function \( q(z) \) with peak at \( z = 0 \) and a square-law loss coefficient \( \alpha \) independent of height. The linear loss coefficient is \( \beta = \beta_0 \exp(-1.75z) \), and four values of \( \beta_0 \) are used such that \( G = \beta_0^2/\alpha q_0 \) takes the values \( 1, 4, 9 \). For these four values of \( \beta_0 \), the broken lines represent the functions \( N_\alpha = q/\beta \). The dotted line is \( N_\alpha = (q/\alpha)^{1/2} \) and the full lines are the profiles \( N(z) \) calculated from Eq. (361), normalized to \( N_\alpha = (q_0/\alpha)^{1/2} \). The F1 ledge is most pronounced for \( G = 9 \), and absent for \( G = \frac{1}{4} \).

\[ q = q_0 \cos \chi \] and that \( \beta(z) = \beta_0 e^{-Kz} \), we have

\[
G(\chi) = \beta_0^2 (\cos \chi)^{2K}/\alpha q_0 \cos \chi \\
= G(0) (\cos \chi)^{2K-1}
\]

If the composition of the atmosphere does not vary, the value of \( G \) for \( \chi = 0 \), namely \( G(0) \), should be a constant. As \( \chi \) increases, \( G(\chi) \) decreases and the “splitting” of the layer becomes less pronounced, or even disappears. This is consistent with the observations that, at mid-latitude stations, the F1 layer is most prominent around noon, and is more commonly observed on summer days than in winter. Since \( G(0) \propto 1/q_0 \), and \( q_0 \) varies with the solar cycle, we can also account for the observation that the F1 layer is more prominent at sunspot minimum than at sunspot maximum. Moreover, the F1 layer sometimes appears during a solar eclipse at times when it would not normally be seen, and this may arise from the reduction of \( q_0 \) and consequent increase of \( G \).

In the actual F region, some of our assumptions do not hold well. First, the production function does not conform to the simple Chapman theory.
because the atmosphere is not isothermal, and the radiation is not monochromatic. If these complications are included in the theory of F1-F2 splitting, the details may be modified, but the basic conclusions do not seem to be affected (King and Lawden, 1964). Second, we have to consider a more complex photochemical situation than is described by our basic set of continuity equations, Eqs. (353)-(355), as has been done by Yonezawa and Takahashi (1960) and Shimazaki (1965). By including extra processes, Shimazaki obtains a cubic equation for $N$, in place of Eq. (358), but again it seems that conclusions drawn from the basic theory are not unduly modified. Third, the theory we have given is not precise under conditions when the time derivatives in the continuity equations are appreciable, such as near sunrise, but it should be adequate for midday conditions (Burkard 1957, 1962; Shimazaki, 1965).

Later in the book, we shall see that the photochemical theory can account for some basic observational facts about the E and F1 layers. It also fits the hypothesis of Bradbury (1938), concerning the formation of the F2 layer at about 300 km. Since there is no known solar radiation capable of causing a peak of production at such a height, Bradbury supposed the production peak to lie at a lower height (now known to be the F1 layer) and attributed the upward increase of electron density to a rapid upward decrease of loss coefficient.

This hypothesis is not, by itself, sufficient to account for the formation of the F2 peak. If the linear loss coefficient $\beta$ varies as $e^{-Kz}$ (where $K = H_i/H_\beta$, as defined previously), then the electron distribution far above the production peak is approximately $q/\beta \propto \exp[(K-1)z]$; see Fig. 28. If $K > 1$, as seems certain, then $q/\beta$ increases indefinitely upward, and we must find some explanation for the existence of a peak of $N$. The possibilities include:

(i) Failure of the Chapman formula for $q$ at heights where the ionized/neutral concentration ratio $(N/n)$ is not small.

(ii) Existence of two peaks in $q(h)$, it being possible to satisfy Eq. (310) at more than one height if certain scale height profiles are postulated.

(iii) Existence of an additional loss process, such as radiative recombination, which might dominate at great heights.

(iv) Lack of equilibrium, such that $N$ never approaches the limiting value $q/\beta$.

(v) Action of some transport process (such as diffusion) to limit $N$ at great heights.

Of these possibilities, the first three do not provide a quantitative explanation of the F2 peak, and so it is thought that (v), and perhaps also (iv),
are most likely to control the F2 layer. We shall return to this topic when we discuss transport processes in Chapter IV.

### 3.63 Ion Composition

Up to this point, the theory does not depend critically on the identity of the reactions, but we should now be more specific about the ionic constituents. These have been identified by rocket-borne mass spectrometers, as described in Sec. 2.5, originally by Johnson et al. (1958). Some representative results for day and night are shown in Figs. 29 and 30 (Holmes et al., 1965).

![Fig. 29. Daytime positive ion composition above White Sands, New Mexico (32°N, 106°W). Ionic mass numbers are shown, their probable identifications being 14+ - N+, 16+ - O+, 18+ - H2O+ (possibly a contaminant), 28+ - N2+, 30+ - NO+, 32+ - O2+. The curve marked "Total" is the sum of all ion concentrations; the dotted curve marked N_e is the electron concentration profile determined from an ionogram. The solar zenith angle \( \chi = 60^\circ \) [Holmes et al., (1965)].]

By day, the ions 30+ and 32+ (identified as NO+ and O2+) predominate below 165 km, with ion 16+ (identified as O+) being dominant above this height. This transition between molecular and atomic ion dominance is consistent with the theory outlined in Sec. 3.61. At night, the transition is found at about 220 km (Fig. 30). Other ions are present as very minor constituents.
We have previously shown evidence that the major neutral constituents in the thermosphere are O and N₂, with some O₂. Thus the ions produced by photoionization are O⁺, N₂⁺, and O₂⁺; and since O₂ is a minor constituent in most of the thermosphere, we expect most of the O₂⁺ ions to be produced by "transfer" reactions (i.e., ion-atom interchange or charge exchange), which is certainly true of the NO⁺ ions.

The photochemical regime in the E and F regions has been the subject of very detailed analysis, in which numerous reactions are taken into account (Nicolet and Swider, 1963; Donahue, 1966b). It is possible to obtain quite satisfactory descriptions of the daytime situation, though there exist insufficient data to determine all the parameters uniquely. Some information on temperature dependence of the parameters can be obtained by comparing the altitude variations of ion composition with adopted temperature profiles.

In Table II we give a brief list of exothermic charged-particle reactions which includes those which seem likely to be of greatest importance. In the recombination equations, stars denote the possible production of atoms in excited states. Some reactions which have been omitted are probably unimportant either because they involve only "trace" neutral constituents (such as nitric oxide and atomic nitrogen) or because faster competing reactions...
III. PHOTOCHEMICAL PROCESSES IN THE IONOSPHERE

### TABLE II
PHOTOCHEMICAL REACTIONS IN THE E AND F REGIONS

<table>
<thead>
<tr>
<th>Reaction Type</th>
<th>(rate, q)</th>
<th>(rate coefficient, $\gamma$)</th>
<th>(rate coefficient, $\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photoionization</td>
<td></td>
<td>(Q1)</td>
<td>(R1)</td>
</tr>
<tr>
<td>$\text{O} + hv \rightarrow \text{O}^+ + e$</td>
<td></td>
<td>(Q2)</td>
<td>(R2)</td>
</tr>
<tr>
<td>$\text{N}_2 + hv \rightarrow \text{N}_2^+ + e$</td>
<td></td>
<td>(Q3)</td>
<td>(R3)</td>
</tr>
<tr>
<td>$\text{O}_2 + hv \rightarrow \text{O}_2^+ + e$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer or interchange</td>
<td></td>
<td>(T1)</td>
<td>(T1)</td>
</tr>
<tr>
<td>$\text{O}^+ + \text{O}_2 \rightarrow \text{O}_2^+ + \text{O}$</td>
<td></td>
<td>(T2)</td>
<td>(T2)</td>
</tr>
<tr>
<td>$\text{O}^+ + \text{N}_2 \rightarrow \text{NO}^+ + \text{N}$</td>
<td></td>
<td>(T3)</td>
<td>(T3)</td>
</tr>
<tr>
<td>$\text{N}_2^+ + \text{O} \rightarrow \text{NO}^+ + \text{N}$</td>
<td></td>
<td>(T4)</td>
<td>(T4)</td>
</tr>
<tr>
<td>$\text{N}_2^+ + \text{O}_2 \rightarrow \text{O}_2^+ + \text{N}_2$</td>
<td></td>
<td>(T5)</td>
<td>(T5)</td>
</tr>
<tr>
<td>Dissociative recombination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{O}_2^+ + e \rightarrow \text{O}^* + \text{O}^{**}$</td>
<td></td>
<td>(T1)</td>
<td>(T1)</td>
</tr>
<tr>
<td>$\text{NO}^+ + e \rightarrow \text{N}^* + \text{O}^{**}$</td>
<td></td>
<td>(T2)</td>
<td>(T2)</td>
</tr>
<tr>
<td>$\text{N}_2^+ + e \rightarrow \text{N}^* + \text{N}^{**}$</td>
<td></td>
<td>(T3)</td>
<td>(T3)</td>
</tr>
</tbody>
</table>

Simple scheme for positive-ion formation and decay

\[
\begin{align*}
\text{R}_1 & \quad \text{O}_2^+ \rightarrow \text{O}, \text{O} \\
\text{Q}_1 & \quad \text{T}_1 \quad \text{O} \rightarrow \text{O}^+ \\
\text{T}_2 & \quad \text{NO}^+ \rightarrow \text{N}, \text{O} \\
\text{Q}_2 & \quad \text{T}_2 \quad \text{N}_2 \rightarrow \text{N}_2^+ \rightarrow \text{N}, \text{N}
\end{align*}
\]

may exist. Orders of magnitude of the coefficients $\gamma$ and $\alpha$ have been given in Table I. We do not quote numerical data here, since these are continually being improved, and the discordance among the results of different experiments is gradually reduced. Lists of experimental results have been given by Paulson (1964) and Ferguson et al. (1965).

The formation and decay scheme sketched below Table II is the simplest which could, in principle, account for the observed ion compositions. The relevant continuity equations have been given by Yonezawa and Takahashi (1960) and Shimazaki (1965). If we relate this scheme to the coefficients $\alpha$ and $\beta$ of the first-order theory (Sec. 3.61), we find that the effective $\alpha$ is a complex combination of the coefficients $\alpha$ and $\gamma$; but that the effective $\beta$ in the F2 layer is simply

\[
\beta(h) = \gamma_{\text{T}_1} n[\text{O}_2] + \gamma_{\text{T}_2} n[\text{N}_2]
\]  

(365)
3.7 Airglow and the Ionosphere

There is a long-standing difficulty concerning the rates of reactions (T1) and (T2), which seem to be too fast to account for actual F-region behavior (Bates and Nicolet, 1960; Swider, 1965). With improved experimental data (e.g., Ferguson et al., 1965; Copsey et al., 1966; Warneck, 1967) the difficulties have been reduced but not necessarily removed, as we shall discuss further in Sec. 5.3.

The data of Figs. 29 and 30 show that N$_2^+$ ions are scarce in the E and F regions; it might, therefore, be supposed that they are destroyed so rapidly by the recombination (R3) that the process (Q2) does not contribute to the observed ionization, even though it does contribute to the heat input. However, a quantitative study shows that (R3) is not rapid enough to account for the scarcity of N$_2^+$ ions. For this reason reactions, such as (T3), (T4), (T5), or others, must be important (Norton et al., 1963). We therefore include (T3) in our scheme, since rate coefficient data indicate that it is the most effective. At night, production of O$^+$ by (Q1) ceases, and the transition level between atomic and molecular ions occurs higher than by day. The remaining O$^+$ ions at this level must be supplied either by other processes or by diffusion from higher levels (Walker and McElroy, 1966).

3.7 Airglow and the Ionosphere

Some of the processes that we have described lead to the emission of light in the ionosphere, and thus contribute to the airglow. The airglow is taken to exclude the polar aurora, but to include emissions in the mid- and low-latitude ionosphere due to a variety of causes. Among the causes of airglow are photochemical reactions of neutral and ionized constituents of the atmosphere, notably the production of excited atoms by dissociative recombination reactions, particle bombardment, and electric fields that can accelerate thermal electrons which collide with and excite other particles. Of course, the airglow constitutes a very large field of research, of which only a part concerns ionospheric physics. Comprehensive accounts of the airglow have been given by Bates (1960) and Chamberlain (1961), and the relation between airglow and the ionosphere has been reviewed by L. Thomas (1967).

Most of the observational data on the airglow have been obtained from ground-based detectors, at night or at twilight. Techniques also exist for separating the partly-polarized scattered light of the lower atmosphere from the unpolarized airglow, and thereby measuring the intensity of some strong airglow lines by day. Rocket experiments provide a means of determining the heights at which the various emissions originate, since these cannot be
III. PHOTOCHEMICAL PROCESSES IN THE IONOSPHERE

reliably determined from the ground. Photographs showing the airglow layers have also been obtained by several astronauts in manned spacecraft; see for instance those of Gillett et al. (1964).

Airglow intensity is usually measured in *rayleighs* (R) or *kilorayleighs* (kR); 
1R = 10^6 photons cm\(^{-2}\) column s\(^{-1}\), or 10\(^{10}\) photons m\(^{-2}\) s\(^{-1}\).

Most of the energy of the airglow, especially in the infrared, is contained in the band spectra of hydroxyl, OH. There are also weak band spectra of O\(_2\) and some weak continuum radiation; hydrogen lines; and metallic features such as the sodium D lines, 5890/5896 Å. These radiations originate at around 70 to 100 km altitude, and are excited by photochemical processes which do not seem very relevant to the ionosphere.

The First Negative system of N\(_2^+\) comprises the (0 – 0) band at 3914 Å, the (0 – 1) band at 4278 Å, and higher order bands. Observations of the 3914 Å radiation in twilight can be used to obtain data on the N\(_2\) content of the upper atmosphere (Chamberlain and Sagan, 1960). Apart from photo-ionization, these radiations are produced during the ionization of N\(_2\) by fast electrons or protons. For particles in the keV range, it is often supposed that about 1 ionization in 50 leads to emission of a 3914 Å photon, and that each ionization consumes 35 eV, so that the conversion efficiency of particle energy to 3914 Å radiation is about 0.002 (Chamberlain, 1961; Rees, 1963; Dalgarno, 1964a). In the absence of aurora, 3914 Å emission is extremely weak; observations in the Soviet Union and Canada at magnetic latitudes 60–65° give a background intensity of about 10 R, from which is deduced an upper limit of corpuscular energy input of 0.3 erg cm\(^{-2}\) s\(^{-1}\) (0.3 mW m\(^{-2}\)). This does represent an appreciable contribution to ionospheric ionization, in comparison to the few erg cm\(^{-2}\) s\(^{-1}\) provided by solar EUV radiation. However, at rather lower latitudes, the upper limits that are placed on the 3914 and 4278 Å nightglow lead to the very low limit of 10\(^{-2}\) erg cm\(^{-2}\) s\(^{-1}\) (10\(^{-2}\) mW m\(^{-2}\)) for corpuscular input (Dalgarno, 1964a).

3.71 THE FORBIDDEN LINES OF ATOMIC OXYGEN AND NITROGEN

The visible airglow lines of atomic oxygen and nitrogen are “forbidden” by the selection rules of atomic physics, and result from transitions out of long-lived states. The lines of greatest interest correspond to the following transitions; the energies above ground states and the lifetimes of the upper states are also shown:

| Oxygen green line: | 1S_0 \rightarrow 1D_2 (4.2 eV) | (2.0 eV) 5577 Å 0.74 s |
| Oxygen red lines: | 1D_2 \rightarrow 3P_2,3P_1 (2.0 eV) | 0 eV 6300/6364 Å 110 s |
| Nitrogen lines: | 2D_{5/2,5/2} (2.4 eV) \rightarrow 4S_{5/2} (0 eV) | 5198/5201 Å 26 hr |
For oxygen, the transitions to the third level of the ground state \( ^3P_0 \) are too weak to be of interest, as is the direct transition between \(^1S\) and \(^3P\) at 2972 Å. The ratio of the 6300 and 6364 Å intensities is fixed, namely 3:1, and we quote only the intensity of the 6300 Å line. Typical values for the night airglow are given by Chamberlain (1961) as 50–100 R for 6300 Å, and 250 R for 5577 Å; the nitrogen lines are extremely weak. The 6300 Å dayglow amounts to a few kilorayleighs.

Rocket observations, first made by Heppner and Meredith (1958), show that the 6300 Å emission originates above 160 km altitude, whereas the 5577 Å emission is mostly located between 85 and 120 km. This lower range of altitude is consistent with the hypothesis of Chapman (1931c) that the green line is excited by the three-body collision reaction

\[
O + O + O \rightarrow O_2 + O(^1S)
\] (366)

If this hypothesis is correct, then the 5577 Å emission has little connection with ionospheric processes. Nor is such a connection observed; the green line emission is patchy and variable, and shows little correlation with magnetic or ionospheric parameters. From the energy-level scheme given above, however, one might expect 5577 Å emission to be followed by 6300 Å emission at the same altitude. The long lifetime of the \(^1D\) state, 110 s, provides the reason why this does not happen. At the levels where the 5577 Å radiation originates, collisions are so frequent that the \(^1D\) atoms are almost always deactivated by collision before they radiate a red line photon.

During the day, excited oxygen atoms are produced by photodissociation of \( O_2 \) by radiation short of 1750 Å (Sec. 1.4). The resulting contribution to the 6300 Å dayglow might amount to 10 kR (Brandt, 1958), though the measured dayglow intensities appear to set an upper limit of 1 kR, and it is presumed that collisional deactivation is responsible for this lack of 6300 Å emission (Dalgarno and Walker, 1964).

A further important mechanism for the production of excited atoms is dissociative recombination. In Table II, Sec. 3.6, we listed three such reactions; the energies liberated in these reactions are as follows:

\[
\begin{align*}
\text{O}_2^+ + e & \rightarrow O + O + 7.0 \text{ [eV]} \quad (R \ 1) \\
\text{NO}^+ + e & \rightarrow N + O + 2.8 \text{ [eV]} \quad (R \ 2) \\
\text{Ne}^+ + e & \rightarrow N + N + 5.8 \text{ [eV]} \quad (R \ 3)
\end{align*}
\]

In principle (R1) could excite both the \(^1D\) and \(^1S\) states; but from the virtual absence of 5577 Å emission from 150 km, we conclude that only the \(^1D\) state is actually excited. The excitation of the \(^1D\) state by (R2), though
energetically possible, violates the conservation of spin (Dalgaro and Walker, 1964); though Peterson et al. (1966) suggest that subsequent reactions can lead to the production of \(^1\)D atoms. Both (R2) and (R3) can lead to the production of excited \(^2\)D nitrogen atoms, though because of the excessively long lifetime of this state (26 hr), collisional deactivation is far more probable than radiation, even at F-region heights. This is no doubt the reason for the extreme weakness of the 5198/5201 Å emission.

Since most of the red line airglow originates in the F region (being partly excited by dissociative recombination), one would expect its intensity to be related to F-region parameters. According to the first-order theory of F-region photochemistry, described in Sec. 3.6, the emission from the F region should be proportional to \(\int \beta N \cdot dh\), evaluated throughout the region. A more detailed theory, including deactivation and other processes, has been given by Chamberlain (1961) and Peterson et al. (1966). In practice, a good correlation between the red line and the F region is found only at low magnetic latitudes, where a relatively bright "tropical arc" of red line emission (Barbier and Glaume, 1962) is found to correspond roughly with enhanced F2 electron concentrations (Sec. 5.45). The correspondence has been detected by means of shipboard airglow observations (Nakamura, 1961). Barbier (1957) has given an empirical formula relating red line intensity to the virtual height and peak electron concentration of the F2 layer, at night in tropical regions. It may be written

\[
J(6300) = K_1 + K_2 (N_m F2) \cdot \exp \left[ \frac{(200-h')F}{H} \right]
\]

in which \(K_1\), \(K_2\), and \(H\) are determined from the data for any given night; they show some variation from night to night. Apart from the "background" term \(K_1\), this expression resembles the expected form of the quantity \(\int \beta N \cdot dh\), though—as we noted in Sec. 2.21—virtual height may not be closely related to the real height of the layer. The values of the coefficient \(K_2\), obtained from the detailed analysis of red line and ionospheric data from Hawaii (Peterson et al., 1966; Peterson and Steiger, 1966), imply that the efficiency of production of red line emission by recombination of F-region ionization is only about 0.1 per ion pair. The component \(K_1\) represents a background, originating from other processes. In mid-latitudes the correlation between red line emission and F2-layer parameters is poor (e.g., Duncan, 1960b), and other processes must be operating. Thermal excitation by electrons probably contributes to the red line dayglow (Dalgaro and Walker, 1964); supporting evidence is found from the predawn enhancement of 6300 Å radiation, attributed to an influx of photoelectrons along geomagnetic field lines when the
3.7 AIRGLOW AND THE IONOSPHERE

sun rises on the conjugate ionosphere (Dalgarno, 1964a; Cole, 1965; Carlson, 1966).

Another interesting phenomenon is the “stable mid-latitude red arc” reported by Barbier (1958), and subsequently found to be confined to a narrow range of magnetic latitudes in both hemispheres. A review of the properties of this phenomenon has been given by Roach and Roach (1963). The arcs are thought to be situated at 300–500 km height, and may possibly be related to F-region perturbations. One instance of correlation with the outer radiation belt was reported by O'Brien et al. (1960). The photochemical reactions described earlier cannot account for all the properties of the arcs, and it may be that electron excitation is the prime cause, perhaps by electric fields (Megill and Carleton, 1964).
4.1 The Transport Term in the Continuity Equation

In this chapter we shall study some of the processes that lead to motion of the ionization. For many purposes, we can think of the electrons and ions as constituting a gas—a plasma—which is a minor constituent of the atmosphere. Like the other gases, the plasma is subject to gravity and collisional forces, but it is also acted upon by electric and magnetic forces. To study the effects of these forces, we have to construct separate equations of motion for the ions and electrons. We shall then find that, with the aid of some reasonable assumptions, we can derive a plasma drift velocity \( V \). The quantity \( NV \) represents the flux of electrons (or ions) due to transport, and its divergence represents the resulting rate of loss per unit volume and unit time. This is the transport term in the continuity equation of Sec. 3.1, which we write as

\[
\frac{\partial N}{\partial t} = q - l(N) - \text{div}(NV) \tag{400}
\]

This is not unlike the heat conduction equation (120) of Sec. 1.33. Before we attempt to solve this equation, or even to evaluate the transport term, it is worth discussing its nature. As it stands, the equation contains derivatives with respect to space and time. Although quite rapid horizontal motions may exist, they do not necessarily contribute much to the term \( \text{div}(NV) \) because the horizontal gradients of \( N \) and \( V \) are usually much smaller than vertical gradients. Horizontal variations generally involve scale distances of hundreds or thousands of kilometers, but vertical scales are only tens of kilometers. So we can often retain just the vertical contribution to the transport term, and write

\[
\text{div}(NV) = \frac{\partial}{\partial h} (NW) \tag{401}
\]
where \( W \) is the upward drift velocity. Exceptions to this rule may arise either when horizontal gradients are especially large (as near sunrise) or when there are special conditions to limit the importance of vertical motions (as in the equatorial ionosphere).

Negative ions are scarce in the parts of the ionosphere where transport processes have any importance (except perhaps in the nighttime E region). To a good approximation, we can ignore their presence. The concentrations of positive ions and electrons can then be equated; because, even when electric polarization charges are developed, the difference between negative and positive charged particle concentrations is only about one part in \( 10^{10} \) (Johnson and Hulburt, 1950). The rate of change of this charge is also insignificant, so that if an electric current flows it must be nondivergent. Assuming every ion to carry charge \( +e \), we can specify the following condition for the ionospheric current density \( j \):

\[
\text{div} \, j = e \, \text{div} (N_i V_i - N_e V_e) = 0
\]

(402)

If \( N_i = N_e \), it makes no difference whether the ion or the electron drift velocity is used in the continuity equation (400). Apart from occasions when we need to distinguish between different kinds of positive ion, we can take \( N_i = N_e = N \) and use (400) for electrons or positive ions.

4.11 FORMING THE EQUATIONS OF MOTION

We now describe the various transport processes, and then form the equations of motion for electrons and ions. In the E and F regions the motions of charged particles are more or less controlled by the earth's magnetic field. The constraint can be expressed by including the appropriate Lorentz \((V \times B)\) terms in the equations of motion. At this stage, we shall not need to consider the properties of the magnetic field itself in any detail, but the magnetic flux density \( B \) and the inclination or dip angle \( I \) will appear in many of our equations.

These equations are analogous to the equation of motion of the neutral air, Eq. (123) in Sec. 1.61. In using them, we are assuming that collisions between particles are so frequent that the ions and electrons can both be treated as fluids. This implies that the random thermal velocities (about 1 km s\(^{-1}\) for the ions and 200 km s\(^{-1}\) for the electrons) cancel out so completely that we can ignore their existence. Normally, this assumption is very well justified for thermal particles in the ionosphere.

The important transport processes may be enumerated as follows. First, ions and electrons can be moved by electric fields. The resulting motions and
electric currents depend on the magnetic field and the collision frequencies, which determine the mobility and electrical conductivity of the charged particles (Sec. 4.2). We cannot yet discuss the origin of the electric fields, as they depend on the “atmospheric dynamo” and other processes intimately concerned with geomagnetism (Chapter VII).

Second, the charged particles can be moved by neutral air winds. These produce forces proportional to the differences between the wind velocity $U$ and the charged particle velocities $V_i$, $V_e$, and to the relevant collision frequencies. In the E region, the winds are mainly associated with the tidal motions described in Sec. 1.7; since these winds also produce the “dynamo” action mentioned in the previous paragraph, the physical situation is quite complicated. For this reason, we include both winds and fields in our treatment of particle motion in Sec. 4.2. In the F region the winds are driven by the daily heating and cooling of the atmosphere (Sec. 1.63); the consequences of the “ion-drag” interaction between the air and the ions is discussed in Sec. 4.24.

Third, the daily temperature changes in the thermosphere affect the charged particles as well as the neutral air. Hence, the plasma takes part in the thermal expansion and contraction of the atmosphere (Sec. 4.44).

Fourth, the plasma (like any other gas) tends to diffuse under the action of gravity and of gradients in its own partial pressure. The electrical forces between ions and electrons tend to keep them together, so that both kinds of particles diffuse at the same speed (unless strong electric fields, capable of separating them, are applied to the plasma). This simple type of “plasma” or “ambipolar” diffusion is impeded by collisions of the charged particles with neutral particles, and tends to be constrained by the earth’s magnetic field. Diffusion proceeds rapidly in the F region but more slowly in the lower ionosphere where collisions are more frequent. We develop the equations for diffusion in Sec. 4.3, and discuss their application to the continuity equation in Sec. 4.4. A particular form of diffusion is the interchange of plasma between the ionosphere and the protonosphere; this process involves chemical charge-exchange reactions, and is dealt with in Sec. 4.5.

In the equations of motion, we can write the partial pressures of the ions and electrons as $p_i = N_i k T_i$ and $p_e = N_e k T_e$, following Eq. (100). The collisional terms include parameters $v_{in}$, $v_{en}$, $v_{ei}$ which we will discuss in Sec. 4.12. The accelerations are set equal to zero, being generally negligible for the large-scale motions that we are considering (they might not be negligible for small-scale wavelike motions). Eq. (403) gives the force per particle acting on positive ions, and can be adapted for negative ions by changing the sign
4.1 THE TRANSPORT TERM IN THE CONTINUITY EQUATION

of the electromagnetic term; Eq. (404) applies to electrons:

\[ m_i \frac{dV_i}{dt} = 0 = m_i g - N_i^{-1} \nabla (N_i kT_i) + e (E + V_i \times B) \]

\[ - m_i v_{ia} (V_i - U) - m_e v_{ei} (V_i - V_e) \]

(403)

\[ m_e \frac{dV_e}{dt} = 0 = m_e g - N_e^{-1} \nabla (N_e kT_e) - e (E + V_e \times B) \]

\[ - m_e v_{en} (V_e - U) - m_e v_{el} (V_e - V_i) \]

(404)

It now seems appropriate to explain the collision terms in these equations in more detail, since they are required at many points later in the chapter.

4.12 THE COLLISION TERMS IN THE EQUATIONS OF MOTION

The rigorous treatment of collisions between charged and neutral particles is very complicated (Chapman and Cowling, 1952), and the simple "free path" approach used in this book cannot be expected to be very accurate. We have written the collision terms in Eqs. (403) and (404), the equations of motion, in a style commonly found in the literature. The force per unit volume experienced by particles of a type \( a \), \( F_{ab} \), due to collisions with particles of a type \( b \), is written

\[ F_{ab} = - F_{ba} = n_a m_a v_{ab} (V_b - V_a) \]

(405)

so that the acceleration of particles of type \( a \) due to collisions with particles \( b \) may then be written

\[ \left[ \frac{\partial V_a}{\partial t} \right]_{ab} = v_{ab} (V_b - V_a) \]

(406)

Throughout this chapter we use the term "collision frequency" in the sense implied by Eq. (405). The parameter \( v_{ab} \) does not represent the real frequency of collisions between particles \( a \) and particles \( b \), but it is regarded as a coefficient indicating the rate of transfer of momentum. The terms "effective collision frequency" and "frictional frequency" are sometimes used to describe this parameter. However, some authors use a different collision parameter, which is more closely related to the actual kinetic theory frequency of collision, and replace the mass \( m_a \) by the "reduced mass" \( m_a m_b / (m_a + m_b) \). In the case of ion-neutral collisions, involving particles of like masses, the insertion of "reduced mass" in Eq. (405) leads to a doubling of the collision frequency parameter. This difference may not be too important, in view of the limitations of "free path" theory, but larger differences can arise when collisions between electrons and heavy particles are considered.

We can define a rate coefficient \( K_{ab} = K_{ba} \) to describe the rate of momentum transfer per unit volume. In a plasma containing ions, electrons, and neutrals, there are three such coefficients (namely, \( K_{in}, K_{en}, K_{ei} \)) and six "effective col-
lision frequencies". Writing $N$ for electron or ion concentration, $n$ for neutral concentration, and assuming $m_i = m_n \gg m_e$, we have

$$
\begin{align*}
\nu_{in} &= nK_{in}, & \nu_{ni} &= NK_{in} \\
\nu_{en} &= nK_{en}, & \nu_{ne} &= \left(\frac{m_e}{m_n}\right) NK_{en} \\
\nu_{ei} &= NK_{ei}, & \nu_{ie} &= \left(\frac{m_e}{m_i}\right) NK_{ei}
\end{align*}
$$

(407)

Three of these parameters have already been used in Eqs. (403) and (404). Numerical values can be found from formulas given by Chapman (1956). In the present notation these give

$$
\begin{align*}
K_{in} &\equiv \frac{\nu_{in}}{n} = 2.6 \times 10^{-9} M^{-1/2} \quad \text{[c.g.s.]} = 2.6 \times 10^{-15} M^{-1/2} \quad \text{[m.k.s.]} \\
K_{en} &\equiv \frac{\nu_{en}}{n} = 5.4 \times 10^{-11} T^{1/2} \quad \text{[c.g.s.]} = 5.4 \times 10^{-16} T^{1/2} \quad \text{[m.k.s.]} \\
K_{ei} &\equiv \frac{\nu_{ei}}{N} = \left\{34 + 4.18 \log_{10} \left(\frac{T^3}{N}\right)\right\} T^{-3/2} \quad \text{[c.g.s.]} \\
&\quad = \left\{59 + 4.18 \log_{10} \left(\frac{T^3}{N}\right)\right\} 10^{-6} T^{-3/2} \quad \text{[m.k.s.]} \quad (408)
\end{align*}
$$

The temperature $T$ is assumed equal for all species, and $M$ is the ion and neutral particle mass expressed in a.m.u. Dalgarno (1964b) gives values of $\nu_{in}$, for atomic oxygen, which agree with Chapman’s formula at 700°K but which vary approximately as $T^{0.4}$. A more general formula for $\nu_{en}$ has been given by Dalgarno et al. (1967).

Table III gives some numerical values based on these equations, and on representative ionospheric and atmospheric data for a moderate level of solar activity. Separate “day” and “night” values are shown for $\nu_{ei}$, $\nu_{ie}$, and $\nu_{ne}$, for which the day-to-night change of $N$ produces significant changes. The quantities $\nu_{in}$ and $\nu_{en}$, however, depend primarily on the neutral concentration $n$, whose day-to-night change is too small to be worth including in the Table. We also show, for future reference, the reciprocals of $\nu_{ni}$ and representative mid-latitude values of the magnetic gyrofrequencies $\omega_i$ and $\omega_e$. We omit the quantity $\nu_{ne}$ because it appears to possess no practical importance.

For calculating ionospheric electrical conductivities and in problems of wave propagation, the following quantities are often used:

$$
\begin{align*}
\nu_i &= \nu_{in} + \nu_{ie}, & \nu_e &= \nu_{en} + \nu_{ei}
\end{align*}
$$

(409)

The term “effective collision frequency” is sometimes applied to these quantities. Generally speaking, $\nu_{ie}$ makes a trivial contribution to $\nu_i$, but we see from Table III that $\nu_{ei}$ makes a major contribution to $\nu_e$ in the F region.
<table>
<thead>
<tr>
<th></th>
<th>Height, km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Day</td>
</tr>
<tr>
<td>Neutral gas concentration ( n ) ( \text{m}^{-3}, \text{10}^{-6} \text{cm}^{-3} )</td>
<td>1.2 \times 10^{19}</td>
</tr>
<tr>
<td>Electron concentration ( N ) ( \text{10}^{11} \text{m}^{-3}, \text{10}^{5} \text{cm}^{-3} )</td>
<td>1.7</td>
</tr>
<tr>
<td>Gas temperature ( T ) ( ^{\circ} \text{K} )</td>
<td>210</td>
</tr>
<tr>
<td>Mean molar mass ( M ) ( \text{a.m.u} )</td>
<td>28</td>
</tr>
<tr>
<td>Collision frequencies ( v_{n} ) ( \text{s}^{-1} )</td>
<td>5800</td>
</tr>
<tr>
<td>( v_{e} ) ( \text{s}^{-1} )</td>
<td>92000</td>
</tr>
<tr>
<td>( v_{el} ) ( \text{s}^{-1} )</td>
<td>2300</td>
</tr>
<tr>
<td>( v_{le} ) ( \text{s}^{-1} )</td>
<td>0.045</td>
</tr>
<tr>
<td>( v_{nt} ) ( 10^{-5} \text{ s}^{-1} )</td>
<td>8.3</td>
</tr>
<tr>
<td>( v_{nt}^{-1} ) ( \text{hr} )</td>
<td>3.3</td>
</tr>
<tr>
<td>Gyrofrequencies ( \omega_{i} ) ( \text{rad s}^{-1} )</td>
<td>160</td>
</tr>
<tr>
<td>( \omega_{e} ) ( 10^{8} \text{ rad s}^{-1} )</td>
<td>8.4</td>
</tr>
<tr>
<td>( v_{n}/\omega_{i} )</td>
<td>3.6 \times 10^{1}</td>
</tr>
<tr>
<td>( v_{e}/\omega_{e} )</td>
<td>1.1 \times 10^{-3}</td>
</tr>
</tbody>
</table>

* Separate "day" and "night" values are given only if they differ markedly; otherwise average values are given. The model atmosphere of CIRA (1965) is adopted; \( M \) refers to both neutral particles and ions, and \( T \) to all constituents. The ground level magnetic field is taken as \( \frac{1}{4} \text{ gauss (G)} = 5 \times 10^{-5} \text{ tesla (T, or Wb m}^{-2}\text{), a typical mid-latitude value; in quantities derived from the magnetic field, "day" values are chosen for the other parameters involved. Numerical values have been rounded. \)
The precise relevance of these quantities, however, has to be justified for any particular problem.

4.2 Electrical Conductivity

From the basic equations of motion given in Sec. 4.1, we can determine the motions of electrons and ions in an applied electric field. This analysis leads to formulas for the electrical conductivity of the ionosphere, as obtained by Cowling (1945), Baker and Martyn (1952, 1953), Chapman (1956), and others. The conductivity is anisotropic because of the geomagnetic field.

For any charged particle we shall use the following symbols: mass \( m \), charge \( q (\pm) \), electron charge \( -e \), angular gyrofrequency \( \omega = B|e|/m \) (always positive; we here dispense with the subscript \( H \) used in Chapter II), and collision frequency with neutral particles \( v \). Suffixes \( i \) and \( e \) are used for positive ions and electrons. At present we ignore negative ions, since they probably play no important role at heights where the conductivity is important; that is, in the E region.

We take \( v \) to be the “effective collision frequency” as used in Eq. (405), though for brevity we omit the subscript \( n \) throughout this section. We are not at present interested in gravity, pressure-gradient forces, or electron-ion collisions. Then the basic Eqs. (403) and (404) can be greatly simplified, and we can write for any charged particle acted on by electric and magnetic fields

\[
m \frac{dV}{dt} = eE + eV \times B - mv(V - U)
\]

As before, \( V \) is the charged particle velocity and \( U \) the neutral air velocity. At first, we take \( U = 0 \). If \( E \parallel B \), the motion is unaffected by the presence of the magnetic field. For the more interesting case of \( E \perp B \), the motions are as indicated in Fig. 31 where, for convenience, we have omitted the last term in Eq. (410) and made the assumption (consistent with the way \( v \) is defined) that the particle is brought to rest at equal intervals \( 1/v \). The trajectories are then cycloidal, and are shown for different values of \( v/\omega \). Both ions and electrons move in the direction \( E \times B \) if \( v/\omega = 0 \), and in general, their motion is inclined at an angle \( \tan(\omega/v) \) to the direction of \( E \). In subsequent equations, we may include a nonzero wind \( U \).

4.21 The Motion of a Single Particle

For a single particle we can find the average drift velocities, over periods long compared with both \( 1/v \) and \( 1/\omega \), by setting \( dV/dt = 0 \) in Eq. (410). Hence, in a Cartesian coordinate system \((\xi, \eta, \zeta)\) in which \( B \) is directed along
Fig. 31. Idealized trajectories for electrons and ions subject to an electric field in the plane of the diagram and a magnetic field directly out of the plane. Charged particles are assumed to collide with neutral particles at regular intervals $1/\nu$, and to possess zero velocity after each collision. In order to show both electronic and ionic gyrations, the diagrams are drawn as though $\omega_e/\omega_i = 10$ (instead of order $10^4$ as in reality). All the trajectories refer to equal intervals of time, namely 5 ionic (or 50 electronic) gyroperiods. Numbers in brackets refer to approximate heights, in kilometers, at which the conditions occur. In general, the head of the vector representing $V_i$ or $V_e$ lies on a semicircle.

the $\zeta$-axis

$$0 = F_\zeta - mv\zeta + eV_\zeta B$$
$$0 = F_n - mv\eta - eV_\eta B$$
$$0 = F_\zeta - mv\zeta$$  \hspace{1cm} (411)

where

$$F = eE + mvU$$  \hspace{1cm} (412)

is the applied force. The solution of these equations can be compactly written in terms of a tensor $k$ which satisfies the equation $V = k \cdot F$. It is

$$k = \begin{bmatrix} k_1 & \pm k_2 & 0 \\ \mp k_2 & k_1 & 0 \\ 0 & 0 & k_0 \end{bmatrix}$$  \hspace{1cm} (413)

in which the upper signs apply to positive ions, the lower signs to electrons (or negative ions). Each component of $k$ is the ratio of a velocity to a force,
IV. TRANSPORT PROCESSES IN THE IONOSPHERE

and is therefore a "mobility per unit charge." We find this quantity more useful than the conventional mobility, defined as a ratio of velocity to electric field, since our \( k \) is valid for both electrical and mechanical forces.

Our three quantities \( k_0, k_1, k_2 \) are essentially positive. By using \( \omega = B|e|/m \) we can write each of them in two forms, one including \( B \) explicitly. The names shown in Eq. (414) are the customary names for the corresponding components of mobility (for example, the "direct" mobility is \( e/mv \)):

Direct, longitudinal: \((\parallel B, \parallel E), k_0 = \frac{1}{mv} = \frac{1}{Be} \frac{\omega}{v} \)

Transverse, Pedersen: \((\perp B, \parallel E), k_1 = \frac{1}{mv} \frac{v^2}{v^2 + \omega^2} = \frac{1}{Be} \frac{\nu \omega}{v^2 + \omega^2} \) \hspace{1cm} (414)

Hall: \((\perp B, \perp E), k_2 = \frac{1}{mv} \frac{\omega v}{v^2 + \omega^2} = \frac{1}{Be} \frac{\omega^2}{v^2 + \omega^2} \)

At any given height in the ionosphere, \( \omega/v \) is not the same for ions as for electrons. The height variation of the mobilities for one kind of carrier is

Fig. 32. Mobilities per unit charge for a single species, as a function of reduced height \( z \). It is assumed that the collision frequency \( \nu \propto e^{-z} \) and that the gyrofrequency \( \omega \) is height independent. The level \( z = 0 \) is taken where \( \nu = \omega \). The longitudinal, transverse, and Hall mobilities \((k_0, k_1, k_2)\) are plotted in units of \( 1/Be \). The combination of two such diagrams, one each for electrons and ions, placed at their appropriate heights, gives the conductivity per ion pair, as in Fig. 33.
shown in Fig. 32 in which altitude is measured from the level where \( v = \omega \), and the logarithmic mobility scale is referred to the value \( 1/Be \).

Following Weekes (1957), we can use Eq. (414) to discuss how charged particles move in response to an applied electric field for which \( F = eE \), and a neutral wind for which \( F = mvU \). Only fields and winds that remain constant for periods greatly exceeding \( v^{-1} \) (for both ions and electrons) need be considered.

**Electric Field and Wind Parallel to the Magnetic Field:** An electric field causes a drift \( eE/mv \), opposite in sense for positive ions (\( e^+ \)) and electrons (\( e^- \)). This is an electric current along \( B \) (and is carried mainly by electrons, for which \( k_o \) is greater than for ions).

A wind causes a drift \( F/k_0 = mvU/mv = U \). Hence, at all heights, both ions and electrons are driven along a field line at a speed equal to the wind component parallel to \( B \).

**Electric Field and Wind Perpendicular to the Magnetic Field:** If \( v \gg \omega \) then \( k_1 = k_0 \gg k_2 \). The magnetic field is of no importance; an electric field produces a current parallel to itself, and a wind carries ions and electrons at its own velocity. This situation applies to all charged particles in the lower ionosphere, below about 75 km.

If \( v \sim \omega \) then \( k_1 \sim k_0 \), and both fields and winds produce drift velocities inclined to themselves. This gives rise to the interesting situation that exists in the ionospheric E region.

If \( v \ll \omega \) then \( k_1 \ll k_2 \ll 1/Be \), and \( k_2 \) is almost independent of height. An electric field causes both electrons and ions to drift in the same direction \((E \times B)\) with speed \( E/B \). A wind produces a Hall drift of speed \( (v/\omega)U \), in the direction \((U \times B)\) for positive ions and the opposite direction for electrons. This corresponds to an electric current. This situation applies to all charged particles above about 200 km.

The motions due to an electric field normal to \( B \) have already been illustrated in Fig. 31.

Among the above are three cases in which the drift velocities for ions and electrons are equal, so that we may speak of a plasma drift velocity. These are first, when \( v \gg \omega \) for all particles and a wind \( U \) drives the plasma with its own velocity, irrespective of direction.

Second, at any height, the field-aligned component of plasma drift due to wind may be written vectorially as:

\[
V_w = (U \cdot B) B/B^2
\]  

(415)

Third, when \( v \ll \omega \) for both ions and electrons, an electric field produces a
plasma drift velocity given by

$$V_E = E \times \mathbf{B}/B^2$$  \hspace{1cm} (416)

4.22 Conductivity of the Ionospheric Plasma

To find the direct-current conductivity tensor $\sigma$, we combine the ionic and electronic motions, using the equations

$$j = \sigma \cdot E = Ne(V_i - V_e)$$  \hspace{1cm} (417)

where $j$ is the current density and $N$ the electron and ion concentration. It is customary to define four components of conductivity:

- **Longitudinal or Direct:** $\sigma_0 = Ne^2(k_{0e} + k_{0i})$
- **Transverse or Pedersen:** $\sigma_1 = Ne^2(k_{1e} + k_{1i})$
- **Hall:** $\sigma_2 = Ne^2(k_{2e} - k_{2i})$
- **Cowling:** $\sigma_3 = (\sigma_1^2 + \sigma_2^2)/\sigma_1$

![Fig. 33. Conductivities per ion pair ($\sigma_0, \sigma_1, \sigma_2, \sigma_3$) plotted on a logarithmic scale relative to the value $Ne/B$, as a function of reduced height $z$ for an idealized isothermal model atmosphere. It is assumed that collision frequency $\nu \propto e^{-z}$ and gyrofrequency $\omega$ is independent of $z$, for both positive ions and electrons; also that $\nu_{0e}/\nu_{0i} \approx 1000$ (independent of height), this being a reasonable value for the actual ionosphere. The zero of $z$ is the level where $\nu V_i = \omega_i \omega_e$, and the important levels where $\nu_i = \omega_i$ and $\nu_e = \omega_e$ are, therefore, situated at $z = \pm \frac{1}{2} \ln(1000)$; also shown are the approximate altitudes at which they occur in the actual ionosphere.]}
These conductivities are all positive quantities. The electron and ion contributions to \( \sigma_2 \) are of opposite sign, but as we shall see from Fig. 33, \( k_{2e} \gg k_{2i} \) at all heights in the ionosphere. This means that the Hall current flows in the direction of \( B \times E \), even though the particles drift in the direction of \( E \times B \).

Three of these conductivities appear in the tensor expression for \( \sigma \), which follows from our definitions and the expression for \( k \):

\[
\sigma = \begin{bmatrix}
\sigma_1 & -\sigma_2 & 0 \\
\sigma_2 & \sigma_1 & 0 \\
0 & 0 & \sigma_0
\end{bmatrix}
\] (419)

The Cowling conductivity appears at a later stage in the theory. The vertical variations of these quantities are sketched in Fig. 33, in which the contributions of the ions and electrons are shown in units of \( Ne/B \). The resulting sums are thus conductivities per ion pair, and should be multiplied by the electron concentration profile to obtain the actual vertical variation of each \( \sigma \).

The important levels in the diagram are defined by \( \omega_e = v_e \) (about 80 km in the actual ionosphere), \( \omega_e \omega_i = v_i v_e \) (about 105 km) and \( \omega_i = v_i \) (about 140 km). We assume \( (v_i/\omega_i) \approx 1000(v_e/\omega_e) \), this ratio being relatively independent of height. Peaks of \( \sigma_1/N \) occur at the levels \( \omega_e = v_e \) and \( \omega_i = v_i \), but the former does not give an actual peak of \( \sigma_1 \) because of the very small value of \( N \) at 80 km. Numerical values of the conductivities have been given by Chapman (1956), using the collision frequency formulas we quoted in Eq. (408).

If we wish to include negative ions, we can do so by adding an extra term to each of the equations for \( \sigma_0, \sigma_1, \sigma_2 \). For instance, if \( N \) is the electron concentration and \( \lambda \) the negative ion/electron concentration ratio, we can write

\[
\sigma_0 = Ne^2 \left[ k_{0e} + (1 + \lambda) k_{0e}^* + \lambda k_{0i}^* \right]
\] (420)

With our definition of mobility, all three components are positive for negative ions. It is not thought, however, that negative ions contribute significantly to ionospheric conductivity, because they are only abundant at heights where \( v_i \gg \omega_i \).

At great heights (above the F2 peak) the "direct" conductivity \( \sigma_0 \) given by Eq. (418) becomes very large, as \( v_e \) and \( v_i \) become very small. The conductivity then depends on ion-electron collisions instead of ion-neutral collisions. We can evaluate this modified conductivity \( \sigma_0^* \) from either of Eq. (403) or Eq. (404) by deleting all but the electric field and electron-ion collision terms, and taking components parallel to \( B \). The field-aligned current
density $j$ is then given by

$$\frac{eE}{m_e v_e} = V_i - V_e = \frac{j}{N_e}$$

(421)

Hence,

$$\sigma_0^* = \frac{j}{E} = \frac{N e^2}{m_e v_e} = \frac{e^2}{m_e K_{ei}}$$

(422)

The formulation in terms of the collision parameter $K_{ei}$ (Eq. (408)) is included to show that $\sigma_0^*$ is independent of $N$, except for the slowly-varying logarithmic term in $K_{ei}$.

### 4.23 Layer Conductivities

Baker and Martyn (1953) have discussed the height-integrated "layer" conductivities in the actual ionosphere. These arise because of the limited vertical extent of the conducting layer in the E region, and are used in the dynamo theory of magnetic variations (Sec. 7.4). If an electric field is generated in the ionosphere by dynamo action, the resulting current $\sigma \cdot E$ may contain a vertical component. If so, charges will accumulate at the boundaries of the conducting layer because the current cannot flow into the region of low conductivity. These "polarization" charges will modify the electric field $E$ until the resulting flow is horizontal. Thus, if we impose the condition of zero vertical current, we can eliminate the vertical electric field from the equations, and we find that the $3 \times 3$ tensor $\sigma$ can be replaced by a $2 \times 2$ tensor $\sigma'$, representing the "layer conductivity" whose components depend on the magnetic dip angle $I$. Using coordinates $x, y$ for the magnetic southward and eastward directions, we can write the layer conductivity as

$$\sigma' = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{yy} \end{bmatrix}$$

(423)

where:

$$\sigma_{xx} = \frac{\sigma_0 \sigma_1}{\sigma_0 \sin^2 I + \sigma_1 \cos^2 I} \approx \frac{\sigma_1}{\sin^2 I}$$

(424)

$$\sigma_{xy} = \frac{\sigma_0 \sigma_2 \sin I}{\sigma_0 \sin^2 I + \sigma_1 \cos^2 I} \approx \frac{\sigma_2}{\sin I}$$

$$\sigma_{yy} = \frac{\sigma_2^2 \cos^2 I}{\sigma_0 \sin^2 I + \sigma_1 \cos^2 I} + \sigma_1 \approx \sigma_1$$

The approximations given at the right arise because generally $\sigma_0 \gg \sigma_1$ or $\sigma_2$, but they are not valid near the magnetic equator where $I = 0$, and where we
4.2 ELECTRICAL CONDUCTIVITY

have

\[ \sigma_{xx} = \sigma_0, \quad \sigma_{xy} = 0, \quad \sigma_{yy} = (\sigma_1^2 + \sigma_2^2)/\sigma_1 = \sigma_3 \quad (425) \]

Two consequences follow from these latter results. First, the very high conductivity \( \sigma_{xx} = \sigma_0 \) along the field lines ensures that these lines are approximately electric equipotentials. Second, the east-west conductivity \( \sigma_{yy} \) at the equator is extremely large, being the Cowling conductivity \( \sigma_3 \) which is comparable to \( \sigma_0 \). This highly conducting strip along the magnetic equator carries a large current known as the “equatorial electrojet,” which is confined to the region a few degrees in width, where \( \sigma_0 \sin^2 I \ll \sigma_1 \cos^2 I \).

The simple model of the dynamo region is therefore a relatively thin, horizontally stratified layer. The effects of the currents as observed at the ground can be calculated most conveniently with the use of “integrated layer conductivities,” \( \Sigma_1 \equiv \int \sigma_1 \, dh \) and \( \Sigma_2 \equiv \int \sigma_2 \, dh \), where the integration is made vertically through the conducting region. Since the major part of the conductivity is found in the E region, \( \Sigma_1 \) and \( \Sigma_2 \) are approximately proportional to the peak electron concentration \( N_m E \). At night, they become small because the E layer almost disappears.

4.24 ACCELERATION OF THE NEUTRAL AIR

So far we have envisaged the neutral air to be either stationary or moving with a fixed velocity \( U \). We shall now consider what happens when the neutral air motion is caused by collisions between the neutral particles and charged particles, when the latter are moved by electric fields. This is the converse of the situation discussed above in Sec. 4.21, concerning the motions of charged particles due to neutral air winds. For all practical purposes, we need only consider collisions between ions and neutrals, since collisions between electrons and neutrals have no appreciable effect on the motion of the neutrals. This “ion-drag” is important in the F region.

The equation for acceleration of the air can be adapted from Eq. (406) of Sec. 4.12. Assuming the velocities to be horizontal, and adding a term for the viscous drag of the air (as in Sec. 1.61, where \( \mu \) is the coefficient of viscosity, \( \rho \) is air density), we have

\[
\frac{\partial U}{\partial t} = v_{ni}(V_i - U) + \frac{\mu}{\rho} \frac{\partial^2 U}{\partial h^2} \quad (426)
\]

This is equivalent to the horizontal components of the equation of motion for the neutral air, Eq. (123), though most of the terms given in that equation are either negligible or irrelevant in our present discussion.
For the moment we neglect viscosity \((\mu = 0)\). Then Eq. (426) shows that, if the ion velocity \(V_i\) is constant, the air is accelerated to this velocity with a time constant \(v_{ni}^{-1}\). From Table III, we see that this time is of order hours in the F region, and therefore comparable to the period of the daily large-scale motions of the atmosphere.

To discuss the effect of viscosity, it is simplest to assume a steady state \((\partial U/\partial t = 0)\). Then Eq. (426) can be rewritten as

\[
H_v^2 \frac{\partial^2 U}{\partial h^2} - U = -V_i
\]

where

\[
H_v = (\mu/\rho v_{ni})^{1/2} = (\mu/\rho nK_{in})^{1/2}
\]

If \(V_i\) and \(H_v\) are independent of height and time, the vertical variation of \(U\) can be expressed in terms of the exponential functions \(\exp(\pm h/H_v)\), subject to appropriate boundary conditions. If \(V_i\) and \(H_v\) do vary with height (but not too fast) we may still expect the variation of \(U(h)\) to be related to \(H_v\). This “scale distance” was introduced by Hines (1960) and Rishbeth et al. (1965), and may be interpreted as a minimum scale size within which ion-drag can maintain a large variation of \(U\) against the effects of viscosity. Roughly speaking, viscosity tends to remove rapidly the difference between the velocities \(U(h_1)\), \(U(h_2)\) at two levels \(h_1\), \(h_2\) if \(|h_1 - h_2| \ll H_v\), but does not have much influence on the velocity difference between two levels separated by more than \(H_v\). In discussing whether the neutral air is likely to be accelerated in the F region, we are interested in the values of \(H_v\) and \(v_{ni}^{-1}\) between say 150 and 300 km. Below 150 km, the time constants \(v_{ni}^{-1}\) are many hours, and there is little chance of significant acceleration due to ion-drag; any air motions are likely to be due to other causes, such as tidal winds or gravity waves. Above 300 km, \(H_v\) is so large (about 100 km) that the neutral air moves bodily, with little vertical variation of velocity (this being confirmed by the detailed wind computations described in Sec. 1.63). Between these limits, two situations arise:

\textit{Day:} Typically \(v_{ni}^{-1} \sim 1\) hr and \(H_v \sim 30\) km at 200–300 km. Both quantities are small enough for the neutral air to be fully accelerated by ion-drag, without much hindrance from viscosity.

\textit{Night:} Typically \(v_{ni}^{-1} \sim 5\) hr and \(H_v \sim 100\) km. Both quantities are so large that the acceleration of the neutral air is slow, and seriously affected by viscosity.

We conclude that motions of the F-region plasma can accelerate the neutral
air effectively by day but not by night, the difference being due to large day-to-night variation of the ratio $N/n$ at heights of 150–250 km. This process of acceleration of the air is known as ion-drag or induction drag. Its importance in the ionosphere was first considered by Cowling (1945) and Martyn (1953a). In detailed studies of this process, the air is generally assumed to be immovable vertically, except for transient motions; the electromagnetic forces driving the plasma are not powerful enough to disturb appreciably the normal hydrostatic distribution of the air. If the air moves only horizontally, then when the air velocity and plasma velocity have become equalized by the drag, both velocities must be horizontal. Thus, as pointed out by Hirono and Kitamura (1956) and by Dougherty (1961), ion-drag tends to suppress the vertical component of the motion of the plasma.

Fig. 34a. Electromagnetic drift, without ion-drag, due to an electrostatic field $E$ into the plane of the diagram. The plasma drift velocity $V = (E \times B)/B^2$ has an upward component $V \cos I$ and northward component $V \sin I$.

Fig. 34b. Ion-drag attains a steady state when the neutral air (assumed to move only horizontally) has been accelerated to a northward speed $V \csc I$. This air motion drags the plasma with speed $V \cot I$, along field lines, and the resultant of the plasma motions is equal to the neutral air velocity; i.e., northward with speed $V \csc I$.

Fig. 34c. Air-drag. A southward wind $U$ is assumed to exist in the neutral air. In a steady state, the velocity imparted to the plasma is $U \cos I$ along the field lines; there is an upward component $U \cos I \sin I$ and a southward component $U \cos^2 I$.

Fig. 34d. Plasma diffusion, under the influence of gravity and a vertical pressure gradient, causes a vertical velocity $W_D$ (Eq. 431) in the absence of a magnetic field. With a magnetic field (but no air motion) the plasma is constrained to diffuse only along field lines, with velocity $W_D \sin I$, of which the vertical component is $W_D \sin^2 I$. The horizontal component of ion motion reacts on the air, and if the latter is free to move horizontally, it is accelerated. The diagram portrays the steady-state solution of the appropriate equations of motion, in which the horizontal speeds of the air and plasma are the same ($W_D \cot I$); the vertical plasma speed is $W_D$ and its total drift speed, parallel to $B$, is $W_D \csc I$. 


We can draw sketches to illustrate some situations arising in the F region with an inclined magnetic field. Figure 34 illustrates ion-drag and air-drag, drawn for dip latitude 45°N, at which the geomagnetic field \( B \) is inclined at a dip angle \( I = \arctan \frac{2}{1} = 63° \). It is assumed that \( v \ll \omega \) for both ions and electrons, which drift with very nearly the same velocity, shown by full lines. Dashed lines show neutral air motion; dotted lines indicate geometrical constructions. Not all the results illustrated are obvious, but they can be verified by solving the detailed equations of motion, as Dougherty (1961) has done. Sketch (a) shows the components of the drift velocity \( V \) produced by an electric field \( E \) normal to the magnetic field \( B \), given by Eq. (416) in Sec. 4.21. In Sketch (b) the neutral air is assumed to be fully accelerated by ion-drag, and its horizontal motion drags the ions downward along the field line. The electrons may be assumed to follow because of electrostatic forces. The combination of electromagnetic drift and the drag by the neutral air gives a resulting plasma velocity which is horizontal; it is equal to the neutral air velocity and is shown by the dashed horizontal line. We could say that Sketch (a) applies to night, and Sketch (b) to day, though the actual situations will be intermediate to these. A more quantitative discussion has been given by Kohl (1963).

Sketch (c) shows the situation produced by a horizontal wind \( U \). The ions are dragged along the magnetic field lines, with a velocity equal to the wind component in that direction, \( U \cos I \). Again, the electrons may be assumed to follow the ion motion. The plasma motion has a vertical component \( U \cos I \sin I \), upwards if the wind blows towards the magnetic equator, and downwards if it blows towards the magnetic pole. The horizontal plasma velocity is \( U \cos^2 I \), so that the horizontal wind and plasma velocities differ by \( U \sin^2 I \). Thus, for motions in the magnetic meridian, the drag force depends on the dip angle which must therefore be taken into account in the equation of motion for the neutral air (Sec. 1.61). But winds in the perpendicular direction, normal to \( B \), produce no plasma motion and the drag force is independent of \( I \). (Throughout the above discussion we have neglected the Hall currents produced by the wind, as described in Sec. 4.21 for the case \( v \ll \omega \), since they are a second-order effect.)

The remaining sketch in Fig. 34 relates to plasma diffusion, and will be discussed in Sec. 4.32.

4.3 Plasma Diffusion

The importance of plasma diffusion was suggested by Hulburt (1928) and the mathematical expressions were derived by Ferraro (1945). In early calcu-
lations of its effects, diffusion was treated as a small perturbation (Mariani, 1956). But Yonezawa (1956) and Ratcliffe et al. (1956) discussed in detail the problem of the formation of the F2 peak and showed that diffusion could provide an explanation.

4.31 The Basic Diffusion Equations

The mathematical theory of diffusion has been treated by Chapman and Cowling (1952). Here we give a simple derivation, based on the equations of Sec. 4.11. At first we consider only vertical diffusion, and assume the geomagnetic field to be vertical. Rearranging and simplifying Eqs. (403) and (404), and using \( W \) to denote vertical drift velocity (with suffixes \( i, e, n \) for ions, electrons, neutrals), we have

\[
d(N_i k T_i) / dh = - N_i m_i g + N_i e E - N_i m_i v_{in} (W_i - W) \tag{429}
\]

\[
d(N_e k T_e) / dh = - N_e m_e g - N_e e E - N_e m_e v_{en} (W_e - W) \tag{430}
\]

We further simplify these equations by the following assumptions:

\( m_i \gg m_e, N_i = N_e = N, W_i = W_e = W_p \) (plasma drift velocity), \( W_n = 0 \) (neutral air at rest), \( m_i v_{in} \gg m_e v_{en} \) (collisions with neutral particles important for ions but not for electrons). On adding the equations, the terms in \( E \) vanish (and, because of Newton’s third law, so do the ion-electron collision terms which we omitted from the equations). This means that the electrons interact with the ions only via the electric field. Rearranging to solve for the drift velocity of the plasma

\[
-W_p = \frac{1}{m_i v_{in}} \left\{ \frac{1}{N} \frac{d}{dh} \left[ N k (T_i + T_e) \right] + m_i g \right\} \tag{431}
\]

The theory of diffusion in the ionosphere was originally developed with the assumption that all kinds of particle involved possess the same temperature. In fact the electron and ion temperatures are significantly different, especially by day, in the F region. To take account of this difference, we find it convenient to define the “plasma temperature” \( T_p \) and the “plasma scale height” \( H_p \) by the equations:

\[
T_p = \frac{1}{2} (T_i + T_e); \quad H_p = 2k T_p / m_i g \tag{432}
\]

We can reconcile this definition of \( H_p \) with our usual definition of scale height, \( H = k T / mg \) as in Eq. (103), if we regard the plasma as a gas in which the mean particle mass is \( \frac{1}{2} m_i \), on account of the negligible mass of the
electrons. After some rearrangement, we can write (431) in the form

$$- W_D = D \left[ \frac{1}{N} \frac{dN}{dh} + \frac{1}{T_p} \frac{dT_p}{dh} + \frac{1}{H_p} \right]$$

(433)

where we introduce the "plasma diffusion coefficient" $D$ and define it as $\frac{2kT_p}{m_i v_{in}}$. The collision frequency $v_{in}$ is proportional to the neutral gas concentration $n$ and depends on the gas temperature $T$; we may therefore write

$$D = \frac{k(T_i + T_e)}{m_i v_{in}} = \left(1 + \frac{T_e}{T_i}\right) \left[b(T) / n\right]$$

(434)

We shall assume the ion and neutral gas temperatures to be equal throughout this section. According to simple kinetic theory, the function $b(T) \propto T^{1/2}$; in reality the temperature dependence is more complicated, and probably lies between $T^{1/2}$ and $T^1$ (Ferraro, 1945, 1957). For ions diffusing through their parent gas (as happens in the F2 layer, where the ions are mostly $O^+$ and the neutrals are mostly O), the value of $b$ is somewhat reduced by the occurrence of charge-exchange reactions between the ions and the neutrals, though this reduction is smaller than was once thought (Dalgarno, 1958, 1964b). According to the calculations of Knof et al. (1964) and Dalgarno (1964b), $b \simeq 7 \times 10^{18} \text{ cm}^{-1} \text{ s}^{-1}$ at 1000°C; thus $D = (2 \times 10^{19} / n) \text{ cm}^2 \text{ s}^{-1}$ for the reasonable daytime condition of $T_e / T_i = 2$.

Readers should note that many papers, including those by Ferraro and Dalgarno, give values only of the diffusion coefficient for ions, which is $kT/m_i v_{in}$. The so-called "ambipolar diffusion" coefficient is twice as great (Johnson and Hulburt, 1950; Ratcliffe and Weekes, 1960); it applies to the diffusion of the plasma in the situation $T_i = T_e = T$. The "plasma diffusion" coefficient that we have defined by Eq. (434) is applicable to the more general situation of $T_i \neq T_e$ which occurs in the F region.

Thermal diffusion has been neglected in our analysis, but is included in the more detailed treatment by Chapman and Cowling (1952). Its inclusion would modify the temperature gradient terms in the diffusion equations, but its effects are probably only perceptible for the lightest ions in the ionosphere (Chapman, 1958; Walker, 1967).

The vertical flux of electrons (or ions) due to diffusion is $NW_D$. Its divergence is the contribution made by diffusion to the continuity equation (400), and can generally be expressed in the form

$$- d(NW_D)/dh = D \mathcal{D} N$$

(435)

where $\mathcal{D}$ is a differential operator. We shall only derive $\mathcal{D}$ for the simple
case in which \( T_i = T_e = T \) and \( m_i = m_e \) so that \( H_p \) is just twice the scale height \( H \) of the neutral gas from which the ions are formed. We can express \( \mathcal{D} \) in terms of reduced height \( z \) by taking \( (H \, dH/dh) = dH/dz \). For this simple model, the neutral particle concentration \( n \propto e^{-z} \) so that we can write \( D = D_0 e^z \). Using these relations, Eq. (433) reduces to

\[
-W_D = \frac{D_0 e^z}{H} \left[ \frac{1}{N} \frac{dN}{dz} + \frac{1}{2} \right] \tag{436}
\]

From Eq. (435), the diffusion term in the continuity equation takes the form derived by Ferraro (1945):

\[
D \mathcal{D} N = \frac{D_0 e^z}{H^2} \left[ \frac{d^2N}{dz^2} + \frac{3}{2} \frac{dN}{dz} + \frac{N}{2} \right] \tag{437}
\]

The effect of gravity and the height dependence of \( D \) are responsible for the appearance of terms in \( dN/dz \) and \( N \), in addition to the second derivative \( d^2N/dz^2 \) which is characteristic of diffusion formulas. Shimazaki (1957) has developed the diffusion formulas for the case in which \( H \) varies linearly with height.

Since \( D \) increases exponentially upward, whereas the other coefficients \( q \) and \( \beta \) in the continuity equation decrease upward, at some level diffusion dominates so completely that the continuity equation reduces to \( \mathcal{D} N = 0 \). The solution of this equation is

\[
N = A_1 e^{-z/2} + A_2 e^{-z} \tag{438}
\]

The first term corresponds to diffusive equilibrium, with \( W_D = 0 \), such that the ionization assumes a scale height twice that of the ionizable gas. For the second term, \( W_D \neq 0 \); and this represents a boundary condition of a finite flux of ionization at \( z \to +\infty \), which is upward if the coefficient \( A_2 > 0 \). If ionization is lost or gained by diffusion at the top of the ionosphere (i.e., to or from the magnetosphere), the \( N(z) \) distribution should contain a component of this type.

### 4.32 Diffusion in an Inclined Magnetic Field

Let us now consider the geomagnetic field to be inclined at an angle \( I \) to the horizontal. In the F region, the plasma can only diffuse along the magnetic field line, in which direction the component of gravity is \( g \sin I \). The field-aligned drift velocity \( W_\parallel \) is then given by a modification of Eq. (436),
IV. TRANSPORT PROCESSES IN THE IONOSPHERE

namely

\[- W_{\parallel} = \frac{D_e e^2}{H} \left[ \frac{1}{N} \frac{dN}{ds} - \frac{1}{2} \sin I \right] \]  

(439)

in which \( s \) denotes reduced distance (measured in units of \( H \)); the signs imply that both \( W_{\parallel} \) and \( s \) are reckoned positive northwards, so that \(-dz/ds = \sin I\). The differential operator \( \mathcal{D} \) can now be defined by an equation similar to Eq. (435), namely

\[- d(NW_{\parallel})/H \, ds = D_{\parallel} N \]  

(440)

At mid-latitudes and high latitudes, \( \sin I \) can be treated as constant along any given field line, throughout the vertical extent of the F region. In this case the same expression is obtained as previously, namely Eq. (437), except that the right-hand side contains an additional factor of \( \sin^2 I \). At low latitudes, however, \( I \) varies appreciably along a field line; rather complicated expressions for \( \mathcal{D} \), in which the coefficients are functions of magnetic latitude, can be derived by using the geometry of a dipole field (Kendall 1962; Lyon, 1963).

So far this analysis has assumed the neutral air to be stationary. What happens at mid-latitudes when the air is accelerated horizontally, as considered in Sec. 4.24, is shown in Sketch (d) of Fig. 34. We assume the plasma pressure gradient to be such that the diffusion velocity is directed vertically downwards, and given by \( W_D \) in Eq. (431). The component parallel to \( B \) is then \( W_D \sin I \), of which the vertical component is \( W_D \sin^2 I \).

If the neutral air is free to move horizontally it is accelerated to a speed \( W_D \cot I \). The field-aligned plasma velocity now contains a component due to drag by the air, \( W_D \cot I \cos I \), which adds to the original diffusion component \( W_D \sin I \), to give a total velocity \( W_D \csc I \) along \( B \). The vertical component of this is \( W_D \), just the same vertical velocity as would occur in the absence of the magnetic field (Dougherty, 1961; Kendall and Pickering, 1967).

4.33 Diffusive Equilibrium

The equations for the diffusion of ions and electrons, Eqs. (429) and (430), serve as a convenient starting-point for discussing the equilibrium distribution of charged particles in the topside F region. At these heights, photochemical processes can be neglected, and the ion-neutral collision frequency is so small that the continuity equation reduces to the partial pressure, gravitational, and electric field terms of Eqs. (429) and (430). The situation is interesting because several species of positive ion may be present. We
suppose all ions to be singly charged and to possess the same temperature $T_i$, though this need not be the same as the electron temperature $T_e$. For the $j$th species of ion, and for the electrons, the equations are

$$d(N_j k T_i)/dh = - N_j m_j g + N_j e E \quad (441)$$

$$d(N_e k T_e)/dh = - N_e m_e g - N_e e E \quad (442)$$

Since $\sum N_j = N_e$ so that $\sum (dN_j/dh) = dN_e/dh$, all these equations can be added to eliminate $E$, the small gravitational term $N_e m_e g$ being neglected. As in Eq. (432), let $T_p = \frac{1}{2}(T_i + T_e)$, which we can call the "plasma temperature"; also let $T_e/T_i = \tau$. We now denote the mean positive ion mass by $m_+$, so that $m_+ N_e = \sum N_j m_j$. Then

$$2 d(N_e k T_p)/dh = - N_e m_+ g \quad (443)$$

On dividing throughout by $(-2N_e T_p)$, we obtain an equation from which we can derive the "plasma scale height," $H_p$:

$$- \frac{1}{N_e T_p} \frac{d(N_e T_p)}{dh} = \frac{m_+ g}{k T_i (1 + \tau)} = \frac{m_+ g}{2 k T_p} = \frac{1}{H_p} \quad (444)$$

From this point we neglect the temperature gradient for simplicity. The upward electric field $E$ may be found from Eq. (442); it is given by

$$e E = \tau m_+ g/(1 + \tau) \quad (445)$$

If we further simplify the equations by assuming thermal equilibrium ($T_e = T_i = T$ so that $\tau = 1$), we can find the "scale heights" of the plasma and of each ion species from Eqs. (441) to (444). They are

$$H_p = - N_e \frac{dh}{dN_e} = \frac{2kT}{m_+ g} \quad (446)$$

$$H_j = - N_j \frac{dh}{dN_j} = \frac{kT}{(m_j - \frac{1}{2} m_+) g} \quad (447)$$

If we wish, we can express ionic masses as $M_j$ (in a.m.u.) provided that the Boltzmann constant $k$ is replaced by the gas constant $R$. We see that if only one ionic species is present, its "effective scale height" is twice that of the neutral species of atomic mass $m_+$ (as before). However, a light ion for which $m_j < \frac{1}{2} m_+$ actually has a negative scale height, so that $N_j$ increases upward, as discussed by Mange (1960).

Eqs. (441) and (442) can be solved by a method due to Hanson (1962) and a specimen equilibrium distribution computed for a mixture of $O^+$, $He^+$,
Fig. 35. Idealized distributions of electrons (e) and of O⁺, He⁺, and H⁺ ions, computed by solving Eqs. (441), (442). Electron and ion concentrations are given in terms of the electron concentration \( N_{eo} \) at the level \( z = 0 \), at which height the ionic composition is arbitrarily chosen to be \( N[O^+] = N[He^+] \approx 0.49 \, N_{eo}, \quad N[H^+] = 0.02 \, N_{eo} \). The level at which the He⁺ and H⁺ concentrations are equal is near \( z = 17 \). The unit of reduced height \( z \) is the scale height of neutral atomic oxygen.

and H⁺ ions, is shown in Fig. 35. It is convenient to use the “geopotential height” defined by Eq. (112) in Sec. 1.22, in order to take account of the vertical variation of \( g \). At great heights, the centrifugal force due to the earth’s rotation can be combined with the gravitational force (Angerami and Thomas, 1964).

We can illustrate Eqs. (441) and (442), for a single ion species, by means of a force diagram, Fig. 36. The vertical electric field is essential for equilibrium; it maintains an upward force on the ions and a downward force on the electrons, so that both species conform to the same scale height \( H_p \) in spite of their very different masses. The existence of this field implies the existence of electric charges somewhere, so that charge neutrality cannot be exact, although it does hold to an extremely high degree of approximation. In the diagram all the vectors are vertical. Unless the pressure gradient and electric field vectors are aligned with the gravity vector, equilibrium cannot be maintained without additional forces. We thus conclude that diffusive equilibrium can only exist if there is horizontal stratification (Chandra, 1964). In general, the ionosphere is not horizontally stratified and a more detailed study shows that small horizontal electric currents flow in
4.4 Solving the Continuity Equation

Much of the progress in ionospheric theory has been achieved by obtaining solutions of the continuity equation and comparing them with observed ionospheric behavior. The full continuity equation is too complicated to be solved in general, so in this section we shall discuss the solutions of some simplified versions of it. We start with the simplified equation

\[
\frac{\partial N}{\partial t} = q - l(N) - \partial (NW)/\partial h
\]  

(448)

in which only the vertical component of plasma velocity is included. We
take account of the following quantities, which will often be expressed in terms of reduced height \( z \) measured in units of \( H \) (the scale height of the ionizable gas):

- the production term, as derived in Secs. 3.2 and 3.3;
- the loss term, as discussed in Sec. 3.6;
- the plasma diffusion term, derived in Sec. 4.3;
- the electromagnetic drift velocity, \( V_E = \frac{E}{B} \times B \), due to an applied electric field \( E \) (Eq. 416). If ion-drag is neglected, the upward component is

\[
W_E = \left( \frac{E}{B} \right) \cos I
\]  
(449)

where \( y \) indicates the magnetic eastward component (Fig. 34);

- the air-drag velocity, \( V_w = (U \cdot B) \frac{B}{B^2} \), due to a neutral air wind \( U \), Eq. (415). Its vertical component is

\[
W_w = U_x \cos I \sin I
\]  
(450)

where \( x \) indicates the magnetic southward component (Fig. 34).

Equation (448) is a partial differential equation in \( h \) and \( t \). It may be simplified by following the motion of a "cell" of ionization, using the total derivative appropriate to vertical motion, namely

\[
d/dt \equiv (\partial/\partial t) + \frac{W}{\partial h}
\]  
(451)

in which we have specified \( dh/dt = W \). Insertion of this operator into the continuity equation yields

\[
dN/dt = q - I(N) - N \frac{\partial W}{\partial h}
\]  
(452)

The usefulness of these equations lies in the fact that if \( \partial W/\partial h \) can be neglected, as may be the case for electromagnetic drifts and air drag, Eq. (452) reduces to a purely photochemical equation. However, as the altitude of the "cell" changes, \( q \) and \( I \) vary so that the equation must be solved numerically, in general. Shimazaki (1957) has described the use of this technique for electromagnetic and thermal motions. For thermal motions, the vertical velocity is given by the analysis of Sec. 1.23, Eq. (114). The implicit assumption, that the vertically moving air carries the plasma with it, is justified by the results of Sec. 4.21, except near the magnetic equator.

4.41 The Continuity Equation with Photochemical Terms Only

In Chapter III, we dealt with the situation of photochemical equilibrium, in which the continuity equation reduces to \( q = I(N) \), which may be used to
approximate the daytime situation in the D, E, and F1 layers. In particular, we discussed the Chapman layer and the transition between the “square law” and “linear” types of loss coefficient, which occurs in the F1 layer. If time variation is included in the equation, a first order ordinary differential equation is obtained, which can be solved analytically in some cases, otherwise numerically (Millington, 1932). The most interesting feature of the solutions is the time lag which occurs at noon, between the times of maximum \( q \) and maximum \( N \); this time constant also concerns eclipse phenomena, and the decay of the layer at night. It has been called the “sluggishness of the ionosphere” by Appleton (1953), and is given by

\[
\text{Square Law Loss: } \Delta t = (2\alpha N)^{-1} \\
\text{Linear Loss: } \Delta t = \beta^{-1}
\]

\( \Delta t \) may also be regarded as an order-of-magnitude “lifetime” of any one ion pair. These time lags do not apply if the coefficients \( \alpha \) and \( \beta \) vary with time, as would be the case in an atmosphere undergoing thermal expansion, owing to the change of molecular concentration at fixed heights. If \( \beta \) increases with time, maximum \( N \) can be attained before noon. This is illustrated by the calculations of Norton and Van Zandt (1964), who were able to reproduce observed \( N(t) \) curves, characteristic of the low-latitude F region, by using a model of a thermally expanding atmosphere. Their equations contained only the photochemical terms, as is appropriate for the low-latitude E region and lower F region where vertical motions are thought to be comparatively unimportant.

4.42 “Day Equilibrium” and “Night Stationary” Diffusion Layers

When diffusion is introduced into the continuity equation, as must be done if the F2 peak is to be adequately treated, the continuity equation contains the diffusion operator \( \partial \) which involves \( \partial/\partial h \) and \( \partial^2/\partial h^2 \). Equilibrium solutions of this equation are of some interest, even for the F2 layer because it is true during much of the day that \( \partial N/\partial t \) is smaller than other terms in the continuity equation. At least they provide an insight into the relative importance of different processes.

Equilibrium solutions of the diffusion equation, with electromagnetic drift, were obtained by Yonezawa (1956, 1958). Rishbeth and Barron (1960) solved the equation numerically in a number of cases, using a Chapman-type production function and exponential formulas, \( D \propto e^2 \) and \( \beta \propto e^{-Kz} \), for the diffusion and loss coefficients. We recall from Sec. 3.6 that \( K = 2 \) and \( K = 1.75 \) if the loss reactions depend respectively on molecular oxygen and
molecular nitrogen in a diffusively separated atmosphere, and that $K = 1$ for the simpler—though unrealistic—assumption that all atmospheric gases have the same scale height. More detailed calculations relating to the equilibrium situation have been made by Bowhill (1962).

We can make some generalizations about behavior of a “day equilibrium” layer. In particular, we are interested in the peak electron concentration $N_m$ and the height of the peak, denoted by $h_m$ (in real height) or $z_m$ (in reduced height). The subscript $m$ distinguishes any quantity evaluated at the “peak” or “maximum” of the layer. The generalizations are:

(a) The F2 peak occurs at a level where diffusion and loss are of comparable importance, i.e., where $\beta_m \sim D_m/H^2$.

(b) At the peak the electron concentration is given by $N_m \sim q_m/\beta_m$, just as it would be in the absence of diffusion. (Also $N \sim q/\beta$ at heights below the peak.)

(c) Well above the peak, the electron distribution is exponential, and takes the form $N \propto e^{-z/2}$, as we obtained in Eq. (438).

Fig. 37. Equilibrium electron distributions $N(z)$ for the F2 layer, $z$ being reduced height. In the absence of drift, the peak $z_m$ lies at a level where $\beta \sim D/H^2$ (full curve). A vertical drift $W$ moves the peak to a level $z_m \pm \Delta z$, the displacement $\Delta z$ being of order $WH/D_m$, positive for upward drift and negative for downward drift; the corresponding $N(z)$ curves are shown by broken lines. For all curves, $N \propto e^{-z/2}$ (diffusive equilibrium) at heights well above the peak, and $N \propto q/\beta$ (photochemical equilibrium) at heights well below the peak.
4.4 SOLVING THE CONTINUITY EQUATION

(d) A vertical drift $W$, due to winds or electric fields, alters the level of the peak by approximately $WH/D_m$ scale heights; $N_m$ is still given, very roughly, by (b).

These four points are illustrated by the sketch in Fig. 37. The vertical thickness of an equilibrium layer is typically a few scale heights. The above conclusions are substantially unaltered if the scale height varies with height. They are also substantially unchanged if the electron and ion temperatures are unequal (L. Thomas, 1966a).

When $q = 0$, it is possible to find "shape preserving" solutions of the diffusion equation of the form

$$N(z, t) = N_s(z) e^{-\beta' t}$$

which decay without change of shape, with an "effective decay coefficient" $\beta'$. Usually we are only interested in the most slowly decaying layer, which has the smallest value of $\beta'$. It may be termed a "night stationary" layer, and has the following properties, parallel to those listed above for the "day equilibrium layer":

(a') The F2 peak is situated at a level where $\beta_m \sim D_m/H^2$.

(b') The effective decay coefficient $\beta' \sim \beta_m$, the loss coefficient at the height of the peak.

(c') Well above the peak, the electron distribution takes the form $N \propto e^{-z/2}$.

(d') A vertical drift $W$, due to electric fields or winds, alters the level of the peak by approximately $WH/D_m$ scale heights; the effective decay coefficient is still given very roughly by (b').

It so happens that in the special case investigated by Duncan (1956) and Martyn (1956), namely $\beta \propto e^{-z}$ (or $K = 1$), the "shape preserving" distribution is mathematically identical to the "Chapman alpha layer" [which arose previously in an entirely different context, namely, the equilibrium between photoionization and square-law loss, Eq. (328)]. It also happens that in this case $\beta' = \beta_m$. In other cases, such as $K = 2$ (Dungey, 1956a) and $K = 1.75$ (Hanson and Patterson, 1964), the analysis is more complicated and $\beta'$ depends somewhat on the drift velocity $W$. Generally $\beta' > \beta_m$, so that the decay of the layer corresponds to the value of $\beta$ at a height slightly below the peak. We also find that, for any fixed $K$, the peak of the "night stationary" layer lies a fraction of a scale height above the peak of the "day equilibrium" layer. This is because the production of ionization, which is most rapid below the peak in the daytime layer, ceases at night. Hence, there is a numerical difference between the statements (a) and (a') although these are
qualitatively the same. The actual values of the numerical coefficients, which are required to complete the order-of-magnitude statements we have listed above, are found in the papers referenced above; they have been summarized by Rishbeth (1967b).

It is sometimes convenient to define a "diffusion rate" (inverse time constant), given by the function \( D/H^2 \) which appears in the above discussion. Denoting this by the symbol \( d \), we can specify the level of the F2 peak by the relation

\[
D/H^2 = d \sim \beta
\]  

(456)

Since \( \beta \) and \( d \) vary exponentially with height, in opposite senses, \( \beta/d \) decreases rapidly upwards, by several per cent per kilometer. Therefore an approximate value of this ratio suffices to define the equilibrium height of the peak quite accurately. The quantity \( d^{-1} \) represents a "diffusive time constant," within which we may expect diffusive equilibrium to be established in the upper F region.

4.43 THE TIME-VARYING "PRODUCTION-LOSS-DIFFUSION" MODEL

Time-varying solutions of the full diffusion equation, giving the theoretical diurnal variation of electron concentration, have been obtained by Gliddon and Kendall (1960, 1962, 1964), Briggs and Rishbeth (1961), and Rishbeth (1963). Various additional influences, such as drifts and solar eclipse effects, have been included in some of this work.

Such calculations provide a simple description of the diurnal behavior of a layer controlled by photoionization, linear loss, and diffusion. At sunrise, the electron concentration near the F2 peak increases at a rate which depends primarily on the production rate \( q \), and diffusion and loss play a secondary role. The height of the peak, which corresponds to the F2 peak in the actual ionosphere, falls because of the rapid production of ionization in the lower F region, and reaches a minimum before noon. By this time, diffusion and loss become important, and for a few hours the layer is not far from equilibrium, in the sense that \( \beta N/\partial t \) is small compared to other terms in the continuity equation. Under these conditions, the peak remains near a level where \( \beta \sim D/H^2 \), as in the case of the "equilibrium layer" discussed earlier. Also, at heights up to the peak, the "equilibrium approximation" \( N \approx q/\beta \) is valid, and the maximum value of \( N \) lags after noon by a time of order \( 1/\beta(h) \), as given by Eq. (454). Above the maximum, however, diffusion is so rapid that the variation of \( N \) closely follows the variation of \( N_m \) at all times.

Later in the day, \( N_m \) decreases and \( h_m \) rises as solar control weakens. Then,
after sunset, $h_m$ approaches its "night stationary" level and the layer develops into the "shape preserving" form which decays exponentially with time throughout the night, in the manner of Eq. (455). The characteristic behavior of $N_m$ and $h_m$ in the production-loss-diffusion model is illustrated by Fig. 38. The two cases shown use identical production and loss rates; but the larger diffusion coefficient used in case (1), as compared to case (2), leads to lower heights $h_m$, and consequent smaller values of $N_m$ by day and an increased decay rate $\beta'$ at night. These differences are consistent with the principles set out in Sec. 4.42.

4.44 THE EFFECT OF VERTICAL MOTIONS

The principles outlined in Sec. 4.42 can be used in various ways to discuss the effect of vertical motions on the electron distribution. As regards the

---

**Fig. 38.** Variation of peak electron concentration $N_m$ and the height $h_m$ of the peak with local time, for a production-loss-diffusion model. The dashed curve represents the height of peak production. The same production and loss rates, $q$ and $\beta$, are used for both models, but for curves 2 the diffusion coefficient is taken to be one-tenth of its value for curves 1. The reduction of $D$ leads to increases of both $h_m$ and $N_m$ [after Rishbeth (1963)].
effects of thermal expansion and contraction, it is convenient to use the "following-the-cell" approach described at the start of Sec. 4.4. Within any given "cell" (at the upper and lower boundaries of which, the gas pressures $p, p + \Delta p$ remain constant), the gas concentration just varies inversely with temperature during the course of thermal expansion (cf. Sec. 1.23); the resulting variations of $q, \beta, \text{and } D$ within the cell may then be computed quite simply. This type of approach was used in a discussion of temperature effects on the equilibrium F2 layer electron distribution (Garriott and Rishbeth, 1963). Within certain assumptions the shape of the $N(z)$ profile is unchanged by variations of temperature, but $N_m$ varies as $T^{-1/2}$. As mentioned in Sec. 4.41, temperature variations can greatly affect the diurnal variation of $N$ at fixed heights; however, G. R. Thomas and Venables (1967) show that the shape of the diurnal variation of $N_mF2$ is not so greatly influenced by variations of neutral air temperature.

Temperature variations, however, affect $h_m$. Thermal contraction at night nullifies the slight nighttime increase of $h_m$, which occurs on the simple production-loss-diffusion model (Fig. 38). Hence some other mechanism must be responsible for the observed fact that at mid-latitudes, the F2 peak is higher at night than by day (Section 5.4). A possible explanation is provided by meridional winds in the neutral air (Hanson and Patterson, 1964; Kohl and King, 1967). As shown in Fig. 7, the winds at mid-latitudes blow towards the pole by day and towards the equator by night, which by Eq. (450) is consistent with the sense of the day-to-night variation of $h_m$. Approximate numerical calculations (Rishbeth, 1967b), based on the principles set out in this section, show the effect of winds to be of the required order of magnitude. In addition, electromagnetic drifts could be effective at night, but details of their behavior are not well known.

Later in the book, we shall have occasion to discuss the effects of winds and electric fields in connection with specific topics, such as the equatorial ionosphere. Here, we conclude with a general statement, that vertical transport is likely to be important if it moves the plasma through one scale height during its lifetime $\Delta t$, as given by Eqs. (453) and (454). Making very rough estimates, we can say that $\Delta t$ varies from about one minute in the E layer up to a few hours in the F2 layer; and that typical electric fields might cause vertical velocities $W \sim 20 \text{ m s}^{-1}$. Hence, in the E layer, any particular ion pair travels only 1 km in its lifetime, much less than a scale height; but in the F2 layer it may travel 200 km or several scale heights. In addition, the vertical drifts caused by neutral winds in the F2 layer may well exceed those due to electric fields. We thus conclude that transport by winds and drifts...
must be important in the F2 layer. Since by Eq. (456) the time constants
associated with diffusion and with loss are comparable in the F2 layer, the
criterion $W \sim H\beta$ for determining the importance of drift is consistent with
our previous statements, namely (d) and (d') of Sec. 4.42.

4.5 Diffusion between the Ionosphere and Protonosphere

In this section we extend the discussion of diffusion to consider the pro-
tonosphere, the region overlying the ionosphere. For this purpose we use the
term "ionosphere" to refer to the region where heavy ions, such as O+,
predominate over the light ions He+ and H+.

The distributions of ions and electrons, given by Eqs. (446) and (447) and
illustrated by Fig. 35, refer to steady-state conditions in which diffusion is
dominant. Hydrogen and helium ions are also subject to chemical processes
which can appreciably modify their distribution in the topside ionosphere.
Helium ions are destroyed by charge-exchange reactions with O2 and N2,
analogous to the reactions (T1) and (T2) in Table II which destroy O+ ions.
The reactions are so rapid that He+ ions in the topside ionosphere cannot
be assumed to be in diffusive equilibrium (Bauer, 1966a).

For hydrogen ions (protons) the situation is influenced by the accident
that the first ionization potentials of atomic oxygen and atomic hydrogen
are almost exactly equal. Consequently the reactions

$$H^+ + O \rightleftharpoons H + O^+ \tag{457}$$

can proceed extremely rapidly. They tend to establish a chemical equilibrium
in which the concentrations are related by

$$n[H^+] n[O] = \frac{3}{8} n[H] n[O^+] \tag{458}$$

—the factor $\frac{3}{8}$ being derived from statistical considerations (Hanson and
Ortenburger, 1961). Let us examine the distributions of H+ ions in an atmo-
sphere in which O and O+ are the dominant neutral and ionized species.
It is arithmetically convenient to express these in terms of "reduced height"
applicable to a particle of mass 1 a.m.u., which we will call $\zeta$. The unit of $\zeta$
is thus $RT/g$, by Eq. (103). By the ordinary hydrostatic equations, the con-
centration $n[H] \propto e^{-\xi}$ and $n[O] \propto e^{-16\xi}$. Provided O+ is the dominant ion,
its distribution will be $n[O^+] \propto e^{-8\xi}$, with an "effective" scale height twice
that of O. For equilibrium conditions, from Eq. (458)

$$n[H^+] \propto n[H] n[O^+] / n[O] \propto e^{-\xi} e^{-8\xi} e^{+16\xi} \propto e^{+7\xi} \tag{459}$$
If however the $H^+$ ions are in diffusive equilibrium, but $O^+$ ions are nevertheless so dominant that the mean ion mass can be taken as 16 a.m.u., Eq. (447) implies that the scale height of the $H^+$ distribution corresponds to an "effective" mass of $-7$ a.m.u. Hence, $n[H^+] \propto e^{-7z}$, exactly the same distribution as applies to chemical equilibrium. However, in spite of this accidental similarity in the equilibrium distributions, Hanson and Ortenburger point out that the distinction between the "chemical" and "diffusive" situations becomes important when nonequilibrium processes are considered, such as the transfer of ionization between the ionosphere (heavy ion region) and the protonosphere.

We can locate the "critical level," dividing the regions of "chemical" and "diffusive" control, as follows. If $k_c$ is the rate coefficient of the charge-exchange reaction (457), then the average lifetime of a proton is $\left\{ k_c n [O] \right\}^{-1}$. The diffusion of protons is obstructed mainly by Coulomb collisions with the ambient $O^+$ ions, for which a diffusion coefficient $D$ can be defined, being inversely proportional to the proton-oxygen ion collision frequency (cf. Eq. (434)). The time for establishment of diffusive equilibrium may be taken as $H^2/D$, where $H$ is the atmospheric scale height. Hence, the critical level is defined by the relation

$$k_c n [O] \sim D/H^2$$

(460)

An equivalent expression in terms of the mean free paths associated with the charge-exchange and Coulomb collision processes was given by Hanson and Ortenburger (1961). As Bauer (1966b) has remarked, Eq. (460) is analogous to the "F2 peak" equation $\beta \sim D/H^2$ which defines the boundary between the photochemical and diffusive regimes in the ionosphere (Sec. 4.42). The numerical values are such that the "critical level" defined by Eq. (460) lies far above the F2 peak, at 700 km or higher. This is partly because the cross section for Coulomb collisions is orders of magnitude greater than that for the ion-neutral collisions involved in ordinary plasma diffusion. Because of the large cross section, the diffusion of protons through oxygen ions is slow, even in the "diffusive" region above the "critical level," and we may think of the oxygen ion region as a "diffusive barrier" separating the ionosphere from the protonosphere. Hanson and Ortenburger conclude that the protonosphere cannot respond to short-term (or even daily) changes in the ionosphere. Instead, it behaves rather like a fluid floating upon—but immiscible with—the denser ionospheric plasma.

Some diffusion through the "barrier" does occur, however, and its probable amount has been estimated by Hanson and Patterson (1963, 1964) and
Geisler (1967a). Piddington (1964) has remarked that the protonosphere contains a great deal more plasma than the F region, so that if sufficient plasma can diffuse downwards at night the nighttime F region could be maintained. The difficulty with this idea is that the ionosphere is probably the source of the protonospheric plasma, photoionization and other production mechanisms being rather slow in the outer atmosphere. By day the charge-exchange reaction (457) supplies protons which diffuse upward through the "barrier" into the protonosphere; however, this upward flux is subject to a calculable upper limit. Hanson and Patterson compare the protonosphere to a capacitor, and point out that the rate of discharge of O\(^+\) ions into the night F region, at the expense of the H\(^+\) population, cannot consistently exceed the limited daytime rate of supply of H\(^+\) ions. The limiting flux depends on temperature and on the inaccurately known concentration of neutral hydrogen; it is of order \(10^7 \text{ cm}^{-2} \text{s}^{-1}\) or \(10^{11} \text{ m}^{-2} \text{s}^{-1}\), one order of magnitude less than is required to maintain the night F2 layer (Hanson and Patterson, 1964). This requirement could be reduced if the F2 layer were raised by upward drift, which reduces the rate of decay of the layer (Secs. 4.42, 4.44); such a drift may, indeed, be caused by the neutral-air winds, discussed in Sec. 1.63.

Another possible consequence of diffusion along field lines through the protonosphere is the transfer of plasma between the ionospheres in opposite hemispheres; however, this seems too slow to be significant (Rothwell, 1963; Kohl, 1966). We must note that existing calculations of proton diffusion have not considered how it might be influenced by the presence of helium ions.
5.1 The D Region

Until the last few years, most of the information about the height variation of electron concentration and collision frequency in the lowest ionosphere was derived from the indirect evidence obtained from radio propagation experiments of the types described in Sec. 2.3. Nowadays, additional information is obtainable from rocket experiments. Even so, the D region remains only partially explored and we do not have the widespread data that exist for the E and F regions.

Apart from the normal D-region behavior that concerns us at present, various types of disturbance occur in the D region. We merely mention them here, since we find it more convenient to discuss them elsewhere in the book. Thus, Sec. 6.1 deals with the effects which occur in the sunlit hemisphere during solar flares. Other types of disturbance are observed at high latitudes, such as auroral absorption and the “blackouts” observed on high-frequency communication circuits, these being attributable to charged particles (Sec. 8.3). Another type of high-latitude disturbance is the “polar cap absorption” which may begin a few hours after solar flares, and is caused by energetic solar protons. We mentioned these in Sec. 3.53 because of their interest as regards D-region photochemistry.

In Sec. 5.12 we shall mention a quite different phenomenon, the “winter anomaly” in absorption, which is less clearly related to magnetic disturbance. This will lead us into the discussion of relationships between the ionosphere and meteorological phenomena, notably the “stratospheric warmings” which have received much attention in recent years.
5.11 D-REGION BEHAVIOR

The long-wave experiments give information about reflection heights, which vary with the solar zenith angle. Frequencies around 16 kHz are reflected at 70–75 km by day but from the E layer by night (Bracewell et al., 1951); the data have been interpreted in terms of one or more discrete D layers below the E layer (Bracewell and Bain, 1952). The solar control of the D layer can be illustrated by the following relation between phase height of reflection of VLF waves, at nearly vertical incidence, and solar zenith angle:

\[ h(\chi) = h(0) + H \ln \text{Ch}(\chi) \]  

(500)

where \( h(0) \approx 72 \text{ km} \), and the "scale height" \( H \approx 5 \text{ km} \) (with a pronounced semiannual variation). The change from night to day conditions generally starts about an hour before ground sunrise (Bracewell et al., 1951). We discussed the relevance of sunrise effects to D-region photochemistry in Sec. 3.53; they provide evidence that the D region is basically solar controlled. A survey of data on the nighttime D region, and of the physical information deduced from them, has been given by A. P. Mitra (1957).

From the "full-wave" analysis of long-wave data (Sec. 2.32), it is possible to construct \( N(h) \) profiles which, even if not uniquely determined, should be correct in broad outline. This has been done for Cambridge (52°N) by Deeks (1966); we illustrate a few of his results in Fig. 39. The sunspot minimum profiles show the building up of the layer from sunrise to noon. We do not give seasonal variations in the figure, but Deeks' results for noon at sunspot minimum show that at all heights summer values of \( N \) exceed equinox values, which in turn exceed winter values (this excludes the days of anomalous absorption in winter, discussed below). Figure 39 also shows the solar cycle variation at noon. Above about 73 km, the sunspot maximum values of \( N \) are greatest, probably because of the greater input of solar X-rays and Lyman \( \alpha \); but at lower heights, where the ionization is mainly due to galactic cosmic rays, the sunspot minimum values are greatest. This is in accordance with the expected solar cycle variation of cosmic rays noted in Sec. 3.33.

The term "C layer" is sometimes used for the ionization produced by cosmic rays (though we should mention that, in the early days of ionospheric sounding, this term was applied to a supposed ionized layer in the stratosphere). In Figure 39 we see that \( N = 100 \text{ cm}^{-3} \) at 60 km, and Pierce (1963) has estimated that, at night, cosmic ray ionization and collisional detachment could maintain \( N \approx 10 \text{ cm}^{-3} \) at 50 km. Below 50 km the ionization has very little effect on radio propagation because of the high collision
Fig. 39. D-region $N(h)$ profiles over England at March equinox, deduced from long-wave observations at Cambridge (52°N, 0°E). The full curves are for sunspot minimum, at night, morning, and noon (numbers denote local time); the broken curve is for noon at sunspot maximum. The dashed portions at the top and bottom of each curve are very approximate [after Deeks (1966)].

frequency, and it is generally studied within the field of atmospheric electricity rather than ionospheric physics.

Further evidence of the solar control of the D region is obtained from Al absorption data. In England, from a long series of measurements, Appleton and Piggott (1954) deduced the solar zenith angle and solar cycle dependence of the absorption at vertical incidence, $L$. If $L = A/(f + f_L)^2$ as in Eq. (220), the data give the zenith angle dependence as

$$A \equiv L(f + f_L)^2 \propto [Ch(\chi)]^{-0.8}$$

and for equinox noon ($\chi = 52^\circ$) the solar cycle variation is given approxi-
mately, in terms of the sunspot number $R$, by

$$A = 300(1 + 0.01R) \text{ [dB]}$$

(502)

where $A$, $L$ are in decibels and $f, f_L$ are in megahertz (Sec. 2.33). Lauter and Nitzsche (1967) have used $A3$ absorption data to study the seasonal variations in the D region.

On certain days in winter, the absorption is greatly enhanced. A variety of D-region data suggest that this enhancement is due to extra ionization in the D region, which also causes low echoes on ionograms. The occurrence of the "winter anomaly" has been discussed by Dieminger (1952), Appleton and Piggott (1954), and L. Thomas (1962a). Thomas studied A1 absorption data and the ionospheric parameter $f_{\text{min}}$ (which as mentioned in Sec. 2.33 is a sensitive indicator of enhanced absorption). He found that enhanced absorption occurs in patches of roughly 1000 km in size with (on given days) a negative correlation between different geographical zones, such as Western Europe and North America. The phenomenon is absent in low latitudes, but it is found in southern mid-latitude regions during local winter (Dieminger et al., 1967). We consider its possible explanation in Sec. 5.12.

5.12 RELATIONS BETWEEN THE IONOSPHERE AND LOWER ATMOSPHERE

Many observations have been made in the past which suggested some relationship between ionospheric phenomena and conditions in the lower atmosphere. They provided various clues which were not, in general, systematically followed up. A review of some of these observations was made by Bauer (1958).

In recent years, there has come to light a connection between D-region parameters and stratospheric temperature (Bossolasco and Elena, 1963). Later investigations concentrated on the phenomenon known as a "stratospheric warming," in which the temperature at around the 10 mb level (about 30 km) may rise tens of degrees above its normal seasonal value, the increase lasting for some days. A large stratospheric warming, such as the classic one over Central Europe in February 1952, can be tracked over many thousands of kilometers. It represents a major change in the upper atmospheric circulation. Shapley and Beynon (1965) have demonstrated a statistical association between stratospheric warmings and D-region absorption, while Belrose (1967) has described marked D-region effects (changes of VLF phase height and increased absorption) that accompanied the event of February 1952. At Christchurch (43°S), Gregory (1965) found increases of D-region electron concentration during the stratospheric warming of June
1963. We show his results in Fig. 40. At the time, there occurred some magnetic disturbance; Belrose considers that magnetic activity may indeed be associated with the phenomena in some way.

![Graph](image-url)

**Fig. 40.** D-region changes accompanying a winter stratospheric warming in June and July 1963, at Christchurch, New Zealand (43°S, 173°E). (a) Noon electron concentration $N$ obtained by the partial reflection method. (b) Stratospheric temperature at the 20–30 mb level (about 25 km altitude) [after Gregory (1965)].

As discussed by Belrose (1967), the D-region effects may be due to changes in atmospheric composition, notably trace constituents such as nitric oxide and ozone which can play an important part in D-region photochemistry. Enhancement of the NO concentration had been suggested as the cause of the winter anomaly in the absorption, mentioned above (Mawdsley, 1961); this theory was developed by Sechrist (1967), who found the anomalous absorption to be associated with enhanced mesospheric temperature. The winter anomaly, therefore, might be regarded as meteorological in origin [though we should note that energetic particle precipitation has been proposed as an alternative explanation (Maehlum, 1967)]. If the meteorological explanation is correct, it implies some physical linkage between the iono-
sphere and lower parts of the atmosphere. The effects of the linkage do not seem to be confined to the D region since Beynon and Jones (1965) present statistical evidence of changes at higher levels during stratospheric warmings. In Sec. 5.42, we shall mention effects observed in the F1 layer.

5.2 The E and F1 Layers

The E and F1 layers of the ionosphere are generally regarded as being fair approximations to the idealized Chapman layer. Although the Chapman theory does provide a first-order description of the behavior of the critical frequencies, detailed study indicates that the layers are more complex. There is little doubt that the layers are within the "photochemical regime" described in Chapter III although transport processes can produce measurable perturbations.

5.21 The Normal E Layer

In this section we describe the normal E layer produced by solar ionizing radiation. The term "sporadic E," or Es, is applied to E-region ionization which does not behave in a regular manner. It is generally considered that sporadic E is not closely related to the normal E layer, so we will discuss it separately, in Sec. 6.3. A good review of the normal E-layer phenomena has been given by Robinson (1959).

In applying the theory of photochemical processes to the E layer, it is generally satisfactory to assume \( \partial N/\partial t = 0 \). For a simple equilibrium Chapman layer in an isothermal atmosphere, we found in Sec. 3.22 that the critical frequency is related to solar zenith angle \( \chi \) by the equation

\[
 f_o = 9000 \left[ (q_0/\alpha) \cos \chi \right]^{0.25} \text{ [Hz]} \quad (503)
\]

(c.g.s. units; for m.k.s. the numerical factor is 9). As before \( q_0 \) is the peak production rate for overhead sun (\( \chi = 0 \)) and \( \alpha \) the recombination coefficient. Many investigators assume a relation of the type \( f_o \propto (\cos \chi)^n \), and find the value of \( n \) as \( \chi \) varies with latitude, for a number of stations. A study of E-layer diurnal variations (Tremellen and Cox, 1947) gave \( n = 0.3 \) (or somewhat more) for diurnal variations, but seasonal variations give a result closer to the theoretical value 0.25. A possible reason why \( n > 0.25 \) in the E layer was mentioned earlier; that is, the existence of a positive scale height gradient, \( \Gamma = dH/dh \). In Eq. (330) we deduced that \( n = 0.25 (1 + \Gamma) \), so that the diurnal variation of \( f_o E \) implies that \( \Gamma \approx 0.2 \), reasonably consistent with
models of the thermosphere; the equation is modified if the temperature gradient causes $\alpha$ to vary with height.

Noon values of the critical frequency $f_{0}E$ are found to vary regularly with solar activity, as measured by the sunspot number $R$ or by other parameters (Allen, 1948; Minnis and Bazzard, 1959). If Eq. (503) is used to correct for variations of solar zenith angle, then it is found that approximately

$$\frac{q_{0}(E)}{\alpha(E)} = 180(1 + 0.01 R) \times 10^{8} \text{ cm}^{-6} \quad \text{[or } 10^{20} \text{ m}^{-6}] \quad (504)$$

The recombination coefficient can then be determined by using the value $q_{0} = 4700 \text{ cm}^{-3} \text{ s}^{-1} = 4.7 \times 10^{9} \text{ m}^{-3} \text{ s}^{-1}$ from the peak of curve $E$ in Fig. 24, which applies for $R \approx 60$ (Allen, 1965). From Eq. (504) it is then found that $\alpha = 1.6 \times 10^{-7} \text{ cm}^{3} \text{ s}^{-1} = 1.6 \times 10^{-13} \text{ m}^{3} \text{ s}^{-1}$, which is of the same order of magnitude as laboratory values (see Table I, Sec. 3.4). In contrast, values of $\alpha \sim 10^{-8} \text{ cm}^{3} \text{ s}^{-1}$ have generally been deduced from eclipse observations (Ratcliffe, 1956b), but are not regarded as very reliable, as we shall see in Sec. 6.2. Similar results are obtained by measuring the "sluggishness," i.e., the interval by which the time of greatest $N_{m}E$ lags behind local noon, which is of order 10 min and should equal $(2\alpha N_{m})^{-1}$ (Appleton, 1937, 1953). From the decay of the E layer at night, Titheridge (1959b) finds $\alpha = 2 \times 10^{-8} \text{ cm}^{3} \text{ s}^{-1}$. However, in all these analyses, the E region was assumed to contain only a single ionic constituent. It may be necessary to consider the presence of two ions with different recombination coefficients to explain the observations satisfactorily (Bowhill, 1961a). This could easily arise from the complex photochemistry of the E region, outlined in Sec. 3.63.

Accurate ionosonde data have been used by Robinson (1960) to determine the height of peak production for overhead sun, $h_{0}$; it is found to be about 108 km.

As regards the latitudinal variations of $f_{0}E$, there are departures from the normal Chapman layer behavior which may be caused by electromagnetic movements associated with the quiet-day $(S_{q})$ currents flowing in the E region. This explanation is favored by Beynon and Brown (1959); moreover, Appleton et al. (1955) find a systematic latitude variation of the post-noon lag of $N_{m}E$ which seems to be due to movements rather than to changes in the "sluggishness" time $(2\alpha N_{m})^{-1}$.

We must point out that the vertical structure of the E layer is less simple than the previous discussion might imply. On ionograms, two or more "cusps" often appear, and so it may be difficult to assign a definite E-layer critical frequency. If a sequence of ionograms is studied, however, it is usually found that one particular "cusp" displays a more consistent diurnal
variation than the others, and this is usually tabulated as the "actual" critical frequency. Occasionally, a second cusp, at a higher frequency, is found to vary in a consistent way, and it is then denoted as $f_{cE}E2$. Additional stratifications of a more or less temporary nature are often observed (Ellyett and Watts 1959; Dieminger, 1959; Robinson, 1959). Since the E layer is produced, not by monochromatic radiation but by a variety of wavelengths (including X-rays and extreme ultraviolet lines such as Lyman $\beta$), it is not surprising that a complex structure exists. So far, however, the data are insufficient for any correspondence between individual spectral lines and ionogram features to be established. Although conventional ionograms do not give very detailed information about ionization between the E and F1 layers, rocket $N(h)$ profiles have shown that there is little (if any) decrease of $N$ above the E layer by day, consistent with the lack of a marked "valley" in the $q(h)$ profile of Fig. 24.

After sunset, $N_mE$ decreases to its nighttime value, typically of order $5 \times 10^3 \text{cm}^{-3} (5 \times 10^9 \text{m}^{-3})$, which could probably be maintained by the extreme ultraviolet radiations from the night sky such as Lyman $\alpha$, Lyman $\beta$, and the helium lines (Ogawa and Tohmatsu, 1966; Sagalyn et al., 1967). In addition, sporadic E peaks exceeding $10^4 \text{cm}^{-3}$ sometimes occur. During times of magnetic disturbance a thick "night E layer," not sporadic in character, may be observed, notably at auroral latitudes (G. A. M. King, 1962b). Under very quiet conditions, there may exist deep "valleys" above the nighttime E layer, in which the electron concentration is much less than $N_mE$. Examples have been illustrated by Wakai (1967), who obtained $N(h)$ profiles from the analysis of both "ordinary" and "extraordinary" virtual heights on low-frequency ionograms recorded near Boulder, Colorado (40°N). Under disturbed conditions a "night E layer" appeared at around 150 km.

5.22 THE F1 LAYER

For the F1 layer, the solar cycle variation is quite well described by the formula

$$q_0(F1)/\alpha(F1) = 500(1 + 0.016 R) \times 10^8 \text{cm}^{-6} \quad \text{[or} \quad 10^{20} \text{m}^{-6}\text{]} \quad (505)$$

given by Ratcliffe and Weekes (1960), who took $\alpha = \frac{1}{2} \times 10^{-8} \text{cm}^3 \text{s}^{-1}$ as typically found from eclipse observations. If, however, we take $q_0 = 3700 \text{cm}^{-3} \text{s}^{-1}$ for $R = 60$ (see Fig. 24), we find $\alpha = 4 \times 10^{-8} \text{cm}^3 \text{s}^{-1}$, more in accordance with laboratory values which are of order $10^{-7} \text{cm}^3 \text{s}^{-1}$ at room temperature. Since $\alpha$ may well decrease with increasing temperature (Swider,
1965), we might expect F1-layer values of $\alpha$ to be smaller than either the E-layer values or the laboratory values.

Although the F1 ledge is not always observed on ionograms, its behavior approximates that of a Chapman layer when it does appear. According to Allen (1948), the critical frequency $f_{\alpha}F1$ varies diurnally as $(\cos \chi)^{0.2}$. This is different from the behavior of $f_{\alpha}E$, mentioned earlier, but it is consistent with the theory that the F1 ledge lies at the transition between the domains of the “quadratic” and “linear” loss formulas (Sec. 3.62). Because of the upward decrease of $\beta$, it can be shown that $f_{\alpha}F1$ should vary more slowly with $\chi$ than the Chapman law $(\cos \chi)^{0.25}$, and that the value of $\alpha$ deduced from Eq. (505) is somewhat too small.

According to Allen (1965) (see our Fig. 24), the height of peak production in the F1 layer should be around 150 km. For summer noon at Slough, however, Thomas et al. (1958) found $h_mF1$ (which should lie near the production peak) to be about 185 km (see Sec. 5.41, Fig. 45), but the determination of $h_mF1$ from ionograms may not be particularly accurate.

There is evidence of some geomagnetic control of the F1 layer (Cummack, 1961), which might possibly result from electromagnetic movements such as we discussed in connection with the E layer. In low latitudes, an additional stratification may appear; this is known as the F1$\frac{1}{2}$ layer (Kasuya, 1957; Kotadia, 1963), and may also be seen during eclipses (Sec. 6.2).

5.3 F-Region Rates

In principle, it should be possible to determine the rates of production, loss and diffusion in the F region from the large amount of available data, gathered by the methods described in Chapter II. But although the physics and chemistry of the F region are becoming quite well understood, most of the rates are probably only known to within a factor of two. The reason for this is that the F2 layer is so complex that it is difficult to isolate the effects of any one process. In the lower F region, the analysis can be simplified by the assumption of equilibrium although it is then only possible to determine the relative values of the important parameters such as $q$, $\alpha$, $\beta$, and $D$; the absolute values are obtainable only if time variations are studied. Thus, we have to turn to other sources of information. We have seen in Chapter III that absolute values of $q$ can be obtained from rocket observations of the solar spectrum; even these, however, are somewhat influenced by uncertainties about atmospheric composition. The uncertainties also affect loss rates derived from laboratory measurements of rate coefficients, even though
these measurements have been greatly refined within recent years.

5.3 SOME POSSIBLE WAYS OF DETERMINING RATES

Noon Electron Concentration: We saw previously that critical frequency data for the F1 layer lead to an approximate relation between $q_0/x$ and the sunspot number $R$, Eq. (505). Although we have suggested that $N = q/\beta$ in the F2 layer, at heights up to the peak (Sec. 4.42), the variability of the F2 layer prevents any general formula for $N_mF2$ being given. Allen (1948) finds that at several stations $N_mF2 \propto (1 + 0.02R)$, and provided $q_0$ varies linearly with $R$, (Eq. (505)), this linear variation of $N_mF2$ supports the theory that the rate of electron loss in the F2 layer is proportional to $N$, not to $N^2$ as at lower heights. Minnis (1955) has developed an ionospheric index $I_F$, used for prediction purposes, which resembles in its values the mean sunspot number $R$, but is more closely related to $N_mF2$. Other summaries of data on $N_mF2$ have been given by Yonezawa and Arima (1959) and Minnis and Bazzard (1960).

Shape of F1-F2 Transition Region: By using the theory of Sec. 3.62, values of $\beta^2/2q$ can be obtained by matching observed $N(h)$ or $h'(f)$ curves to theoretical curves (G. A. M. King, 1961b; King and Lawden, 1964, 1966; French, 1966).

Night Decay: Values of $\beta$ for the F2 layer have been deduced from nighttime $N(t)$ variations. After sunset, the decay of $N$ at any height $h$ should initially depend on the local loss coefficient $\beta(h)$. Once the nighttime "stationary" layer becomes established, the decay should depend on the value of $\beta$ at the peak (Sec. 4.42), though electromagnetic and thermal motions complicate the analysis. Ratcliffe et al. (1956) suggested the empirical formula (with $h$ in km)

$$\beta(h) = 10^{-4} \exp\left[\frac{(300 - h)}{50}\right] \text{[s}^{-1}] \tag{506}$$

based on a study of $N(h, t)$ data from several stations. A more detailed analysis of F2-layer data enabled Quinn and Nisbet (1965) to estimate both the loss coefficient $\beta$ and the diffusion coefficient $D$. The height-integrated rate of loss can also be deduced from data on the rate of change of total electron content (Taylor, 1965; Yuen and Roelofs, 1966; Titheridge, 1966b).

Least-squares Analysis of Continuity Equation: $N(h, t)$ data can be used to determine the set of parameters which best satisfies the F-region continuity equation. From nighttime data for Puerto Rico (19°N), Shimazaki (1964) deduced that $\beta = 10^{-4}$ s$^{-1}$ and $D = 5 \times 10^9$ cm$^2$ s$^{-1}$ at 300 km; the method can be extended to estimate vertical drift velocities (Shimazaki, 1966). Chun-Ming Huang (1966) applied least-squares analysis to daytime data for a chain
of stations on the 75°W meridian, obtaining estimates of drift velocity and the values $q = 1100 \text{ cm}^{-3} \text{ s}^{-1}$ and $\beta = (2-4) 10^{-4} \text{ s}^{-1}$ at 300 km, for sunspot maximum (Sept. 1957; $R = 240$).

Layer Thickness: Wright (1962) has used extensive $N(h, t)$ data to study the thickness of the F2 layer below the peak, which is related to the scale height of the ionizable gas and hence to its temperature. Becker (1967) finds that the temperatures deduced in this way generally agree to within 10% with those deduced from satellite drag data.

Eclipses: We have already mentioned that eclipse results do not give reliable values of $\alpha$ in the F1 layer. Similar remarks apply to the methods depending on the “sluggishness” of the layer. For the F2 layer, all these methods are generally useless because of the transport processes. In low latitudes this difficulty appears less serious, and from eclipse data Van Zandt et al. (1960) were able to obtain the results (with $h$ in km):

\begin{align*}
\beta(h) &= 6.8 \times 10^{-4} \exp\left[\frac{(300 - h)}{103}\right] \text{ [s}^{-1}\text{]} \quad (507) \\
q(h) &= 880 \exp\left[\frac{(300 - h)}{186}\right] \text{ [cm}^{-3} \text{ s}^{-1}\text{]} \quad (508)
\end{align*}

Considering the extremely high level of solar activity at the time of the eclipse, the values of $q$ and $\beta$ at 300 km are quite consistent with other data, such as Eq. (506) for $\beta$. We shall consider this below, in Sec. 5.32. But the scale heights in these equations, if associated with $N_2$ (for $\beta$) and O (for $q$), imply a temperature of about 3200°K, probably too high in comparison with values deduced from layer thickness and atmospheric data.

Sunrise: Just after F-layer sunrise, the peak of $q$ lies in the F2 layer, and we expect to have $\frac{\partial N}{\partial t} = q(\chi) = q_0/Ch(\chi)$. Rishbeth and Setty (1961) found values of $q(\chi)$ but, owing to uncertainties in $Ch(\chi)$ and other quantities, this method has not led to accurate values of $q_0$. Better information can be obtained by studying the variations of total electron content, obtained from experiments with moon echoes and geostationary satellites. Data obtained in Hawaii from Syncom 3 show that, on a series of days, the rate of increase of total content is very consistent even though marked day-to-day differences appear at later hours (Garriott et al., 1965). Some data for 1964–1966 are shown in Fig. 41.

The inner ordinate scale in Fig. 41 is most directly related to the observations. It is the integrated production rate of ionization at $\chi = 90°$, for which $Ch(\chi) = 15$. According to the photochemical scheme discussed in Sec. 3.63, the observed ionization results from the photoionization of atomic oxygen (process Q1 in Table II), as $N_2^+$ ions are rapidly lost by processes (R3) or (T3) + (R2). When a model for the neutral atmosphere is assumed,
<table>
<thead>
<tr>
<th>Quantity and symbol</th>
<th>Sunspot min.</th>
<th>Sunspot max.</th>
<th>Units [c.g.s.]</th>
<th>Units [m.k.s.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2 layer (noon)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zürich sunspot number</td>
<td>$R$</td>
<td>0</td>
<td>180</td>
<td>—</td>
</tr>
<tr>
<td>Solar 10 cm flux</td>
<td>$S$</td>
<td>70</td>
<td>225</td>
<td>—</td>
</tr>
<tr>
<td>Gas temperature, 300 km</td>
<td>$T$</td>
<td>900</td>
<td>1700</td>
<td>$^\circ$K</td>
</tr>
<tr>
<td>Production rate, 300 km</td>
<td>$q$</td>
<td>50</td>
<td>750</td>
<td>$10^{-22}$ W m$^{-2}$ Hz$^{-1}$</td>
</tr>
<tr>
<td>Loss coefficient, 300 km</td>
<td>$\beta$</td>
<td>0.4</td>
<td>7</td>
<td>$10^{-4}$ s$^{-1}$</td>
</tr>
<tr>
<td>Diffusion coefficient, 300 km</td>
<td>$D$</td>
<td>7</td>
<td>2</td>
<td>$10^{10}$ cm$^2$ s$^{-1}$</td>
</tr>
<tr>
<td>Peak electron concentration</td>
<td>$N_m$</td>
<td>4</td>
<td>16</td>
<td>$10^{8}$ cm$^{-3}$</td>
</tr>
<tr>
<td>Height of peak</td>
<td>$h_m$</td>
<td>240</td>
<td>320</td>
<td>km</td>
</tr>
<tr>
<td>Total electron content</td>
<td>$\int N dh$</td>
<td>2</td>
<td>7</td>
<td>$10^{13}$ cm$^{-2}$</td>
</tr>
</tbody>
</table>

| F2 layer (midnight) |             |             |                |                |
| Gas temperature, 300 km | $T$   | 750         | 1250           | $^\circ$K      |
| Loss coefficient, 300 km | $\beta$ | 0.06        | 4              | $10^{-4}$ s$^{-1}$ |
| Diffusion coefficient, 300 km | $D$   | 14          | 2              | $10^{10}$ cm$^2$ s$^{-1}$ |
| Peak electron concentration | $N_m$ | 1           | 5              | $10^{8}$ cm$^{-3}$ |
| Height of peak      | $h_m$      | 320         | 380            | km             |

F1 layer (noon), moderate solar activity ($R = 60$): $N_m F1 \sim 3 \times 10^5$ cm$^{-3}$; production rate $q_0 \sim 4000$ cm$^{-3}$ s$^{-1}$; loss coefficients $\alpha \sim \frac{1}{3} \times 10^{-7}$ cm$^3$ s$^{-1}$, $\beta \sim (0.01-0.1)$ s$^{-1}$; shape factor $G = \frac{\beta^2}{q_0} \sim (0.5-50)$. See Eq. (505) for solar cycle variation.

---

the sunrise measurements can be related to the integrated production of both \( \text{O}^+ \) and \( \text{N}_2^+ \) for overhead sun, \( Q_0 \), as shown by the outer left-hand scale in Fig. 41 (due to F. L. Smith, III). Little variability is found during these sunspot minimum years although an increase in 1966 accompanies the increase of solar 10-cm flux. These values of \( Q_0 \) are consistent with the EUV flux measurements of Hinteregger (1965) at a similar level of solar activity (Sec. 3.31).

Fig. 41. Integrated production rates at sunrise, \( \chi = 90^\circ \), for a location near Hawaii (20°N, 160°W), determined from Syncom 3 observations. Points show individual daily values; the curves show 30-day running means of \( Q \) (full-line, left-hand scales) and of solar 10-cm flux \( S \) (broken line, right-hand scale). As almost all of the ions produced are \( \text{O}^+ \), the inner left-hand scale is labelled \( Q_{90} [\text{O}^+] \). Using a neutral atmosphere model very similar to CIRA 1965, the total ion production at \( \chi = 0^\circ \), including both \( \text{O}^+ \) and \( \text{N}_2^+ \), may be computed from the sunrise measurement. These total values, about 25 times larger than \( Q_{90} [\text{O}^+] \), are read on the outer left-hand scale [F. L. Smith, III, Stanford University, dissertation; *J. Geophys. Res.* 73, 7385, (1968)].

5.32 VALUES OF F-REGION PARAMETERS AND THEIR VARIATIONS

Table IV gives rough values of F-region parameters. The total electron content data are average values from Bhonsle *et al.* (1965), illustrated in Fig. 42. Though more refined data are becoming available, the general trend shown in the Table may well hold good. It will only be possible to compute definite values when we know more about the variation of solar ionizing
5.3 F-REGION RATES

radiation and atmospheric composition with solar activity, and about the temperature dependence of various rate coefficients.

Let us now consider the values of \( q, \beta, \) and \( D \) at a fixed height, say 300 km. As solar activity increases, the noon value of \( q \) increases rapidly, because the flux of ionizing radiation and the concentration of atomic oxygen both increase (see Eq. (306); at noon the optical depth \( \tau \ll 1 \) at 300 km). If we write \( \beta = \sum \{ \gamma n [M] \} \) and \( D \propto b/n \) (see Eqs. (365) and (434)), we see that the increase of molecular gas concentration \( n [M] \) and of total gas concentration \( n \) leads to an increase of \( \beta \) and decrease of \( D \). At a given level of solar activity, the day-to-night change of gas concentration causes \( \beta \) to be smaller, and \( D \) to be larger, at night than by day. These changes of \( \beta \) and \( D \) are modified, though probably not greatly, by temperature dependence of the parameters \( \gamma \) and \( b \). The data given in Table IV include a plausible (but hypothetical) \( \sqrt{T} \) dependence of both parameters.

Some values of \( \beta (300) \) are plotted in Fig. 43 as a function of neutral gas temperature. They are derived both from experimental observations and from empirical models, and in some cases the values of \( T \) have had to be estimated from appropriate data on solar activity. To provide the simplest possible illustration of how \( \beta \) might be expected to vary with \( T \), we have plotted the gas concentration \( n [N_2] \) at 300 km, taken from the CIRA (1965) models for noon. The scales on the diagram are such that the \( n [N_2] \) curve represents \( \beta \)

![Fig. 42. Solar cycle variation of total electron content, measured at Stanford, California (37°N, 122°W) by satellite Faraday rotation observations; the satellites used are indicated on the figure [Bhonsle et al. (1965)].](image)
Fig. 43. Loss coefficient $\beta$ at 300 km altitude as a function of neutral gas temperature. The relation between temperature and 10-cm solar flux, according to the CIRA (1965) models, is shown by the top scales; one of these applies to midnight, the other to noon. The right-hand scale and the full curve show the variation of neutral N$_2$ concentration with temperature at noon, also taken from the CIRA models; it is assumed that $\beta = 10^{-15} n(N_2)$. The plotted values of $\beta$ are as follows:

Values deduced from empirical models of F2 layer: *: Noon values from Table IV [Rishbeth (1964)]. Y: Estimated from model [Yonezawa and Takahashi (1960)]. O: Noon, equatorial F2 model [Norton and Van Zandt, (1964)]. ○: Midlatitude E and F layer model, 1000 hours local time [Norton et al. (1963)].

Values deduced from experimental observations: ♦: Eclipse observation, 0900 hours local time [Van Zandt et al. (1960)]. Δ: Decay of night F2 layer, summer [Quinn and Nisbet (1965)]. ▼: Decay of night F2 layer, winter [Quinn and Nisbet (1965)]. ———— Decay of night F2 layer [RSST: Ratcliffe et al. (1956)]. For RSST and point-the appropriate temperatures have been estimated from the level of solar activity prevailing at the time of the observations.

if we take $\gamma = 10^{-12}$ cm$^3$ s$^{-1}$ (independent of temperature) and ignore any contribution due to O$_2$. This value of $\gamma$ is somewhat smaller than is obtained from contemporary laboratory experiments. However, we can conclude that
5.4 F-REGION ANOMALIES

the observational data are consistent with the idea of a temperature-dependent $\beta$. We shall discuss the winter night values in Sec. 5.43.

Analysis of nighttime F2-layer data yields values of $D(300)$ in the range $(2-5) \times 10^9 \text{ cm}^2 \text{s}^{-1}$ at moderate levels of solar activity (1959–1961), according to Shimazaki (1964) and Quinn and Nisbet (1965). Shimazaki (1966) has commented on the discrepancy between these results and the much larger theoretical values of $D(300)$. (See for instance those in Table IV, which are computed from Eq. (434) using the data of Dalgarno (1964b) for $v_{in}$ and reasonable values of other parameters.) If the theoretical values are correct, the nighttime F2 layer would be expected to decay more rapidly than it is observed to do; we shall return to this problem in Sec. 5.43.

We should also enquire how the available data fit the observed parameters of the F2 peak. For this purpose, we use the equations of the “day-equilibrium” layer, Sec. 4.42. Using suffix $m$ to denote quantities at the peak, and rough numerical values from Rishbeth and Barron (1960):

$$N_m = 0.75 q_m / \beta_m, \quad \beta_m = 0.6 D_m / H^2 \quad (509)$$

Given a model atmosphere, comprising vertical distributions of temperature and gas concentrations, it is possible to find the height $h_m$ at which the second equation is satisfied, and then compute $N_m$ from the first equation. This approach has been tried by Rishbeth (1966), using data available at sunspot minimum. To obtain values of $N_m$ and $h_m$ compatible with ionospheric data, it is necessary to use values of the coefficients $\gamma$ much smaller than the laboratory values, or to adopt a model atmosphere containing much less $N_2$ and $O_2$ than does the CIRA (1965) model. Movements due to winds or electric fields, though neglected in this calculation, are not expected to affect this conclusion in essence.

5.4 F-Region Anomalies

Early investigations of the F2-layer critical frequency revealed that it does not behave at all in the manner of a simple Chapman layer. As a typical sample of the available data, we show in Fig. 44 the variation of $f_{oF2}$ throughout two complete years, at sunspot minimum (1953) and sunspot maximum (1958), at Lindau, Germany ($52^\circ N$). These data show many of the phenomena we shall discuss. Some of the outstanding problems of F2-layer morphology are:

(1) behavior of the F2 peak ($N_mF2, h_mF2$) and the diurnal behavior of electron concentration at fixed heights;
Fig. 44. Variations of critical frequency $f_{o}F2$ throughout the sunspot minimum year 1953 and sunspot maximum year 1958, at Lindau, Germany (52°N, 10°E), as a function of Middle European Time (MET). The data are grouped in 27-day solar rotations. Numerals I—XII indicate the first days of each month; vertical bars indicate "sudden ionospheric disturbances" [W. Dieminger, Max-Planck-Institut für Aeronomie, Lindau].
(2) the seasonal, annual, and semiannual anomalies;
(3) the maintenance of the F layer at night;
(4) the polar F layer;
(5) the equatorial anomaly.

Each of these problems will be discussed in turn and some of the proposed explanations advanced, even though they may be only tentative. Further discussion has been given by Rishbeth (1967a). We should state that our remarks apply to the average behavior of the undisturbed F2 layer. More than the other layers, the F2 layer is subject to day-to-day variations of perhaps 20%, and may even change appreciably from one hour to the next. Some description of this variability is given in a review of F2-layer morphology by Wright (1962). We discuss some of the phenomena of F-region storms in Sec. 8.4, so we shall not deal with them here.

5.41 DIURNAL VARIATIONS OF THE F2 PEAK

In Fig. 45 we reproduce the variations of $N_m F_2$ and $h_m F_2$ for quiet days at Slough for a number of months of low solar activity (1953) and moderately high activity (1950), given by Thomas et al. (1958). These display some of the anomalies we shall discuss. Figure 45 shows the seasonal anomaly (for January 1950) and also reveals that $N_m F_2$ is fairly sharply peaked about noon in winter, but is rather irregular in other seasons. The total content data of Fig. 46, which refer to the Southern hemisphere, show morning and evening maxima in most summer months (e.g., December 1965).

The morning and evening maxima are inconsistent with the simple production-loss-diffusion model of Sec. 4.43, but they may be connected with temperature changes (Sec. 4.44). Evans (1965a) has attributed the evening maximum to the decrease of electron (rather than neutral) temperature though it appears that this explanation can hold only if $\beta$ is strongly dependent on $T_e$ (G. R. Thomas and Venables, 1967). Because $T_e$ is strongly influenced by conduction of heat along field lines, its variations could depend not only on local solar irradiation but also on conditions at the magnetically conjugate point (Sec. 6.5). This provides a possible explanation of the observation of Eyfrig (1963) that stations where the magnetic declination is easterly tend to differ in the diurnal variation of $N_m F_2$ from stations with westerly declination.

An entirely different explanation of the morning and evening maxima of $N_m F_2$ has been suggested by Kohl and King (1967), who attribute them to vertical drift due to winds. At mid-latitudes, the winds at 300 km possess a poleward component between about 0900 hours and 1800 hours local time.
Fig. 45. Average diurnal variations of $N_m F_2$, $h_m F_2$, $h_m F_1$, $h_m E$ for the 10 International Quiet Days in 6 representative months at Slough, England (52°N, 1°W). $R$ is the mean Zürich sunspot number [after Thomas et al. (1958)].

(Fig. 7), which causes a downward drift and hence a reduction of $N_m F_2$. Any subsequent recovery in the evening depends on the detailed time variation of the winds (Geisler, 1967b; Rishbeth, 1967b). The relative importance of the “winds” and “thermal expansion” effects on the diurnal variation of $N_m F_2$ remains to be settled.

As regards $h_m F_2$, the heating and cooling of the atmosphere alters the height at which $D/H^2 \sim \beta$, near which $h_m F_2$ should occur (Sec. 4.42). Although the drop of $h_m F_2$ at sunrise, and its subsequent rise (Fig. 45), are consistent with the theoretical behavior outlined in Sec. 4.43, the difference between noon and midnight $h_m F_2$ (generally 100 km) is much greater than would be expected from theory, especially when the day-night temperature changes are allowed for (Rishbeth, 1964). The meridional winds could be the cause of this, since they produce a strong upward drift at night and a downward drift by day (Kohl and King, 1967; Rishbeth, 1967b). At lower
5.4 F-REGION ANOMALIES

Fig. 46. Diurnal variation of total electron content at Auckland, New Zealand (34°S, 176°E), monthly median values for June 1965 to January 1966, obtained by means of the geostationary satellite Syncom 3 [after Titheridge (1966b)].

latitudes, the day-night difference is smaller, and changes sign near the equator where $h_m F_2$ is higher by day than by night. These facts are unexplained, though some kinds of movements might be responsible. Temperature variations cause $h_m F_2$ to increase generally with increasing solar activity, by about 0.4 km per unit increase of sunspot number $R$ (J. O. Thomas, 1962; Stubbe, 1964; Rishbeth, 1964).

5.42 ANNUAL, SEMIANNUAL, AND SEASONAL ANOMALIES

These anomalies have been known for a long time, and were discussed by Berkner et al. (1936) and many others. A detailed analysis of mid-latitude $f_o F_2$ data was made by Yonezawa and Arima (1959) and Yonezawa (1959), who extracted the magnitudes of the different components of anomalous behavior. In discussing these components, we do not imply that they necessarily have distinct physical causes.

Over the world as a whole, there exists an annual variation of $N_m F_2$, which is about 20% greater in December than in June. This compares with a 6% variation of solar ionizing flux, greatest in January, due to the variation of sun-earth distance. In addition, there is a seasonal or winter anomaly, in that noon $N_m F_2$ tends to be greater in winter than in summer. This is a daytime phenomenon, and vanishes at night. The winter anomaly does not exist at sunspot minimum, but becomes apparent as solar activity increases, as may be seen by comparing the 1953 and 1950 curves in Fig. 45. Croom et al.
(1960) showed that the winter anomaly is most marked within a certain range of latitudes. In the Northern hemisphere, this range appears related to magnetic coordinates, being centered on the locus of dip angle $I = 70^\circ$. But in the Southern hemisphere, the distribution of $N_mF_2$ seems to depend more on geographic latitude (Sato and Rourke, 1964). Rishbeth and Setty (1961) found that, just after layer sunrise, $dN/dt$ in the F2 layer is much greater in winter than in summer; and that the increase of $N$ starts at a greater solar zenith angle ($\chi \approx 97^\circ$) in winter than in summer ($\chi \approx 93^\circ$). This seasonal sunrise anomaly can be seen in the curves of $N_mF_2$, Figs. 44 and 45.

A graph of noon total electron content versus sunspot number (Fig. 42) shows not only the winter anomaly but also the high values at equinox, which represent the semiannual variation. This variation remains evident at sunspot minimum (Yeh and Flaherty, 1966; Yuen and Roelofs, 1967), unlike the winter anomaly. The semiannual variation in $N_mF_2$ (Yonezawa, 1959) is comparable in size to the annual variation. It is prominent in lower latitudes and the Southern hemisphere, but it is masked by the winter anomaly at those times and places where the latter is most prominent. The maxima of the semiannual variation occur in April and October, and coincide remarkably with those of the semiannual variation in neutral air temperature (Jacchia, 1963). The semiannual variation has also been found in $h_mF_2$ (Becker, 1966), being no doubt associated with the temperature variation.

Complete explanations of these anomalies are still lacking. A hypothesis of seasonal composition changes in the atmosphere—specifically, of the neutral O/N$_2$ concentration ratio—was put forward by Rishbeth and Setty (1961) and Wright (1963). If the proportion of molecular gas in the F region were decreased, the value of $\beta$ would be reduced. At sunrise, when the peak of $q$ is in the F2 layer, the proportion of ionizing solar radiation absorbed by N$_2$ (and not producing observable ionization) would be reduced. Since at sunrise $dN/dt = q$, the increase of $q$ could be connected with the seasonal sunrise anomaly. Other evidence comes from the study of the F1–F2 transition region, as in Sec. 3.62 (G. A. M. King, 1961a). The parameter $G = \beta^2/\sigma q$ is larger in summer than in winter, the difference being greater than is attributable to changes of $\chi$ (Eq. 358). From this, a seasonal change of $\beta$ is deduced (King and Lawden, 1966), this being the factor most likely to be variable.

On the other hand, Fig. 45 shows that the evening decrease of $N_mF_2$ is at least as rapid in winter as in summer; moreover the values of $\beta$ deduced from the evening decrease of total electron content are greater in winter than in summer (Titheridge, 1966b). These observations suggest that a change of $\beta$
cannot be the only factor influencing the seasonal variations. Other possible factors include winds and seasonal temperature changes in the neutral atmosphere, and diffusion of ionization along magnetic field lines in the protonosphere. The last of these appears too slow to be significant (Sec. 4.5), while the roles of the others have yet to be established.

The existence of neutral composition changes in the F region is still hypothetical. Such changes would be produced if there occurred a change in the O/N$_2$ ratio at the turbopause near 100 km, where diffusive separation of the neutral gas begins (Sec. 1.4); or if the altitude of the turbopause varied because of atmospheric disturbances (G. A. M. King, 1961a, 1966). It has been suggested that large-scale meridional circulations of air in the mesosphere and lower thermosphere give rise to changes of the O/N$_2$ ratio at the turbopause (F. S. Johnson, 1964; G. A. M. King, 1964). This seems consistent with the theory of seasonal changes of mesospheric temperature and atomic oxygen content in polar regions (Young and Epstein, 1962), mentioned previously in Secs. 1.31 and 1.62. Indeed, Bullen (1964) has associated abrupt increases of $\beta$ at Cape Hallett (72°S) with stratospheric warmings in the Antarctic; French (1966) and King and Lawden (1966) have shown other examples of abrupt seasonal changes of $\beta$ in southerly latitudes. Bellchambers and Piggott (1960) have remarked on the suddenness of the change between characteristic “summer” and “winter” patterns of diurnal $f_{\nu}$F2 behavior, at Halley Bay (76°S). A similar change of behavior, taking place within about one month in spring and autumn, is discernible at Lindau (52°N); see the 1958 data in Fig. 44.

5.4.3 Maintenance of the Night F2 Layer

At night $N_m$F2 decreases rather irregularly. This decrease may not be sustained throughout the night, especially on long winter nights when $N_m$F2 appears to reach a “base level” of order 10$^5$ cm$^{-3}$, about which it fluctuates. This behavior can be seen in Figs. 44 and 45. A similar “base level” may be seen in some moon-echo and satellite observations of total electron content (Evans and Taylor, 1961) (see also Figs. 15 and 46). In the hours preceding F-layer sunrise, there may be noticeable increases or decreases; instances can be found in Figs. 44 and 46. Some of these pre-sunrise phenomena may be attributable to an influx of photoelectrons along geomagnetic field lines, when the conjugate point is illuminated by the sun (Evans, 1965b; Carlson, 1966). These photoelectrons could raise the electron temperature, but they do not possess sufficient energy to produce much new ionization (see Sec. 6.5).

The problem of how the “base level” is maintained is linked to the problem
of F-region rates. We saw in Sec. 5.32 that if $q = 0$ at night, the F2-layer variations imply values of $\beta$ (e.g., the “winter night” data in Fig. 43) or of $D$ that are smaller than would be expected on other grounds. We have also noted that the F2 peak is higher at night than by day at mid-latitudes (Fig. 45), and explained this by upward drift due to winds. Although this drift does tend to preserve the layer, by reducing the effective loss rate $\beta'$ (Sec. 4.42), it cannot halt the decay completely, so we deduce that the “base level” is maintained by some production or influx of ionization. A flux of at least $10^8$ (ion pairs) cm$^{-2}$ s$^{-1}$ is needed in the absence of drift or winds (Hanson and Patterson, 1964; Yonezawa, 1965). The upward drift due to winds, however, lowers this requirement to (say) $3 \times 10^7$ cm$^{-2}$ s$^{-1}$, not excessively greater than the flux of order $10^7$ cm$^{-2}$ s$^{-1}$ available by diffusion from the protonosphere (Sec. 4.5).

Alternatively, ionization might be produced within the night F2 layer. According to Byram et al. (1961) and Ogawa and Tohmatsu (1966), photo-ionization by extreme ultraviolet radiation in the night sky is quite insufficient to maintain the F2 layer; but a possible alternative is corpuscular ionization, probably by soft electrons or protons of $10^2-10^3$ eV energy (Antonova and Ivanov-Kholodnii, 1961). Although severe limits can be placed on the input of charged particles because of the 6300 Å airglow they would cause (Nicolet and Swider, 1963; Dalgarno, 1964a), the data do not exclude the possibility of a production rate of $10^8$ cm$^{-2}$ s$^{-1}$, which would contribute appreciably to the F region, but would result in only about 100 R of 6300 Å emission, compatible with observation. The existence of such production is still hypothetical; see Krassovsky et al. (1965).

5.44 The Polar F Layer

The morphology of the high-latitude F region is very complicated. Even at the geographic South Pole, where the diurnal variation of $\chi$ is negligible, there are diurnal changes of electron concentration which must be due to movement processes (Knecht, 1959), and there are many other phenomena indicating the importance of movements (Piggott and Shapley, 1962). When data from many high-latitude stations are considered, it is found that the variations of $f_oF2$ show appreciable dependence on Universal Time (Duncan, 1962; Sato and Rourke, 1964). A possible explanation of the pattern of variations of $N_mF2$ and $h_mF2$ has been developed by King et al. (1967a); it is that the neutral air winds (Fig. 7) blow across polar regions from the day hemisphere toward the night hemisphere, producing vertical drifts in the F region which depend on the orientation of the geomagnetic field (Sec. 4.24).
The resulting effects in the F region are then controlled by both geographic and geomagnetic factors, as the observations imply. There are pronounced seasonal changes in high latitudes, as we mentioned in Sec. 5.42.

The maintenance of the F-region ionization during the polar winter presents a problem which resembles that of the nighttime F2 layer (Sec. 5.43), but it is more severe. It has been discussed by Piggott and Shapley (1962). It has been demonstrated (e.g., L. Thomas, 1966b) that solar photionization is an adequate source, as long as $\chi < 103^\circ$. At the midwinter pole, $\chi = 113^\circ$ so some other source is needed. This might be corpuscular ionization; alternatively, Rastogi (1960) and Hill (1963) consider that the polar ionosphere is maintained by horizontal transport from lower latitudes, produced by electric fields. Very large drift speeds of several kilometers per second are required, in order to transport the plasma sufficiently far during its lifetime of a few hours, and it is arguable whether the required electric fields exist (Duncan, 1964). A discovery of considerable interest is that the polar F region appears to be sharply divided from the mid-latitude F region by a “trough”; see Sec. 5.52.

5.45 THE EQUATORIAL F2 LAYER

The F2 layer at low latitudes is very peculiar. Sometimes, electron concentrations are greater at midnight than at noon. Norton and Van Zandt (1964) succeeded in explaining daytime $N(h, t)$ curves in terms of photochemical processes only, taking into account the diurnal temperature changes in the atmosphere. Vertical diffusion is neglected because ionization cannot diffuse across the geomagnetic field lines. It can, however, diffuse along the field (i.e., horizontally at the equator) and this may affect the latitude distribution of ionization. However, Hirono and Maeda (1955) show that electromagnetic drift could also influence the diurnal variation of electron density.

A plot of noon values of $N_mF2$ as a function of latitude, shows a pronounced “trough” centered on the magnetic dip equator, with “crests” 15° to 20° to the north and south (Appleton, 1946; Bailey, 1948). The trough is also found in plots of $N$ at fixed heights below (Croom et al., 1959) and above (J. W. King et al., 1964) the F2 peak (Figs. 47 and 48). Studies by Rastogi (1959), Lyon and Thomas (1963), and Rao and Malthotra (1964) show that the anomaly exists during most of the day, being most pronounced around sunset, but disappears after midnight. The anomaly shows rather different features in different longitudes, and different parts of the sunspot cycle.

It was suggested by S. K. Mitra (1946) that the anomaly might be due to the diffusion of ionization away from the equator, causing an accumulation
Fig. 47. Variation of $N_m F_2$ and of electron density (electron concentration) at fixed heights with magnetic dip, for noon on magnetically quiet days in September 1957. The zero level for each curve is indicated on the left [after Croom et al. (1959)].

Fig. 48. Latitude variation of electron density (electron concentration) at fixed heights, determined by means of the Alouette I topside sounder satellite above Singapore ($1^\circ N$, $104^\circ E$) at 1234 local time, 15 September 1963. A magnetic field line is shown [King et al. (1967d)].
of electrons to the north and south. To test this suggestion theoretically, it is necessary to use a form of the diffusion equation which takes account of the geometry of the geomagnetic field (Sec. 4.32). Time-varying solutions of this equation show that Mitra's mechanism can only produce a "trough" which is much "shallower" and "narrower" in latitude than the real equatorial trough (Baxter, 1967).

An alternative theory, due to Martyn (1947) and Duncan (1960a), makes use of electromagnetic drift. On this theory, eastward electric fields exist during the day in the equatorial region, thereby producing an upward plasma drift (Eq. (449)). The plasma lifted in this way then diffuses down the magnetic field lines and away from the equator, in much the same way as in Mitra's theory. The combination of electromagnetic drift ($\mathbf{\text{E}}\mathbf{\times}\mathbf{B}$) and diffusion ($\mathbf{\text{D}}\mathbf{\cdot}\mathbf{B}$) gives rise to a "fountain-like" pattern of plasma motion, and indeed the term "fountain effect" has been used in this connection (Fig. 49). Thus, in the Mitra theory the "crests" of the anomaly are fed by diffusion from the region high above the equator, where the production rate is very small; but on the Martyn theory the plasma is drawn from lower levels, near the F2 peak, where the production rate is larger. Thus, we would expect Martyn's

![Fig. 49. The "equatorial fountain." Vector plot of electron flux in the meridian plane, in a theoretical steady-state model of the equatorial F region. The motions are due to the combined action of plasma diffusion along magnetic field lines and electromagnetic drift across field lines, produced by an assumed distribution of eastward electric field. The magnetic field lines are shown every 200 km above the equator [Hanson and Moffett (1966)].](image-url)
mechanism to be more effective, provided that the necessary electric fields exist. The analysis of Hirono and Maeda (1955) suggests that they can indeed be produced by dynamo action in the lower ionosphere, as we shall consider in Sec. 7.4. The upward drifts by day are expected to be of order 10 m s\(^{-1}\). Furthermore, equilibrium solutions of the continuity equation, including diffusion and drift, show that such drifts are sufficient to produce an equatorial "trough" very like what is observed (Bramley and Peart, 1965; Hanson and Moffett, 1966).

At most times, except for periods near the equinoxes, the equatorial trough is observed to be asymmetrical about the equator, to a greater extent than can be attributed to variations of solar zenith angle. Hanson and Moffett (1966) showed that asymmetries could be produced by horizontal winds blowing across the trough, of order 50 m s\(^{-1}\).

Some small perturbations of the low-latitude F2 layer can be related to local lunar time (McNish and Gautier, 1949; Rastogi, 1961), and are probably caused by electromagnetic drifts associated with lunar tides (Sec. 7.4).

5.5 The Topside Ionosphere

The topside ionosphere, above the F2 peak, has been explored by satellites carrying topside sounders, such as Alouette, and direct measuring instruments. From these we have obtained a broad picture of the spatial distributions of electron concentration and temperature, and (to a lesser extent) of ion composition (Sec. 5.51). Information about the topside at particular sites is also obtained by ground-based techniques, such as incoherent scatter and rocket experiments.

Considering that we expect the structure of the topside ionosphere to be determined by transport processes, it is not surprising that the geomagnetic field plays an important part. Magnetic effects are particularly evident in the equatorial region, where the topside electron distribution is closely related to the low-latitude structure of the F2 layer (Sec. 5.45); this is illustrated by the Alouette data plotted in Fig. 48. Alternatively, the data can be displayed as contours of electron concentration in height and latitude (Lockwood and Nelms, 1964). In the morning these take the form of a "dome," centered on the magnetic equator; later in the day, when the F2 layer "equatorial trough" develops, the topside ionosphere contains a "ledge" of ionization coincident with the magnetic field line linking the "crests" of \(N_mF2\). At fixed heights above the "ledge," the daytime electron concentration, plotted versus latitude, displays a shallow maximum at the equator. This is shown by the
650 km curve in Fig. 48, and also appears at 1000 km, as found by the satellite data of Brace et al. (1967).

It should be pointed out that most satellite topside data have been obtained with satellites in high-inclination orbits. As a result, the latitude variations of electron concentration (or other parameters) are well determined; but local-time variations have to be obtained by putting together data obtained over periods of months, during which conditions can change appreciably.

5.51 ION COMPOSITION

We would expect the positive ion composition to change with increasing height, from the oxygen ions which populate the F region to the light ions of the heliosphere and protonosphere. Profiles of log $N(h)$, obtained by topside sounders or other means, do generally show a change of slope similar to that of the theoretical profile of Fig. 35. From the topside sounder data one can estimate the “plasma scale height” of Eq. (446), namely $[-N_e dh/dN_e]$, and use it to study the variations of temperature and mean ionic mass, even though this approach does not determine all the parameters uniquely. For instance, Watt (1965) deduced that at mid-latitudes, the transition between oxygen ions and light ions typically occurs around 500–600 km at night and up to 1000 km by day. This day-to-night change may just be due to daily thermal expansion and contraction. In addition, the transition height increases with increasing latitude, which implies that at a fixed height, the mean ionic mass increases with latitude (Bauer and Blumle, 1964), consistent with the ion composition results obtained from Ariel I (Bowen et al., 1964a). Nelms and Lockwood (1967) suggest that the changes in shape of $N(h)$ profiles with increasing latitude show an increase of ion temperature, as well as a decrease of light ion concentration. They note that similar conclusions were obtained from the VLF data from Alouette (Barrington et al., 1965).

It is uncertain whether the light ions are really in diffusive equilibrium below 1000 km, because of the importance of photochemical processes (Sec. 4.5). The $\text{He}^+$ and $\text{H}^+$ distributions measured in rocket mass-spectrometer experiments by Taylor et al. (1963) do not seem compatible with diffusive equilibrium; but on the other hand, Brace et al. (1967) consider that their global data on $T_e$ and $N_e$ at 1000 km altitude do show the existence of diffusive equilibrium in the topside ionosphere.

In low latitudes, diffusive equilibrium can only be expected to apply to the electron distributions along magnetic field lines, not to the vertical $N(h)$
profiles. From a study of field-aligned electron distributions, derived from Alouette data for sunspot minimum (1963), Rishbeth et al. (1966) located the O\(^+\)–He\(^+\) transition at around 600 km. However, Farley et al. (1967) obtained rather different composition results with the scatter radar equipment at Jicamarca, near the dip equator; in 1965/66 a transition from O\(^+\) to H\(^+\) was found at 600–700 km altitude at night, and at over 900 km by day, there being only 10\% or less of He\(^+\) at any height. In comparing the results obtained by different methods, one has to take account of differing conditions, particularly solar-cycle changes which may have a marked influence on the abundances of light ions.

5.52 Ionospheric “Troughs” and the “Plasmapause” in the Magnetosphere

Very sharply bounded “troughs” of electron density have been detected by topside sounders and other satellites. They have been described by Muldrew (1965), Sharp (1966), Calvert (1966), and Nelms and Lockwood (1967). They are best observed at night, and are found at magnetic dip latitudes of about 60–70°, and are detectable at heights as low as the F2 peak. More than one trough may be present at any one time. At the edge of a trough, the electron density at 1000 km may change by a factor of two in about 10 km, and sometimes even steeper horizontal gradients are observed. The plasma scale height, associated with the vertical gradient of log \(N\), does not show abrupt changes. It has been suggested that the trough (or the trough furthest from the pole, when two or more troughs exist) may be regarded as the boundary between the “mid-latitude” and “polar” regions of the ionosphere, though just how they are produced and maintained has not been determined. During magnetic disturbances the troughs move towards the equator.

To relate the structure of the topside to electron density variations in the magnetosphere, we assume that the plasma is in diffusive equilibrium along any magnetic field line. By integrating Eq. (444) along a field line, we can find an equation linking the electron density \(N_B\) at some base height in the ionosphere (say 1000 km), and the electron density \(N_A\) at the equatorial apex of the field line. In terms of geopotential height \(h^*\):

\[
\frac{N_B T_{PB}}{N_A T_{PA}} = \exp \int_{h_B^*}^{h_A^*} \frac{dh^*}{H_p(h^*)} \tag{510}
\]

If \(H_p\) varies smoothly with position, not only in height but also between neighboring field lines, then it should be possible to relate the latitudinal
variation of $N_B$—as found from satellite data—to the radial variation of $N_A$, as found from whistler data (Carpenter and Smith, 1964) or incoherent scatter (Bowles, 1963). This relationship between $N_A$ and $N_B$ has been investigated by Angerami and Thomas (1964) and Thomas and Dufour (1965). Of special interest to this relationship is the "plasmapause" or "knee," where the radial decrease of $N_A$ becomes markedly steeper than it is close to the earth (Angerami and Carpenter, 1966; Carpenter, 1966). As we shall see when we discuss the structure of the magnetosphere (Sec. 7.54), the plasmapause appears to be linked by magnetic field lines to the troughs we have discussed in the present section.
VI

SOME IONOSPHERIC PHENOMENA

6.1 Solar Flare Effects

Solar flares are customarily viewed in Hα light (6563 Å) as a temporary brightening of a small portion of the solar chromosphere. The areas are associated with sunspots, plage, filaments, and high magnetic field gradients. Accelerated motion pictures of limb prominences and flares provide a spectacular picture of the great turbulence occurring in the solar chromosphere and corona. Occasionally, an explosive flare will generate shock waves observed to propagate across the disk in the chromosphere. From the standpoint of the ionosphere and radio propagation, there are a number of terrestrial effects observed which provide information about the flare mechanism.

The general classification "sudden ionospheric disturbance", or SID, includes many of these phenomena which accompany solar flares (Dellinger, 1937). Occurring nearly simultaneously with the optical Hα flare are:

**SWF**: Short wave fadeout or "Type I" absorption. Waves reflected from the E and F layers are strongly attenuated by enhanced ionization in the D region.

**SCNA**: Sudden cosmic-noise absorption. A fairly large flare may cause a 2 dB increase in absorption or even more at frequencies of 18 to 25 MHz, which pass completely through the ionosphere without reflection (Shain and Mitra, 1954; Bhonsle, 1960; Jayaram and Chin, 1967).

**SFD**: Sudden frequency deviation of signals from a stable HF transmitter, reflected from the E or F1 layers. The deviation is due to a changing phase path, including a lowering of reflection height, as extra ionization is produced.
SPA: Sudden phase anomaly. The phase advance of long waves reflected at oblique incidence indicates a drop of several kilometers in the height of reflection (Bracewell et al., 1951; Bracewell and Straker, 1949).

SEA: Sudden enhancement of atmospherics recorded at frequencies around 20 kHz (Bureau, 1937; Ellison, 1953). Like the SPA, this indicates a change of VLF propagation conditions.

SFE: or crochet: (Magnetic) solar flare effect. A disturbance of the geomagnetic field, attributed to increased ionospheric conductivity, which leads to increased current flow.

These effects indicate an increase of electron concentration in the D region, perhaps by a factor of ten in a major flare. Such an increase was indeed measured by the partial reflection technique (Belrose and Cetiner, 1962). The fact that ionospheric effects begin at the time of visual sighting of the flare implies that they are due to electromagnetic, rather than corpuscular, radiation. The fact that ionization is produced at lower heights than usual shows that the ionizing radiation possesses a smaller absorption cross section than the radiation responsible for the normal D layer. Satellite measurements have now shown that the total solar X-ray emission can easily rise by several orders of magnitude in the wavelength range $2 < \lambda < 10$ Å in a class 2 flare (Culhane et al., 1964). We should also remember that this large increase is generated in an area of perhaps 1/1000 of the solar disk. In contrast, the Lyman $\alpha$ increase may be only a few percent (Lindsay, 1963; Hallam, 1964). This clearly identifies the X-ray enhancement as responsible for the D region SID effects.

With regard to the magnetic crochet, we recall that the conductivity per ion pair is so small below 80 km that huge changes of electron concentration would be required to enable sizable electric currents to flow. Hence, the currents producing the crochet probably flow in the E region, and might be connected with the increase of $f_o E$ which has been reported on occasions (Taubenheim, 1957). But information on electron concentration changes above the D region during solar flares is rather incomplete, largely because of absorption in the D region.

The sudden frequency deviation has been investigated by Chan and Villard (1963), Davies and Baker (1966), Donnelly (1967) and others. In this experiment, the frequency of a signal from a very stable transmitter, such as WWV, is monitored at some distance after one or more F region reflections. A solar flare produces extra ionization, thus altering the phase path length of the
transmissions and inducing a small, transient frequency change. Although the extra D region ionization produces large absorption effects discussed above, it contributes very little to the phase path changes measured in a SFD. These changes are due to extra ionization deposited in the E–F1 regions by longer wavelength X-rays and EUV (Kanellakos et al., 1962). SFD measurements at two or more frequencies can be used to distinguish between a phase path reduction due to a decrease in the refractive index (as in the E region) and that due to a lowered reflection height (as produced by extra ionization in the F1 layer).

![Diagram of solar flare effects on 21 May 1967]  

**Fig. 50.** Solar flare effects on 21 May 1967, showing changes of total electron content; optical brightness at the center of the Hα line for (a) the main section of the flare and (b) a smaller explosive section; sudden frequency deviation; increase of solar 10-cm flux; absorption (short-wave fadeout); and sudden phase anomaly [Garriott et al. (1967)].
6.2 Eclipse Phenomena

F2 region effects have also been detected. Knecht and Davies (1961a, b) found that \( N_mF2 \) was increased at widely separated stations following some flares. More recently, observations of ATS-1, a geostationary satellite above the Pacific Ocean, have revealed substantial changes in the total electron content occurring simultaneously at five different stations. One such event for 21 May 1967 is shown in Fig. 50, along with correlative evidence of some of the other effects discussed. The figure shows that the increase of about \( 2 \times 10^{16} \text{ m}^{-2} \) occurred in approximately two minutes. To explain this increase, the soft X-ray and EUV flux must increase by one or more mW m\(^{-2}\) for an interval of two to ten minutes, depending on the wavelength involved (Garriott et al., 1967).

Although not included in the phenomena of a SID, radio noise bursts closely accompany \( H\alpha \) and X-ray flare events. Various types of radio noise are thought to be produced by bremsstrahlung, synchrotron emission and plasma oscillations in the solar atmosphere (Wild, 1964; Smerd, 1964). On some occasions, 10.7-cm radio noise shows an increase of two to three orders of magnitude in a few minutes or less, coincident with an X-ray and optical flare.

Solar flare phenomena were observed in an extreme form during the famous event of 23 February 1956. A few minutes after the start of the flare, which was accompanied by an intense SID on the sunlit side of the earth, a worldwide increase of cosmic rays was observed and strong D-region disturbances occurred in the dark hemisphere produced by the precipitation of energetic particles. Many papers on this event have appeared, and a summary of the observations has been given by Bailey (1959).

6.2 Eclipse Phenomena

On those occasions when the sun is partially or totally eclipsed by the moon, changes are observed in the electron distribution in the ionosphere. This is particularly true in the D, E, and F1 regions, but sometimes the effects extend up to the F2 peak. These events provide an opportunity to calculate the loss and production terms in the continuity equation, and they provided early evidence that the E and F1 layers are produced by electromagnetic—not corpuscular—radiation (Appleton and Chapman, 1935). Although variations of absorption have been observed during eclipses (Busch et al., 1956), implying changes in the D region, most studies relate to the higher layers. More recently D-region ion concentration profiles obtained in Greece during the 20 May 1966 eclipse have provided evidence of the increasing importance
of X-rays relative to Lyman $\alpha$ radiation as the solar activity increases, for ionization between 70 and 80 km (Bowling et al., 1967).

When considering the E and F1 regions, we may start with the continuity equation applicable to the peak of the layer, all movements being neglected:

$$\frac{\partial N_m}{\partial t} = q_m - \alpha N_m^2$$

**Fig. 51.** Idealized curves showing the variations of peak production rate $q_m$ and peak E-region electron concentration $N_m$ for a hypothetical total eclipse occurring before noon, and of $q_m$ on a control day. The times of the four optical contacts are: $t_1, t_2, t_3, t_4$; and $t_a, t_b$ are the times of maximum and minimum $N_m$.

thermal atmosphere, but during the eclipse we must multiply this by a factor $\varphi(t)$ equal to the uneclipsed fraction of the sun's disk (as seen from the appropriate height), assuming the ionizing radiation is emitted uniformly across the disk. Thus $q_m = \varphi q_0 \cos \chi$, and this function, as well as the peak electron density $N_m$, is shown in Fig. 51 for a hypothetical total eclipse.

At second contact $t_2$, the production function is zero, and we may calculate the recombination rate as

$$\alpha = -\frac{\partial N_m}{\partial t} N_m^2$$

in which the implicit assumption has been made that the height of $N_m$ does
not change during the eclipse. At the E-region peak, Minnis (1955) has found \( \alpha = 1.5 \times 10^{-8} \text{ cm}^3 \text{s}^{-1} \) using this method.

It is also possible to assume a value for \( \alpha \) and predict the shape of \( N_m \) versus time for the eclipse, looking especially for the time separation between third contact \( t_3 \) and \( t_b \), the electron density minimum. From the "sluggishness" relation (453) the time difference is expected to be a few minutes, but the observations do not show any very consistent pattern (Ratcliffe, 1956b).

It now appears very probable that \( \alpha \) has been underestimated in these eclipse analyses because some of the EUV and X-radiation originates in the solar corona and is not occulted by the moon. Elwert (1958) has calculated that 10 to 20% of the X-radiation may come from above the photosphere. In several rocket flights, J. D. Purcell, R. Tousey, and M. J. Koomen have obtained exposures of the full solar disk and corona between 170-500 \( \text{Å} \). On one occasion, the emission at two solar radii was as intense as 10% of the average over the quiet disk equatorial zone. Since \( q \) is usually a much larger term than \( \partial N/\partial t \) in the E-region continuity equation, the electron production remaining at the time of total optical eclipse could cause the value of \( \alpha \) deduced by using Eq. (601) to be too small. If as much as 10% of the ionizing radiation should remain at totality, the more correct value of \( \alpha \) computed from Eq. (600) would be substantially larger than that obtained by using Eq. (601).

Furthermore, the radiation has been found to be much more intense above active regions on the sun, so the production function cannot be the smoothly varying geometrical function originally assumed. Above a "plage" of average brightness, Friedman (1964) has found the soft X-ray brightness to be 4-25 times greater than that of the quiet disk. Similar results are shown in the EUV spectroheliograms obtained by J. D. Purcell, D. L. Garrett, and R. Tousey (Friedman, 1964). To illustrate the nonuniformity of the solar X-radiation, we show in Fig. 52 an X-ray photograph obtained by J. H. Underwood and W. S. Muney. The enhanced brightness above active regions is most striking, as well as limb brightening at the longer wavelengths.

The eclipse measurements of \( \alpha \) become more consistent with results from entirely different methods when a two-ion mixture is considered (Bates and McDowell, 1957; Bowhill, 1961a). If the two ions have distinctly different values of \( \alpha \), one ion decays more rapidly, resulting in an "effective" recombination coefficient which diminishes with time. Bowhill suggests that the two ions are \( \text{NO}^+ \) and \( \text{O}_2^+ \) and that the latter may decay most rapidly. If so, the \( \text{NO}^+ / \text{O}_2^+ \) concentration ratio should increase after sunset, which seems to be borne out by the rocket ion spectrometer results of Figs. 29 and 30 at
Fig. 52. X-ray photographs of the sun, taken about 1820 hours UT on 3 October 1967 by a glancing-incidence telescope carried in an Aerobee rocket launched from White Sands Missile Range, New Mexico. Each of the photographs was taken through a different filter, the wavelength ranges being shown beneath each picture. The most prominent features of the photographs are the two belts of X-ray active regions in the Northern and Southern hemispheres. These correspond well with the belts of activity visible in Hα. Note that the small ring-shaped features are due to the telescope being slightly out of focus. Limb brightening is visible at the longer wavelengths. The complex structure on the top-right (northwest) limb in exposure (b) is probably due to a large coronal condensation. The jetlike feature extends out about 0.3 solar radii [NASA photograph, obtained by J. H. Underwood and W. S. Muney of Goddard Space Flight Center].

E and F1 region heights. However, evidence is lacking that the recombination coefficients actually differ to the extent required by this theory.

Above the F1 layer, the height variation of loss coefficient can lead to formation of an additional stratification during eclipses, as discussed by Ratcliffe (1956b). This is termed the "eclipse F1½" layer, and is shown in Fig. 53. Munro and Heisler (1958) have pointed out difficulties of interpreting eclipse ionograms, arising from the possibility of oblique echoes, while
Gledhill (1959) has also discussed the "valley" that may form above the F1 layer during eclipses.

Eclipse results in the F2 region are more difficult to interpret than those at lower altitudes. Near the peak, little or no effect may be observed and movements resulting from diffusion or electromagnetic drift may also be important. A successful analysis was made by Van Zandt et al. (1960) for the October 1958 eclipse at Danger Island (Fig. 53). They made several simplifying assumptions which appear to be rather well justified by the results. They have first assumed that temperature changes and all movement terms are negligible. Since Danger Island is at a magnetic dip latitude of only 11°, vertical diffusion is certainly slow. Also, their calculations have been made only at altitudes greater than 280 km, where the loss of electrons is believed to be proportional to the electron concentration ($\beta N$). Furthermore, these altitudes are all well above the production peak, so the production rate at any fixed height is approximately independent of $\cos \chi$ for control days; on the eclipse day, $q$ need only be multiplied by the un eclipsed fraction of the sun's disk. The radiation principally responsible for F2 ionization is in the EUV range, and it is probably more uniformly distributed across the disk than is the E-region ionizing radiation. This minimizes one of the principal difficulties encountered at lower altitudes. Within these assumptions, the con-
tinuity equation for the F2 region can be written

\[ \frac{\partial N(h,t)}{\partial t} = q(h) \left[ \frac{\varphi(h,t)}{N(h,t)} \right] - \beta(h) \] \hspace{1cm} (602)

Since \( q(h) \) is relatively constant well above the production peak, as discussed earlier, and \( \beta(h) \) should be nearly constant if temperature changes are negligible, Eq. (602) is of the form \( y = qx - \beta \). The values of \( y \) and \( x \) can be determined experimentally as a function of time at any fixed altitude, and we may expect a graph of \( y \) versus \( x \) to be approximately linear, with a slope of \( q(h) \) and intercept \( \beta(h) \). Van Zandt \textit{et al.} made these calculations at each 10 km height interval between 290 and 450 km and obtained rather good consistency throughout the range. Their results for \( q \) and \( \beta \) have already been given in Sec. 5.31.

The 20 July 1963 eclipse was observable at totality from the Millstone Hill Observatory, Massachusetts, U.S.A. (latitude 43°N). This fortunate occurrence permitted the altitude profiles of electron concentration and electron and ion temperature to be measured every 15 min throughout the eclipse, using the incoherent scatter technique (Evans, 1965c). The electron temperature showed a large decrease centered on the time of totality, its minimum value at 350 km being 1100°K as compared to 2200°K at the same time on the control days. The decrease of \( T_e \) was about 1000°K at all heights above 350 km. These observations thus show that the electron temperature responds rapidly to variations in the input heat flux due to solar photons. Decreases of 100°K-300°K also occurred in the ion temperature, which, however, remained less than \( T_e \) at all times.

These temperature variations produced a most interesting change in the electron concentration profile. The rapid temperature (and scale height) decrease approaching totality caused the ionosphere to contract, and \( N_{mF2} \) to \textit{increase} from about \( 3.6 \times 10^5 \text{ cm}^{-3} \) just prior to the eclipse, to a value of \( 4.9 \times 10^5 \text{ cm}^{-3} \) at totality. This rapid diffusion would not be expected at the lower latitude of Danger Island although a small increase is observable in Fig. 53. The total electron content continually diminished as \( N_{mF2} \) increased, just as one would expect, owing to downward diffusion and recombination at lower altitudes. This picture of F-region morphology during the eclipse shows the futility of calculating loss rates at latitudes and altitudes where diffusion is so important. Furthermore, Evans (1965d) has discussed a number of earlier eclipse observations, and concludes that increases of \( N_{mF2} \) occur at stations of high magnetic dip (\( I \geq 60° \)), provided the eclipse is at least 90% total at F1-layer heights. He attributes the increase to a decrease
6.3 SPORADIC E

6.31 PROPERTIES OF SPORADIC E LAYERS

Ionosondes often detect dense layers or patches of ionization in the E region, at heights of 100 km to 120 km, which do not seem to be related to the normal daytime E layer. This phenomenon is known as "sporadic E" or "Es" because it does not show any regular behavior. Sometimes Es appears in sheets which completely hide the overlying F layer. At other times, it may be patchy and partially transparent to waves reflected from higher layers. Some examples of Es, though rather inconspicuous, can be found in Fig. 9.

In their comprehensive survey article, Thomas and Smith (1959) adopt the IGY definition of a sporadic E reflection as an E-region echo characterized by one or more of the following:

1. random time of occurrence,
2. partial transparency (echoes also obtained from higher layers),
3. variation of penetration frequency with transmitter power,
4. virtual height independent of frequency.

Much of the work on Es has necessarily consisted of classification and tabulation of data. For this purpose, several types of Es have been defined, which are mostly distinguished by the appearance of the Es traces on ionograms; sketches of these are shown in the Thomas and Smith review cited above. A more recent review, containing many original papers, has been published by Smith and Matsushita (1962); a further group of papers appears in the journal *Radio Science*, being the record of a conference on mid-latitude Es.

Sporadic E is studied with the aid of vertical incidence ionograms, oblique
incidence ionograms and fixed frequency recordings. Backscatter methods are also useful, particularly if a steerable antenna is used so that the motions of individual Es patches can be followed (Peterson et al., 1959; Shearman and Harwood, 1960). Rockets have been flown through Es layers, and data obtained from propagation and probe techniques have indicated the presence of very thin layers of dense ionization at heights of 100 to 130 km (Jackson and Seddon, 1958). There is some evidence that these layers occur at a series of “preferred” heights, at intervals of about 6 km (Pfister and Ulwick, 1958; Gringauz, 1958).

The parameters used to describe Es include the critical frequencies $f_{o}Es$ and $f_{x}Es$. Es traces often do not show the usual “cusps” and sometimes a “top frequency” at which reflections are obtained, $f_{t}Es$, is used (Bibl et al., 1955). In many cases it may be assumed that $f_{t}Es = f_{x}Es$. The “blanketing frequency” $f_{b}Es$ is the lowest frequency at which echoes from higher layers are received through the Es layer. Generally $f_{b}Es < f_{t}Es$, so that the Es layer is partially transmitting over a (small) range of frequency. Sometimes the Es layer completely blankets all layers above it, and thus prevents observation of the F layer. Another parameter is the virtual height $h'Es$ which is usually independent of frequency over most of the trace, in which case it should be nearly equal to the actual height of the base of the Es layer.

Many surveys of Es occurrence do not distinguish between different types, but merely assess the frequency with which some arbitrary criterion, such as $f_{t}Es > 5$ MHz, is met. In temperate latitudes, occurrence of this condition has a probability of about 50% around noon in summer, and there are lesser peaks of occurrence near noon in winter and midnight in summer. In addition, variations with longitude may exist.

“Patches” of Es may be tens or hundreds of kilometers in extent, and drift with velocities of order 50 m s$^{-1}$. There seems to be little correlation between mid-latitude Es and magnetic activity. Some connection between Es and thunderstorms has been suggested (Rastogi, 1957).

In high latitudes, one type of Es is associated with aurora (Knecht, 1956). Sometimes thick, comparatively long-lived layers of ionization are observed at night, especially in auroral latitudes (G. A. M. King, 1962b), but we have already noted that such layers are defined as “night E” since they do not meet the criteria for sporadic E. Other types of high-latitude Es show interesting patterns of geographical and temporal distribution (L. Thomas, 1962b), which suggest that direct particle bombardment is responsible for production of Es ionization; we return to this phenomenon in Sec. 7.54.

Near the magnetic equator, a distinctive type of “equatorial” Es is ob-
6.3 SPORADIC E

served, which is patchy and transparent to waves reflected from higher layers. It is strongly associated with the "electrojet" current (Secs. 4.23, 7.43) which flows along the magnetic equator by day (Skinner and Wright, 1957). The equatorial Es irregularities appear to be aligned with the geomagnetic field (Bowles et al., 1960), and may be caused by plasma instabilities arising from the flow of the large electrojet current (Farley, 1963b).

Apart from these special high-latitude and equatorial circumstances, it is difficult to see how sources of ionization could be sufficiently localized to produce Es. Any such source would have to operate for the observed lifetime of Es patches, which may persist for some hours, since the mean lifetime of ions in the E region is only a few minutes. Consequently, most attention has been given to explanations involving a redistribution of existing ionization rather than the production of extra ionization, such as the "wind-shear" theory of Whitehead (1961). This theory has been quite successful in explaining some properties of mid-latitude Es layers, and we shall describe it below. Because it depends on magnetic field parameters, the theory may explain the connection between Es occurrence and the strength of the horizontal geomagnetic field component (Heisler and Whitehead, 1960). It has been shown that gravity waves in the neutral atmosphere may be the source of the wind shears necessary to this theory of Es (Axford, 1963). It does not seem that ordinary turbulence could produce the strong nonuniformities of ionization present in Es.

6.3.2 THE WIND SHEAR THEORY OF MID-LATITUDE SPORADIC E

The observations of horizontal winds in the neutral air in the E region suggest that large vertical shears (velocity gradients) are often present. Whitehead (1961, 1967) has developed the theory that the shears can redistribute E-region ionization and lead to the formation of the sporadic layers. The theory depends on the vertical motions of ionization that occur when a horizontal wind blows across the magnetic field; we have already discussed (Sec. 4.1) how vertical motions are more effective than horizontal motions in producing changes in the electron distribution.

Figure 54 illustrates a situation in which a vertical shear of the east-west wind produces an Es layer. It is an extreme case, in which the wind actually reverses with height. The wind drags with it the positive ions, for which the collision frequency (with neutrals) exceeds the gyrofrequency, so that $v_i > \omega_i$. The ions experience a Lorentz $(V \times B)$ force, and are driven at an angle to the wind velocity, so that they accumulate within the shear. The electrons are quite unaffected by the neutral wind, since $v_e \ll \omega_e$, but are
VI. SOME IONOSPHERIC PHENOMENATA

**Fig. 54.** Idealized illustration of the wind-shear mechanism in the E region. The plane of the diagram represents an east-west vertical plane at the magnetic equator, seen from the south. The ions are driven by horizontal winds and deflected by the magnetic field, so as to move in the directions shown. The electrons are constrained by electrostatic forces to follow the ion motion, so that ionization accumulates within the shear. The angle $\theta$ is given by $\tan \theta = \omega_i/v_i$ which increases upward [after Hines (1964b)].

constrained to move along the magnetic field lines. They move in such a way as to neutralize the space charge set up by the ion motion, so that the accumulated layer consists of neutral ionization. In discussing the production of this layer, it is sufficient to consider only the ion motions, which are governed by the tensor mobilities we introduced in Sec. 4.21, Eqs. (413) and (414).

An eastward wind $U_y$, blowing normal to B, exerts a force $F_y = m_i v_i U_y$ on the ions. This produces a "transverse" drift $k_1 F_y$ in its own direction, and a "Hall" drift $k_2 F_y$ in the direction normal to itself and to B. The resulting motion shown in Fig. 54 is thus inclined to the wind at an angle $\theta$ given by

$$\tan \theta = k_2/k_1 = \omega_i/v_i$$

(603)

The vertical component of the ion motion is

$$w_i = k_2 F_y \cos I = \frac{\omega_i v_i}{v_i^2 + \omega_i^2} U_y \cos I = U_y \frac{\omega_i}{v_i} \cos I$$

(604)

since $\omega_i^2 \ll v_i^2$. (Hines (1964b) has clarified the difficulties about signs which arose in Whitehead's original analysis.)

An earlier suggestion by Dungey (1956b) depended on the vertical motion produced by a north-south wind, instead of an east-west wind. Such a wind would produce a force $F_x$, resulting in components of ionic drift $k_0 F_x$ parallel to B, and $k_1 F_x$ in the magnetic meridian normal to B. The resulting vertical
drift is then found to be \((k_0 - k_1) F_x \sin I \cos I\). In the E region \(k_0\) and \(k_1\) are virtually equal for ions; the small difference between them involves the very small ratio \(\omega_i^2/\nu_i^2\), as compared to the ratio \(\omega_i/\nu_i\) appearing in (604). At F-region heights this drag would be more effective, corresponding to sketch (c) in Fig. 34, but it is dubious whether suitable wind shears exist.

To investigate the changes in electron distribution resulting from a shear in the vertical velocity given by Eq. (604), we assume that electrostatic forces ensure that the electrons move vertically with the ions, so that we may take \(w = w_i = w_e\) as the vertical plasma velocity. For brevity let \(N' = \partial N/\partial h\) and \(w' = \partial w/\partial h\). Then the E region continuity equation may be written as

\[
\frac{\partial N}{\partial t} = q - \alpha N^2 - Nw' - wN'
\]  

(605)

At equilibrium, at the peak of the layer \((N' = 0)\), we have a quadratic equation for \(N\), which gives

\[
N = \frac{1}{2\alpha} \left[ (w'^2 + 4\alpha q)^{1/2} - w' \right]
\]  

(606)

If we assume that in a strong shear \(w'^2 \gg 4\alpha q\) we find that the extreme values of \(N\) are given by

\[
\begin{align*}
\text{max}(w' < 0), & \quad N_{\text{max}} = -w'/\alpha \\
\text{min}(w' > 0), & \quad N_{\text{min}} = q/w'
\end{align*}
\]  

(607)

We thus have the rather surprising result that the maximum electron concentration in the Es layer is independent of \(q\), for this special case.

From Eq. (607), we see that the peak of \(N\) should occur where \(w'\) has its greatest negative values. In relating this to the shear \(\partial U_y/\partial h\), we must note that \(\nu_i\) is height-varying in Eq. (604). Apart from this complication, we can associate a sporadic E layer with a region in which the eastward component of wind decreases rapidly upwards \((\partial U_y/\partial h \text{ large and negative})\), even though \(U_y\) itself may be small in such a region. In more complete calculations, such as those of Axford (1963), it is shown that the layer approaches the steady state with a time constant equivalent to \((w')^{-1}\).

Although plasma diffusion is too slow at E-region heights to have much influence on the normal E layer, it can be important in sporadic E layers because of the very steep gradients of electron concentration. Thus for any accurate calculation, diffusion terms of the form \(D \partial^2 N/\partial h^2\) must be included in Eq. (605). However, it appears that at present, there is difficulty in explaining the slab-shaped \(N(h)\) profile that Es layers are observed to possess. Gleeson and Axford (1967) suggested that this shape could be explained if the electron temperature were enhanced within an Es layer, but observational support for this idea is lacking.
VI. SOME IONOSPHERIC PHENOMENA

The wind-shear theory has been quite controversial (Layzer, 1964), but it is supported by some observational evidence (MacLeod, 1966). In general, the closer the wind measurements are related in space and time to the Es observations, the better the correlation is found to be. Wright et al. (1967) have made a particularly detailed comparison between ionosonde Es data and the winds measured by a series of shots fired from the Barbados gun (Sec. 1.82).

The wind-shear theory, however, meets a serious quantitative difficulty. Values of \( w' \) deduced from observation are of order \( 10 \, \text{m s}^{-1} \, \text{km}^{-1} \) or \( 10^{-2} \, \text{s}^{-1} \). If we take \( \alpha = 10^{-7} \, \text{cm}^3 \, \text{s}^{-1} \) (Table I), we find from Eq. (607) that \( N_{\text{max}} \sim 10^5 \, \text{cm}^{-3} \) only, no greater than the electron concentration in the normal E layer, and much less than observed values of \( N_m \text{Es} \) which may exceed \( 10^6 \, \text{cm}^{-3} \). This difficulty could be overcome by supposing that Es layers are composed of long-lived ions, such as metallic ions. Indeed, there is now mass-spectrometer evidence that Es layers do contain metallic ions (Young et al., 1967). The recombination coefficients for such ions are much smaller than the dissociative recombination coefficients of the NO\(^+\) and O\(_2\)\(^+\) ions which exist in the normal E layer, and probably correspond to radiative recombination, for which \( \alpha \sim 10^{-12} \, \text{cm}^3 \, \text{s}^{-1} \) (see Table I). In this case \( w'/\alpha \sim 10^{10} \, \text{cm}^{-3} \), much greater than observed values of \( N_m \text{Es} \), so the electron concentration is presumably limited by plasma diffusion or by an alternative loss process, which might be electron attachment to neutral molecules. We may therefore expect quite a complex situation to exist in an Es layer produced by wind shears, involving production of metallic ions by photoionization or detachment processes, loss by attachment or recombination, and transport of electrons and ions by winds and plasma diffusion.

If attachment is the dominant loss process, the loss term in Eq. (605) takes the form \(- a(h) N\) instead of \(- \alpha N^2\), and it is no longer true that \( N_{\text{max}} \) is independent of \( q\), as in Eq. (607). The attachment coefficient \( a(h) \) can be deduced from studies of the decay of meteor trails. Although the meteor observations mostly relate to altitudes of 80-100 km, values in the range \( 10^{-3} \, \text{s}^{-1} \) to \( 10^{-2} \, \text{s}^{-1} \) can be estimated for typical Es heights of 105-110 km; two-body attachment is probably more important than three-body attachment at these heights (Greenhow and Hall, 1961).

6.4 Ionospheric Irregularities

So far in this book, we have dealt only with the large-scale or "background" distribution of ionization. Superimposed on this background, small-
scale irregularities of ionization seem to exist at every level in the ionosphere. The basic properties to be measured are their size and shape, intensity (i.e., fractional deviation of electron density), and their direction and speed of motion. The measurements are likely to be influenced by instrumental selection. For instance, a spectrum of irregularities may exist at a particular level, of which only certain components will be recorded by any particular equipment. Some information about irregularities is given by standard ionograms or $h'(f)$ curves, and by continuous recordings of phase heights at fixed frequencies, or $h'(t)$ curves. These are particularly useful for the detection of large “travelling ionospheric disturbances,” to be discussed later in this section.

Irregularities can be detected by their ability to scatter radio waves. Signals at frequencies of 50–100 MHz, too high to be reflected from a smooth ionosphere even at oblique incidence, may be received 1000 km from the transmitter, having been scattered from irregularities in the D region at 80 to 90 km (Bailey et al., 1955). This forward scatter is of practical importance, but most of the scientific study of the irregularities themselves has been conducted with more specialized techniques.

Though small-scale irregularities are a normal feature of the ionosphere, there seems to be no generally accepted theory of their formation. Below the turbopause—at around 100 km—turbulent air motions might produce irregularities. This might establish an irregular pattern of electric fields, which could cause irregular drifts at greater heights where the ionization is not directly influenced by air motion (Dagg, 1957). However, the required electrical coupling along geomagnetic field lines, though good for large-scale fields, is limited for small-scale fields (Farley, 1959). Gravity waves seem the most probable explanation of the large-scale travelling disturbances (Sec. 6.44), and it may be that smaller scale motions of the neutral air play some part in producing irregularities. Another possibility is that instability may arise in parts of the ionosphere where $N$ increases upwards (Liu and Yeh, 1966); the production of equatorial Es irregularities by plasma instability has already been mentioned (Sec. 6.31).

6.41 The Fading Method of Measuring Ionospheric Drifts

The pulse reflection method of Mitra (1949) generally uses a transmitter working at a frequency of a few megahertz, and three receivers, arranged in a triangle about one wavelength apart, to record the fluctuations in amplitude of the signal reflected from the ionosphere. This fading is usually attributed to irregularities situated near—or somewhat below—the height of
reflection (though in principle it might be caused by irregularities at any lower height). By choosing suitable radio frequencies, it is possible to study the F layer or the daytime E layer, and the method has also been applied to nighttime Es layers; see the review by Briggs and Spencer (1954). Using c.w. signals from distant transmitters at about 1 MHz, Beynon and Goodwin (1967) have studied movements in the D region and lower E region, while Sprenger and Schminder (1967) used signals in the 100–300 kHz range to study the night E region. An alternative method uses a radio star as the source; the fading of the signals recorded by spaced receivers (using frequencies in the range 30–100 MHz) is then mainly due to irregularities in the F region.

Figure 55 sketches an idealized situation, in which the signals at three receivers undergo fading owing to the drift of ionospheric irregularities in the direction shown. If the pattern of irregularities drifts without changing, and its scale size greatly exceeds the receiver spacing, all three records would have the same form, but with systematic time delays between the appearance

Fig. 55. (a) Idealized signal amplitudes obtained from three receivers of a Mitra drift experiment, showing time delays. (b) The assumed layout, in which A is the apex of a right-angled triangle with sides AX = x, AY = y. The dashed lines indicate three positions of a cross section of the amplitude pattern, which takes time t_x to drift from A to X and time t_y from A to Y.
of different features. Thus for receivers \( A \) and \( X \) we should have a set of equal delays, \( t_x' = t_x'' = \cdots \); and for receivers \( A \) and \( Y \) we should have \( t_y' = t_y'' = \cdots \). In practice the patterns do not remain unchanging, so that the delays \( t_x', t_x'' \) (etc.) are not identical. However, it is generally possible to obtain median values \( t_x, t_y \) from which the drift speed \( V \) and azimuth \( \theta \) can be obtained by solving the equations (where \( AX = x, AY = y \))

\[
\begin{align*}
t_x &= (x \cos \theta) / V, \\
t_y &= (y \sin \theta) / V
\end{align*}
\]  

(608)

A special recorder was devised by Phillips to measure directly the time delays between the signals received at a pair of spaced antennas. More complex systems for recording drifts have been devised; for instance Haubert and Doyen (1966) have constructed a square array of 36 antennas. By means of a 36-element visual display, the drift motions can be seen by eye or recorded on film. An 89-element array has been built near Adelaide, South Australia.

For the reflection method, we should expect the velocity \( V \) of the pattern observed at the ground to be twice the actual drift velocity in the ionosphere (though for the radio-star method, the two velocities are equal). This simple idea has been questioned by Wright and Fedor (1967), who compared drift speeds \( (V) \), observed at the ground from reflections from night Es layers, with simultaneous neutral wind speeds \( (U) \) determined by chemical releases from gun-launched probes. In the E region, the neutral air and the ionization should move with the same speed. The observations gave better support to the hypothesis that \( V = U \) than to the expected result that \( V = 2U \); the existence of this discrepancy, and possible reasons for it, are under further investigation.

The simple "time-delay" approach is only valid if the ionospheric irregularities possess certain statistical properties. It has largely been superseded by a more sophisticated analysis which treats fading statistically (Briggs et al., 1950; Phillips and Spencer, 1955; Bowhill, 1961b). First, it should be noted that the wavefield at the ground is formed by diffraction from an irregular "screen" in the ionosphere. This "screen" consists of irregularities which impose phase variations on the wavefront of the downcoming radio wave. If the variations are "deep," in the sense that the phase deviations exceed one radian, the pattern at the ground is (in some circumstances) smaller in scale than the screen, but it may retain the general shape and orientation of the irregularities (Hewish, 1951; Fooks and Jones, 1961). For a "shallow" screen, in which the phase changes are less than a radian, the amplitude pattern on the ground should be very similar in size and shape to the structure of the irregularities in the ionosphere.
To describe the shape and size of the irregularities in the screen, use is made of the autocorrelation function of the electron density distribution. In a two-dimensional Cartesian system, this function \( \rho(\xi, \eta) \) measures the average correlation between the amplitudes at the two points \((x, y)\) and \((x + \xi, y + \eta)\) for all \(x\) and \(y\). Then \( \rho = 1 \) at the origin \((\xi = \eta = 0)\), and decreases with increasing radial distance. The correlation may depend on direction, and it is generally assumed that the irregularities can be described in terms of a “characteristic ellipse,” whose radius in any direction is the distance in which the correlation falls to the value 0.5. The shape, size, and orientation of this ellipse provide information about the ionospheric irregularities.

The pattern of irregularities may change as it moves. If it does, its persistence can be described by a time correlation function \( \rho(\tau) \), which measures the average correlation between the patterns existing at two instants separated by interval \(\tau\). The ratio of the scale sizes of the space correlation function \( \rho(\xi, \eta) \) and the time correlation function \( \rho(\tau) \) gives a “velocity” parameter \( V_c \); Briggs et al. (1950) and Phillips and Spencer (1955) discuss its definition, and how it is derived from actual data. If \( V_c = 0 \), the pattern drifts without change. If \( V_c \) is comparable to the drift velocity \( V \) of the pattern, it implies that random changes in the pattern contribute to the observed fading as much as does the bodily drift, which appears to be the situation usually encountered.

All these quantities can be determined from sets of fading records by a complicated procedure known as “full correlation analysis.” This has been applied by Fooks and Jones (1961) who compare the results with those of a simpler analysis developed by Yerg (1956) and the straightforward “time-delay” approximation described earlier. They further discuss the circumstances under which the simpler methods can give meaningful results. The time-delay method seems to lead to overestimates of drift velocity. Moreover, if the irregularities are elongated, the simple time-delay method tends to emphasize motions normal to the axis of elongation, and thus give incorrect results concerning the direction of drift.

The reflection method of Mitra has been used to study drifts in many parts of the world. Most of the data have been analyzed only by simple procedures such as the time-delay method. Deduced drift velocities are typically a few tens of meters per second, but the details of their magnitudes and directions are quite complicated and consistent patterns do not always emerge. Survey articles, with references, have been given by Shimazaki (1959a), Briggs (1962), Rawer (1963), and Rao and Rao (1963).
The irregularities in the E and lower F region are often quoted as being a hundred meters or so in dimension. Probably there exists a distribution of excess electron density which contains a broad spectrum of spatial periods. When "correlation ellipses" are computed, the ratios of their major and minor axes are generally found to lie in the range from 1.5 to 2. The irregularities are thus not greatly elongated, and the directions of elongation do not seem to follow a very clear-cut pattern. On the other hand, F2-layer irregularities are generally found to be elongated in the direction of the geomagnetic field, particularly in equatorial latitudes, in auroral latitudes, and in the topside ionosphere.

The ratio $V_e/V$ (which, as previously stated, measures the relative importance of random change and bodily drift in producing fading), is sometimes small and sometimes large, but its median value is near unity. Typical irregularities are thought to contain a fractional deviation of electron density, $\Delta N/N$, of a few percent. Fooks (1962) has quoted a figure of 2% for irregularities in the E layer and lower F layer by day.

There has been much discussion as to (i) whether the drifts measured by the fading method represent real motions of the background ionization, and (ii) what relation the drifts bear to neutral air motions. No simple solution to these problems seems to exist; however, regarding (i), Clemmow and Johnson (1959) and Kato (1965) showed that under certain circumstances an irregularity aligned with the magnetic field drifts without change of shape under the influence of an applied electric field. Regarding (ii), the discussion of Sec. 4.21 leads us to expect neutral winds and ionization drifts to be closely related in the lower ionosphere, but very little related in the F region.

6.42 Radio-Star Scintillation

The irregularities principally responsible for radio-star scintillations are located in the F region, at heights of 250 km and above. On occasion they are above the level of the F2 peak and are therefore inaccessible to conventional ground-based sounding, although they can be studied by satellite techniques such as beacon transmissions and the topside sounder.

Early observations of radio-star scintillations at 81 MHz showed that records obtained from spaced receivers showed good correlation for spacings up to a few kilometers. For a spacing of 200 km, however, the occurrences of scintillation at the two sites are related, but the scintillations themselves are quite uncorrelated (Smith et al., 1950). This indicated a terrestrial, and presumably ionospheric, origin of the scintillations. By using diffracting screen theory, Hewish (1952) was able to estimate an altitude of 400 km and
dimensions of several kilometers for irregularities producing scintillations at 45 MHz. Steady drift velocities of 100 m s\(^{-1}\) or more were found. Later work on nighttime scintillation has tended to confirm Hewish's conclusions in general, but some scintillation at low elevation angles has been associated with Es ionization (Wild and Roberts, 1956) and E-region irregularities (Chivers and Greenhow, 1959).

In order to make a statistical study of the occurrence of scintillation, it is necessary to assign numerical indices to describe the amount of scintillation. This process tends to be somewhat subjective though it is possible for individual observers to achieve consistency in the analysis of a long series of records. Prolonged series of observations, extending over several years, have been made in Britain (Briggs, 1964; Chivers, 1960) and in West Africa (Koster and Wright, 1960). Scintillations occur mainly at night, but are sometimes observed by day. It is not clear whether the apparent seasonal variations observed in Britain are real, or result from the dependence of scintillation on the zenith angle of the radio source. But it is clear that scintillations are more common at sunspot maximum than at sunspot minimum although this is the opposite of the solar-cycle dependence of spread F. It has also been found from satellite observations that the regions producing scintillations are quite dependent on magnetic latitude (Yeh and Swenson, 1959; Kent, 1961); scintillations are especially pronounced when the signals pass through the ionosphere at auroral latitudes.

### 6.43 Spread F

Spreading of F-region traces on ionograms has been divided into two main types. One is "range spreading", in which two or more traces with different virtual heights are seen at frequencies well below \(f_oF2\), and which seems to be prevalent at lower latitudes. The other is "frequency spreading," in which the high-frequency ends of the traces are branched or blurred. This is the type most commonly seen at higher latitudes, and the range \(\Delta f\) of frequencies covered by the "spreading" is related to the fractional deviation of electron concentration within the irregularities (i.e., \(\Delta f / f_oF2 \approx \frac{1}{2} \Delta N/N\)). Indices to measure spread F have been defined empirically and used for statistical studies, including comparisons with the empirical indices used for radio-star scintillations. Published bulletins of ionospheric data contain information about the occurrence of spread F, but scaling practices are known to vary, and care must be exercised in the use of such data for detailed statistical studies. On account of their availability, they have been widely used for studies of worldwide occurrence of spread
F. Among these may be mentioned the work of Shimazaki (1959b) and Singleton (1960, 1962).

From such studies, it has been found that spread F occurs most frequently in equatorial and auroral latitudes, less frequently in temperate and polar regions. Spread F occurrence is positively correlated with magnetic disturbance at high latitudes but negatively correlated at low latitudes. At low latitudes the phenomenon appears to be closely connected with the increase of height of the F layer after sunset. It has been suggested that there is a causal connection between these phenomena. Singleton further finds that probability of occurrence is negatively correlated with F2-layer critical frequency. This appears to hold also for the solar-cycle variation at temperate latitudes. Thus spread F is less common at sunspot maximum than at sunspot minimum, opposite to the behavior of radio-star scintillation. For this reason, Briggs (1964) thinks that at sunspot maximum the greatest degree of irregularity occurs above the F2 peak, where it cannot produce effects on bottomside ionograms. This seems to be borne out by a comparison of topside and bottomside data (Dyson, 1967).

Ionograms obtained from the Alouette topside sounder have also given considerable information about the distribution of spread F. Calvert and Schmid (1964) found that it occurred most frequently at low latitudes, during the night and mid-morning, and at high latitudes, at any hour. It was found to be less common at mid-latitudes. More details of the distribution of topside spread F have been given by King et al. (1967c).

Much study has been devoted to the detailed form of spread F echoes on ionograms, both bottomside and topside, and the way in which they are related to the geometry of the irregularities. This work entails wave propagation calculations for an irregular medium. Considerable success has been achieved in relating field-aligned irregularities associated with equatorial spread F to the “ducts” which guide the radio waves transmitted by Alouette. In circumstances when multiple reflections occur, waveguide calculations have proved fruitful in explaining features observed on ionograms (Calvert and Cohen, 1961; Pitteway and Cohen, 1961; Muldrew, 1963).

6.44 Travelling Ionospheric Disturbances

Large irregularities are occasionally detected as distortions of ionogram traces. A particular type, known as a “travelling ionospheric disturbance” (TID), produces “kinks” in the “o” and “x” traces near the F2 cusps on an ionogram, and these gradually move down toward lower frequencies (Munro, 1950, 1958; Rawer, 1959). These are believed to be wavefronts of distur-
bances, many hundreds of kilometers in extent, which move at speeds of order $5-10 \text{ km min}^{-1}$, or several hundred kilometers per hour. They have been observed to travel for horizontal distances of up to 3000 km. The apparent downward motions seen on ionograms can be accounted for if the wavefronts are supposed to be inclined to the vertical. By comparing observations made at two Australian stations 1000 km apart, Heisler and Whitehead (1961) found the group velocity of travelling disturbances to be $10 \text{ km min}^{-1}$, roughly twice the phase velocity which was deduced from observations at a single station. As mentioned in Sec. 1.64, Hines (1960) has identified the travelling disturbances as atmospheric gravity waves, while Wickersham (1965) found most of the ionospheric data to be consistent with this interpretation.

The disturbances cause "tilts" in the F2 layer, the motions of which can be studied by a direction-finding technique, using pulse signals at a few megahertz (Bramley, 1953). These motions are also visible on high-frequency backscatter sounders which survey the ionosphere for thousands of kilometers about the transmitting/receiving location (Tveten, 1961). The equipment transmits millisecond r.f. pulses at frequencies above the F2-layer critical frequency and then receives energy backscattered from the ground, after an intervening ionospheric reflection. No echoes are received within the "skip zone," but returns at greater distances from all azimuths are observed with a rotating, directive antenna. The motions of large-scale travelling disturbances cause a "ripple" to sweep across the echo pattern during the several hours in which the disturbance traverses the area surveyed by the sounder.

More recently, the incoherent scatter sounder at the Arecibo Ionospheric Observatory (Thome, 1964) has been used to detect oscillations in the electron density. Excursions of $\pm 10\%$ or $\pm 20\%$ are found at altitudes between 150 and 600 km with periods between 20 min and several hours. The disturbance motion can be demonstrated even better by tilting the radar beam off vertical and then rotating it about the vertical direction, much as with conventional ground backscatter equipment. Velocities of $3-6 \text{ km min}^{-1}$ are measured.

Satellites may also be used to study large-scale irregularities in the ionosphere. By observing the Faraday rotation of signals from Explorer 7, Titheridge (1963) obtained statistical information on the sizes of irregularities, which ranged from 5 to 500 km. Large irregularities, exceeding 50 km in horizontal scale, generally formed part of a sequence, suggesting that the disturbances were wavelike in character. Geostationary satellites provide further scope for studying these phenomena.
6.5 Electron and Ion Temperatures

6.51 The Lack of Thermal Equilibrium

During the early sixties it became known that in the ionosphere the electron temperature $T_e$ sometimes differs from the temperatures of ions and neutrals, $T_i$ and $T_n$. The differences have been mentioned at several places in the book. They became particularly evident from the direct measurements obtained with rockets and satellites, which gave values of $T_e$ much higher than plausible values of $T_n$ appropriate to the prevailing conditions. In addition, the spectra of incoherent scatter signals frequently could be interpreted only in terms of values of $T_e/T_i > 1$. Some of the pertinent data have been reviewed by Bourdeau (1963) and Evans (1967a).

In this section we do not attempt to discuss the data in much detail, but will briefly describe the relevant theory. We wish to bring together the several aspects of electron temperatures which have been mentioned elsewhere in the book.

It is now clear that the lack of thermal equilibrium is a normal feature of the F region, certainly by day and perhaps by night. At lower levels, collision frequencies are sufficiently large to ensure good thermal contact between the electrons, ions and neutrals, and it is theoretically unlikely that temperature differences can be maintained in the D region or even in the E region. However, this does not mean that differences can never exist.

The basic reason for the lack of equilibrium is that the energy carried by ionizing photons or charged particles generally exceeds the energy required for ionization. Much of the excess is carried away by the photoelectrons. Unless the time taken for these fast "superthermal" photoelectrons to lose excess energy to the neutral gas is very much shorter than their lifetime before recombination, the average electron energy exceeds that of other particles. It is only meaningful to speak of an electron temperature if the superthermal electrons share their energy with other electrons, so that the electron gas possesses a Maxwell velocity distribution, or something very close to it. This does appear to be the case for ionospheric electrons though sometimes there may exist a "high energy tail" of superthermal electrons, particularly in the upper ionosphere and protonosphere where collisions are so infrequent that the electrons do not become thermalized within their lifetime.

6.52 Thermal Balance Equations

The equations that describe the heat balance of the electrons and ions resemble those applicable to the neutral gas, that we discussed in Sec. 1.3.
The equations also have some resemblance to the production, loss and transport equations for the ionization, which we discussed in Chapters III and IV.

The physical processes involved have been discussed by several authors, notably Hanson (1963) and Dalgarno et al. (1963). To start with, the total rate of input of solar energy is given by the equations of Sec. 1.32. It is then necessary to investigate the photoionization processes in more detail, to derive the energy distribution of the resulting photoelectrons. The energetic photoelectrons are slowed down to thermal energy partly by inelastic collisions with neutral particles, partly by Coulomb collisions with ambient ions, and partly by Coulomb collisions with ambient electrons; this last process provides the heat input to the electron gas. If the photoelectrons travel only a short distance (much less than a scale height) before losing their energy, the heating is said to be "local." Above 300 km the heating may largely be due to photoelectrons produced at some other altitude, since electrons with energies of only a few electron volts can travel large distances before becoming thermalized; they may even escape from the ionosphere, travel along geomagnetic field lines and deposit their energy in the opposite hemisphere. This process of "nonlocal" heating has been discussed by Geisler and Bowhill (1965a, b). As an alternative to photoionization, the initial energy source may be fast particles entering the atmosphere, in which case complex calculations are needed to compute the electron heating function for an energy spectrum of incident particles.

We shall suppose that a function $Q_e(h, t)$ can be evaluated to describe the heat input to the electron gas, whether by solar photoionization or by energetic particles. Since the thermal capacity per unit volume of the electron gas is $\frac{1}{2}kN_e$, the rate of rise of electron temperature is given by

$$dT_e/dt = \frac{2Q_e}{3kN_e}$$

(609)

where $k$ is Boltzmann's constant. An extra term should appear in Eq. (609) if $N_e$ is varying with time (Banks, 1967); however, it is usually small.

In constructing equations to describe the heat balance, we shall as usual employ subscripts $e, i, n$ to denote electrons, ions, neutrals, respectively. A suffix $j$ is used to distinguish between different neutral species (concentrations $n[j]$); $M_i$ denotes ion mass in a.m.u. Numerical values are obtained from Spitzer (1956), Hanson (1963) and other papers to be cited, and are expressed in the c.g.s. units customarily used in this subject.

For the moment we neglect heat conduction. The rate of cooling of the
electron gas can be written

\[- \frac{dT_e}{dt} = \frac{N_i (T_e - T_i)}{17 M_i T_e^{3/2}} + \sum_j u_{en}^{(j)n} n_j [j] (T_e - T_n) \tag{610}\]

The first term on the right of this equation represents heat loss to the ions. It has to be replaced by a summation if more than one ion species is present. The other term represents the heat lost to the neutral gases, the items in the summation referring to the neutral species O, O₂, N₂, and He. Hanson (1963) and Dalgarno et al. (1963, 1967) give data on the coefficients we denote by \(u_{en}\); they take account of elastic collisions between electrons and neutral particles, as well as rotational and vibrational excitation of the molecular gases.

On combining Eqs. (609) and (610) we can solve for \(T_e\). If we take account only of cooling by ions in Eq. (610), as may be a reasonable first-order approximation, we obtain a balance equation

\[Q_e / N_e^2 = 7.6 \times 10^{-6} (T_e - T_i) / M_i T_e^{3/2} \text{ [eV cm}^{-3} \text{ s}^{-1}] \tag{611}\]

For a fixed value of \(T_i\), the right-hand side of Eq. (611) has a maximum value when \(T_e = 3 T_i\). If the ratio \(Q_e / N_e^2\) is increased beyond this maximum, the electrons cannot lose heat fast enough to balance the rate of input, and \(T_e\) is then limited by the cooling due to neutral gases and by thermal conduction (which we consider shortly). Although Eq. (611) neglects these other processes and is therefore very approximate, we shall see in Sec. 6.53 that it does account for some features of the observational data.

We can also compute the ion temperature \(T_i\). We shall only give equations for a single ion species although they can be generalized to include several species. The ions are heated by Coulomb collisions with the electrons, at a rate \(Q_i\) which is equal to the ionic cooling rate for electrons used in Eq. (610). They are cooled by elastic collisions with neutral particles, and also by charge-transfer collisions with their parent atoms. The rate of cooling depends on coefficients \(u_{in}^{(j)}\); values of these coefficients for different neutral species \((j)\) have been given by Brace et al. (1965) and by Dalgarno and Walker (1967). We then have

\[Q_i = 7.6 \times 10^{-6} (N_e N_i / M_i) (T_e - T_i) T_e^{-3/2} \text{ [eV cm}^{-3} \text{ s}^{-1}] \tag{612}\]

\[L_i = N_i \sum_j u_{in}^{(j)n} n_j [j] (T_i - T_n) \tag{613}\]

If there is only a single ion species then of course \(N_i = N_e\). We can then set
So far, we have assumed equilibrium and neglected heat conduction. The assumption of equilibrium is quite reasonable, since the electron gas responds very rapidly to changes of heating, as pointed out by da Rosa (1966). On the other hand, thermal conduction is very important in limiting the electron temperature. Assuming equilibrium, we can write the thermal conduction equation in the form

$$Q_e - L_e + \frac{d}{dh} \left( \kappa \frac{d T_e}{dh} \right) = 0$$

(614)

in which $\kappa$ is the thermal conductivity. In the upper ionosphere, heat conduction in the electron gas takes place only along field lines, and is controlled by electron-ion collisions only. In this case, the thermal conductivity is that of a fully ionized gas (Spitzer, 1956; Geisler and Bowhill, 1965b); it depends strongly on $T_e$, but it is independent of $N_e$. If $I$ denotes the magnetic dip angle, then

$$\kappa = \kappa_{ei} = 7.7 \times 10^5 \sin^2 I \times T_e^{5/2} \left[ \text{eV cm}^{-1} \text{s}^{-1} \text{K}^{-1} \right]$$

(615)

This expression is inadequate below 200 km, because there the conductivity is controlled by electron-neutral collisions (Banks, 1966). We can introduce another conductivity parameter $\kappa_{en}$, dependent on electron-neutral collisions, and obtain the combined conductivity from the formula

$$\frac{1}{\kappa} = \frac{1}{\kappa_{ei}} + \frac{1}{\kappa_{en}}$$

(616)

Dalgarno et al. (1967) take

$$\kappa_{en} = 3.2 \times 10^7 \sin^2 I \times T_e N_e / v_{en} \left[ \text{eV cm}^{-1} \text{s}^{-1} \text{K}^{-1} \right]$$

(617)

but use a more general expression for $v_{en}$ than the formula we gave in Sec. 4.12.

Numerical solutions of the conduction equation (614) have been obtained by Geisler and Bowhill (1965b), da Rosa (1966), and Dalgarno et al. (1967). The obvious effect of thermal conduction is to smooth out the $T_e(h)$ profile,
particularly under conditions when $T_e$ is large and the thermal conductivity correspondingly large. If the heat flux is assumed to vanish at the top of the ionosphere, there must be an isothermal region in which $T_e$ is independent of height. This is essentially true above 400 km for most of the theoretical curves in the papers just cited.

If a gradient of $T_e$ exists along field lines at great heights, there must be a flux of heat between the ionosphere and protonosphere. The exchange of energy between these two regions has been described by Geisler and Bowhill (1965a). In particular, they show that heat drains from the protonosphere after sunset, at first rapidly; within about an hour, the reduction of electron temperature reduces the thermal conductivity and the flow becomes much slower. At sunrise, ionospheric electron temperatures build up to their daytime values within about two hours, as computed by da Rosa (1966). The increase of $T_e$ begins at a solar zenith angle $\chi \approx 110^\circ$, in contrast to the production of ionization which does not become significant until $\chi \leq 100^\circ$.

Up to about 250 km, the thermal coupling between ions and neutrals is good enough to ensure that $T_i = T_n$. At greater heights the coupling becomes worse, whereas the thermal coupling between ions and electrons improves. Consequently, $T_i > T_n$ and at much higher heights $T_i \rightarrow T_e$. This behavior is shown in the profiles of Fig. 56, taken from the results of Dalgarno et al. (1967), who computed complete diurnal variations of $T_e$ and $T_i$. The broken curve marked $T'_e$ is computed from the equation $Q_e = L_e$, and therefore neglects conduction; the diagram thus shows the way in which conduction smooths out the electron temperature profile. The experimental data plotted in the graph are those of Evans (1965a), obtained by the incoherent scatter technique.

6.53 Some Phenomena Concerned with Electron and Ion Temperatures

The large amount of observational data on electron and ion temperatures include rocket data at a number of sites; satellite data, offering global coverage but limited height and local time resolution; and incoherent scatter experiments which can give detailed information at a limited number of sites. We refer to the review articles by Evans (1967a) and Gringauz (1967). In this section, we wish to discuss some points of interest which emerge from the data.

Figure 57 shows some vertical profiles of electron temperature obtained at Fort Churchill by Brace et al. (1963). At Wallops Island, a mid-latitude station, $T_e$ is greater by day than by night. On a disturbed day, $T_e$ is enhanced.
and approaches the values observed at Fort Churchill, an auroral station, at which quiet day and disturbed day values are similar. However, it has not been established that magnetic disturbance generally increases the electron temperature at mid-latitude stations by day, though it may do so by night. Thus, the increase of electron temperature at Wallops Island might be due to the proximity, under disturbed conditions, of the auroral region. Since Willmore (1965) has reported a decrease of $T_e$ during storms, the pattern of electron temperature variations during magnetic disturbance is not yet well determined (Gringauz, 1967).

The excess of $T_e$ over $T_i$ is observed at heights down to about 130 km (Evans, 1967b; Carru et al., 1967). It is not expected that significant enhancements of $T_e$ should exist at lower altitudes, except possibly in sporadic E layers (Sec. 6.32) or in high latitude regions subject to intense particle bom-
6.5 ELECTRON AND ION TEMPERATURES

Fig. 57. Electron temperature profiles obtained on five rocket flights from Wallops Island, Virginia (38°N, 75°W) and Fort Churchill, Canada (59°N, 92°W) [Brace et al. (1963)].

... bombardment. In the daytime F region, $T_e/T_i$ generally lies between 2 and 3. At sunset, and during a solar eclipse, $T_e$ is observed to decrease rapidly, and the thermal contraction produced thereby might cause the observed increases of $N_e$, described in Secs. 5.41 and 6.2 (Evans, 1965a, d).

Near sunrise the ratio $Q_e/N_e^2$ in Eq. (611) becomes large, and $T_e$ might then attain values of several thousand degrees (Dalgarno and McElroy, 1965). In reality $T_e$ is limited by thermal conduction (da Rosa, 1966); however, large values of $T_e$ are observed near the magnetic equator, where the gradient of $T_e$ is small in the direction of the geomagnetic field lines and conduction is, therefore, rather ineffective in cooling the electron gas (Farley et al., 1967).

Because the electron cooling rate depends strongly on electron concentration [Eq. (611)], we might expect variations of $T_e$ to be inversely related to those of $N_e$. On the whole, such a relationship is borne out by the data...
(Willmore, 1965; Brace et al., 1967). From world-wide data for 1000 km altitude, obtained by the Explorer XXII satellite, Brace et al. discovered an equatorial minimum in the latitude variation of $T_e$; they suggested that this might be due to the presence of a relatively large amount of neutral hydrogen, which is more effective than other neutral constituents in cooling the electron gas.

The available data suggest that the ratio $T_e/T_i$ decreases from day to night in the mid-latitude F region, but remains in excess of unity. This implies the existence of a heat source. Willmore (1964) used Ariel I satellite data to plot the worldwide distribution of heat input, which was found to bear some relation to the fluxes of low energy particles observed by Savenko et al. (1963). More recently, Nathan and Seaton (1966) and Nathan (1966) have shown that a small particle flux (of order $10^{-2}$ erg cm$^{-2}$ s$^{-1}$), consistent with limits set by airglow data, suffices to give the heat inputs required by Willmore's data and the Explorer XVII observations (Brace et al., 1965). These calculations, however, neglect the heat gain in the topside ionosphere by conduction from the protonosphere, and the simultaneous loss by conduction to the base of the thermosphere. It is well established that some increases of electron temperature during darkness can be attributed to the influx of photoelectrons along field lines from the sunlit conjugate point, increases of red line airglow emission also being observed (Cole, 1965; Carlson, 1966; Evans, 1965b).

We may mention one further indication of the interaction of the ionosphere and protonosphere. Contrary to the theoretical prediction that $T_e$ should be independent of height, Evans' profiles generally show a positive gradient $dT_e/dh$ at great heights (usually more pronounced than in the example shown in Fig. 56). Evans (1967d) suggests that heating of the protonosphere, due to escaping ionospheric photoelectrons, has been underestimated by previous theories; it seems that such heating might maintain the electron temperature at about 5000°K in the protonosphere, as suggested by some experimental data (Serbu and Maier, 1966).
7.1 Introduction

Any discussion of the ionosphere would be incomplete without consideration of the terrestrial magnetic field. We have already encountered examples of this. For instance, many of the experimental techniques described in Chapter II depend on, or are influenced by, the effect of the magnetic field on radio wave propagation. In addition, we discussed in Chapter IV how the magnetic field controls the motion of ionization. In evaluating all these effects, it is necessary to have an adequate description of the main properties of the earth's field, such as its direction and intensity. We discuss this topic in Sec. 7.2.

In addition to the main geomagnetic field, which originates within the earth, there exist short-term variations such as we describe in Sec. 7.3. Some of these are fairly regular, being related in period to the solar and lunar day. They are attributed to motions of the upper atmosphere, which produce electromotive forces and electric currents by dynamo action (Sec. 7.4). Others constitute the disturbances known as magnetic storms, and are associated with a whole range of ionospheric and geophysical phenomena, which are reviewed in Chapter VIII.

The geomagnetic field pervades the region around the earth, extending to distances of several earth radii. This is called the magnetosphere although we have previously used other terms, such as protonosphere and exosphere, when discussing different aspects of it (Fig. 1). In our description of the magnetosphere (Sec. 7.5) we can only outline the picture, emerging from contemporary data, of its relationship to the ionosphere. The relationship is especially pertinent to high-latitude and storm phenomena, but we wish to avoid the complexity which attends any deep discussion of these subjects.
7.11 Historical Notes

One of the earliest experiments related to geomagnetism was described in 1600 by Queen Elizabeth's physician, William Gilbert, and it provided considerable insight into the nature of the earth's magnetic field. As related by S. Chapman (1951), Gilbert "cut a spherical piece of the naturally magnetized mineral called lodestone, and examined the distribution of direction of the magnetic force over its surface by means of tiny magnetized needles freely pivoted. He saw that the distribution of dip agreed with what was known of the earth's field. Hence, he concluded that the earth itself is a great magnet, similar to his magnetized sphere except in size." We can now show mathematically that the external field of a uniformly magnetized sphere is identical to that of a dipole of appropriate magnitude placed at the center of the sphere. Since we find, as did Gilbert, that this does provide a reasonable first-order fit to the observed field near the earth's surface, we consider the dipole field in more detail a little later.

A major advance was made in 1722 by G. Graham, who was a clockmaker in London. He discovered the regular diurnal variations of the compass needle and also the larger perturbations which occur at times of magnetic storms. Further work by A. Celsius and J. C. Wilcke showed that magnetic disturbances are quite widespread and well correlated with the visible auroras. It was still another century before H. Schwabe, in 1843, noted the periodicity of the sunspot number, and it was later established that this periodicity is reproduced in the frequency of magnetic disturbance.

Further historical details are given in the standard work on geomagnetism by Chapman and Bartels (1940). We have already mentioned (Sec. 2.1) the historical association of the "atmospheric dynamo" concept of Balfour Stewart with the ionosphere; the theory was further developed by A. Schuster and others. Early in the present century, K. Birkeland and C. Störmer began their studies of the motions of charged particles in the geomagnetic field, which are of concern to cosmic-ray physics. Later, Chapman and Ferraro (1931) published their first studies of the interaction of solar particle streams and the earth's field. From these beginnings have grown the modern concepts of solar-terrestrial physics and magnetic storm theory.

7.2 Representation of the Main Geomagnetic Field

The strength of the earth's field is measured at numerous magnetic observatories located all over the globe, now supplemented by satellite-borne in-
7.2 REPRESENTATION OF THE MAIN GEOMAGNETIC FIELD

Instruments. The standard components recorded on ground-based magnetograms are usually: (a) northward \( X \); eastward \( Y \); downward \( Z \); or (b) easterly declination \( D = \arctan (Y/X) \); horizontal intensity \( H = (X^2 + Y^2)^{1/2} \); vertical intensity \( V = |Z| \). Sometimes the dip angle or inclination, \( I = \arctan (Z/H) \), is also given. The Cartesian coordinate system differs from those conventionally used for the ionosphere (south, east, up) and in meteorology (east, north, up).

The mathematical representation of the geomagnetic field is treated by Chapman and Bartels (1940). A very convenient way to describe the field is by means of a "potential function." Since the currents which produce the main field of the earth do not flow across the earth’s surface (measured atmosphere-to-ground currents average about \( 10^{-12} \) A \( m^{-2} \), which are too small to produce significant magnetic fields), the field is "curl-free" at the surface, \( \mathbf{V} \times \mathbf{B} = 0 \), and it, therefore, may be obtained from a scalar potential, \( \mathbf{B} = - \mathbf{\nabla} V \). Since \( \mathbf{V} \cdot \mathbf{B} = 0 \) also, we see that the potential function \( V \) satisfies Laplace’s equation. It is generally written as the sum of a series, of which each term is a product of three functions, each containing only one of the variables, colatitude \( \theta \) increasing southward, longitude \( \lambda \) increasing eastward and radial distance \( r \) increasing outward:

\[
V = a \sum_{n=1}^{\infty} \left( \frac{a}{r} \right)^{n+1} \sum_{m=0}^{n} P_n^m(\cos \theta) \left[ (g_n^m) \cos m\lambda + (h_n^m) \sin m\lambda \right]
\]  

(700)

In this equation \( P_n^m(\cos \theta) \) is the Schmidt function (a modified form of associated Legendre polynomial) of degree \( n \) and order \( m \), and \( a \) is the earth radius. We have ignored a radial solution proportional to \( (r/a)^n \), which restricts our results to the region \( r > a \) and assumes that the currents producing the main field are inside the earth’s surface. This assumption has been verified to within 0.1%, which is as close to zero as may be determined from currently available magnetic data. The index \( n = 0 \) is missing because it would correspond to an unrealistic magnetic monopole; also \( P_n^m(\cos \theta) = 0 \) for \( m > n \). We see that the index \( n \) determines the order of the magnetic multipole; e.g., \( n = 1 \) gives a dipole potential, \( n = 2 \) gives a quadrupole potential, and so on.

When we consider a single magnetic dipole at the earth’s center, oriented along the geographic axis, these equations become very simple. In this case \( n = 1, m = 0 \), and \( P_1^0(\cos \theta) = \cos \theta \). Thus

\[
V = a g_1^0 \left( \frac{a}{r} \right)^2 \cos \theta
\]

(701)

We may obtain the magnetic field components by differentiation. They
are \( Z = (\partial V/\partial r) \), radially downward; \( X = (1/r) (\partial V/\partial \theta) \), northward; \( Y = -[(1/r) \sin \theta] (\partial V/\partial \lambda) \), eastward. For the potential of Eq. (701):

\[
\begin{align*}
X &= - g_1^0 (a/r)^3 \sin \theta \\
Y &= 0 \\
Z &= - 2g_1^0 (a/r)^3 \cos \theta
\end{align*}
\]

(702)

The total field is given by

\[
(X^2 + Y^2 + Z^2)^{1/2} = g_1^0 (a/r)^3 (1 + 3 \cos^2 \theta)^{1/2}
\]

(703)

The dip angle \( I \) is reckoned positive when the field points downwards; it is given by

\[
\tan I = \frac{Z}{(X^2 + Y^2)^{1/2}} = 2 \cot \theta
\]

(704)

for the simple dipole field. The value of \( g_1^0 \) may be identified with the southward component of magnetic field at the equator and at \( r = a \); for the earth, \( g_1^0 \approx -0.30 \) G. Variations of the field are generally expressed in units of \( 1 \gamma = 10^{-5} \) G = \( 10^{-9} \) T (tesla or Wb. m\(^{-2}\)).

It is impossible to fit the earth's field very well with a single dipole term as we have considered above. As the next best approximation, we can let \( n = 1 \) and \( m = 1 \), which generate two additional terms in the potential expression. We find

\[
P_{11}^1 (\cos \theta) = \sin \theta, \quad V = a (a/r)^2 \{ g_1^0 \cos \theta + (g_1^1 \cos \lambda + h_1^1 \sin \lambda) \sin \theta \}
\]

(705)

The first term in the bracket corresponds to Eq. (701), and the second term provides new contributions to each of \( X \), \( Y \), and \( Z \). Physically, these could be generated by two additional orthogonal dipoles placed at the center of the earth but with their axes in the plane of the equator. Their effect is to incline the total dipole term to the geographic pole by an amount

\[
\alpha = \arctan \left[ \frac{(g_1^1)^2 + (h_1^1)^2}{(g_1^0)^2} \right]^{1/2}
\]

(706)

Thus, the potential function (705) could be produced by a single dipole inclined at an angle \( \alpha \) to the geographic pole and with an equatorial field strength \( [(g_1^0)^2 + (g_1^1)^2 + (h_1^1)^2]^{1/2} \). For the earth, the single dipole equivalent is tipped by about 11.5° toward 70°W longitude and has an equatorial strength of about 0.312 G.

For \( n > 1 \), additional multipoles contribute to \( V \). Finch and Leaton (1957) have computed the first 48 terms in Eq. (700) (up to \( m = n = 6 \)) for epoch 1955. The dipole term predominates, and the higher order multipoles are not only smaller in magnitude but also diminish more rapidly with altitude.
7.3 VARIATIONS OF THE GEOMAGNETIC FIELD

However, even this model cannot give any representation of the local variations of the geomagnetic field as only features whose scale exceeds about one-sixth of the earth's radius can be accurately described by formulas involving only \( m \leq n \leq 6 \). Later, Cain et al. (1965) included magnetic data obtained from the Vanguard 3 satellite in a more complex evaluation of the main field, and we may expect this use of satellites to increase. Hendricks and Cain (1966) have tabulated coefficients up to \( m = n = 9 \).

For the purpose of ionospheric experiments, such as satellite Faraday rotation observations, better results can be obtained by using "local models," with individual values for each station (Golton, 1963), than by using models derived from worldwide data.

A significant improvement can be obtained, while retaining a simple mathematical form, by using "eccentric dipole coordinates" (Cole, 1963). In this model, the dipole is moved away from the earth center (about 400 km toward the western Pacific) to obtain a better worldwide fit to the magnetic field. The actual magnetic field configuration is very complex and there are various anomalies, notably in the South Atlantic region.

It is customary to define two kinds of latitude connected with geomagnetism (Chapman, 1963a):

Dipole Latitude (formerly "geomagnetic latitude"), referred to the "centered dipole" approximation of Eq. (705);

Dip Latitude (formerly "magnetic latitude"), \( \arctan \left( \frac{1}{2} \tan I \right) \), defined from the actual magnetic dip angle \( I \) in accordance with Eq. (704).

Another type of coordinate system has been found very useful in the discussion of magnetospheric phenomena, especially the motions of trapped energetic particles (Sec. 7.52). This is the \( (B, L) \) system of McIlwain (1961), in which \( B \) is the field intensity. For the special case of a dipole field, the coordinate \( L \) is constant on any "magnetic shell," which is a surface of revolution of a field line about the dipole axis. The value of \( L \) is the geocentric distance of the shell in the equatorial plane, measured in units of the earth's radius \( R_E \). The shell intersects the earth's surface at the dipole latitude given by (arc sec \( \sqrt{L} \)), by the geometry of a dipole field. For the actual geomagnetic field, \( L \) is defined in terms of an integral, and the angle (arc sec \( \sqrt{L} \)) is sometimes termed "invariant latitude" \( \Lambda \). Hence, \( L \cos^2 \Lambda = 1 \).

7.3 Variations of the Geomagnetic Field

7.31 The Quiet-Day Variations

We have already noted that the magnetic elements \((X, Y, Z)\) undergo a
daily oscillation about their mean values. Figure 58 shows the solar daily magnetic variation, or \( S_q \), for quiet days in the 1922–1933 equinoctial periods at various latitudes (Vestine et al., 1947). The more obvious features include the phase reversal in \( \Delta Y \) and \( \Delta Z \) (but not \( \Delta X \)) across the equator; and the large increase in \( \Delta X \), associated with the equatorial electrojet. From data such as these, an appropriate scalar magnetic potential can be evaluated by the methods explained in Chapter XX of Chapman and Bartels (1940).

Fig. 58. The \( S_q \) variation in the magnetic components \( X, Y, Z \) at various latitudes from equinox data, 1922–1933. The scale at the right indicates a range of 50 \( \gamma \) [after Vestine et al. (1947)]. (The deviations \( \Delta X', \Delta Y' \) are measured with respect to geomagnetic north and east, respectively.)
Another convenient way to represent the source of the regular magnetic variations is by a current sheet, flowing on the surface of a sphere above the earth. This surface is usually assumed to be at an altitude of about 100 km where the maximum conductivity is found. The height-integrated current density can be expressed directly in terms of the magnetic field components, using Ampère's circuital relations. If horizontal variations can be neglected, the southward ($x$) and eastward ($y$) components of magnetic flux density $B$ and integrated layer current density $J$ are related by the equations (where $f$ is a number):

$$-f \mu_0 J_y = B_x, \quad f \mu_0 J_x = B_y \quad (707)$$

Note that according to the sign convention for geomagnetism $\Delta Y = B_y$ but $-\Delta X = B_x$. On simple theory the numerical factor $f = \frac{1}{2}$, but in reality $f$ depends on the contribution of currents induced in the ground by the overhead current systems. Detailed analysis shows that ground currents cause about one-third of the observed magnetic effect (corresponding to $f = \frac{1}{3}$), though the proportion is not the same for the different magnetic components (Chapman and Bartels, 1940). If the ground were a perfect conductor, and complications due to horizontal variations and the earth's curvature were neglected, the ground and ionospheric currents would be equal and opposite, and would contribute equally to the observed magnetic variations. We should then have $f = 1$. We must always remember that the convenient current sheet representation, in which all currents are horizontal, may not correspond to the actual current flow.

The current system required to produce the external part of the $Sq$ variation (i.e., neglecting earth currents), for equinox and sunspot minimum, is shown in Fig. 59. Current flows parallel to the streamlines with 10,000 A between each line. The main features of Fig. 58 can be related to the current system of Fig. 59 by inspection. We remember that an eastward sheet current above the ground produces a northward magnetic field at the surface, and so forth. Therefore, between the latitudes of about $\pm 40^\circ$, the eastward daytime current produces a positive $\Delta X$ component, as can be seen at Alibag and Honolulu. At higher latitudes, however, the phase reverses to a negative $\Delta X$ (southward) by day. The transition occurs near the current focus at the latitudes of Tucson and Watheroo. The current sheet does not include the equatorial electrojet required to explain the large variation at Huancayo, but a current belt along the magnetic dip equator can be added to produce this feature. As we shall see in Sec. 7.43, the equatorial electrojet originates from the high conductivity existing along the dip equator.
The $\Delta Y$ variation implied by Fig. 59 should be eastward before about 1100 hours local time, westward in the afternoon for northern latitudes, and reversed in the Southern hemisphere; this is also observed in Fig. 58. The $\Delta Z$ variations are not understood so easily, since they must be produced by currents removed from the point in question because overhead currents produce horizontal variations only. The vertical variations are most pronounced near the current focus because the counter-clockwise current flow produces an upward magnetic field near the focus in the Northern hemisphere, while a positive $\Delta Z$ is produced in the Southern hemisphere near Watheroo. At night, the current densities are much smaller owing to the reduced conductivity in the E region.

Even in quiet periods the $Sq$ pattern shows a day-to-day variability of magnitude and position (Mayaud, 1965). In high latitudes the $Sq$ variations (Fig. 58) do not correspond to the simple current system of Fig. 59; additional currents exist even under quiet conditions, and are sometimes denoted by $Sq^p$ (Sec. 8.2).

A magnetic variation can be found with the fundamental period of a lunar day, about 24.8 hr. The magnitude of the lunar ($L$) variation is much smaller than the solar ($Sq$) variation, being of the order of $\pm 4\gamma$. Also, the predominant mode is a *semidiurnal* variation as is found in ocean tides, rather than the predominant diurnal period in $Sq$. 

---

*Fig. 59.* The overhead current system corresponding to the external part of $Sq$ for sunspot minimum, equinoctial conditions. Numbers denote $10^3$ A, so that the current flowing between any two neighboring streamlines is $10^4$ A [After Chapman and Bartels (1940) and Vestine (1960)].
7.32 The Disturbance Variations

The earth's magnetic field is frequently disturbed, and undergoes variations other than the quiet $S_q$ and $L$ variations. We have already mentioned effects accompanying solar flares, in Sec. 6.1. If the field is severely disturbed, a magnetic storm is said to occur. Often storms begin abruptly with a *storm sudden commencement* (SSC or just SC) and a fairly well defined pattern of behavior ensues. Many storms start gradually, without an SC; and it sometimes happens that an impulsive change like an SC is observed in the magnetic field, but that no storm ensues. This is called a *sudden impulse* (SI).

To describe the changes which occur during SC storms, different "disturbance variation" functions have been defined. They are evaluated by means of statistical analysis of data from many observatories and many storms. They seek to distinguish variations which depend on local time from those that depend on storm time, which is the time elapsed from the commencement of the storm. We emphasize that these functions are merely convenient numerical descriptions of the observations, and do not necessarily correspond to physical processes. We cannot discuss the actual current systems without some knowledge of the magnetosphere, so we shall return to this topic later, in Sec. 8.2.

The disturbed value of a magnetic element (we will use $X$ as an example, but any of the other elements may be treated similarly) is defined by $\Delta X = X_{\text{obs}} - X_q$, where $X_{\text{obs}}$ is the observed value, and $X_q$ is the quiet-day value including the regular $S_q$ and $L$ variations. The values of $\Delta X$ depend not only on "storm time" $\tau$ (the number of hours elapsed after the commencement of the storm) but also on the colatitude $\theta$ and longitude $\lambda$ of the observing station. The disturbed value is usually split into two parts: (a) the $Dst$ or *storm time* variation, which is the average value of $\Delta X$ around a circle of constant latitude and (b) the $SD$ solar daily disturbance or the $DS$ disturbance local-time inequality. All three of these quantities vary with storm time. We think of $DS$ as a "snapshot" showing the variation of $\Delta X$ with longitude at a particular $\theta$ and $\tau$.

The deviation $\Delta X(\tau, \theta, \lambda)$ is a component of the disturbance field $D$, which is divided into parts:

$$D = Dst + DS; \quad \bar{D} = Dst + SD$$

(708)

in which $\bar{D}$ is an average of $D$ over 24 hr. The $SD$ description was first developed when it was thought that the longitude variation of $D$ maintained a fixed orientation in solar time. It is now known that the phase and amplitude
of the longitude (or local time) variation changes during the course of a storm, and the quantity $DS$ has, therefore, been defined (Akasofu and Chapman, 1964). These authors consider that "$SD$ is merely the mean of $DS$ over a day," but it would appear that $SD$ could be the average of $DS$ over any convenient interval. These relations may be expressed as

$$D = Dst(\tau) + \sum_n c_n \sin(n\lambda + \epsilon_n)$$

(709)

where

$$DS = \sum_n c_n \sin(n\lambda + \epsilon_n)$$

(710)

The evaluation of these coefficients at any latitude requires either the observation of many storms of comparable intensity at a single station or the observations of several stations at the same latitude for a single storm. Ideally, both of these approaches lead to the same result; but differences occur in practice owing to local variations, the different characteristics of individual storms, and so on.

The form of the $SD$ and $Dst$ variations has been extensively studied by Sugiura and Chapman (1960), who analyzed data from 26 observatories for 346 storms which occurred from 1902 to 1945.

In a typical magnetic storm, the SC is followed by an "initial" or "positive" phase lasting a few hours. During this time the geomagnetic field intensity is increased, the "stormtime" component $Dst(H)$ being positive. This is probably due to the compression of the geomagnetic field by the solar plasma. After a few hours the main phase sets in, in which $Dst(H)$ becomes negative and the field is depressed for a day or so. This reduction can be represented as the field of a "westward ring current" which is oppositely directed to the main geomagnetic field. $Dst(H)$ reaches its greatest negative values about 18–24 hr after the SC (Fig. 60). There are often irregular fluctuations of field during the initial and main phases, but within a day or two the storm passes to the recovery phase and the field returns smoothly to normal, with an exponential time constant of about one day.

The amplitude of the $Dst$ magnetic field is greatest in low latitudes, and decreases polewards. This is opposite to the field of the current systems associated with the $DS$ disturbance local-time inequality. This field is most intense at auroral latitudes, where it exceeds $Dst$, but decreases towards low latitudes, where $Dst$ is the dominant storm variation. In general, the variation observed at any station is a complicated combination of $DS$ and $Dst$, which depends on the local time at which the SC occurs.

In the early stages of a storm, the $DS$ system changes rapidly, but after
some hours it settles into a pattern which remains nearly fixed with respect to the sun, and gradually decays as the storm progresses. Once these patterns are established, the harmonic components of the local-time variation \( SD \) at any station maintain almost constant phases though the amplitudes decay gradually. The \( SD \) system appears to possess a mid-latitude focus, as does the \( Sq \) system, though \( SD \) and \( Sq \) differ in phase. Even for moderately strong storms the \( SD \) variation exceeds \( Sq \) at mid-latitudes.

### 7.33 Indices of Activity

The earth's magnetic field is seldom quiet, even when there are no storms. Several empirical indices have been developed to describe the amount of variability at any given time. We briefly describe some of these, and it is useful to mention also the conventional parameters used to describe solar activity.

Values of the most important indices are published in the CRPL-FB (now IER-FB) series publications of the U.S. Environmental Science Services Administration (formerly Central Radio Propagation Laboratory), Boulder, Colorado, and in the Journal of Geophysical Research. A description of the magnetic indices has been given by J. Bartels (1957). The international and “planetary” indices are evaluated at the Meteorologisch Instituut, De Bilt, Netherlands.

At any one station, for each three-hour period of every Greenwich day, the range of variation of each of the Cartesian field components \( X, Y, Z \) is
measured; the greatest of the three ranges is called the amplitude $a$ (in gammas). The solar quiet-day ($Sq$) and lunar ($L$) variations are removed before the range is obtained from the records. The amplitudes $a$ for twelve representative observatories are combined to give a "planetary" value for any three-hour period, the average for a Greenwich day being $Ap$.

The index $K$ runs from 0 to 9, and is related to the amplitude by a quasi-logarithmic scale which is chosen individually for each station, according to its activity. Typically the qualification for $K = 9$ is a disturbance of 300 $\gamma$ for a low-latitude (but not equatorial) station, 500 $\gamma$ for a station in mid-latitudes, and 2000 $\gamma$ for a station in the auroral zones.

The $K$-indices for the twelve observatories are combined to yield a "planetary" figure $Kp$ (also eight per day), determined to an accuracy of one-third of a unit, running as: 0, 0+, 1-, 1, 1+, 2-, ..., 8, 8+, 9-, 9. The maximum disturbance 90 has only been recorded on a few occasions. $Kp$ values are available for the Polar Year 1932/33 and for 1940 onwards. The diagrams originated by J. Bartels at Göttingen (reproduced in C.R.P.L. Bulletins) show its variation over several months at a time. The twelve "$Kp$ observatories" are nearly all in the Northern hemisphere, and the "planetary" indices may not reliably indicate conditions at other stations.

The oldest index is the magnetic character figure $C$ assigned to describe the "storminess" of a given day at one station on an arbitrary scale of 0: 1: 2. Its average for many stations, $Ci$, is computed to one place of decimals, and thus can range from 0.0 to 2.0. Though the assignment of $Ci$ is somewhat subjective, the index is useful and its values date back to 1905.

A similar number $Cp$, ranging from 0.0 to 2.0, is designed to resemble the classical character figure $Ci$. It is derived from $Kp$, on a scale which is adjusted to make $Cp$ fit $Ci$ as closely as possible.

A special "quarter-hourly" index $Q$ was introduced during the International Geophysical Year (1957/58) for studies of high-latitude phenomena. It is computed for each fifteen minute period at stations whose geomagnetic latitude exceeds 58°. It is based on variations of the horizontal field only, and is thus sensitive to variations in local ionospheric currents.

Values of $Kp$ are used as a basis for the selection of the International Quiet and Disturbed Days in each month: the "five quietest" and "ten quietest" (which include the five) IQD's and the five IDD's.

For convenience, we also mention here some commonly used solar indices. Solar activity is specified in terms of the 10.7 cm (2800 MHz) radio noise flux, $S$. Measurements of $S$ at Ottawa by A. E. Covington extend back to 1948. Its values range from 70 units at sunspot minimum to about 250 units
at an intense sunspot maximum, the unit being $10^{-22}$ W m$^{-2}$ Hz$^{-1}$. An older index of activity is the Wolf or Zurich sunspot number $R_z$, obtained from daily solar observations, and derived by counting sunspots and spot groups. Values extend in an unbroken sequence back to 1749. There is also an American sunspot number $R_A$. A correlation between $R_z$ and $f_0E$ has been established statistically over two solar cycles. This can be used to deduce, from measured critical frequencies, an “equivalent ionospheric sunspot number” $R_E$ as a measure of solar ionizing radiation. Minnis and Bazzard (1960) also define a monthly index $I_{F2}$, based on the critical frequency $f_0F2$ and enumerated on a scale resembling that of $R_z$. The statistical relationship between $S$, $R_z$ and $I_{F2}$ has been examined by Joachim (1966).

Because of the relation between solar and geomagnetic phenomena, it is sometimes convenient to measure time in terms of “solar rotations,” with a synodic period (i.e., relative to the sun-earth line) of about twenty-seven days. Solar rotations are numbered in R. C. Carrington’s series, of which No. 1 began on 9 November 1853.

7.4 Dynamo Theory

The dynamo theory was first suggested by Balfour Stewart in 1882 to account for the daily variations of the geomagnetic field, and developed quantitatively by A. Schuster in 1908. Although there are serious difficulties with some aspects of it, the theory seems basically correct and we shall review it here. A good account has been given by Hines (1963) and more recently by Maeda and Kato (1966). To recapitulate and extend the outline of tidal theory given in Sec. 1.7:

(1) The sun and moon produce tidal forces in the atmosphere, the periods being fractions of the solar day (24 hr) and lunar day (24.8 hr).

(2) These forces set up standing waves in the atmosphere which result in (primarily horizontal) air motions. The atmosphere responds differently to forces of different periods.

(3) The motion of the air across the geomagnetic field induces electro-motive forces, which drive currents at levels where the electrical conductivity is appreciable (principally in the E region), thus causing the solar quiet-day and lunar magnetic variations.

(4) Because of the vertical and horizontal variations of conductivity, currents cannot flow freely in all directions; polarization charges are thereby set up, modifying the flow of current.
(5) The electrostatic fields associated with these charges are transmitted to the F region, via the highly-conducting geomagnetic field lines, where they cause electromagnetic drifts.

Stages (1) and (2) come within the field of atmospheric dynamics and may be thought of as the "mechanical" side of the dynamo theory; we have dealt with it in Sec. 1.7. The "electrodynamic" side comprises stages (3) to (5).

7.41 THE BASIC THEORY

Our discussion of charged-particle motions and electrical conductivity in Sec. 4.2 enables us to calculate the currents induced by neutral air winds. For both ions and electrons, the steady drift velocity $V$ due to a wind $U$ can be found from Eq. (410) by equating the collisional force $mv(U - V)$ to the Lorentz force $\mathbf{e}V \times \mathbf{B}$. As discussed in Sec. 4.21, the resulting drifts are different for ions and electrons, so that a current flows. Another way of describing this process, which leads to the same result, is that the wind $U$ blows across the geomagnetic field, inducing a field $U \times \mathbf{B}$. This induced field drives a current $\sigma \cdot U \times \mathbf{B}$, where $\sigma$ is the tensor conductivity of Eq. (419).

The current produced by the wind alone need not satisfy the equation $\text{div } \mathbf{j} = 0$. At any point where $\text{div } \mathbf{j} \neq 0$, electric charge accumulates so that the ionosphere becomes polarized. An electrostatic polarization field $-V\Phi$ is set up ($\Phi$ denoting electric potential) and adjusts itself until the current flow is both nondivergent and horizontal. The appropriate conductivity is then the layer conductivity $\sigma'$ given by Eq. (423), and the induced and polarization electric fields add to give a total field $E_t$. The current is then

$$\mathbf{j} = \sigma' \cdot \mathbf{E}_t = \sigma' \cdot (U \times \mathbf{B} - V\Phi)$$

(711)

In most existing analyses of the problem, $U$ and $\mathbf{B}$ are assumed independent of height throughout the conducting region, and the equation for the height-integrated current is written in terms of the integrated conductivities mentioned in Sec. 4.23; it is

$$\mathbf{J} = \Sigma' \cdot (U \times \mathbf{B} - V\Phi)$$

(712)

Another condition results from the large conductivity $\sigma_0^*$ [given by Eq. (422)] existing along field lines through the magnetosphere, namely that the electric potential must be virtually equal at all points on a given field line above the dynamo region. Any potential difference which might arise between magnetically conjugate points at the top of the dynamo layer tends to be removed by a flow of current along the field line linking the points. We shall see in Sec. 7.44 that this flow is so small that it does not signifi-
cantly violate the statement that the ionospheric current flow \( J \) must be horizontal (except possibly at auroral latitudes).

The large conductivity parallel to \( B \) also causes the electrical coupling of the E and F regions, as suggested by Martyn (1947) and investigated by Farley (1959, 1960) and others. It is believed that E-region electrostatic fields are accurately mapped into the F region, even for scale sizes down to a few kilometers. The electric field acts on the ionization in the F region, causing it to drift with the velocity previously derived in Eq. (416), namely

\[
V = E \times B / B^2
\]  

(713) in which \( E = -\nabla \Phi \) represents the electrostatic polarization field. At F-region heights, as discussed in Sec. 4.3, a wind \( U \) does not move the plasma across the magnetic field, so there is no induced component of electric field.

This electrodynamic drift action in the F region has been likened to an electric motor running without load, as the back e.m.f. \( V \times B \) set up by the drift of ionization across the field lines just balances the applied electrostatic field. Actually, a small current flows because the electron and ion drift velocities are not precisely equal, implying that some energy is expended in collisions and in overcoming gravity; this is because the ion-neutral collision frequency is not entirely negligible in the F region. Apart from this second-order effect, the vertical component of the F-region drift velocity is produced by the eastward electric field component \( E_y \) and is given by \( W = (E_y/B) \cos I \).

Dynamo theory, as condensed into Eqs. (711) to (713), is very idealized. In practice, there are numerous difficulties, both in the assumptions made in the theory and in obtaining numerical values of the parameters. We shall review the procedure and point out some difficulties. Perhaps the most basic difficulty is that the "thin sheet" approach to dynamo theory fails if the wind velocity varies considerably with height through the conducting regions. There are strong indications that it does, and that the height variations may be too irregular to enable any meaningful "average" velocity to be defined. In these circumstances, the use of height-integrated currents and conductivities is questionable, but the extension of the theory to take account of height variations within the dynamo layer has proved excessively difficult.

7.42 DYNAMO CALCULATIONS

The dynamo calculations can be approached from two standpoints. One is to solve the equations on a local scale, inserting observed values of the local magnetic variations and comparing them with other data. The most that can be said in this regard is that limited agreement is sometimes ob-
maintained between the E-region data, meteor winds, and ground-level pressure oscillations; but that it is difficult to relate F-layer drifts to the other data. To some extent, this may be due to the difficulty of separating the electric field into its "induced" and "electrostatic" components. The other approach is to assume that a steady state exists and to solve the equations on a global scale to obtain the layer current density $J$, electrostatic potential $\Phi$ and wind velocity $U$. Not only must these quantities satisfy the dynamo equations at all points of the conducting layer but also the current continuity equation, $\text{div} J = 0$, must be satisfied everywhere. Fortunately, it then becomes possible to separate the "induced" and "electrostatic" components by procedures which make use of the identity curl grad $\Phi = 0$. In this approach some dynamical constraint on the wind velocity must be assumed. It is generally assumed that the wind velocity $U$ can be derived either from a scalar potential (the terms "scaloidal" and "irrotational" being used to describe this case) or from a vector potential ("toroidal" or "solenoidal"). The latter approach, however, seems to lead to difficulties at the poles of rotation. In addition, the Coriolis forces arising from the earth's rotation introduce considerable complications, so they are often neglected.

All calculations are based on models for the tensor conductivity $\Sigma$. Since the E-region electron concentration changes by a factor of 30 (or more) between day and night, $\Sigma$ undergoes a variation which is predominantly diurnal, but which contains higher harmonics. Usually the zero level of the magnetic variations (i.e., the value of field in the absence of any ionospheric current) is uncertain to within several gammas. This causes a corresponding uncertainty $\Delta J$ (say) in ionospheric current density. Even though $\Delta J$ is not a function of local time, the resulting uncertainty in electric field, $\Delta E = \Sigma^{-1} \cdot \Delta J$, varies diurnally. This must affect the conclusions regarding the relative importance of the diurnal and other components.

The situation is different for lunar tides because, unlike the sun, the moon has very little effect on the ionospheric electron concentration, so that the conductivity does not undergo a lunar variation. Provided the data are averaged in such a way as to remove the solar effects, the predominant lunar variation should be the semi-diurnal gravitational tide.

The calculations of Baker (1953) and Fejer (1953) start with the pressure variations and compute the resulting electromagnetic effects. They neglect the diurnal variation of conductivity and therefore obtain only semi-diurnal magnetic effects and drifts from the pressure variations, assumed semidiurnal. Several of the Japanese workers start with the magnetic $Sq$ data and proceed in the reverse direction (e.g., Maeda, 1955; Hirono and Kitamura, 1956;
Kato, 1956). The computed E-region winds turn out to be predominantly diurnal, whereas the observational data (meteor trails, E-region drifts, barometric oscillations) suggest that semidiurnal variations dominate at mid-latitudes (Sec. 7.45).

Considering the uncertainties that arise in discussing the dynamo region itself, it is not too surprising that the extension of the theory to F-region drifts gives few firm conclusions. This is partly because F-layer drift experiments are not easy to interpret. A number of F2-layer phenomena have been attributed to electromagnetic drifts; we have given reasons for supposing that drifts might be important though they are limited by ion-drag (Secs. 4.24, 4.44). Not much success has yet been achieved in relating solutions of the continuity equation, with electromagnetic drift included, to observed F2-layer behavior, except for the equatorial F2 layer (Sec. 5.45). Some E-layer phenomena are probably due to the (relatively small) effects of electromagnetic drift (Sec. 5.2).

7.43 The Equatorial Electrojet

We have previously referred to the equatorial electrojet, the eastward current which flows along the dip equator by day and produces the large variation of magnetic field at stations like Huancayo (Fig. 58). The basic reason for the existence of the electrojet is the large value of Cowling conductivity at the dip equator (Baker and Martyn, 1953; see Eq. (425)). To a first order, the experimental data seem qualitatively consistent with this theory. According to this theory, the electrojet current may be regarded as part of the $Sq$ current system; there should, therefore, exist a close relationship between electrojet intensity and the worldwide magnetic $Sq$ variation, but this is not found in the data (Osborne, 1966). Hence, this "conductivity enhancement" theory of the electrojet may be too simple, and there may be other mechanisms that contribute to the electrojet phenomena.

Rocket-borne magnetometers have detected the electrojet current sheet at altitudes between about 90 and 130 km (Singer et al., 1952; Maynard and Cahill, 1965a,b; Davis et al., 1967). Davis et al. also detected a small westward electrojet current at night. This reversed nighttime electrojet has also been detected by measurements of horizontal motions in the E region, obtained with the Jicamarca radar system (Balsley, 1966). We should, indeed, expect the electrojet to be present at night since the geometrical factors causing enhanced conductivity are present no less than by day, even though the electron concentration is small and the current correspondingly weak.
Last, we should mention that the large current densities in the daytime electrojet are thought to give rise to plasma instabilities, hence causing equatorial sporadic E (Farley, 1963b; see Sec. 6.31).

### 7.44 Electrical Coupling via the Magnetosphere

It has been shown that magnetic field lines above the ionosphere may be regarded as very good electrical conductors (Farley, 1960). This implies that magnetically conjugate points in the ionosphere are electrically coupled and must be at virtually the same potential (Hines, 1963).

If there were no coupling by field lines, the potential at magnetically conjugate points could be quite different, on account of the differences between the geographic and magnetic axes of the earth. At the solstices, and in certain regions of the earth, the differences would be most pronounced. When coupling exists, however, the potential is almost equalized by the flow of current along field lines. Thus the equation \( \text{div} \mathbf{J} = 0 \) no longer holds for the horizontal currents in the ionosphere, and field-aligned currents must be included in the dynamo equations (Maeda and Murata, 1965; Van Sabben, 1966; Block, 1967).

According to Van Sabben, the total current flowing through the magnetosphere may be an appreciable part of the total current flowing in the Sq system. The current density in the magnetosphere is quite small, however; Maeda and Murata quote a typical value of \( 10^{-12} \, \text{A m}^{-2} \). Owing to the convergence of magnetic field lines, the field-aligned current density just above the ionosphere is greater than it is far out in the magnetosphere, but it is, nevertheless, very much smaller than the horizontal current density in the ionosphere (of order \( 10^{-6} \, \text{A m}^{-2} \)). Hence, although the total current vector is not quite horizontal (Dougherty, 1963), its inclination is actually extremely small. In auroral latitudes, considerable currents flow into the ionosphere along field lines, according to models proposed by Fejer (1964) and Boström (1964), and the field-aligned current densities may be greater than at mid-latitudes.

To drive the magnetospheric currents, small electric fields must exist parallel to the magnetic field lines. Reid (1965) estimates the fields associated with the mid-latitude current systems to be only of order \( 10^{-10} \) to \( 10^{-8} \, \text{V m}^{-1} \) at heights well above the dynamo region (say above 250 km altitude). A typical mid-latitude field line is \( \sim 10^4 \) km long, so the potential drop along it is only \( 10^{-3} \) to \( 10^{-1} \, \text{V} \). Comparing this with the total voltage developed in the dynamo system, say 10 kV, we can justify the assumption that magnetospheric field lines are electric equipotentials.
7.4 OBSERVATIONAL DATA

It is useful to keep in mind the orders of magnitude of the various parameters of importance to the dynamo theory. Some of these are listed in Table V.

Barometric oscillations have been analyzed in detail, but the comparison between ground-level and dynamo region conditions involves calculation of the tidal amplification factor and phase variations, using an atmospheric model such as the example shown in Fig. 8 (Sec. 1.72). As we have discussed earlier, this has not been a particularly fruitful approach to the determination of ionospheric winds.

Drifts of meteor trails can be observed by radar techniques, and give winds

| TABLE V |
| SOME ORDERS OF MAGNITUDE OF QUANTITIES IN DYNAMO THEORY |

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure oscillations</td>
<td>1 mb at ground, or 1 part in 1000. At 100 km, probably about 1 part in 10.</td>
</tr>
<tr>
<td>Wind velocities at 100 km</td>
<td>50 m s⁻¹</td>
</tr>
<tr>
<td></td>
<td>5 × 10⁻⁵ T (Wb m⁻²)</td>
</tr>
<tr>
<td>Total geomagnetic flux density, (B)</td>
<td>0.5 G</td>
</tr>
<tr>
<td></td>
<td>5 × 10⁴ γ</td>
</tr>
<tr>
<td>Daily (S_q) variation at ground</td>
<td>20 γ</td>
</tr>
<tr>
<td>Currents (total in each vortex of (S_q) system)</td>
<td>6 × 10⁴ A</td>
</tr>
<tr>
<td>Current density, (j)</td>
<td>10⁻⁶ A m⁻²</td>
</tr>
<tr>
<td>Layer current density, (J)</td>
<td>10 A km⁻¹</td>
</tr>
<tr>
<td>Maximum potential difference within system</td>
<td>10⁴ V</td>
</tr>
<tr>
<td>Induced electric field, (E)</td>
<td>10⁻³ V m⁻¹</td>
</tr>
<tr>
<td>Vertical drift velocity of F region ionization, (W)</td>
<td>10 m s⁻¹</td>
</tr>
<tr>
<td>Daytime integrated conductivities</td>
<td>(\sigma_1 , dh \approx 10 \text{ mhos} )</td>
</tr>
<tr>
<td>(nearly all contributed by E region)</td>
<td>(\sigma_2 , dh \approx 20 \text{ mhos} )</td>
</tr>
<tr>
<td></td>
<td>(\sigma_3 , dh \approx 200 \text{ mhos} )</td>
</tr>
</tbody>
</table>

a For collision frequencies see Table III in Sec. 4.12

in the neutral air at 80–110 km height, but the data are limited to a few locations. A long series of measurements has been made at Jodrell Bank (53 °N) where the amplitude of the 12-hr oscillation is somewhat greater than that of the 24-hr wind (Greenhow and Neufeld, 1961); whereas at Adelaide
(35°S) the 24-hr component is more pronounced (Elford, 1959) (Sec. 1.71).

**E-layer drifts** have been measured at various stations (Sec. 6.41). Early results were reviewed by Briggs and Spencer (1954) and an analysis of IGY data given by Shimazaki (1959a); another survey has been given by Briggs (1962). Strong semidiurnal components are found, whose vectors generally rotate clockwise with advancing time in the Northern hemisphere, and counter-clockwise in the Southern hemisphere. Diurnal components comparable to the semidiurnal ones have been reported, but since the normal E layer is absent at night, their detection necessitates inclusion of data from nighttime Es layers, which may not be strictly comparable with the normal E-layer data.

**F-layer drifts** are also reviewed in the papers quoted above. When the velocities are subjected to harmonic analysis, the diurnal component is found to predominate. Often there are fairly abrupt changes near sunrise and sunset, which perhaps arise from the changes of E-region conductivity accompanying the formation and disappearance of the E layer.

**Magnetic variations** of the **D, H, Z elements** on quiet days follow a regular latitudinal pattern and can be represented by overhead current systems (see Figs. 58 and 59). The determination of the ionospheric currents is, however, influenced both by ground currents and by the problem of the unknown zero-level.

**Rocket measurements of ionospheric currents** are made by means of sensitive magnetometers, which record the change of magnetic field as the rocket penetrates a current sheet. This technique was originally used to locate the equatorial electrojet (Singer et al., 1952), but the weaker currents of the mid-latitude Sq system can now be measured (Burrows and Hall, 1965; Davis et al., 1965). The measured currents agree with estimates made from ground magnetograms though Burrows and Hall do not find close agreement between observed and theoretical current distributions.

**The electric field** is the principal unknown in all the calculations. If it could be computed or measured with confidence, the applicability of dynamo theory would be much better established. Unfortunately, it is extremely difficult to measure the electric field *in situ*. The field is of order 1 mV m⁻¹, whereas a variable field typically 100 times greater is developed around space vehicles by their motion across the geomagnetic field, and by photoemission from their surfaces, and other causes. However, the motions of barium clouds—if interpreted as electromagnetic drifts—do possess the expected order of magnitude, and may give a measure of ionospheric electric fields (Haerendel et al., 1967).
7.5 The Magnetosphere

7.51 The Extent of the Magnetosphere

Our picture of the magnetosphere has evolved enormously since the launching of the first space probes. In this book, we are interested in its possible relationships with the ionosphere. The following brief account is based on contemporary ideas; we have drawn on the review papers of the 1966 Belgrade Symposium (King and Newman, 1967), especially the papers of Ness, Dungey, Obayashi, and O’Brien. In Fig. 61 we present a rough sketch of the magnetosphere, viewed from the equatorial plane.

The main geomagnetic field extends, in roughly dipole fashion, to distances of a few earth radii. Beyond this it is influenced by the flow of charged particles known as the solar wind (see Parker, 1967), which causes the geo-

![Fig. 61. Rough sketch of the earth’s magnetosphere, viewed from the equatorial plane. Solar wind flow is supersonic where shown by full arrows, but it is subsonic and disturbed in the apex of the sheath, where shown by dashed arrows. Some typical geomagnetic field lines are shown, though the extent to which they merge with the neutral sheet is ill defined. Black spots indicate two positions where neutral points may exist on the magnetopause. The domain of auroral particles is stippled, and the domain of stably trapped particles is shaded.](image-url)
magnetic field to be confined within a boundary called the magnetopause. Outside the magnetopause, on the sunward or day side, exists the shock front or earth's "bow wave" where the supersonic flow of the charged particles is interrupted and slowed down. Inside the apex of the shock front the flow is disturbed and subsonic although supersonic flow is resumed further downstream.

On the day side, the magnetopause lies at about $10R_E$ (earth radii). On the night side, the magnetic field is drawn out into a long tail (Piddington, 1960) which extends certainly to the moon's orbit at $60R_E$ and possibly very much further. The pattern of field lines within the tail is still largely unknown. Some field lines may join with a weak general interplanetary field (Dungey, 1961), in which case the tail is not completely closed. A region of somewhat enhanced plasma density exists in the tail, and contains a thin "neutral sheet" in which the magnetic field is very weak, and across which the field reverses in direction. This sheet is possibly linked by field lines to the auroral regions. Neutral points may also exist on the magnetopause, as shown in Fig. 61.

The magnetosphere is permeated by hydromagnetic waves and electromagnetic VLF radiations, which have provided a great deal of information about its physical condition.

In Secs. 7.52 and 7.53, we discuss the control exerted by the geomagnetic field on the motions of energetic particles and thermal plasma. Lastly, in Sec. 7.54, we shall consider the ways in which the structures of the magnetosphere and ionosphere are linked.

7.52 Trapped Particle Radiation

Within the magnetosphere are situated the radiation belts. On the basis of early satellite measurements, Van Allen and Frank (1959) delineated two radiation belts, the inner at $L \approx 1.5$ and the outer at $L \sim 3$ to 4 (where we use the $L$ coordinate of McIlwain (1961); see Sec. 7.2). The properties of the belts depend so much on the type and energy of the particles considered, that any experimental result is influenced by characteristics of the particle detectors and of the satellite orbit used. This makes it very difficult to summarize the properties of the belts in any simple but meaningful way. It appears, however, that the inner belt contains relatively constant fluxes of protons with energies of tens and hundreds of MeV. Beyond $L = 2$ the flux of protons is smaller, and the outer belt largely consists of energetic electrons from tens of keV to a few MeV. The fluxes in the outer belt are very variable, and their variations appear to be connected with magnetic disturbances. The ring current which causes the main phase $Dst$ decrease of magnetic field
during storms is believed to be carried by low-energy protons (up to about 10 keV) and by some electrons, trapped at $L \sim 4$ (Frank, 1967). The trapping region extends to about $L \sim 6$ (Fig. 61); beyond it, the field configuration does not permit particles to remain in stably trapped orbits, and the auroral particles which populate this region do not belong to the radiation belts proper. The properties of the radiation belts have been reviewed by Hess et al. (1965), Schardt and Opp (1967) and O'Brien (1967).

The particles trapped within the radiation belts execute complex trajectories. These can roughly be described as a combination of three types of motion, which are sketched in Fig. 62. The numerical values we quote below

\begin{center}
\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{magneto.png}
\caption{Sketch indicating the motions of energetic particles trapped on a magnetospheric field line. Particles gyrate about the lines of magnetic induction $\mathbf{B}$ in the sense shown by the small loops. For particles of either kind, the "guiding center" (instantaneous center of gyration) "bounces" back and forth along the field line between the mirror points $M_1$, $M_2$ and "drifts" in longitude as indicated by the hollow arrows. The longitude drift is partly due to the outward decrease of $|\mathbf{B}|$ since at any point on the particle trajectory the radius of gyration is infinitesimally larger on the "outer" part of the gyration (further from the earth's center) than it is on the "inner" part; and partly to the curvature of the field lines.}
\end{figure}
\end{center}

(Hess et al., 1965) refer to 1 MeV particles at 2000 km altitude at the magnetic equator, though some of the parameters are not very energy-dependent:

1. **Gyration around a magnetic field line** at the cyclotron period or gyro-period ($2\pi m/eB$), $7\mu s$ for electrons and $4$ ms for protons, the radius of gyration being $0.3$ km and $10$ km respectively.
(2) **Bouncing along a field line**, between the mirror points (one in each hemisphere) at which the particle is reflected in the converging magnetic field. The bounce period is 0.1 s for electrons, 2 s for protons.

(3) **Drifting azimuthally around the earth**, with a period of revolution of 50 min for electrons, which drift eastward; 30 min for protons, which drift westward.

Each type of motion is characterized by an "adiabatic invariant" quantity, constant for any one particle unless the particle is subjected to a perturbing force which varies with a period comparable to the characteristic period of the motion. In the case of the gyration, the invariant is the magnetic moment of the particle; the other invariants are written in terms of integrals of the motion. A trapped particle, in the course of bouncing and drifting, remains on an "L-shell" (surface of constant \(L\)) in the region where the field is not grossly distorted from the dipole form, but where the distortion is severe (as for \(L > 5\)) this simple rule fails (Roederer, 1967).

Particles are removed from the radiation belts by collisions with neutral particles (Walt and MacDonald, 1964). Such collisions are extremely rare above 2000 km altitude, but it is possible for low-frequency electromagnetic waves to perturb the motions of particles at such high altitudes and lower their mirror points, thereby causing loss. As a given particle drifts in longitude, its mirror height may vary because of variations in the geomagnetic field, so that the chance of loss also varies. It has been suggested that in the South Atlantic geomagnetic anomaly where the field is abnormally weak, the loss of radiation belt particles is especially rapid. This loss might lead to observable phenomena, such as enhanced airglow and heating and ionization in the ionosphere, and may even influence the particle distribution in the belts to a detectable extent (Vernov et al., 1963; Gledhill and Van Rooyen, 1962; Gledhill et al., 1967). In one instance a stable mid-latitude arc of 6300 Å airglow emission appeared to be related to the outer radiation belt (O'Brien et al., 1960). However, the connections between the radiation belts and ionospheric phenomena have yet to be fully determined.

### 7.5.3 Hydromagnetic Motions in the Magnetosphere

We now wish to introduce some hydromagnetic concepts used in modern theories of magnetic storms, which provide an alternative way of looking at the motions of the plasma. The application of these concepts to the magnetosphere has been reviewed by Hines (1964a) and by Dungey (1967). Since the geomagnetic field lines above the E region are (almost) electric equipotentials, the electrostatic fields which cause F-layer drifts must also permeate the
magnetosphere. At these levels the fields \( B \) and \( E \) and the plasma drift velocity \( V \) satisfy the "freezing-in" equations

\[
\begin{align*}
0 &= E + V \times B, & V_\perp &= E \times B / B^2 \\
\end{align*}
\]

in which \( V_\perp \) is the component of motion normal to \( B \). The second of these has already appeared as the F-region motor equation (713). The "freezing-in" theorem makes no statement about motions parallel to \( B \), which are controlled by gravity, pressure gradients, and so forth, as discussed in Chapter IV, Sec. 4.1.

The "freezing-in" theorem is of very general application to the magnetosphere, stellar atmospheres, and ionized interstellar matter. It is valid whether or not the electric field \( E \) is purely electrostatic (i.e., derivable from a potential \( \Phi \)), provided its time variation is slow enough and the electrical conductivity is great enough. Basically, the theorem states that the ionized matter and magnetic field move in such a way that

(i) a set of charged particles, if connected by a field line, remain so connected at all times;

(ii) any given tube of plasma bounded by a set of field lines always encloses the same magnetic flux.

The proof of these statements is somewhat involved, and has been given by Dungey (1958) and Hines (1964a). Since a "line of force" is an abstraction and not a physical object, its "motion" is somewhat arbitrary. In elementary magnetostatics, "lines of force" are visualized as being fixed in space; yet the result (i) just quoted entitles us to identify a magnetic field line by the material situated on it, so that its motion is taken to be that of the material. We then think of the effect of a transverse electrostatic field as a translation of the magnetic field line itself, with velocity \( E \times B / B^2 \), in which all charged particles take part. In addition, each individual particle is subject to the forces which cause the various types of motion illustrated in Fig. 62 of Sec. 7.52. These forces become more important with increasing particle energy, and dominate the motions of magnetospheric particles above say 1 keV. Below about 1 keV the motion is dominated by the \( E \times B / B^2 \) drift. Thus, as a working rule we might regard 1 keV as a rough limit of energy at which the idea of "freezing-in" holds for particles in the magnetosphere.

The translational motion of magnetic field lines can take place in such a way that \( B \) does not change with time. If the electric field is purely electrostatic, this must always be the case, because then

\[
\frac{\partial B}{\partial t} = - \text{curl } E = \text{curl } \text{grad } \Phi = 0
\]
There is then a "circulation" or "convection" in three dimensions (Gold, 1959), which is possible because the presence of the nonconducting lower atmosphere enables the field lines in the upper atmosphere to move independently of those in the solid earth. This "convection" is illustrated by Fig. 63, which shows two different positions of a "tube" bounded by convecting field lines.

![Diagram](image)

**Fig. 63.** The relation between ionospheric currents and magnetospheric motions. An E-region current, as in (a), is mainly carried by a flow of electrons in the opposite direction, as in (b). Plasma at higher levels moves so as to remain on the same flux tube at all times; two positions that a typical flux tube occupies in the course of the circulation are shown. The diagram may be taken to represent the pattern of $Sq$ currents, as viewed from the sun [Hines (1965c)].

The motions of field lines are opposed by frictional collision forces acting on the ions in the E region, since $v_i > \omega_i$ below about 140 km and the ions are not "frozen in" to the magnetic field. But since $v_e \ll \omega_e$ in the dynamo region, the electrons are "frozen in," and do take part in the motion of the field lines. Thus, the ions and electrons in the F region drift with the same velocity as the electrons in the E region, apart from small differences due to the vertical variation of $\mathbf{B}$. All have the cycloidal types of motion shown in Fig. 31 for the case $v \ll \omega$ and drift at right angles to the electric field. In the E region, the relative motions of ions and electrons represent an electric current and the ion-neutral collisions represent joule loss. This is just another way of looking at the motions from which the conductivities were derived.

This description implies that any current flowing in the E layer is accompanied by motions of ionization in the F region and magnetosphere. Conversely, if any motion of ionization normal to $\mathbf{B}$ is set up in the magneto-
sphere by external causes, it will produce corresponding motions of the 
F-region ions and electrons and will set up electron currents in the E region, 
which will dissipate energy. The circulatory motion in the magnetosphere, 
however, is frictionless because of the absence of collisions, but has to take 
place in such a way that the field lines are merely interchanged and the field 
is not distorted in shape. The magnetospheric circulation does not directly 
produce any magnetic variations at the ground; these result indirectly from 
the E-region currents and induced earth currents, in which the energy is 
eventually dissipated.

We can use these ideas to discuss the lower boundary of the magnetos- 
sphere. The F region can be regarded as part of the magnetosphere, in that 
v < ω for all charged particles. We could even extend the magnetosphere 
downwards to include the E region, as far as electron motions are concerned. 
Another possible (though physically different) criterion for the base of the 
magnetosphere would be the equality of magnetic and neutral gas pressures. 
Typically in the ionosphere, the magnetic pressure \( B^2/2\mu_0 \approx 10^{-3} \text{ Nm}^{-2} \), 
which exceeds the gas pressure above about 130 km (Fig. 2).

In their original discussion of the subject, Axford and Hines (1961) envis- 
aged a hydromagnetic circulation of the whole magnetospheric plasma. This 
might be set up during magnetic storms by frictional interaction between the 
solar wind and the magnetosphere though other driving mechanisms may 
exist. A general circulation is present even in quiet conditions because the 
type of linkage shown in Fig. 63 implies that any ionospheric current system 
must be accompanied by magnetospheric motion. According to our sketch 
of the magnetosphere (Fig. 61), mid- and low-latitude field lines convect 
within a closed volume, but high-latitude field lines are connected with more 
distant regions and may penetrate the geomagnetic tail. In principle, the 
convection patterns in the magnetosphere could be mapped by tracing the 
electron motions in the dynamo region, provided the configuration of the 
geomagnetic field is known (Taylor and Hones, 1965).

7.54 Ionospheric and Auroral Phenomena in Relation to the 
Magnetosphere

Various upper atmospheric phenomena are linked with the structure of the 
magnetosphere shown in Fig. 61, with the populations of energetic particles 
contained in this structure, and with the circulation of the thermal plasma.

In Fig. 61 we showed a domain of "auroral" particles which lie outside 
the stable trapping regions populated by the Van Allen radiation. This 
domain intersects the earth's surface in two bands which encircle the two
geomagnetic poles. These bands may be identified with the "auroral ovals" which are the loci of precipitation of the energetic particles causing aurora. The ovals are closest to the geomagnetic poles on the day side and furthest from them on the night side, their approximate limits being 78° and 68° dipole latitude under quiet conditions. In the Northern hemisphere, the ovals correspond quite well to the area marked with triangles in Fig. 64, which we discuss later. The orientation of the auroral ovals is roughly fixed with respect to the sun. As the earth rotates, each oval appears to revolve about the geomagnetic poles and, therefore, passes over different regions of the

Fig. 64. An idealized representation of the two main zones of auroral particle precipitation in the northern hemisphere, where the average intensity of the influx is indicated very approximately by the density of symbols and the coordinates are geomagnetic latitude and geomagnetic time. The "discrete" events are represented by triangles (which closely correspond to the undisturbed "auroral oval") and the "diffuse" events are indicated by the dots [Hartz and Brice (1967)]. (Geomagnetic time is local solar time measured with respect to geomagnetic longitude.)
earth. In particular, let us consider the midnight sector of the oval, which is the sector furthest from the pole, and is where bright auroras most often occur (see Fig. 64). As the oval revolves, this sector traces out a band, two or three degrees broad, approximately centered on dipole latitude 68°. This is the auroral zone. During magnetic disturbance the auroral ovals expand in area, and the auroral zone moves to a lower latitude (Fel'dshteyn, 1963; Akasofu, 1966).

The distribution of auroras is connected with the distributions of other high-latitude phenomena. The ionospheric phenomena include radio black-outs (auroral absorption) and high-latitude sporadic E, whose distributions are obtained from riometer, radio propagation and ionosonde data. The occurrence of these phenomena shows marked maxima at certain times of day, which vary from station to station. On polar maps plotted in magnetic coordinates, the “isochronic” lines joining points at which the maxima occur simultaneously (in Universal Time) are found to be spiral in form (Agy, 1954, 1960; L. Thomas, 1962b). A similar spiral pattern has been found for magnetic activity at high latitudes (Nikol’skiy, 1961), for auroral activity, and for hydrogen Hα airglow; the data have been summarized by Nagata (1963).

Originally, these patterns were interpreted as Störmer spirals, which are the loci of precipitation of charged particles originating in a distant source and deflected by the geomagnetic field (Störmer, 1917). This interpretation is now abandoned and the patterns are considered to be related to the structure of the magnetosphere. Indeed, Akasofu identifies the two main “spirals” as sectors of the auroral oval. Hartz and Brice (1967) have combined data on many high-latitude phenomena, and plotted them on a polar map in terms of dipole latitude and local geomagnetic time (i.e., longitude with respect to the direction of the sun); see Fig. 64. On this diagram separate “spirals” do not appear, but Hartz and Brice find that the phenomena fall into two classes located in the two different zones postulated by Piddington (1965). One class, “discrete” events marked by triangles on the map, includes bright auroral displays; magnetic bays (“polar substorms”); ionospheric phenomena such as auroral absorption, sporadic E and spread F; and certain types of micropulsation, VLF emission and particle precipitation. Such phenomena are attributed to intense, rapidly fluctuating and spatially limited “splashes” of precipitated particles. These events occur in a band which is very similar (if not identical) to the auroral oval, normally ranging in latitude from 78° to 68° but expanding towards lower latitudes during disturbances. The other class, “diffuse” events marked by dots on the map, includes various steady or slowly varying auroral, ionospheric and VLF
phenomena, which are attributed to a steady "drizzle" of electrons (> 40 keV) from the outer radiation belt. The "drizzle" has been measured by particle detectors on satellites, such as Alouette I (McDiarmid and Burrows, 1964). According to Fig. 64, the "diffuse" events occur in a zone centered at about 65°; they are most frequent at about 0800 hours local geomagnetic time, whereas the "discrete" events are most frequent around 2200 hours. Hartz and Brice state that a similar pattern, but with the zones shifted poleward by about 5°, applies to the Southern hemisphere. We must remember that statistical studies of phenomena do not necessarily display the conditions existing at any one instant of time.

If the above description is correct, the particles responsible for the phenomena in the auroral oval are not stably trapped; they may come from the tail region, or even from outside the magnetosphere via the tail. The low-latitude boundary of the auroral oval should correspond to the outer (high-latitude) boundary of the trapped particle domain.

Another feature of the magnetosphere is the "plasmapause" or "knee" in the distribution of thermal plasma. From observations of whistlers, it is possible to deduce the radial distribution of electron concentration in the equatorial plane, as reviewed by Carpenter and Smith (1964). The rapid outward decrease observed at about 4 earth radii is now thought to be part of a magnetically aligned structure (Fig. 65), the electron concentration

![Fig. 65. Idealized meridian cross section of the magnetosphere near 1400 hours local time. The shaded region shows the location of the high-density plasma inside the plasmapause or "knee" (electron concentration of order 100 cm\(^{-3}\)). The region outside the plasmapause (electron concentration of order 1 cm\(^{-3}\)) is linked by magnetic field lines to the ionospheric "troughs." The dashed part of the boundary shows the low-altitude region in which the structure of the "knee" is not well known. The position of the plasmapause corresponds to moderate but steady magnetic activity (\(K_p = 2-4\)) [Carpenter (1966)].]

\[\text{Plasmasphere}\] 2 ~100 electron cm\(^{-3}\)
\[\text{Plasmapause}\] 4 6 8
\[\text{Plasma trough}\] ~1 electron cm\(^{-3}\)
being much greater within the plasmapause than outside it. In the Figure, the plasmapause is shown at $L = 4$, a position typical of slight magnetic disturbance, with $Kp \sim 3$ (Angerami and Carpenter, 1966; Carpenter, 1966). Under very quiet conditions, the plasmapause expands and may be found at $L \sim 6$; under very disturbed conditions, it contracts and may be found at $L < 3$.

It has been suggested that the plasmapause marks the outer boundary of the region in which the plasma rotates with the earth (Nishida, 1966; Brice, 1967). The magnetic tubes of force in this region undergo the type of circulation pictured in Fig. 63, on which is superimposed the earth's rotation. The circulation outside the plasmapause follows a different pattern and does not rotate with the earth. The changes in position of the plasmapause may then represent a change in the relative strengths of the inner and outer circulations.

The ionospheric "troughs" described in Sec. 5.52 are situated on (or near) the magnetic field lines which delineate the plasmapause, and seem to be linked with the plasmapause in some way.

The features we have described in this section seem to possess the general property that they move towards lower latitudes and smaller $L$-values during magnetic disturbances. This applies both to features associated with the thermal plasma (the "plasmapause" and ionospheric "troughs") and to those associated with energetic particles, such as the different manifestations of the "auroral oval."
8.1 Synopsis of Storm Effects

The events known as magnetic storms comprise a very complicated sequence of phenomena. Our account is necessarily condensed, and most of our statements are subject to reservations of one sort or another. Theories of storms are tentative, and individual storms vary so much that exceptions may be found to almost any description of the observed phenomena. Many of the phenomena that we shall mention occur at times of slight magnetic disturbance, as well as during the specific events called “storms.”

Magnetic storms are initiated by solar disturbances. Most severe storms have a sudden commencement (SC) which takes place one or two days after a solar chromospheric flare (Jenkins and Paghis, 1963), but many storms cannot definitely be attributed to any particular flare. Large flares are commonest at sunspot maximum, and SC storms are most frequent then. Another type of storm, common during the declining part of the solar cycle and generally weaker than the SC type, is not associated with flares. These storms often start gradually with no SC and tend to recur at intervals of 27 days, the period of solar rotation as viewed from the earth. Sometimes in a sequence of recurrent storms, some have an SC and some do not. These phenomena have been attributed to periodic immersion of the earth in particle streams emitted, more or less continuously, from active areas on the sun which became known as (magnetic) $M$ regions. Since $M$ regions could not positively be identified with visual phenomena on the solar disk, even though they must persist for some months, they remained somewhat hypothetical. It has been suggested (Dessler and Fejer, 1963; Piddington, 1964) that $M$-region storms are associated with certain features in the interplanetary magnetic field. These rotate with the sun, hence causing a 27-day recurrence, but there need be no association between the storms and visible
features on the solar disk. Space probes such as IMP-1 have indeed detected ordered structures in the interplanetary field, rotating with the sun and retaining their general shape throughout a few rotations (Wilcox and Ness, 1965). Their general features can be correlated with the measured plasma velocities in the solar wind and with variations of the geomagnetic $Kp$ index. Results from other space probes confirm that geomagnetic disturbances are connected with disturbances of the solar wind.

Table VI shows some storm phenomena, with notes on their durations and possible causes; some of these are reasonably well established but others are quite speculative. The Table includes, for completeness, the solar-flare phenomena which we have mentioned previously in Sec. 6.1. It also brings together a number of items dealt with in other parts of the book. Any numerical values quoted are intended to be mere orders of magnitude, characteristic of strong (but not exceptional) SC storms. Timewise, the phenomena fall into two categories. Those listed under (1) and (2) in Table VI mostly accompany the visual flare, and involve electromagnetic radiation or relativistic particles. Effects (3) to (8), which comprise the storm proper, depend on the much slower propagation of clouds of solar plasma. A typical delay time of 1.5 days between the flare and the SC implies a speed of order $10^3$ km s$^{-1}$ from the sun to the earth. But solar cosmic-ray increases and PCA events listed under (2) begin not long after the flare and seem to be due to solar protons traveling at velocities of up to $10^5$ km s$^{-1}$.

Much of the interest in geomagnetic storms is centered on the aurora, cosmic rays, and the radiation belts. These are included under (4) and (5) in Table VI, but we do not propose to discuss them further here. The mid-latitude red arcs have been referred to in Sec. 3.7. The relationship between micropulsation phenomena and the state of the magnetosphere has been discussed in a review by Troitskaya (1967).

The increase of satellite drag during storms is almost certainly due to an increase of atmospheric temperature. A detailed analysis of this effect has been given by Jacchia (1963, 1965) and Jacchia et al. (1967). It is found that the temperature in the upper thermosphere increases by 1.0 to 1.5°K per unit increase of the geomagnetic parameter $Ap$; the rate of increase being greater in auroral latitudes than in lower latitudes.

In the rest of this chapter, we look at two aspects of storms in more detail: the electric current systems in Sec. 8.2, and ionospheric effects in Secs. 8.3 and 8.4.
<table>
<thead>
<tr>
<th>Disturbance</th>
<th>Outline of phenomenon</th>
<th>Timing and duration</th>
<th>Possible cause</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Solar flare</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Solar Phenomena (Sec. 6.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optical</td>
<td>Eruption in solar chromosphere</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enhanced Hα emission over fraction ($&lt; 10^{-3}$) of solar disk</td>
<td>$\frac{1}{2}$ hr</td>
<td></td>
</tr>
<tr>
<td>Ultraviolet</td>
<td>Significant enhancements</td>
<td>All effects start approximately simultaneously</td>
<td></td>
</tr>
<tr>
<td>X-rays</td>
<td>Strong emission in 1–10 Å range</td>
<td>(Complicated line dependences and frequency drifts)</td>
<td>Varies with type</td>
</tr>
<tr>
<td>Radio frequencies</td>
<td>Several types of emissions covering all observable wavelengths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Type IV, Type II, Type III)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Geophysical phenomena</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sec. 6.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sudden ionospheric disturbance (SID)</td>
<td>Effects in sunlit hemisphere</td>
<td>Accompanies visual flare, persists $\sim \frac{1}{2}$ hr</td>
<td>D region ionization increased by X-rays</td>
</tr>
<tr>
<td></td>
<td>strong absorption; anomalous VLF reflection; effects in F region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic crochet</td>
<td>Sudden disturbance of field ($\sim 20 y$)</td>
<td>Accompanies visual flare</td>
<td>Increase of ionospheric currents because of enhanced conductivity</td>
</tr>
<tr>
<td>Solar cosmic rays</td>
<td>Latitude dependent increase</td>
<td>Starts about $\frac{1}{2}$ hr after flare</td>
<td>Protons $\leq 100$ MeV</td>
</tr>
<tr>
<td>Polar cap absorption (PCA)</td>
<td>Intense radiowave absorption in magnetic polar regions</td>
<td>Starts few hours after flare</td>
<td>Protons 1–10 MeV</td>
</tr>
</tbody>
</table>
### Magnetic storm

**Interaction of low energy plasma with the earth's field**

**Impulsive increase of field (10 \( \gamma \))**

**Impulsive: simultaneous (to within ~ 1 min) over the earth**

**Impact of plasma "front" on boundary of magnetosphere**

<table>
<thead>
<tr>
<th>Segment</th>
<th>Event Description</th>
<th>Duration</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sudden commencement (SC)</td>
<td>Impulsive increase of field (10 ( \gamma ))</td>
<td>2-6 hr</td>
<td>Compression of field by impinging plasma</td>
</tr>
<tr>
<td>Initial phase</td>
<td>Field increased; ( Dst(H) \sim +20 ( \gamma )</td>
<td>1-2 days</td>
<td>Orbiting charged particles in magnetosphere give &quot;ring current&quot;</td>
</tr>
<tr>
<td>Main phase</td>
<td>Field depressed; ( Dst(H) \sim -50 ( \gamma )</td>
<td>1-2 days</td>
<td>Orbiting charged particles in magnetosphere give &quot;ring current&quot;</td>
</tr>
<tr>
<td>Recovery phase</td>
<td>Field returns exponentially to normal</td>
<td>Time constant ~ 1 day</td>
<td>Particles removed by collisions (charge exchange) with neutral atoms</td>
</tr>
<tr>
<td>Polar substorm</td>
<td>Magnetic bays at high latitudes (( \sim 500 ( \gamma )), produced by auroral electrojet and associated with aurora, become more frequent and intense</td>
<td>Duration, ~ 1 hr, tends to recur at ~ 3 hr intervals</td>
<td>Possibly related to phenomena in the geomagnetic tail</td>
</tr>
</tbody>
</table>

**Forbush cosmic ray decrease**

<table>
<thead>
<tr>
<th>Event Description</th>
<th>Starts within ~ 1 hr of SC</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Galactic) cosmic ray intensity temporarily reduced</td>
<td></td>
<td>Particle trajectories perturbed by the distortion of the interplanetary magnetic field, due to the solar plasma stream</td>
</tr>
<tr>
<td>Disturbance</td>
<td>Outline of phenomenon</td>
<td>Timing and duration</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Radiation belts (Sec. 7.52)</td>
<td>Outer Van Allen belts initially depleted during storm, then replenished</td>
<td>Large variations in outer belt can occur within hours</td>
</tr>
<tr>
<td>(5) Aurora</td>
<td>Expansion of &quot;auroral oval&quot;</td>
<td>Complicated sequences of phenomena occur; some phenomena observed to occur within hours.</td>
</tr>
<tr>
<td>Mid-latitude red arcs (Sec. 3.71)</td>
<td>Enhanced [O I] 6300 Å emission in bands some thousands of kilometers long oriented along magnetic latitudes, ~ 500 km wide, and situated at ~ 400 km altitude.</td>
<td>Stable for many hours</td>
</tr>
<tr>
<td>Magnetospheric phenomena</td>
<td>Changes in thermal plasma plasmaopause moves inward from ( L = 5 ) to ( L = 3.5 ) and magnetic field changes occur within few hours.</td>
<td>Stable for many hours</td>
</tr>
</tbody>
</table>
### 8.1 SYNOPSIS OF STORM EFFECTS

<table>
<thead>
<tr>
<th>(7) Atmospheric Effects</th>
<th>Drag on satellites increases because density at great heights is increased by a rise of thermospheric temperature</th>
<th>Persists for some days</th>
<th>Absorption of hydromagnetic waves; corpuscular heating; joule heating</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Sec. 1.33)</td>
<td>Starts soon after SC (time resolution in data is few hours)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Ionospheric effects</td>
<td>Enhanced absorption, $E_d$ ionization, following regular spatial distributions</td>
<td>Strong universal-time and local-time dependences</td>
<td>Precipitation of electrons of few kilo-electron volts</td>
</tr>
<tr>
<td>Blackout, storm $E_d$ at high latitudes (Secs. 7.54, 8.3)</td>
<td>Diurnal variation of VLF phase is abnormal $f_0F2$ increased during first day of storm then generally depressed, but sometimes increased especially in low latitudes</td>
<td>Persists for &gt; 10 days</td>
<td>Possible chemical changes in atmosphere</td>
</tr>
<tr>
<td>D region storm effect and after-effect (Sec. 8.3)</td>
<td>Effects last for many days, with strong daily variations</td>
<td>Temperature changes; increase of loss coefficient; electromagnetic movements</td>
<td></td>
</tr>
<tr>
<td>Mid- and low-latitude F region changes of electron density (Sec. 8.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Solar flare phenomena are included for the sake of completeness. Numerical magnitudes represent a moderately strong storm.*
8.2 Storm Current Systems

In Sec. 7.32 we expressed the deviation of the geomagnetic field from its normal quiet-day value as the sum of a component \( D_{st} \) (a function of storm time but independent of longitude) and a longitude variation \( DS \); see Eq. (708). This representation of the data is purely mathematical, and we now wish to consider the actual current systems that produce these variations.

Just as in the mathematical formula \( D = D_{st} + DS \), the disturbance field \( D \) may be evaluated for any component or element of the geomagnetic field, the contribution of induced ground currents being included. Following Chapman (1963b), we write

\[ D = DCF + DR + DP \]  

(800)

in which \( DCF \) denotes the disturbance due to solar corpuscular flux, \( DR \) the disturbance due to ring current, and \( DP \) the disturbance due to polar current systems. Sometimes an extra term \( DT \) is added to represent the effect of currents flowing in the geomagnetic tail; but since the size of such tail currents and their relation to the other systems are not well established, we confine our discussion to the three terms proposed by Chapman. A review of the various contributions has been given by Obayashi (1967); we show in Fig. 66a a sketch of an idealized magnetogram for a mid-latitude station during a strong storm, and indicate the contributions of the various current systems. The nomenclature of Eq. (800) has become quite widely used, though variants of it exist, and the subdivisions we mention below are only tentatively defined.

\( DCF \) is the corpuscular flux effect, ascribed to neutral streams of ionized particles emitted from the sun. In their pioneer work on geomagnetic storm theory, Chapman and Ferraro (1931–1933) solved idealized mathematical problems relating to this effect, and showed that the compression of the geomagnetic field by the particle stream could produce the “initial phase” of the magnetic storm, in which the magnetic field at ground level is increased \( (D_{st}(H) > 0) \). The \( DCF \) currents mostly flow at the boundary of the geomagnetic field, which we now call the magnetopause (Fig. 61). Nowadays the process is generally described in hydromagnetic terms; the solar stream is actually a cloud of plasma, which may be regarded as an enhancement of the “solar wind.” The arrival of this plasma in the vicinity of the earth sets up an impulsive disturbance which is transmitted through the magnetosphere in the form of hydromagnetic waves, and is observed as the “storm sudden commencement” (SC). The differences in the time of occurrence of
the SC at different places, which may be a few minutes, are associated with the propagation of these waves.

$DR$ is the disturbance attributed to a westward ring current in the magnetosphere. To a first order the ring current is symmetrical about the geomagnetic axis, and its field gives the "main phase" decrease ($Dst(H) < 0$). Any asymmetry in the ring current contributes to $DS$. The ring current has been tentatively subdivided into two components, sketched in Fig. 66a, though the physical difference between them has still to be determined:

$DR1$ is most intense in great storms. Its maximum intensity may occur within a few hours of the SC, and it disappears within a day.

---

Fig. 66a. Schematic magnetogram showing the storm variations of the horizontal field $H$ at a station in mid-latitudes ($30^\circ$-$50^\circ$). The broken line shows the suggested variation of the "ring current" field $DR$, which is tentatively divided into components $DR1$ (predominant in the first day of the storm) and $DR2$ (which decays more slowly). The deviations from this curve are produced by the "corpuscular flux" component, which follows the sudden commencement SC; and the "polar" field $DP$, which includes two negative $DP1$ bays.

Fig. 66b. Idealized current systems, looking down from the north geomagnetic pole, for the polar current systems $DP1$ (auroral electrojet) and $DP2$ (twin vortex); and the composite $DS$ pattern. The outer boundary of each diagram represents $30^\circ$N dipole latitude; numbers denote local time [after Obayashi (1967)].
DR2 is a weaker but longer lived feature, with a decay time of one or two days. It is the principal contributor to Dst in weak storms.

Anticipating the discovery of the radiation belts, Singer (1957) suggested that the ring current might be carried by energetic particles trapped in the manner described in Sec. 7.52. Evidence for the existence of these ring current particles was obtained by the OGO 3 satellite during magnetic storms in June and July 1966 (Frank, 1967). Large increases occurred during the storm in the fluxes of protons (0.2 to 50 keV) between $L = 3$ and $L = 6$, with a maximum at around $L = 3.5$; a minor contribution to the ring current appeared to be made by electrons in a similar energy range. Magnetic field observations made by a satellite indeed suggest that the DR current is centered at about $L = 3.5$ (Cahill, 1966). The decay of the DR currents may be largely controlled by the removal of trapped protons in charge-exchanging collisions with ambient neutral hydrogen atoms, as proposed by Dessler and Parker (1959).

DP is the disturbance due to ionospheric currents, principally flowing in polar regions, though some DP current flows in the mid-latitude ionosphere, and there may be substantial flows along magnetospheric field lines. The DP field has a strong local-time variation and is the principal contributor to DS in high latitudes. Two divisions of DP have been tentatively proposed (Fig. 66b):

DP1 produces magnetic “bays,” which are rather abrupt magnetic disturbances lasting an hour or two, and which can be positive or negative. Two large negative bays are shown in Fig. 66a. Bays occur during quiet times but are more intense and frequent during storms, and are most intense at auroral latitudes, where they may exceed 1000 $\gamma$ in size. They are part of the “polar substorms” which are connected with intense auroral phenomena. The DP1 current system comprises an intense westward “electrojet” along the nighttime part of the auroral oval (Sec. 7.54) with return currents either side. An eastward electrojet may exist at times in a different part of the oval.

DP2 is a twin-vortex system, the return currents of which extend to low latitudes. It is present in all phases of the storm, though its intensity and orientation may change. The current flows are not concentrated enough to be regarded as “electrojets.” A weak form of DP2 exists even at quiet times; it is known as $Sq^p$ and appears at high latitudes in the $Sq$ variations of Fig. 58.

The idealized DS current system for the storm local-time variation, also shown in Fig. 66b, is made up partly of DP2 and partly of a statistical aver-
aging of $DP1$ substorms, probably with some contribution from $DR$. For discussions of the relationship between these different patterns see Akasofu (1966), Cole (1966), and Obayashi (1967); this subject is still controversial.

To explain the $DP$ current systems, one must postulate localized enhancements of either electric field or conductivity, or both (Nagata and Fukushima, 1952; Cole, 1960). The conductivity can be enhanced by ionization produced by charged particle precipitation, such as certainly occurs during the substorms associated with the $DP1$ electrojets, these being part of the “discrete” particle phenomena of Sec. 7.54, Fig. 64. The particle precipitation may well be caused by phenomena within the geomagnetic tail, possibly some kind of plasma instability or magnetic field-line merging (Atkinson, 1966; Axford, 1967; Speiser, 1967).

Given sufficient increases in conductivity, the $DP$ currents might possibly be produced by dynamo action of the type described in Sec. 7.4. This would seem to require very strong localized wind velocities of up to 1 km s$^{-1}$ (Cole, 1960; Boström, 1964). Alternatively, Boström shows that the electrojets might be produced by magnetospheric electric fields of order 1 mV m$^{-1}$; he shows how the geometry of the electric field and current systems depends on the magnetospheric conductivity.

Kern (1962) and Fejer (1963) have suggested that the $DP$ (or $DS$) currents are driven by mechanisms which produce separation of charge in the magnetosphere, and which depend on the geometry of particle motions. Subject to the requirement of electrical neutrality in the magnetosphere as a whole, there could exist some imbalance of charge in the trapped energetic particle population described in Sec. 7.52, and an opposite imbalance in the thermal particles which participate in the circulation described in Sec. 7.53. Since the orbits of these two populations are different, space charges may develop and will then lead to neutralizing current flows along field lines and through parts of the ionosphere. When this principle is incorporated in “dynamo” calculations of the kind discussed in Sec. 7.42, the resulting magnetic variations are found to resemble $DS$ (Fejer, 1964; Gottlieb and Fejer, 1967). This mechanism is probably more relevant to the $DP2$ system than to $DP1$.

8.3 Storm Effects in the Lower Ionosphere

It is well known that radio communications are affected during magnetic storms, the terms “fadeout” and “blackout” being commonly used. Though the difficulties in ionospheric propagation are partly caused by the decreased F2-layer critical frequencies (to be discussed in Sec. 8.4), they are partly due
to D-region absorption. The effects are most severe in high latitudes, particularly around the auroral zone.

We shall discuss the high-latitude phenomena first. They fall into various categories, of which the first three were originally classified by Reid and Collins (1959):

I. The "short-wave fadeout" (SWF), part of the "sudden ionospheric disturbance" (SID) which accompanies solar flares (Dellinger, 1937). This is observed all over the sunlit hemisphere, so it is not particularly a "high-latitude" phenomenon (Sec. 6.1).

II. The "auroral zone absorption" or "auroral blackout" which occurs during polar substorms and the main phase of geomagnetic storms. It is associated with certain types of aurora, and its positional distribution is related to the "auroral oval" discussed in Sec. 7.54.

III. The "polar cap absorption" (PCA) which begins a few hours after solar flares (and thus well before the storm SC) and persists for some days; pronounced day-to-night variations occur at places where the sun rises and sets. Maps showing the development of PCA have been published by Obayashi (1959) and Obayashi and Hakura (1960). We have already mentioned the interest that has been taken in PCA events, in relation to D-region aeronomy, particularly concerning negative ions (Sec. 3.5).

IV. Additionally, and more recently, Bailey and Pomerantz (1965) have identified absorption events due to "relativistic electron precipitation," at subauroral latitudes during magnetic disturbances. These are attributed to ionization produced near 70 km by electrons of order 1 MeV energy (whereas PCA is attributed to MeV protons).

D-region storm effects at mid-latitudes can be detected by their influence on radio waves reflected from the ionosphere. In particular, storms affect the amplitude of low frequency (LF) waves of 30–300 kHz (the A3 technique, Sec. 2.33) and the phase and amplitude of very low frequency (VLF) waves of 10–30 kHz (Sec. 2.32). At the time of a magnetic storm the "primary" effects occur; they include rapid fading of LF waves and rapid phase fluctuations of VLF waves. The "primary" effects disappear within a day or two, and are followed by the "after-effect" which begins about two days after the storm and lasts for ten or more days. The "after-effect" causes severe LF absorption and an abnormal diurnal phase variation of VLF waves. The VLF phenomena are illustrated by Fig. 67, due to J. S. Belrose, which shows the diurnal variations of phase of 16 kHz waves observed 90 km from the transmitter. The dashed lines show the pattern of diurnal variation which
normally occurs at the same season and the same part of the solar cycle. A detailed study of the "after-effect" has been made by Lauter and Knuth (1967) using A3 absorption data obtained in Europe over a period of some years. The increases of absorption are confined to latitudes above 45°, and are most frequent in summer. Lauter and Knuth suggest that the extra D-region ionization is produced by energetic electrons, which are stored in the outer radiation belt (at $L \approx 3$) during a storm and are subsequently precipitated into the D region over a period of days. Phenomena similar to the
storm "after-effect" have also been observed in the D region after a nuclear explosion (Brady et al., 1964).

The most striking storm phenomenon in the E region is the "storm sporadic E" which occurs at moderately high latitudes. Its distribution has been studied by L. Thomas (1962b), and appears to be related to the auroral oval discussed in Sec. 7.54. The thick "night E" layers discussed by G. A. M. King (1962b) also occur in auroral regions at times of magnetic disturbance.

As regards the normal mid-latitude E layer, Sato (1957b) and Brown and Wynne (1967) have found that \( f_{\text{OE}} \) is decreased by a few percent during storms. The virtual height \( h'\text{E} \) shows slight changes (about \( \frac{1}{2} \text{ km} \)) from its usual behavior. These effects are attributed to electromagnetic drift produced by the storm electric field systems. In support of this explanation, Brown and Wynne have resolved the E-layer variations into \( S\omega \) and \( S\delta \) components, in the manner described in Sec. 7.3 for magnetic variations, and have found a resemblance in form to the magnetic \( S\delta(H) \) variation.

### 8.4 Storm Effects in the F Region

#### 8.41 Observed Phenomena

In contrast to the rather elusive effects of solar flares (Sec. 6.1), changes in the F layer during a storm are profound. The critical frequency \( f_{\text{OF}} \) is not greatly affected; Sato (1957b) and Brown and Wynne (1967) found a slight decrease of about 0.3 MHz, which implies a reduction of about 10% in \( N_m\text{F1} \). We may take this to indicate that the production rate \( q \) is not greatly changed during a storm, as we might indeed expect. But the reduction of F2-layer electron concentration sometimes results in the F1 "cusp" becoming visible on ionograms at times when it would not normally be seen. This may be interpreted in terms of the "F-layer splitting" theory of Sec. 3.62, as an increase of the parameter \( G = \beta^2/\omega q \), probably due to an increase of loss coefficient \( \beta \) (G. A. M. King, 1962a, 1967). King has found that \( G \) generally increases when the magnetic disturbance \( K\)-figure is increased.

During storms, the \( h'(f) \) curves for the F2 layer are greatly changed. The critical frequency \( f_{\text{OF}} \) is usually reduced at mid-latitudes and the virtual height \( h' \) is greatly increased, sometimes attaining values of 600 km or more. The increase was originally thought to indicate real changes in the height of the F2 layer, but this interpretation was shown to be false when the \( N(h) \) profiles were computed; the real height of the F2 layer is actually the same, to within about \( \pm 20 \text{ km} \), on storm days and quiet days (Thomas and Robbins, 1958; Becker, 1964). The increase of \( h'\text{F2} \) is due to group retarda-
8.4 STORM EFFECTS IN THE F REGION

Fig. 68. $Dst$ variations of $N_mF2$ in each of eight latitude zones for strong and weak magnetic storms with sudden commencements. The ordinate is the approximate percentage deviation from the quiet day behavior versus storm time (in hours). The zone number is shown in parenthesis between the applicable geomagnetic latitudes: 60°(1) 55°(2) 50°(3) 45°(4) 40°(5) 30°(6) 20°(7) 10°(8) − 10° [Matsushita (1959)].
tion occurring below the height of reflection (Sec. 2.21), and should be regarded only as a sensitive indicator of F-layer disturbance, not possessing any quantitative meaning. Studies of F2-layer effects during magnetic disturbance include correlations of ionospheric parameters with magnetic indices, such as $Kp$ and $Ap$. Generally, it is found that F2-layer disturbance lags behind magnetic disturbance, by about a day (e.g., J. W. King, 1961).

The storm changes in $f_0F2$ can be analyzed in terms of the $SD$ and $Dst$ variations originally applied to magnetic parameters. This was done by Martyn (1953b,c), and a more extensive analysis was carried out by Matsushita (1959), who used data for 109 storms at 38 observatories situated in eight latitude zones. Matsushita’s $Dst$ curves are reproduced in Fig. 68. In geomagnetic dipole latitudes 45–60°, $Dst(NmF2)$ is positive for the first few hours and then becomes negative, reaching a greatest depression of 30% in $NmF2$ (for strong storms) about 24 hr after the SC. In dipole latitudes 20–45°, the positive phase lasts longer and the negative phase is weaker (about 10% depression of $NmF2$). In low latitudes, there is no negative phase and $Dst$ is weakly positive (5 to 10% in $NmF2$) throughout the storm. The local-time ($SD$) components of the variation display complicated changes of amplitude and phase, especially during the earlier parts of the storm, and gradually die away in the recovery phase. Brown and Wynne (1967) have also computed $SD$ variations of F1-layer parameters.

These results are statistical averages for a number of storms. Individual ionospheric storms can be either “positive” or “negative,” according to whether $NmF2$ is increased or decreased from mean quiet day values. “Positive” storms appear to be more common in winter (Sato, 1957a). “Negative” storms generally possess an initial positive phase, so that the curves of $Dst(NmF2)$ possess a superficial resemblance to magnetic $Dst(H)$ curves (e.g., Fig. 60). This resemblance does not necessarily imply any actual connection; the magnetic $Dst$ variation is attributed mainly to the $DR$ ring current well outside the ionosphere. We may note that the positive phase lasts longer in the $NmF2$ variation than in the magnetic field variation. Appleton and Piggott (1952) showed that the form of the $NmF2$ variation depends on the local time of the start of the storm. Following this, L. Thomas and Venables (1966) find that the major changes of $NmF2$, in mid-latitudes, depend on the local time at which the main phase of the magnetic storm begins (i.e., when $Dst(H)$ decreases through zero). If this occurs at night, an immediate depression of $NmF2$ ensues, but not if it occurs by day.

Satellite and moon-echo observations show that the total electron content $\int N dh$ decreases during storms, and with increase of $Kp$ generally (Ross,
Fig. 69. Ionospheric variations during a magnetic storm, which started at about 1300 hours on 12 November 1960 (marked by a bar). Local time runs from 0000 to 2400 hours and the $K$-index is a measure of magnetic disturbance during each 3-hr period (0 = quiet; 9 = very disturbed). Circled points are hourly values of peak electron density (concentration) $N_mF_2$ at Slough (52°N, 1°W); V-marks are upper limits of $N_mF_2$ (unit $10^{12}$ m$^{-3}$ or $10^{13}$ cm$^{-3}$). The full line shows the total electron content, $n_T = \int N dh$ (unit $10^{17}$ m$^{-3}$ or $10^{18}$ cm$^{-3}$), measured by lunar radar at Jodrell Bank (53°N, 2°W) during times when the moon was suitably placed [Taylor (1961)].
The reduction of total electron content was extremely pronounced during a severe storm in November 1960 (Taylor, 1961), as shown in Fig. 69. Both $N_mF_2$ and $\int N dh$ increase more slowly after sunrise on the storm day than on quiet days. The important point that is obtained from the total content observations is that the storm effects in the F2 layer cannot be interpreted in terms of a mere redistribution of ionization.

At moderately high latitudes (dipole latitude about 60°, $L=4$) Gledhill et al., (1967) have found correlations between F-region disturbances and energetic particle precipitation. The disturbances are largely detected by their effect on $h'F_2$, so the altitude at which they occur is not precisely defined.

8.42 POSSIBLE EXPLANATIONS

F-region storm effects present many unsolved problems. We can do no more than review briefly some possible explanations, which include “thermal expansion,” “temperature dependent rates,” “composition change,” and “electromagnetic drift” theories. Very possibly, all of these contribute to F-layer storm phenomena.

Early papers (Berkner et al., 1939) attributed the depression of electron concentration to thermal expansion of the F region, this being largely suggested by the increase of F2-layer virtual height, which is now known to be misleading. A more recent discussion has been given by Matuura (1963).

Thermal expansion of the F region would be expected to cause some thickening of the layer, accompanied by some decrease in $N_mF_2$; there should be an increase of $h_mF_2$ (Sec. 4.44). Thus, if thermal expansion were the main cause of storm effects, there should be a correlation between the storm changes of $N_mF_2$ and $h_mF_2$ which does not, in fact, exist. Hence, although satellite drag studies have shown that the atmosphere is indeed heated during storms, this explanation seems inadequate (G. A. M. King, 1966).

In order to explain large decreases of $N_m$, it has been suggested by Yonezawa (1963) that the loss coefficient is enhanced because of a postulated temperature dependence of the ion-atom interchange coefficients $\gamma$ (Table II, Sec. 3.63). It, indeed, has been found that the rate coefficient for reaction (T2), between $O^+$ and $N_2$, is roughly proportional to gas temperature at thermal energies (Bohme et al., 1967); and that it increases rapidly with increasing electron temperature (Schmeltekopf et al., 1967). During magnetic disturbance the electron temperature has been found to increase at night in the mid-latitude F region, but it is not yet clear whether it consistently does so by day (Sec.
A temperature dependence of the diffusion coefficient $D$ could help to produce storm effects, but the data on $D$ (Sec. 4.31) do not predict any very strong dependence.

We can see from our discussion of F-region loss processes [Sec. 3.63, Eq. (359)] that the loss coefficient $\beta$ could alternatively be increased by changes of atmospheric composition which increase the proportion of molecular gases in the F region (Seaton, 1956). Such a change might conceivably occur if dynamic processes in the vicinity of 100 km were enhanced such as to increase turbulence and raise the upper limit of mixing of $O_2$ and $N_2$ with atomic oxygen (G. A. M. King, 1962a). Subsequent papers (G. A. M. King, 1966, 1967) discuss evidence that mixing is caused by atmospheric waves, which are generated by disturbances in auroral regions. The increase of $\beta$ then decreases $N_mF_2$, as well as causing the increase of $G = \beta^2/\alpha q$ observed in the F1 layer, mentioned earlier in this section.

The drift theory of Martyn (1953a) attributes storm effects to electromagnetic movements, produced by the electric fields associated with the storm current systems. Maeda and Sato (1959) attributed the occurrence of positive and negative storms at different latitudes to a latitude-dependent combination of quiet-day and storm ($Sq$ and $DS$) drift velocities.

The drift calculations can be divided into three main stages, which have much in common with the dynamo theory scheme (Sec. 7.4) and suffer from much the same uncertainties:

(a) Computation of ionospheric storm current systems from the magnetic data. This requires assumptions about the contribution of induced ground currents to the magnetic effects.

(b) Calculation of electric fields driving these currents, by inserting conductivities, and then the derivation of F-region drift velocity by assuming transmission of the electric field up the geomagnetic field lines. On the simplest theory, this velocity is $E \times B/B^2$.

(c) Insertion of the drift velocity into the continuity equation and calculation of the resulting change of electron concentration. The vertical component of drift would be expected to be the most important, but its effects are likely to be modified by ion-drag and the consequent neutral air motion (Sec. 4.24).

So far the drift theory has not led to very satisfactory explanations of storm phenomena. The problems arising in (a) and (b) are very similar to those of the dynamo calculations; one simplification exists, in that the storm electric fields are possibly purely "electrostatic," with no "induced" component due
to dynamo action. In addition to the electromagnetic drift velocity, there may be changes of the vertical drift velocities caused by the neutral winds in the F region, the winds presumably being influenced by changes of atmospheric temperature during storms. Regarding (c), the complexity of the F-layer continuity equation makes it difficult to compute the changes of electron concentration, even when the drifts are known. However, one can summarize the probable effects of drifts quite simply, by using the ideas of Secs. 4.43 and 4.44. An upward drift transports ionization into a region of lower loss, thus prolonging the average lifetime of ionization. Thus, during the day, upward movements would certainly increase the total electron content $\int N \, dh$, and would probably increase $N_m F_2$. A downward movement would have opposite effects. At night, once a "stationary" layer were established, upward or downward movement would alter the effective rate of decay of the layer.

We can see that theories of F-region storms are still quite speculative, and there is not yet any detailed explanation of the complex phenomena. In spite of the great progress achieved in recent years, the problems of magnetic and ionospheric storms will remain of interest for many years to come. Indeed, the same can be said for upper atmosphere physics as a whole, as we have tried to show in this book.
List of Principal Symbols

This list is not exhaustive. It omits many symbols which appear temporarily in the course of a derivation, or as constants or coefficients in an equation. Some commonly used subscripts are given; however, chemical and spectroscopic symbols, and abbreviations for units, are not given. For a few basic quantities, numerical magnitudes are given (mostly as published by the U.S. National Bureau of Standards). Although many symbols have more than one meaning, only one meaning is generally used within any one chapter. If a quantity occurs as a vector or tensor, the fact is indicated in parentheses in this list; the components and scalar magnitudes of such quantities are generally denoted by the same symbol in ordinary type. (Bold face type is used for vectors and bold face, sans serif for tensors in the text, but not in this list.)

\[ A \]
Thermal conductivity parameter of neutral gas; cross section of space vehicle; radio wave absorption parameter; amplitude of received radio signal \((A_o, A_x \text{ for } o \text{ and } x \text{ modes})\); magnetic disturbance index \((Ap—\text{planetary})\)

\[ B \]
Magnetic flux density (vector)

\[ C \]
Region of ionosphere (cosmic ray layer)

\[ C_i \]
Speed of sound; magnetic character figure \((C_i, C_p, \text{ etc.})\)

\[ C_D \]
Drag coefficient

\[ D \]
Region of ionosphere

\[ D \]
Coefficient of plasma (ambipolar) diffusion; drag force; magnetic or ionospheric disturbance variation \((DS—\text{solar}, Dst—\text{storm time}, DCF—\text{corpuscular flux}, DR—\text{ring current}, DP—\text{polar}, DT—\text{geomagnetic tail}); \text{magnetic declination}\)

\[ \partial \]
Differential operator in plasma diffusion theory

\[ E \]
Region of ionosphere \((Es—\text{sporadic E})\)

\[ E \]
Electric field (vector)

\[ F \]
Region of ionosphere \((F1, F1\frac{1}{2}, F2 \text{ layers})\)

\[ F \]
Force acting on a particle (vector)

\[ F_{ab} \]
Force per unit volume due to collisions between particles \(a, b\) (vector)

\[ F_\infty \]
Energy flux of ionizing radiation at top of atmosphere

\[ G \]
Gravitational constant, \(6.670 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}\); electron cooling parameter in wave interaction; \(F1 \text{ layer shape parameter } \beta^2/qa\)

\[ G_p, G_v \]
Rate of gain of heat per unit volume, at constant pressure and at constant volume

\[ H \]
Scale height \((H_v—\text{scale height associated with viscous drag}); \text{ horizontal intensity of geomagnetic field}\)

\[ I \]
Magnetic dip angle; intensity (photon flux) of ionizing radiation \((I_o—\text{value outside atmosphere})\)

\[ I_{F2} \]
Ionospheric F2 index

\[ J \]
Airglow intensity; height-integrated electric current density (vector)
LIST OF PRINCIPAL SYMBOLS

$K$ Wavenumber (vector); ratio of scale heights of F region production and loss functions; radio wave absorption coefficient per unit electron concentration ($K_o$, $K_x$ for o and x modes); magnetic disturbance index ($K_p$—planetary)

$K_{lb}$ Rate coefficient for collisions between particles $a$, $b$

$L$ Rate of loss of heat; radio wave absorption; magnetic or ionospheric lunar variation; McIlwain magnetic coordinate

$M$ Particle mass in atomic mass units; solar "magnetic" region

$M_E$ Mass of earth

$N, N_e$ Electron concentration, alternatively known as electron density ($N_m$—peak electron concentration)

$N_+, N_-$ Positive and negative ion concentrations

$N$ $M/m = 6.0225 \times 10^{26} \text{(kg mole)}^{-1} = 6.0225 \times 10^{23} \text{(g mole)}^{-1}$ (Avogadro’s number)

$P_r, P_t$ Received and transmitted power in incoherent or Thomson scatter

$P_{m}^{in}$ Schmidt function in magnetic potential theory

$Q$ Heat production rate; height-integrated electron production rate

$R$ Gas constant, 8314 J ($\text{K}^{-1}$)$\text{(kg mole)}^{-1}$; mean sunspot number ($R_x$—Zürich; $R_A$—American); reflection coefficient ($R_o$, $R_x$ for o and x modes)

$R_E$ Radius of earth, mean value 6371 km

$R_{mn}$ Radial function in tidal oscillation

$S$ Solar 10 cm flux density; magnetic or ionospheric solar variation ($S_q$—quiet day; $S_D$—disturbance variation)

$T$ Temperature; orbital period; period of oscillation ($T_s$—solar; $T_L$—lunar)

$U$ Neutral air velocity (vector)

$V$ Plasma drift velocity (vector); magnetic potential; vertical component of geomagnetic field ($=|Z|$)

$V_c$ Apparent drift velocity due to random change of moving ionospheric irregularities

$W$ Vertical component of drift velocity ($W_D$—vertical component of plasma diffusion velocity), positive upwards

$X$ Magnetoionic parameter $\omega_n^2/\omega^2$; northward component of geomagnetic field

$Y$ Magnetoionic parameter $\partial H/\partial Y (Y_L, Y_T)$—longitudinal and transverse components of $Y$; eastward component of geomagnetic field

$Z$ Magnetoionic parameter $v/\omega$; downward component of geomagnetic field

$a$ Attachment coefficient; earth radius in spherical harmonic formula; magnetic disturbance amplitude

$b$ Plasma diffusion parameter $nD$

$c$ Velocity of light in free space, 2.9979 $\times 10^8 \text{ m s}^{-1}$

$c_p, c_v$ Specific heat at constant pressure and at constant volume

$d$ Plasma diffusion rate $D/H^2$

$e$ Base of natural logarithms, 2.7183

$e$ Electron charge, $-1.6021 \times 10^{-19} \text{ C}$

$f$ Wave frequency ($f_s, f_x, f_e$—critical frequencies for magneto ionic modes; $f_N$—plasma frequency; $f_0$—value of $f_N$ at space vehicle; $f_{il}$—gyrofrequency; $f_{il}$—longitudinal component of $f_{il}$)

$g$ Acceleration due to gravity (vector), sea level mid-latitude value 9.807 m s$^{-2}$
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{nm}$, $h_{nm}$</td>
<td>Coefficients in spherical harmonic expansion of magnetic potential function</td>
</tr>
<tr>
<td>$h$</td>
<td>Height ($h'$—virtual height; $h^*$—geopotential height or height measured from base of ionosphere)</td>
</tr>
<tr>
<td>$h_{mn}$</td>
<td>Equivalent depth of tidal oscillation</td>
</tr>
<tr>
<td>$i$</td>
<td>Imaginary operator $\sqrt{-1}$; angle of incidence</td>
</tr>
<tr>
<td>$j$</td>
<td>Electric current density (vector)</td>
</tr>
<tr>
<td>$k$</td>
<td>Boltzmann's constant, $1.3805 \times 10^{-23}$ J (°K)$^{-1}$; mobility (tensor)</td>
</tr>
<tr>
<td>$k_x$, $k_z$</td>
<td>Horizontal and vertical wavenumbers</td>
</tr>
<tr>
<td>$l$</td>
<td>Rate of loss of electrons and ions</td>
</tr>
<tr>
<td>$m$</td>
<td>Particle mass in general; electron mass, $9.1091 \times 10^{-31}$ kg (sometimes denoted by $m_e$)</td>
</tr>
<tr>
<td>$m_s$</td>
<td>Satellite mass</td>
</tr>
<tr>
<td>$n$</td>
<td>Neutral gas concentration ($n[X]$—concentration of species $X$); exponent of $\cos \chi$ variation of critical frequencies</td>
</tr>
<tr>
<td>$p$</td>
<td>Neutral gas pressure, or partial pressure</td>
</tr>
<tr>
<td>$q$</td>
<td>Rate of production of electrons and ions</td>
</tr>
<tr>
<td>$r$</td>
<td>Radial geocentric distance; position coordinate (vector)</td>
</tr>
<tr>
<td>$r_e$</td>
<td>Classical electron radius, $2.817 \times 10^{-15}$ m</td>
</tr>
<tr>
<td>$s$</td>
<td>Distance along path, e.g., of radio or light ray (normally occurs as path element, $ds$)</td>
</tr>
<tr>
<td>$t$</td>
<td>Time in general, or local time ($t_z$, $t_y$—time delays in drifts measurements)</td>
</tr>
<tr>
<td>$u_{ab}$</td>
<td>Coefficients specifying the rate of loss of heat from particles $a$ to particles $b$</td>
</tr>
<tr>
<td>$v$</td>
<td>Velocity, e.g., of space vehicle or of cell of air (vector)</td>
</tr>
<tr>
<td>$w$</td>
<td>Vertical velocity, specifically that of ions and electrons due to wind shear (also $w' = \partial w / \partial h$)</td>
</tr>
<tr>
<td>$x$</td>
<td>Cartesian coordinate, positive southwards; ratio of geocentric distance to scale height, $r/H$, in Chapman integral</td>
</tr>
<tr>
<td>$y$</td>
<td>Cartesian coordinate, positive eastwards</td>
</tr>
<tr>
<td>$z$</td>
<td>Reduced height (measured in units of scale height)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Recombination coefficient</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Linear loss coefficient in F region ($\beta'$—effective loss coefficient for night F2 layer)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Ion-atom interchange rate coefficient; ratio of specific heats of gas, $c_p/c_v$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Detachment coefficient</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Charge carried by particle; heat liberated per unit of radiation absorbed</td>
</tr>
<tr>
<td>$\varepsilon_0$</td>
<td>Electric permittivity of free space, $8.8542 \times 10^{-12}$ F m$^{-1}$</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>Reduced height in terms of neutral hydrogen scale height</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Ionizing efficiency; spatial coordinate in autocorrelation function</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Angle between radio wave normal and magnetic field; colatitude; azimuth of drift motion</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Radio wave absorption coefficient ($\kappa_o$, $\kappa_x$ for $o$ and $x$ modes); rate coefficients of detachment reactions; thermal conductivity ($\kappa_{ab}$—related to species $a$, $b$)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Longitude; negative ion ratio $N_-/N_e$; wavelength ($\lambda_m$, $\lambda_o$—radio wavelength in ionospheric medium and in free space)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Coefficient of viscosity of air; phase refractive index ($\mu'$—group refractive index)</td>
</tr>
</tbody>
</table>
LIST OF PRINCIPAL SYMBOLS

$\mu_0$ Magnetic permeability of free space, $1.2566 \times 10^{-6} \text{ H m}^{-1}$

$\nu$ Collision frequency ($\nu_{ab}$—between particles of species $a, b$; $\nu_M$—for mono-
energetic electrons with neutral particles)

$\xi$ Spatial coordinate in autocorrelation function

$\pi$ $3.14159...$

$\rho$ Gas density; autocorrelation function; apparent reflection coefficient of
radio waves; photodetachment coefficient

$\sigma$ Absorption cross section ($\sigma_e$—Thomson cross section for incoherent scatter);
electrical conductivity (tensor)

$\tau$ Optical depth; temperature ratio $T_e/T_i$; electron cooling time in wave
interaction; storm time; time variable in autocorrelation function

$\phi$ Latitude; eclipse obscuration function; radio phase ($\phi_0, \phi_x$—for $o$ and $x$
modes)

$x$ Solar zenith angle; angle between zenith and radio ray path; velocity diver-
gence in tidal function ($x_s$—solar; $x_L$—lunar)

$\omega$ Wave angular frequency ($\omega_N$—plasma angular frequency; $\omega_H$—angular
gyrofrequency; $\omega_L$—longitudinal component of $\omega_H$). Note: angular gyro-
frequency is denoted simply by $\omega$ in the treatment of mobility, conductivity,
etc., where wave frequency is not involved.

$\Gamma$ Scale height gradient $dH/dh$

$\Delta$ Deviation of any quantity from its undisturbed value ($\Delta t$—time lag or
"sluggishness" of ionosphere)

$\Theta_{mn}$ Hough (colatitudinal) function in tidal oscillation

$\Lambda$ Magnetic invariant latitude

$\Sigma$ Height-integrated conductivity (tensor)

$\Phi$ Heat flux (vector); electrostatic potential

$\Psi$ Tidal potential

$\Omega$ Angular velocity of earth (vector); Faraday rotation angle

Commonly Used Subscripts

0, 1, 2, 3 Direct (or longitudinal); Pedersen (or transverse); Hall; Cowling components
of mobility or electrical conductivity

$x, y, z$ Cartesian coordinates (generally south, east, up)

$\xi, \eta, \zeta$ Cartesian coordinates ($\zeta$-axis aligned with magnetic field)

$e, i, n, p$ Electron; ion; neutral gas; plasma (used for temperature, mass, partial
pressure, etc.)

$+, -$ Positive and negative (ions)

$a, b$ Any two species of particle in general

$m, n$ Mode numbers (latitudinal, azimuthal) of tidal functions, or of spherical
harmonic components

$o, x, z$ Ordinary; extraordinary; $z$-mode in magnetoionic theory (critical frequen-
cies, refractive indices, absorption and reflection coefficients, etc.)

0, 1, $m$ Level (in atmosphere) of peak electron production for overhead sun; peak
electron production, in general; peak electron concentration

$i, j, n$ Used generally in connection with summations


BARRINGTON, R. E., BELROSE, J. S., and NELMS, G. L. (1965). Ion composition and temperatures at 1000 km as deduced from simultaneous observations of a VLF plasma resonance and topside sounding data from the Alouette 1 satellite, J. Geophys. Res. 70, 1647-1664.

BARRON, D. W., and BUDDEN, K. G. (1960). The numerical solution of differential equa-


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES 283


DOUGHERTY, J. P. (1961). On the influence of horizontal motion of the neutral air on
REFERENCES


EVANS, J. V. (1965b). Cause of the mid-latitude winter night increase in $f_o F_2$, *J. Geophys. Res.* 70, 4331-4345.


REFERENCES


GARRIOTT, O. K. (1960). The determination of ionospheric electron content and distri-


REFERENCES


REFERENCES

JACCHIA, L. G. (1965). The temperature above the thermopause, in "Space Research"


REFERENCES


KOHL, H., and KING, J. W. (1967). Atmospheric winds between 100 and 700 km and
REFERENCES


London, J. (1967). The average distribution and time variation of ozone in the strato-


Mawdsley, J. (1961). Air density variations in the mesosphere, and the winter anomaly
REFERENCES 297


REFERENCES


REFERENCES


REFERENCES


SATO, T. (1957a). Disturbances in the ionospheric F2 region associated with geomagnetic...
REFERENCES

storms. II. Middle latitudes, J. Geomag. Geoelect. 9, 1–22.


AUTHOR INDEX

Numbers in italics refer to the pages where the complete reference is cited.

A

Adams, G. W., 112, 275
Agy, V., 249, 275
Aikin, A. C., 102, 111, 275, 298
Ainsworth, J., 42, 295
Akasofu, S-I., 230, 249, 261, 275
Allan, R. R., 31, 294
Allan, C. W., 14, 19, 98, 99, 100, 101, 102, 103, 166, 168, 169, 171, 275
Allan, E. F., Jr., 43, 302
Angerarni.T. J., 148, 189, 275
Antonova, L. A., 105, 182, 275
Aono, Y., 81, 275
Appleton, E. V., 48, 64, 65, 151, 162, 163, 166, 183, 193, 266, 275
Arima, Y., 169, 179, 308, 309
Atkinson, G., 261, 276
Axford, W. I., 201, 203, 247, 261, 276, 288

B

Bailey, A. D., 79, 80, 112, 298
Bailey, D. K., 102, 104, 183, 193, 205, 262, 276
Bailey, V. A., 68, 79, 276
Bain, W. C., 41, 63, 161, 276, 279
Baker, D. M., 191, 284
Balsley, B. B., 86, 201, 237, 276, 279
Bandeen, W. R., 27, 41, 305
Banks, P. M., 214, 216, 276
Barbier, D., 124, 125, 276
Barnett, M. A. F., 48, 275
Barrington, R. E., 60, 68, 70, 187, 276, 298
Barron, D. W., 63, 85, 151, 175, 276, 286, 302
Barry, G. H., 51, 287
Bartels, J., 48, 222, 223, 226, 227, 228, 231, 277, 282
Barth, Ch. A., 109, 277
Barthle, R. C., 77, 285
Bartman, F. L., 27, 41, 277, 305
Bateman, R., 205, 276
Bauer, S. J., 72, 81, 82, 157, 158, 163, 187, 277
Baxter, R. G., 185, 277
Bazzard, G. H., 166, 169, 233, 297
Becker, W., 170, 180, 264, 268, 277
Bedinger, J. F., 43, 278
Bellchambers, W. H., 181, 278
Belrose, J. S., 60, 67, 163, 164, 187, 191, 263, 276, 278, 298
Bennett, W. H., 80, 278
Berkner, L. V., 179, 268, 278
Beynon, W. J. G., 163, 165, 166, 206, 278, 303
Bhonsle, R. V., 74, 171, 172, 173, 190, 278
Bibl, K., 200, 278
Biondi, M. A., 105, 109, 276
Blackband, W. T., 74, 278
Blamont, J. E., 21, 278
Block, L. P., 238, 278
Blumle, L. I., 60, 81, 82, 187, 277, 287
Bohme, D. K., 268, 278
Bosolasco, M., 163, 278
Boström, R., 238, 261, 278
Bourdeau, R. E., 80, 213, 278, 279
Bowen, P. J., 98, 187, 279
Bowen, W. A., Jr., 237, 240, 304
Bowhill, S. A., 73, 74, 112, 152, 166, 195, 207, 214, 216, 217, 279, 288, 297
Bowles, K. L., 83, 84, 85, 189, 201, 279, 306

311
Bowling, J. S., 194, 279
Boyd, R. L. F., 187, 279
Brace, L. H., 60, 80, 187, 199, 215, 217, 219, 220, 279, 298, 304, 305
Bracewell, R. N., 63, 161, 191, 279
Bradbury, N. E., 117, 279
Brady, A. H., 264, 279
Bramley, E. N., 212, 280
Brandt, J. C., 24, 123, 280
Branscomb, L. M., 109, 280
Breit, G., 48, 50, 280
Brice, N. M., 60, 248, 249, 251, 280, 289, 290
Briggs, B. H., 154, 206, 207, 208, 210, 211, 240, 280
Brinton, H. C., 187, 305
Broglio, L., 46, 280
Brown, G. M., 166, 264, 266, 278, 280
Brown, J., 20, 303
Browne, I. C., 73, 280
Buckley, R., 81, 280
Budden, K. G., 48, 55, 62, 63, 161, 191, 276, 278, 280
Bull, G. V., 37, 44, 204, 298, 308
Bullen, J. M., 181, 280
Bureau, R., 191, 280
Burgess, B., 75, 280
Burkard, O., 117, 280, 281
Burke, M. J., 67, 278
Burrows, J. R., 250, 296
Burrows, K., 237, 240, 281, 284
Busch, R., 193, 200, 278, 281
Butler, S. T., 40, 281, 304
Byram, E. T., 24, 182, 281, 295

C

Cahill, L. J., Jr., 237, 260, 281, 297
Cahn, J. H., 140, 302
Cain, D. L., 78, 294
Cain, J. C., 225, 281, 290
Calvert, W., 60, 81, 188, 211, 281, 290
Carignan, G. R., 80, 199, 217, 219, 279, 298, 304
Carleton, N. P., 125, 297
Carlson, H. C., 125, 181, 220, 281
Carpenter, D. L., 189, 250, 251, 275, 281
Carru, H., 85, 86, 218, 281
Cetiner, E., 191, 278
Chalmers, J. A., 47, 281
Chamberlain, J. W., 23, 121, 122, 123, 124, 281
Chan, L. K., 191, 192, 281, 292
Chandra, S., 148, 149, 281, 282
Chaney, L. W., 41, 277
Chapman, S., 1, 15, 23, 48, 89, 92, 94, 97, 123, 129, 130, 132, 137, 143, 144, 193, 222, 223, 225, 226, 227, 228, 230, 231, 258, 275, 282, 305
Charney, J. G., 25, 282
Chin, G. Y., 190, 292
Chivers, H. J. A., 210, 282
Chubb, T. A., 24, 98, 182, 281, 295
Chudakov, A. E., 244, 306
Chun-Ming Huang, 169, 282
Clemesha, B. R., 41, 282
Clemmow, P. C., 209, 282
Cohen, R., 201, 211, 279, 281, 300
Cole, K. D., 14, 125, 220, 225, 261, 282, 283
Colegrove, F. D., 19, 283
Collins, C., 262, 301
Cook, G. E., 46, 283
Cook, G. R., 99, 283
Cooper, G., 122, 288
Copsey, M. J., 121, 283
Cowling, T. G., 23, 129, 132, 141, 143, 144, 282, 283
Cox, J. W., 165, 306
Craig, R. A., 25, 27, 37, 40, 283
Crawford, F. W., 81, 283
Crombie, D. D., 264, 279
Croom, S. A., 179, 183, 184, 283
Culhane, J. L., 191, 283
Cummack, C. H., 17, 168, 283

D

Dagg, M., 205, 283
Dalgarno, A., 99, 109, 122, 123, 124, 125, 130, 144, 175, 182, 214, 215, 216, 217, 218, 219, 220, 279, 283, 284
Daniels, W. E., 225, 281
da Rosa, A. V., 74, 76, 171, 172, 173, 192, 193, 216, 217, 219, 278, 284, 288
AUTHOR INDEX 313

Davies, K., 48, 182, 191, 193, 284, 294
Davies, M. J., 76, 192, 193, 288
Davis, M. J., 76, 192, 193, 288
Davis, T. N., 237, 240, 284
Deeks, D. G., 63, 161, 162, 284
de Jager, C., 21, 278
Delinger, J. H., 190, 262, 284
de Mendonca, F., 74, 76, 284, 288
Denison, J. S., 21, 99, 100, 121, 174, 299
Dessler, A. J., 252, 260, 284
Detwiler, C. R., 14, 284
Dieminger, W., 31, 163, 167, 284
Ditchburn, R. W., 99, 284
Donahue, T. M., 24, 119, 284
Donley, J. L., 80, 81, 82, 277, 279
Donnelly, R. F., 191, 284
Dougherty, J. P., 84, 85, 141, 142, 146, 238, 284, 285, 286
Doupnik, J. R., 57, 60, 285
Doyen, G., 207, 290
Drake, F. D., 78, 294
Drazin, P. G., 25, 282
Dufour, S. W., 189, 305
Duncan, R. A., 124, 153, 182, 183, 185, 285
Duncey, J. W., 153, 202, 242, 244, 245, 285
Dyson, P. L., 211, 285

F

Farley, D. T., Jr., 85, 86, 188, 201, 205, 219, 235, 238, 286
Faucher, G. A., 41, 286
Fedor, L. S., 204, 207, 308
Fejes, F., 113, 120, 121, 268, 286, 287, 303
Fejer, J. A., 68, 81, 84, 236, 238, 252, 261, 284, 286, 287, 289, 290
Fel’dshteyn, Ya. L., 249, 287
Fenwick, R. B., 1, 287
Ferguson, E. E., 113, 120, 121, 268, 286, 287, 303
Ferraro, V. C. A., 143, 144, 145, 149, 222, 258, 282, 287
Finch, H. F., 224, 287
Fiocco, G., 41, 287
Fish, R. A., 24, 292
Fitzenreiter, R. J., 60, 81, 82, 277, 287
Fjeldbo, G., 78, 287
Fjeldbo, W. C., 78, 287
Flaherty, B. J., 180, 308
Fooks, G. F., 183, 207, 208, 287, 294
Fooks, J., 73, 289
French, A. G., 169, 181, 287
Friedman, H., 24, 98, 100, 182, 195, 281, 287, 295
Fukushima, N., 261, 298

G

Gallagher, P. B., 77, 285
Gardner, F. F., 66, 287
Garrett, D. L., 14, 284
Garriott, O. K., 10, 71, 74, 76, 77, 156, 170, 171, 172, 173, 192, 193, 267, 278, 284, 287, 288, 295
Gartlein, C. W., 125, 244, 299
Gautier, T. N., 186, 296
Geisler, J. E., 28, 29, 30, 31, 159, 178, 214, 216, 217, 288
Ghosh, S. N., 43, 278
Gibbons, J. J., 63, 288
Gillett, F. C., 122, 288
Gilman, G. L., 263, 303
Glaume, J., 124, 276
Gledhill, J. A., 197, 244, 268, 288
Gleeson, L. J., 203, 288
<table>
<thead>
<tr>
<th>Author Name</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gliddon, J. E. C.</td>
<td>149, 288</td>
</tr>
<tr>
<td>Goe, G. B.</td>
<td>60, 281</td>
</tr>
<tr>
<td>Gold, T.</td>
<td>246, 288</td>
</tr>
<tr>
<td>Golden, P. D.</td>
<td>120, 121, 287</td>
</tr>
<tr>
<td>Goldberg, R. A.</td>
<td>149, 282</td>
</tr>
<tr>
<td>Golomb, D.</td>
<td>43, 302</td>
</tr>
<tr>
<td>Golton, E.</td>
<td>76, 225, 288, 289</td>
</tr>
<tr>
<td>Goodwin, G. L.</td>
<td>206, 278</td>
</tr>
<tr>
<td>Goody, R. M.</td>
<td>13, 298</td>
</tr>
<tr>
<td>Gorchakov, E. V.</td>
<td>244, 306</td>
</tr>
<tr>
<td>Green, A. L.</td>
<td>48, 289</td>
</tr>
<tr>
<td>Green, J. L.</td>
<td>188, 219, 286</td>
</tr>
<tr>
<td>Green, J. S. A.</td>
<td>40, 289</td>
</tr>
<tr>
<td>Greenhow, J. S.</td>
<td>36, 85, 204, 210, 239, 282, 289</td>
</tr>
<tr>
<td>Gregory, J. B.</td>
<td>163, 164, 289</td>
</tr>
<tr>
<td>Gringauz, K.</td>
<td>72, 200, 217, 218, 289</td>
</tr>
<tr>
<td>Gurnett, D. A.</td>
<td>60, 289</td>
</tr>
<tr>
<td>Haerendel, G.</td>
<td>44, 240, 289</td>
</tr>
<tr>
<td>Hagfors, T.</td>
<td>289</td>
</tr>
<tr>
<td>Hagg, E. L.</td>
<td>60, 289</td>
</tr>
<tr>
<td>Hakura, Y.</td>
<td>262, 299</td>
</tr>
<tr>
<td>Hall, J. E.</td>
<td>73, 204, 289</td>
</tr>
<tr>
<td>Hall, L. A.</td>
<td>98, 289, 291</td>
</tr>
<tr>
<td>Hall, S. C.</td>
<td>74, 267, 295</td>
</tr>
<tr>
<td>Hall, S. H.</td>
<td>240, 281</td>
</tr>
<tr>
<td>Hallam, K. L.</td>
<td>191, 289</td>
</tr>
<tr>
<td>Hanson, W. B.</td>
<td>13, 19, 24, 42, 147, 153, 156, 158, 159, 182, 185, 186, 214, 215, 283, 289, 290, 303</td>
</tr>
<tr>
<td>Hargreaves, J. K.</td>
<td>73, 280</td>
</tr>
<tr>
<td>Harmschmacher, E.</td>
<td>193, 281</td>
</tr>
<tr>
<td>Harp, R. S.</td>
<td>81, 283</td>
</tr>
<tr>
<td>Harris, I.</td>
<td>11, 17, 18, 105, 290</td>
</tr>
<tr>
<td>Hertz, T. R.</td>
<td>248, 249, 290, 298</td>
</tr>
<tr>
<td>Harwood, J.</td>
<td>200, 303</td>
</tr>
<tr>
<td>Haselgrove, J.</td>
<td>58, 168, 177, 178, 305</td>
</tr>
<tr>
<td>Hasted, J. B.</td>
<td>108, 109, 268, 278, 297</td>
</tr>
<tr>
<td>Haubert, A.</td>
<td>207, 290</td>
</tr>
<tr>
<td>Heikila, W. J.</td>
<td>81, 290</td>
</tr>
<tr>
<td>Heisler, L. H.</td>
<td>196, 201, 212, 290, 298</td>
</tr>
<tr>
<td>Helm, H.</td>
<td>183, 294</td>
</tr>
<tr>
<td>Hendricks, S. J.</td>
<td>225, 281, 290</td>
</tr>
<tr>
<td>Henry, R. J. U.</td>
<td>99, 284</td>
</tr>
<tr>
<td>Heppner, J. P.</td>
<td>123, 240, 284, 290</td>
</tr>
<tr>
<td>Hess, W. N.</td>
<td>243, 290</td>
</tr>
<tr>
<td>Hewish, A.</td>
<td>207, 209, 290</td>
</tr>
<tr>
<td>Hibberd, F.</td>
<td>267, 290</td>
</tr>
<tr>
<td>Hill, G. E.</td>
<td>183, 290</td>
</tr>
<tr>
<td>Hines, C. O.</td>
<td>14, 18, 22, 25, 26, 28, 31, 33, 34, 38, 40, 140, 202, 212, 233, 238, 244, 245, 246, 247, 276, 290, 291, 300</td>
</tr>
<tr>
<td>Hinteregger, H. E.</td>
<td>14, 21, 98, 172, 289, 291</td>
</tr>
<tr>
<td>Hirao, K.</td>
<td>81, 275</td>
</tr>
<tr>
<td>Hirono, M.</td>
<td>141, 183, 186, 236, 291</td>
</tr>
<tr>
<td>Hirsh, A. J.</td>
<td>114, 115, 291</td>
</tr>
<tr>
<td>Hoffman, J. H.</td>
<td>20, 42, 299</td>
</tr>
<tr>
<td>Holmes, J. C.</td>
<td>20, 42, 118, 119, 204, 291, 292, 299, 309</td>
</tr>
<tr>
<td>Hones, E. W., Jr.</td>
<td>247, 305</td>
</tr>
<tr>
<td>Horowitz, R.</td>
<td>43, 295, 298</td>
</tr>
<tr>
<td>Huch, W. F.</td>
<td>122, 288</td>
</tr>
<tr>
<td>Hugill, J.</td>
<td>81, 290</td>
</tr>
<tr>
<td>Hulburt, E. O.</td>
<td>48, 89, 127, 142, 144, 149, 291, 292</td>
</tr>
<tr>
<td>Hultqvist, B.</td>
<td>111, 291</td>
</tr>
<tr>
<td>Hunt, B. G.</td>
<td>112, 291</td>
</tr>
<tr>
<td>Hunt, D. C.</td>
<td>19, 291</td>
</tr>
<tr>
<td>Huxley, L. G. H.</td>
<td>68, 291</td>
</tr>
<tr>
<td>Hynek, D. P.</td>
<td>85, 297</td>
</tr>
<tr>
<td>Istomin, V. G.</td>
<td>21, 80, 291</td>
</tr>
<tr>
<td>Ivanov-Kholodnii, G. S.</td>
<td>105, 182, 275</td>
</tr>
<tr>
<td>Jacchia, L. G.</td>
<td>18, 45, 180, 253, 291, 292</td>
</tr>
<tr>
<td>Jackson, J. E.</td>
<td>72, 81, 82, 200, 277, 292</td>
</tr>
<tr>
<td>Jayaram, B.</td>
<td>190, 292</td>
</tr>
<tr>
<td>Jean, A. G.</td>
<td>264, 279</td>
</tr>
<tr>
<td>Jenkins, D. C.</td>
<td>225, 281</td>
</tr>
<tr>
<td>Joachim, M.</td>
<td>233, 292</td>
</tr>
<tr>
<td>Johnson, C. Y.</td>
<td>20, 42, 80, 118, 119, 204, 291, 292, 299, 309</td>
</tr>
<tr>
<td>Johnson, F. S.</td>
<td>13, 19, 23, 24, 181, 283, 292</td>
</tr>
<tr>
<td>Johnson, M. A.</td>
<td>209, 292</td>
</tr>
<tr>
<td>Johnson, M. H.</td>
<td>127, 144, 149, 292</td>
</tr>
<tr>
<td>Jones, E. S. O.</td>
<td>165, 278</td>
</tr>
<tr>
<td>Author</td>
<td>Pages</td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td>Jones, I. L.</td>
<td>37, 207, 208, 287, 292</td>
</tr>
<tr>
<td>Jones, L. M.</td>
<td>41, 277</td>
</tr>
<tr>
<td>Joseph, J. H.</td>
<td>24, 292</td>
</tr>
<tr>
<td>Kaiser, T. R.</td>
<td>47, 292</td>
</tr>
<tr>
<td>Kamiyama, H.</td>
<td>102, 292</td>
</tr>
<tr>
<td>Kanellakos, D. P.</td>
<td>192, 292</td>
</tr>
<tr>
<td>Kasuya, I.</td>
<td>168, 293</td>
</tr>
<tr>
<td>Kato, S.</td>
<td>37, 38, 40, 209, 233, 237, 293, 296</td>
</tr>
<tr>
<td>Kellogg, W. W.</td>
<td>14, 293</td>
</tr>
<tr>
<td>Kelso, J. M.</td>
<td>56, 293</td>
</tr>
<tr>
<td>Kendall, P. C.</td>
<td>146, 149, 154, 288, 293</td>
</tr>
<tr>
<td>Kenrick, G. W.</td>
<td>48, 293</td>
</tr>
<tr>
<td>Kent, G. S.</td>
<td>41, 210, 282, 293</td>
</tr>
<tr>
<td>Kern, J. W.</td>
<td>261, 293</td>
</tr>
<tr>
<td>King, G. A. M.</td>
<td>14, 57, 117, 167, 169, 171, 180, 181, 200, 264, 268, 269, 293</td>
</tr>
<tr>
<td>King, J. W.</td>
<td>28, 29, 56, 177, 178, 182, 183, 184, 199, 211, 241, 266, 293, 294</td>
</tr>
<tr>
<td>King-Hele, D. G.</td>
<td>31, 45, 294</td>
</tr>
<tr>
<td>Kirby, R. C.</td>
<td>205, 276</td>
</tr>
<tr>
<td>Kitamura, T.</td>
<td>141, 236, 291</td>
</tr>
<tr>
<td>Kilore, A.</td>
<td>78, 294</td>
</tr>
<tr>
<td>Knecht, R. W.</td>
<td>81, 83, 182, 193, 200, 294, 306</td>
</tr>
<tr>
<td>Knoebel, H. W.</td>
<td>73, 297</td>
</tr>
<tr>
<td>Knof, H.</td>
<td>144, 294</td>
</tr>
<tr>
<td>Knuth, R.</td>
<td>263, 295</td>
</tr>
<tr>
<td>Kochanski, A.</td>
<td>33, 294</td>
</tr>
<tr>
<td>Kockarts, G.</td>
<td>23, 294</td>
</tr>
<tr>
<td>Kohl, H.</td>
<td>28, 29, 142, 156, 159, 177, 178, 182, 294</td>
</tr>
<tr>
<td>Koster, J. R.</td>
<td>210, 295</td>
</tr>
<tr>
<td>Kotadia, K. M.</td>
<td>168, 295</td>
</tr>
<tr>
<td>Krassovsky, V. I.</td>
<td>182, 295</td>
</tr>
<tr>
<td>Kreplin, R. W.</td>
<td>98, 295</td>
</tr>
<tr>
<td>Kupperian, J. E., Jr.</td>
<td>24, 295</td>
</tr>
<tr>
<td>LaGow, H. E.</td>
<td>42, 295</td>
</tr>
<tr>
<td>Lange, I.</td>
<td>226, 307</td>
</tr>
<tr>
<td>Laport, L.</td>
<td>226, 307</td>
</tr>
<tr>
<td>Lauter, E. A.</td>
<td>66, 163, 263, 295</td>
</tr>
<tr>
<td>Lawden, M. D.</td>
<td>117, 169, 171, 180, 181, 293</td>
</tr>
<tr>
<td>Lawrence, R. S.</td>
<td>74, 76, 267, 295</td>
</tr>
<tr>
<td>Layzer, D.</td>
<td>204, 295</td>
</tr>
<tr>
<td>Leaton, B. R.</td>
<td>224, 287</td>
</tr>
<tr>
<td>Legg, A. J.</td>
<td>184, 199, 211, 294</td>
</tr>
<tr>
<td>Leibach, H.</td>
<td>65, 295</td>
</tr>
<tr>
<td>Lenfeld, G. M.</td>
<td>66, 299</td>
</tr>
<tr>
<td>Levy, G. S.</td>
<td>78, 294</td>
</tr>
<tr>
<td>Lindsay, J. C.</td>
<td>98, 191, 295</td>
</tr>
<tr>
<td>Lindzen, R. S.</td>
<td>28, 29, 38, 40, 295</td>
</tr>
<tr>
<td>Little, C. G.</td>
<td>65, 66, 71, 209, 288, 295, 299, 304</td>
</tr>
<tr>
<td>Liu, C. H.</td>
<td>205, 295</td>
</tr>
<tr>
<td>Liu, V. C.</td>
<td>41, 277</td>
</tr>
<tr>
<td>Lockwood, G. E. K.</td>
<td>60, 186, 187, 188, 293, 298</td>
</tr>
<tr>
<td>Loewenthal, M.</td>
<td>85, 286</td>
</tr>
<tr>
<td>Logachev, Yu. I.</td>
<td>244, 306</td>
</tr>
<tr>
<td>London, J.</td>
<td>3, 295</td>
</tr>
<tr>
<td>Long, A. R.</td>
<td>60, 305</td>
</tr>
<tr>
<td>Lovell, A. C.</td>
<td>209, 304</td>
</tr>
<tr>
<td>Lütt, R.</td>
<td>44, 240, 289</td>
</tr>
<tr>
<td>Lusignan, B.</td>
<td>65, 296</td>
</tr>
<tr>
<td>Lyon, A. J.</td>
<td>146, 166, 183, 275, 296</td>
</tr>
<tr>
<td>McAfee, J. R.</td>
<td>24, 284</td>
</tr>
<tr>
<td>McClure, J. P.</td>
<td>188, 219, 286</td>
</tr>
<tr>
<td>McDiarmid, I. B.</td>
<td>60, 250, 296, 298</td>
</tr>
<tr>
<td>MacDonald, W. M.</td>
<td>244, 307</td>
</tr>
<tr>
<td>McDowell, M. R. C.</td>
<td>195, 277</td>
</tr>
<tr>
<td>McElroy, M. B.</td>
<td>121, 130, 214, 215, 216, 217, 218, 219, 283, 284, 307</td>
</tr>
<tr>
<td>Mellwain, C. E.</td>
<td>225, 242, 296</td>
</tr>
<tr>
<td>McKibbin, D. D.</td>
<td>42, 303</td>
</tr>
<tr>
<td>MacLeod, M. A.</td>
<td>204, 296</td>
</tr>
<tr>
<td>McNish, A. G.</td>
<td>186, 296</td>
</tr>
<tr>
<td>Maeda, H.</td>
<td>183, 186, 236, 291, 296</td>
</tr>
<tr>
<td>Maeda, K.</td>
<td>233, 238, 269, 296</td>
</tr>
<tr>
<td>Maehlum, B.</td>
<td>164, 296</td>
</tr>
<tr>
<td>Maier, E. J. R.</td>
<td>220, 303</td>
</tr>
<tr>
<td>Malthotra, P. L.</td>
<td>183, 301</td>
</tr>
<tr>
<td>Mange, P.</td>
<td>19, 22, 147, 296, 299</td>
</tr>
<tr>
<td>Manning, E. R.</td>
<td>43, 278</td>
</tr>
<tr>
<td>Maple, E.</td>
<td>237, 240, 304</td>
</tr>
<tr>
<td>Mariani, F.</td>
<td>143, 296</td>
</tr>
<tr>
<td>Author</td>
<td>Pages</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>Martyn, D. F.</td>
<td>68, 132, 138, 141, 153, 185, 235, 237, 266, 269, 276, 296</td>
</tr>
<tr>
<td>Mason, E. A.</td>
<td>144, 294</td>
</tr>
<tr>
<td>Massey, H. S. W.</td>
<td>105, 114, 277</td>
</tr>
<tr>
<td>Matsushita, S.</td>
<td>199, 265, 266, 296, 304</td>
</tr>
<tr>
<td>Matsuura, N.</td>
<td>268, 296</td>
</tr>
<tr>
<td>Mawdsley, J.</td>
<td>164, 296</td>
</tr>
<tr>
<td>May, B. R.</td>
<td>46, 63, 276, 297</td>
</tr>
<tr>
<td>Mayaud, P.</td>
<td>226, 237, 297</td>
</tr>
<tr>
<td>Maynard, N. C.</td>
<td>237, 297</td>
</tr>
<tr>
<td>Mayr, H. G.</td>
<td>187, 220, 279</td>
</tr>
<tr>
<td>Mead, G. D.</td>
<td>243, 290</td>
</tr>
<tr>
<td>Meadows, E. B.</td>
<td>20, 118, 292, 297</td>
</tr>
<tr>
<td>Mechtly, E. A.</td>
<td>73, 297</td>
</tr>
<tr>
<td>Megill, L. R.</td>
<td>108, 109, 112, 125, 140, 268, 275, 278, 297, 302</td>
</tr>
<tr>
<td>Meredith, L. H.</td>
<td>123, 290</td>
</tr>
<tr>
<td>Metzger, P. H.</td>
<td>99, 283</td>
</tr>
<tr>
<td>Millington, G.</td>
<td>151, 297</td>
</tr>
<tr>
<td>Millman, G. H.</td>
<td>85, 297</td>
</tr>
<tr>
<td>Minnis, C. M.</td>
<td>166, 169, 195, 233, 297</td>
</tr>
<tr>
<td>Mitra, A. P.</td>
<td>65, 161, 190, 297, 303</td>
</tr>
<tr>
<td>Mitra, S. K.</td>
<td>183, 297</td>
</tr>
<tr>
<td>Mitra, S. N.</td>
<td>205, 297</td>
</tr>
<tr>
<td>Miyazaki, S.</td>
<td>81, 275</td>
</tr>
<tr>
<td>Moffett, R. J.</td>
<td>185, 186, 214, 215, 283, 290</td>
</tr>
<tr>
<td>Moorcroft, D. R.</td>
<td>85, 297</td>
</tr>
<tr>
<td>Morrissey, J. F.</td>
<td>41, 286</td>
</tr>
<tr>
<td>Muldrew, D. B.</td>
<td>188, 211, 297</td>
</tr>
<tr>
<td>Munro, G. H.</td>
<td>34, 196, 211, 297, 297</td>
</tr>
<tr>
<td>Murata, H.</td>
<td>238, 296</td>
</tr>
<tr>
<td>Murgatroyd, R. J.</td>
<td>13, 27, 298</td>
</tr>
<tr>
<td>Murphy, A. C.</td>
<td>264, 279</td>
</tr>
<tr>
<td>Murphy, C. H.</td>
<td>37, 44, 204, 298, 308</td>
</tr>
<tr>
<td>Murray, W. A. S.</td>
<td>73, 280</td>
</tr>
<tr>
<td>Nagata, T.</td>
<td>249, 261, 298</td>
</tr>
<tr>
<td>Nagy, A. F.</td>
<td>80, 298</td>
</tr>
<tr>
<td>Nakada, M. P.</td>
<td>243, 290</td>
</tr>
<tr>
<td>Nakamura, T.</td>
<td>124, 298</td>
</tr>
<tr>
<td>Narcisi, R. S.</td>
<td>79, 112, 298</td>
</tr>
<tr>
<td>Nathan, K. V. S. K.</td>
<td>220, 298</td>
</tr>
<tr>
<td>Nelms, G. L.</td>
<td>60, 186, 187, 188, 276, 295, 298</td>
</tr>
<tr>
<td>Nertney, R. J.</td>
<td>63, 288, 298</td>
</tr>
<tr>
<td>Ness, N. F.</td>
<td>253, 307</td>
</tr>
<tr>
<td>Nesterov, V. E.</td>
<td>244, 306</td>
</tr>
<tr>
<td>Neufeld, E. L.</td>
<td>36, 239, 289</td>
</tr>
<tr>
<td>Newman, W. S.</td>
<td>241, 294</td>
</tr>
<tr>
<td>Newton, G. P.</td>
<td>43, 298</td>
</tr>
<tr>
<td>Ney, E. P.</td>
<td>122, 288</td>
</tr>
<tr>
<td>Nicolet, M.</td>
<td>11, 16, 19, 20, 21, 23, 43, 92, 102, 105, 109, 111, 119, 121, 182, 277, 294, 298, 299, 301</td>
</tr>
<tr>
<td>Niemann, H.</td>
<td>199, 304</td>
</tr>
<tr>
<td>Nier, A. O.</td>
<td>20, 42, 299</td>
</tr>
<tr>
<td>Nikol'skiy, A. P.</td>
<td>249, 299</td>
</tr>
<tr>
<td>Nisbet, J. S.</td>
<td>169, 174, 175, 301</td>
</tr>
<tr>
<td>Nishida, A.</td>
<td>251, 299</td>
</tr>
<tr>
<td>Nitzsche, P.</td>
<td>163, 295</td>
</tr>
<tr>
<td>Norberg, W.</td>
<td>27, 41, 305</td>
</tr>
<tr>
<td>Norman, K.</td>
<td>98, 194, 279</td>
</tr>
<tr>
<td>Obayashi, T.</td>
<td>81, 258, 261, 262, 299</td>
</tr>
<tr>
<td>O'Brien, B. J.</td>
<td>125, 243, 244, 299</td>
</tr>
<tr>
<td>Ochs, G. R.</td>
<td>201, 279</td>
</tr>
<tr>
<td>Ogawa, T.</td>
<td>167, 182, 299</td>
</tr>
<tr>
<td>Olatunji, E. O.</td>
<td>184, 294</td>
</tr>
<tr>
<td>Ong, P. P.</td>
<td>268, 278</td>
</tr>
<tr>
<td>Opp, A. G.</td>
<td>243, 303</td>
</tr>
<tr>
<td>Ortenburger, I. B.</td>
<td>157, 158, 290</td>
</tr>
<tr>
<td>Osborne, D. G.</td>
<td>237, 299</td>
</tr>
<tr>
<td>Oya, H.</td>
<td>81, 299</td>
</tr>
<tr>
<td>Pack, J. L.</td>
<td>62, 300</td>
</tr>
<tr>
<td>Paghis, I.</td>
<td>252, 292</td>
</tr>
<tr>
<td>Parker, E. N.</td>
<td>241, 260, 284, 299</td>
</tr>
<tr>
<td>Parthasarathy, R.</td>
<td>66, 299</td>
</tr>
<tr>
<td>Patterson, T. N. L.</td>
<td>23, 24, 153, 156, 158, 159, 182, 277, 290, 299, 299</td>
</tr>
<tr>
<td>Paul, A. K.</td>
<td>58, 299</td>
</tr>
<tr>
<td>Paulson, J. F.</td>
<td>120, 300</td>
</tr>
<tr>
<td>Pawsey, J. L.</td>
<td>66, 287</td>
</tr>
<tr>
<td>Peart, M.</td>
<td>186, 280</td>
</tr>
<tr>
<td>Peterson, A. M.</td>
<td>200, 300</td>
</tr>
<tr>
<td>Peterson, V. L.</td>
<td>124, 300</td>
</tr>
<tr>
<td>Petit, M.</td>
<td>85, 86, 218, 281</td>
</tr>
<tr>
<td>Pfister, W.</td>
<td>200, 300</td>
</tr>
<tr>
<td>Phelps, A. V.</td>
<td>62, 300</td>
</tr>
<tr>
<td>Phillips, G. J.</td>
<td>207, 208, 280, 300</td>
</tr>
</tbody>
</table>
AUTHOR INDEX 317

Pickard, G. W., 48, 293
Pickering, W. M., 146, 149, 293
Piddington, J. H., 159, 242, 249, 252, 300
Pierce, E. T., 161, 300
Piggott, W. R., 55, 62, 63, 64, 65, 67, 70, 162, 163, 182, 183, 266, 275, 278, 300, 306
Pineo, V. C., 85, 297
Pisarenko, N. F., 220, 244, 303, 306
Pitteway, M. L. V., 33, 63, 211, 300
Pokhunkov, A. A., 20, 300
Pomerantz, M. A., 262, 276
Poppoff, I. G., 102, 105, 109, 300, 307
Posakony, D. J., 74, 76, 267, 295
Pounds, K. A., 98, 191, 279, 283
Pratt, D. S., 200, 300
Pratt, R., 182, 294
Priester, W., 10, 17, 18, 43, 105, 290, 298
Pritchard, A. G., 166, 275
Procunier, R. W., 41, 286
Purcell, I. D., 14, 24, 284, 300
Quinn, T. P., 169, 174, 175, 301
Raitt, W. J., 187, 279
Rao, B. R., 208, 301
Rao, G. L. N., 183, 208, 301
Rastogi, R. G., 183, 186, 200, 301
Rawer, K., 55, 193, 200, 208, 211, 278, 281, 300, 301
Reber, C. A., 20, 21, 43, 301
Reddy, B. M., 187, 220, 279
Reed, K. C., 184, 199, 294
Rees, M. H., 102, 122, 301
Reid, G. C., 112, 238, 262, 301
Rieger, E., 44, 240, 289
Rishbeth, H., 10, 140, 151, 154, 155, 156, 170, 171, 174, 175, 177, 178, 179, 180, 188, 280, 288, 301, 302
Roach, F. E., 125, 244, 299, 302
Roach, J. R., 125, 302
Robbins, A. R., 58, 168, 177-179, 183, 184, 264, 283, 305
Roberts, J. A., 210, 308
Robinson, B. J., 165, 166, 167, 302
Roederer, J. G., 244, 302
Roelofs, T. H., 169, 180, 309
Rose, G., 163, 284
Rosenberg, N. W., 22, 37, 43, 302
Ross, W. J., 74, 76, 266, 267, 290, 302
Rothwell, P., 302
Rourke, G. F., 180, 303
Sagalyn, R. C., 79, 167, 302
Sagan, C., 122, 281
Salpeter, E. E., 84, 85, 302
Sandford, M. C. W., 41, 276
Sanford, P. W., 98, 191, 279, 283
Sato, T., 180, 264, 266, 269, 296, 302, 303
Savenko, I. A., 220, 244, 303, 306
Sayers, J., 121, 283
Schaeffer, E. J., 20, 303
Schardt, A. W., 243, 303
Schiff, H. I., 113, 286
Schmeltekopf, A. L., 113, 120, 121, 268, 286, 287, 303
Schmerling, E. R., 56, 57, 58, 60, 143, 169, 174, 285, 301, 303
Schmid, C. W., 211, 281
Schmidtke, G., 98, 291
Schminder, R., 206, 304
Schweizer, W., 98, 289
Schwentek, H., 66, 163, 284, 303
Scott, W. E., 226, 307
Seaton, M. J., 220, 269, 298, 303
Seaton, S. L., 179, 268, 278
Sechrist, C. J., 109, 164, 303
Seddon, J. C., 71, 72, 200, 292, 303
Sen, H. K., 62, 303
Serbu, G. P., 220, 303
Setty, C. S. G. K., 143, 169, 170, 174, 180, 301, 302
Shain, C. A., 65, 190, 297, 303
Shapley, A. H., 163, 182, 183, 300, 303
Sharp, G. W., 42, 188, 303
Shavrin, P. I., 220, 244, 303, 306
Shawhan, S. D., 60, 289
Shearman, E. D. R., 200, 303
Shefov, N. N., 182, 295
Sherman, F. S., 41, 286
Shimazaki, T., 117, 120, 145, 150, 169, 175, 208, 211, 240, 303, 304
Shinn, D. E., 207, 208, 280
Silberstein, R., 48, 304
Singer, S. F., 237, 240, 260, 304
Singleton, D. G., 211, 304
Skinner, N. I., 201, 304
Slowey, I., 18, 253, 292
Small, K. A, 40, 281, 304
Smith, C. R, 187, 305
Smith, D., 121, 283
Smith, E. K., 199, 304, 305
Smith, P. G., 209, 304
Smith, F. L., 76, 77, 170, 288
Smith, L. G., 73, 112, 279, 297
Smith, P. A, 183, 211, 294
Smith, R. L., 60, 189, 250, 281, 289
Smullin, L. D., 41, 287
Spencer, M., 206, 207, 208, 240, 280, 300
Spencer, N. W., 43, 80, 199, 215, 217, 219, 220, 279, 304
Speiser, T. W., 261, 304
Spitzer, L., 216, 304
Sprenger, K., 206, 304
Stark, C. N., 41, 286
Steele, F. K., 264, 279
Steiger, W. R., 124, 300
Sterling, D. L., 188, 219, 286
Sterne, T. E., 45, 304
Stewart, A. L., 99, 284
Störmer, C., 249, 304
Stolarik, J. D., 237, 240, 284
Stonehocker, G. H., 171, 174, 197, 306
Straker, T. W., 63, 161, 191, 279
Stroud, W. G., 27, 41, 305
Stubbbe, P., 179, 305
Suchy, K., 200, 278
Sugiura, M., 230, 231, 305
Sullivan, W. P., 167, 302
Sutcliffe, H. K., 85, 289
Sutton, O. G., 26, 305
Swenson, G. W., 74, 210, 308
Swider, W., Jr., 105, 119, 121, 167, 182, 299, 305
T
Taeusch, D. R., 199, 304
Takahashi, H., 117, 120, 174, 309
Taubenheim, J., 191, 305
Taylor, G. N., 169, 181, 267, 268, 286, 305
Taylor, H. A., Jr., 187, 305
Taylor, H. E., 247, 305
Thomas, G. E., 24, 284
Thomas, G. R., 156, 177, 305
Thomas, J. A., 199, 305
Thomas, J. O., 55, 56, 58, 60, 143, 148, 168, 169, 174, 177, 178, 179, 183, 184, 189, 264, 275, 283, 301, 305
Thomas, L., 121, 153, 163, 183, 200, 249, 266, 296, 305, 306
Thome, G. D., 212, 306
Thrane, E. V., 60, 62, 63, 67, 68, 70, 276, 300, 306
Titheridge, J. E., 57, 58, 166, 169, 179, 180, 212, 306
Titus, P., 27, 41, 305
Tohmatsu, T., 167, 182, 299
Tolstoy, I., 34, 306
Torr, D. G., 244, 268, 288
Torr, M. R., 244, 268, 288
Tousey, R., 14, 24, 284, 300
Townsend, J. W., 20, 297
Tremellen, K. W., 165, 306
Troitskaya, V. A., 253, 306
Truttse, Yu. L., 182, 295
Tuve, M. A., 48, 50, 280
Tveten, L. H., 212, 306
U
Ulwick, J. C., 200, 300
V
Van Allen, J. A., 125, 242, 244, 299, 306
Vanderslice, J. T., 144, 294
Van Rooyen, H. O., 244, 288
Van Sabben, D., 238, 306
Vasseur, G., 85, 86, 218, 281
Venables, F. H., 156, 177, 266, 305, 306
Verniani, F., 253, 292
Vernov, S. N., 244, 306
Vestine, E. H., 226, 228, 307
Vice, R. W., 68, 287
Villard, O. G., Jr., 76, 191, 192, 193, 281, 288, 292
Volland, H., 18, 28, 307

W
Wakai, N., 167, 307
Waldteufel, P., 85, 86, 218, 281
Walker, J. C. G., 121, 123, 124, 130, 144, 215, 216, 217, 218, 283, 284, 307
Walt, M., 244, 307
Warneck, P., 121, 307
Warren, E. S., 59, 307
Watanabe, K., 21, 98, 291
Watkins, C. D., 85, 289
Watson-Watt, R. A., 48, 307
Watt, T. M., 187, 307
Webber, W., 103, 307
Weekes, K., 8, 39, 63, 135, 144, 161, 167, 191, 279, 301, 307
Wells, H. W., 179, 268, 278
Westover, D., 60, 305
Whitehead, J. D., 201, 212, 290, 307
Whitten, R. C., 102, 105, 109, 300, 307

Wickersham, A. F., Jr., 212, 307
Widdel, H. U., 163, 284
Wilcox, J. M., 253, 307
Wild, J. P., 193, 210, 307, 308
Wilkes, M. V., 25, 34, 35, 37, 39, 94, 96, 307, 308
Wilkins, E. M., 13, 292
Willmore, A. P., 98, 187, 191, 194, 218, 220, 279, 283, 308
Wright, J. W., 44, 58, 170, 177, 180, 204, 207, 298, 299, 308
Wright, R. W. H., 41, 201, 210, 282, 295, 304
Wyller, A. A., 62, 303
Wynne, R., 264, 266, 280

Y
Yeh, K. C., 74, 180, 205, 210, 295, 308
Yerg, D. G., 208, 308
Yonezawa, T., 11, 98, 117, 120, 143, 151, 169, 174, 179, 182, 308, 309
Young, C., 14, 181, 309
Young, J. M., 118, 119, 204, 291, 309
Yuen, P. C., 76, 77, 169, 170, 180, 288, 309

Z
Zimmerman, S. P., 33, 309
SUBJECT INDEX

A
Absorption of radio waves, 53, 63–66, 162, 163, 191, 192, 254, 262, 263
A1 method (pulse absorption), 64, 162, 163
A2 method (cosmic noise), 64–66
A3 method (c.w. field strength), 64, 66, 163
deviative, 64
nondeviative, 50, 64
solar-cycle variation, 162, 163
Absorption coefficient, for radio waves, 50, 64, 67
Absorption cross section, for ionizing radiation, 89, 90, 98, 99
Acceleration of air by ionization, 139–142, see also Ion-drag
Acoustic waves, 33
Adiabatic invariant, 244
Aftereffect, D region, 257, 262–264
Air-drag, 141, 142, 150, see also Ion-drag
Airglow, 3, 12, 89, 105, 121–125, 182, 220, 244
dayglow, 123, 124
oxygen green line, 122, 123
oxygen red line, 122–125, 220
relation to F region, 124, 125
rocket and spacecraft observations, 121–123
stable mid-latitude red arc, 125, 244, 256
tropical arc, 124
Ambipolar diffusion, see Plasma diffusion
Appleton–Hartree equation, 49, 61, 62, 63
Artificial clouds, 31, 36, 43, 44, 240
Artificial heating of ionosphere, 86
Artificial satellites, see Satellites
Associative detachment, 107, 108, 112
Atmosphere, rotation of, 31
Atmospheric composition, 19–22, 42, 43
variations of, relation to ionosphere, 164, 180, 181, 269
Atmospheric dynamics, 25–34
Atmospheric nomenclature, 1–5
Atmospheric tides, 25, 28, 34–40, 233
Atmospheric waves, 25, 31–34, 269
Attachment, 88, 107, 110, 111, 204
radiative, 107, 108
three-body, 107, 108
Aurora, 200, 249, 253, 256
Auroral absorption, 160, 249, 262
Auroral electrojet, 255, 260, 261
Auroral oval, 248–251, 256, 262
Auroral particles, 241, 242, 243, 247, 248
Auroral zone, 14, 248, 249, 262
B
Backscatter, 200, 212 see also Incoherent scatter
Balloons, 2, 40
Barbier formula, 124
Barium ion clouds, 44, 240
Barometric equation, 5, 23, see also Hydrostatic equation
Base level of ionization, in F region, 181, 182
Blackouts, 160, 249, 257, 261, 262
Blanketing by sporadic E, 53, 200
Bow wave of magnetosphere, 242
C
C layer, 48, 161
Chapman layer, 92–94, 153, 165, 166, 168
Chapman’s grazing incidence function, 15, 94–97
Chapman’s production function, 92–94
Character figure, magnetic, 232
Characteristic ellipse, 208
Charge transfer, 112
Charged particles, motions in electric and magnetic field, 132-136, 242-246
Chemical transport, 13, 14, 28
Chemiluminescent trails, 37, 43, 44
Chirp sounder, 51
Chromosphere, 190, 254
Circulation of magnetic field lines, 245-247, 251, 256
Collision frequency, 49, 61, 62, 70, 129-132
effective, 62, 129, 130, 132
of monoenergetic electrons, 62, 67, 70
Collisional deactivation, 123, 124
Collisional detachment, 107, 108, 112, 161
Composition, see Atmospheric composition, Ion composition
Conducting layer, 47
Conduction, see Heat conduction
Conductivity
electrical, 128, 132, 136-139, 191, 233-239, 261
components of tensor, 136
layer (integrated), 138, 139, 234, 235, 239
thermal, 3, 16, 33, 216, 217
Conjugate points
electrical coupling of, 234, 238
photoelectron travel between, 124, 125, 181, 214, 220
thermal coupling of, 177
Contaminants, 28, 36, 43, 44
Continuity equation, 31, 37, 88, 110, 113, 126, 127, 145, 149-151, 169, 170, 193, 194, 195, 237, 269, 270
Conversion coefficient, for long waves, 63
Coriolis force, 26, 27, 30, 32
Corpuscular flux, 258, 259
Corpuscular heat source, 18
Corpuscular ionization, 87, 88, 102-105, 106, 122, 182, 183
Cosmic radio noise, 65
Cosmic rays, 102-104, 161, 193, 253, 255
Cospar International Reference Atmosphere (CIRA), 11, 12
Cowling conductivity, 136, 137, 139, 237
Critical frequency
defined, 53
eclipse variations of, 197
ionograms, seen on, 52-54, 59
normal ionospheric layers, 165, 166, 168, 175, 176
storm variations of, 264-266
Critical level, 158
Crochet, 191, 254
Cross-modulation, 68-70, 83, see also Wave interaction
Current systems, see Electric currents
Cusp, on ionogram, 53, 54, 115, 166, 167, 211, 264

D

D region, 3, 48, 160-165
absorption of radio waves, 60-66, 162, 163
aftereffect of storms, 257, 262-264
chemical processes, 105-113
continuity equations, 110, 111
disturbances, 160, 257, 261-264
eclipse effects, 193, 194
electron distributions, 83, 161, 162, 164
experimental techniques for studying, 60-70, 72, 73, 81, 82, 83
ion composition, 79, 80, 111-113
irregularities, 205, 206
long wave reflection, 62, 63, 161, 262, 263
negative ions, 106, 107, 110-113
normal behavior, 161-163
production, 102, 103
relation to lower atmosphere, 163-165
solar cycle variation, 161-163
solar flare effects, 190-192, 254, 262
storm effects, 257, 261-264
winter anomaly, 160, 163, 164
Day equilibrium layer, 151-154, 175
Declination effect, F region, 177
Detachment, 107-112
associative, 107, 108, 112
collisional, 107, 108, 112, 161
photodetachment, 88, 107, 108, 110, 112
Deviative absorption, 64
Differential absorption, 72, 73
Differential Doppler technique, 72
Diffusion of ionization, 128, 142–149,
see also Plasma diffusion
between ionosphere and protonosphere, 128, 157–159, 181
during eclipse, 197, 198
equatorial ionosphere, 183, 185, 186
formation of F2 peak, 117, 143, 151–155
nighttime ionosphere, 121
Diffusion equations, 143–149, 151–155
Diffusion coefficient, 144, 169, 171, 173, 175
temperature dependence of, 144, 173, 269
Diffusion rate, 154
Diffusive barrier, protonosphere, 158, 159
Diffusive equilibrium, see also Diffusive
separation
attainment of, 22
of light gases, 23, 24
of plasma, 146–149, 152, 154, 158, 187, 188
Diffusive separation, 4, 6, 19–22
Direct conductivity, 136–138
Direct measurements, in ionosphere, 78–82
comparison with other measurements, 81, 82, 83
Direct mobility, 134
Dispersion equation for atmospheric
waves, 32, 33
Dissociation of molecular gases, 3, 12, 14, 19–21
Dissociative recombination, 88, 105, 106, 108, 113, 114, 120, 121, 123, 124, 204, see also Recombination coefficients
Disturbance variations of geomagnetic
field, 229–231, 258–261
Diurnal bulge, 18, 28
Doppler shift
Lyman alpha, 24
radio waves, 70–72, 74–76, 86
Drag
on artificial satellites, 18, 42, 45–46, 253, 257
on sphere, 41
Drag coefficient, 41, 45
Drift, see also Drift velocity, Dynamo
type, Electromagnetic drift
due to winds, 128, 135, 150, 177, 178, 182, 186, 234
effect in E layer, 237, 264
in F2 layer, 151–157, 269, 270
vertical, importance of, 126, 127
Drifts of ionospheric irregularities, 36, 37, 237
observed by Mitra method, 205–209
radio-star method, 209, 210
relation to winds, 207, 209
Drift velocity of plasma, 126–128, 135, 136, 150, 234, 235, 239, 240
Dumbbell electrostatic probe, 80
Dynamo theory, 34, 47, 233–240, 261, 269
coupling of E and F regions, 235
electric fields produced by dynamo ac-
tion, 128, 186, 221, 233–235, 261

E
E region (layer), 3, 48, 165–167
chemical processes, 105–109, 113, 117–121
conductivity, electrical, 132–139, 233–236, 239
continuity equation, 113, 114, 150, 151
critical frequency, 52–54, 165–167
drifts, 206, 209, 237, 240
eclipse effects, 193–195, 197
electrical coupling to F region, 235, 246, 247
electromagnetic movements, 166, 237, 264
height, 166, 178
ion composition, 79, 80, 85, 118–121, 204
irregularities, 205, 209, 210
night, 167, 200, 264
normal behavior, 165–167
production, 100–102
solar cycle variation, 166
solar flare effects, 190–192
storm effects, 264
transport processes, 127, 132, 201–204
E2 critical frequency, 167
Es, see Sporadic E
SUBJECT INDEX 323

Earth currents, 227, 247

Eclipses, solar
- effects on ionosphere, 116, 193–199
- production and loss rates determined from, 166, 167, 170, 194–198
- temperature variations during, 198, 199, 219

Eddy transport, 3, 12, 13

Effective collision frequency, 62, 129, 130, 132

Effective decay coefficient, 153

Effective recombination coefficient, 111, 120, 195

Electric currents, see also Dynamo theory
dynamo generation of, 233–236
- F region, 148, 149
- field-aligned, in magnetosphere, 238
- joule heating by, 12, 14
- quiet day systems, in ionosphere, 166, 227, 228, 246
- rocket measurements of, 240
- solar flare effect, 191, 254
- storm systems, 253, 258–261

Electric fields
- airglow excitation by, 121, 125, 256
diffusive equilibrium of plasma, 146–149
- irregular, 205
- magnetospheric, 238, 244–246
- measurements of, 44, 240
- motions due to, 127, 128, 132–136, 141, 142, 150–154, 156, 157, 185, 186
- produced by dynamo action, 34, 47, 186, 233–235, 239, 261

Electrical conductivity of ionosphere, see Conductivity

Electrodynamic drift, see Electromagnetic drift

Electrojet
- auroral, 255, 259, 260, 261
equatorial, 86, 139, 201, 237, 238, 240

Electromagnetic movements, see Drift,
- Drift velocity, Dynamo theory, Electromagnetic drift

Electron concentration, see also D, E, F regions
determination from ionograms, 50–60
different measurements compared, 81, 82, 83
direct measurements by rockets and satellites, 78–82
-ground-based measurements, for lower ionosphere, 60–70
-incoherent scatter measurements, 82–86
-propagation experiments for measuring, 70–78

Electron cooling parameter, 69, 70

Electron cooling rate, 214, 215

Electron density, see Electron concentration

Electron gyrofrequency, see Gyrofrequency

Electron hybrid resonance, 60

Electron loss rates, 169–171, 173, 174

Electron production rates, 99–105, 170–172, see also Production function

Electron temperature, 213–220
eclipse variations, 198, 199, 219
equations relating to, 213–217
-excess over gas temperature, 13, 213
-measurement of, 79–81, 84, 85
-phenomena related to, 177, 217–220
-relation to plasma diffusion, 143, 144, 147, 149
-storm variations, 218, 268

Electrostatic analyzer, 79, 80

Electrostatic fields, 234, 235, 245, 269, see also Polarization fields

Equation of balance, see Continuity equation

Equation of motion
-of neutral gas, 26, 28, 29, 31, 139, 140
-of plasma, 126–129, 132

Equatorial electrojet, 86, 139, 201, 237, 238, 240

Equatorial F2 layer, 183–186

Equilibrium layer, 151–154

Equivalent depth, 38

Equivalent height, see Reflection height

Es, see Sporadic E
Escape temperature, 23
EUV, see Ultraviolet
Excitation of atoms and molecules, 123, 124, 215
Exosphere, 3, 22–24, 29, 30
Exospheric temperature, 18
Extraordinary mode, 49, 50
Extraordinary wave, ionograms, 52–54, 58
Extreme ultraviolet, see Ultraviolet

F

F region (layer), 3, 48, 168–186
airglow, in relation to, 123–125, 182
anomalies, 175–186
chemical processes, 105–109, 113, 114, 119–121
diffusion in, 141, 142–149, 151–157
drifts, 206, 209, 234, 235, 239, 240
electric fields associated with diffusion, 146–149
electrical coupling to E region, 235, 246, 247
ion composition, 84, 85, 113, 114, 118–121
irregularities, 205, 206, 209–212
linkages with magnetosphere, 145, 188, 238, 246, 247
with protonosphere, 158, 159, 181, 217, 220
production, 100–102, 170–172, 197, 198
rates, 168–175
splitting into F1 and F2 layers, theory of, 114–117
temperatures of ions and electrons, 80, 84, 85, 213–220
transport processes, 127, 128, 139–149, 203
values of parameters, 171
F1 layer, 3, 48, 165, 167, 168, 171
continuity equation, 113, 114
critical frequency, 52, 53, 168
eclipse effects, 116, 193–197
gemagnetic control, 168
height, 168, 178
normal behavior of, 167, 168
solar cycle variation, 167
solar flare effects, 190, 192
storm effects, 264, 266, 269
F1½ layer, 168, 196
F1–F2 transition region, shape of, 114–117, 169, 180
F2 layer, 3, 48, 168–186
anomalies, 175–186
continuity equation, 149–157
critical frequency, 52–54, 59, 175, 176
diurnal variations, 177–179
eclipse effects, 170, 197–199
effect of temperature changes, 155–156
of vertical drifts, 151–153, 155–157, 269, 270
equatorial, 183–186
formation of peak, 117, 118, 143, 151–154
height, 177–179, 264, 268
high latitude, 182, 183
nighttime, 181, 182
solar cycle variation, 169, 171
solar flare effects, 192, 193
storm effects, 257, 264–270
sunrise behavior, 170, 172, 180
thickness, 76, 170
traveling disturbances, 211, 212
Fading method, see Drifts, Mitra method
Faraday rotation, 50, 71, 73–76, 85, 212
First Negative bands, 122
Flare, see Solar flare
Following motion, differentiation, 26, 150, 156
Forbidden lines, 122–125
Forbush cosmic-ray decrease, 255
Force diagram, 148, 149
Fountain effect, 185
Free path theory, 129
Freezing-in, 245–247
Frictional frequency, 129
Full correlation analysis, 208
Full wave analysis, 62, 63, 161

G

Geomagnetic field, 5, 221–251, see also
Magnetosphere, Storms
activity, 231, 232
components (elements) of, 223
coordinates defined, 225
current sheet representation, 227, 228
dip angle, 127, 223, 225
dipole approximation, 223, 224
disturbance variations, 229–231
effect on winds, 26, 29
ground current contribution, 227
heat current influenced by, 177, 216, 219
hydromagnetic motions in, 244–247
latitudes, 225
mathematical representation of, 222–225
motions of ionization controlled by, 127, 132–136, 141, 142, 145, 146, 183, 185, 186, 188, 201–203
potential function, 223, 224
quiet day variations, 225–228, 233, 239, 240
sporadic E connected with, 201, 202
trapping of particles in, 242–244
Geomagnetic storms, 252–270, see also
  Storms
Geomagnetic tail, 242, 255, 256, 258, 261
Geopotential height, 7, 8, 148, 188
Geostationary satellites, 71, 76, 77, 170, 179, 193, 212
Geostrophic winds, 27, 31
Gravitational tides, 35, 236
Gravity waves, 12, 14, 18, 22, 25, 31–34, 37, 39, 201, 205, 212
Grazing-incidence function, 15, 94–97
Grenade, 41
Ground currents, see Earth currents
Gun launchings, 43, 44
Gyrofrequency, 49, 130–132

H
Ha line, 190, 192, 193, 249, 254
h′(f) curves, 51–55, 115, 169, 264
Hall conductivity, 136
Hall current, 137, 142
Hall drift, 135, 202
Hall mobility, 134
Heat balance
  of electrons and ions, 213–217
  of neutral gas, 11–19, 28
Heat conduction, 3, 11, 15–19, 216, 217, 219
Heat input to electron gas, 213, 214
Heat loss, 12, 214–216
Heat production, 11–16
Heat transport, 12–14
Heating during magnetic disturbance, 14, 18, 43, 253, 257
Heliosphere, 3, 187
Helium, in atmosphere, 3, 23, 24, 81, 148, 157, 159, 187, 188
Helium lines, He I, He II, 99, 100, 167
Historical notes, 2, 34, 35, 47, 48, 222
Hybrid Doppler–Faraday method, 75, 76
Hybrid resonance, 60
Hydrogen, in atmosphere, 3, 23–24, 81, 122, 148, 157–159, 187, 188, 220
Hydromagnetic motions, 244–247
Hydromagnetic waves, 12, 242, 257, 258
Hydromagnetic storms, 252–270, see also
  Storms
Integrated conductivity, see Conductivity
Interchange reactions, 106–109, 119–121
Internal wave, 32
International Disturbed Days, 232
International Quiet Days, 232
Ion-atom interchange, 106, 108, 113, 115, 120, 121, 268
Ion composition
  D region, 111–113
  E and F regions, 113–114, 118–121
  measurements, 60, 79–81, 84, 85
  protonosphere, 157–159
  relation to diffusive equilibrium, 147, 148, 158
  topside ionosphere, 187, 188
Ion-drag, 26, 128, 139–142
  effect on air motions, 26, 30, 31, 33, 39
  limiting of vertical plasma drift, 141, 142, 269
Ion-ion (ionic) recombination, 105, 106, 110
Ion temperature, 213–220
  compared to electron and gas temperature, 217–220
  eclipse variations, 198
equations relating to, 215, 216
measurement of, 79, 81, 84, 85
relation to plasma diffusion, 143, 144, 147
Ion trap, 80, 82
Ionization limits for atmospheric gases, 14, 99, 100
Ionization rate coefficient, 89
Ionizing efficiency, 89, 98
Ionograms, 51–60
Ionosonde, 50, 51, 58, 59
Ionosphere, see also D, E, F regions
defined, 3
discovery of, 47, 48
Martian, 71, 77, 78
named, 48
regions (layers) of, 48
Ionospheric currents, see Dynamo theory,
Electric currents
Ionospheric index, \( I_{Fs} \), 169, 233
Ionospheric irregularities, 36, 204–212
  cause of, 205
correlation functions, 208, 209
random changes of, 208, 209
shape and size, 209, 212
Irregularities, small-scale, see Drifts,
Ionospheric irregularities
Isopycnic level, 9, 10

J
Joule heating, 12, 14
Joule loss, 33, 246

K
Kinematic viscosity, 29, see also Viscosity
Knee, see Plasmapause

L
Laboratory measurements of reaction rates, 109, 120, 121, 168, 174
Lamination method of ionogram reduction, 55–57, 60
Langmuir probe, 79, 80
Lasers, 41
Latitude, related to geomagnetic field, 225
Layer conductivity, 138, 139, 234
Limiting thermospheric temperature, 17, 18
Linear loss coefficient, 114–115, 120,
  169–171, 173–175, 269
eclipse determination of, 170, 197, 198
temperature dependence of, 173–175, 268
Local heating, 214
Long waves, 61–63, 161, 162, 191, see also VLF radio waves
Longitudinal conductivity, 136, see also
  Direct conductivity
Longitudinal mobility, 134, 135
Loss
  of heat, 12, 214–216
  of ionization, see Loss processes
Loss coefficient, F region, see Linear loss coefficient
Loss formulas, linear and square law,
  113, 114, 151, 168
Loss processes, 88, 105–109
Lower ionosphere, see D region, E region
Lunar magnetic variations, 34
Lunar radar, see Moon echoes
Lunar tides, 186, 236
Lyman \( \alpha \), 19, 24
  D region production, 102, 103, 106,
  109, 161, 194
  nighttime, 24, 167
  solar flares, 191
Lyman \( \beta \), 99, 100, 106, 167
Lyman \( \gamma \), 99, 100

M
M region, 252
Magnetic bays, 249, 255, 259, 260
Magnetic disturbances, see Storms
Magnetic elements, see Geomagnetic field components
Magnetic field, 5, 221–251, see also Geomagnetic field
Magnetic field lines, in magnetosphere,
  238, 241, 242, 244–247
Magnetic shell, 225, 244
Magnetic storms, 229–231, 252–270, see also Storms
Magnetic variations
  disturbances, 229–231
  quiet days, 225–228, 240
Magnetoionic theory, 48–50
  generalized, 62, 67
Magnetopause, 5, 17, 241, 242, 258
Magnetosphere, 5, 241–251, 255, 256, 261
  relation to ionosphere, 145, 188, 189, 247–251
Main phase of magnetic storm, 230, 255, 259
Manned spacecraft, 122
Mars, 71, 77, 78
Martyn's theorem, 62, 66
Mass spectrometer
  results obtained from, 20, 21, 112, 113, 118, 119, 187, 204
  types of, 42, 43, 79, 80
Matrix method of ionogram reduction, 56, 57
Maxwellian velocity distribution, 23, 213
Mesopause, 3, 13, 14, 19, 27, 28
Mesosphere, 3, 13, 27, 28, 31, 164, 181
Metallic ions, 112, 204
Metastable states, 107, 108, 110, 112
Meteor trails
  decay of, 204
  winds deduced from, 28, 31, 36, 37, 236, 237, 239, 240
Micropulsations, 249, 253, 256
Mid-latitude red arc, see Airglow
Mitra method, 205–209, see also Drifts
Mixing, of atmospheric gases, 19, 22, 269
Mobility, 133–135
  components of, 134
Model atmosphere, 10, 11
Molecular conduction, 12, 13, 16, 18
Molecular diffusion, 22, 24
Molecular viscosity, 12, 26, 33, see also Viscosity
Moon echoes, 71, 73, 74, 170, 266, 267
N
  \(N(h)\) profiles, 55–60, 82, 83, 161, 162, 169, 264
  Negative ions, 88, 106–108, 110–113
Neutral atmosphere, 1–46
  dynamics of, 25–40
  experiments for studying, 40–46
  heat balance of, 11–19
  heating during magnetic disturbance, 14, 18, 43, 253, 257
  structure of, 5–11, 19–24
  variations of temperature, 12, 18
Neutral points, 241, 242
Neutral sheet, 241, 242
Night E layer, 167, 200, 264
Night F2 layer, 181, 182
Night stationary layer, 153–155, 270
Nitric oxide, role in D region, 102, 106, 109, 110, 164
Noctilucent clouds, 31
Nondeviative absorption, 50, 64
Nonlocal heating, 214
Nuclear explosions, 44, 264
O
  OI, see Airglow, Oxygen red line
  Optical depth, 15, 16, 17, 90, 91
    level of unit optical depth, 91, 98, 99
  Ordinary mode, 49, 50
  Ordinary wave, ionograms, 51–54
  Ozone, 2–4, 13, 40, 110, 112
  Ozonosphere, 3
P
  Partial reflections, 61, 66–68, 70, 83
  Particles, see also Corpuscular ionization, Trapped particles
    airglow and, 121, 122, 182
    depth of penetration into atmosphere, 102, 104
    heating due to, 12, 220
    ionization produced by, 87, 102–105, 257, 261, 262, 263
    ionospheric effects of, 164, 193, 268
    precipitation patterns, 248–250, 262
    Peak production, height of, 91, 92, 94, 166, 168
    Pedersen conductivity, 136, 137
    Pedersen mobility, 134
    Penetration frequency, see Critical frequency
Perfect gas law, 5
Photochemical equilibrium, 88, 94, 150, 151, 152
Photochemical processes, 87–125
Photochemical regime, 88, 119, 165
Photodetachment, 88, 107, 108, 110, 112
Photodissociation, 12, 14, see also Dissociation
Photoelectrons, 16, 213, 214, 220
travel along magnetic field lines, 124, 125, 181, 214, 220
Photoionization, 11, 12, 87, 89–102, 106, 120, 214, see also Production rate
Planetary fly-bys, 77, 78
Planetary magnetic indices, 232
Planetary radar, 84
Planetary waves, 25
Plasma diffusion, 128, 141–157, 203, 204
Plasma diffusion coefficient, 144
Plasma drift velocity, see Drift velocity
Plasma frequency, 49, 54, 60
Plasma instability, 201, 205, 238, 261
Plasma scale height, 143, 147, 187, 188
Plasma temperature, 143, 147
Plasmapause, 189, 250, 251, 256
Polar cap absorption (PCA), 65, 66, 112, 160, 254, 262
Polar current systems, 258–261
Polar F layer, 182, 183
Polar substorms, 249, 255, 260, 261
Polarization, of radio waves, 50, 63
Polarization charges, in ionosphere, 127, 138, 233, 234
Polarization fields, 149, 234, 235
see also Electrostatic fields
Polynomial method of ionogram reduction, 55, 57, 60, 66
Positive ions, see also Ion composition atomic/molecular ratio, 114
D region, 112, 113
Pressure gauges, 42
Pressure oscillations, 35, 236, 239
Probes, see Direct measurements, Space probes
Production
of heat, 11–16, 214, see also Heating of ionization, 87, 89–106
Production-conduction model, of thermosphere, 15, 18, 19
Production function, for heating, 15, 16, 214
for ionization, 89–98
Production-loss-diffusion model, of F region, 154, 155
Production rates
D region, 102, 103
E region, 100–102
F region, 100–102, 170–172, 197, 198
Profiles, defined, 1
Propagation experiments using rockets and satellites, 70–78, 82, 83
Protonosphere, 3, 128, 157–159, 187
relation to ionosphere, 157–159, 217, 220
Pulse sounding, 50–60

Q
Quasi-longitudinal approximation, 49, 50, 74
Quiet-day variation, of geomagnetic field, 225–228, 233, 239, 240, see also Electric currents

R
Radiation belts, 125, 242–244, 250, 253, 256, 260
Radiative recombination, 88, 105, 106, 117, 204
Radio stars, see Drifts, Scintillations
Rate coefficients, typical values of, 106, 107
Real height, determination from ionograms, 55–60
Recombination, 105, 106
dissociative, 88, 105, 106, 108, 113, 114, 120, 121, 123, 124, 204
ionic, 105, 106, 110, 113
radiative, 88, 105, 106, 117, 204
three-body, 105, 106
Recombination coefficients
from ionospheric observations, 166–168, 194–196
in sporadic E layers, 204
Recovery phase of magnetic storm, 230, 255
Reduced height, 6, 8, 9
Reflection coefficient
for long waves, 63
for partial reflections, 67, 68
Reflection height
for long (VLF) waves, 62, 63, 161, 191, 262, 263
for short (HF) waves, 190, 192
Refractive index, 49, 55
Relativistic electron precipitation, 262
Resonance probe, 81
Resonance theory of tides, 35, 40
Resonances, in plasma, 60, 81
Ribbon microphone, 42
Ring current, 230, 255, 258–260
Riometer, 65, 66
Rockets, description of experiments, 41–44, 70–73, 78–82
airglow measurements by, 121, 122
ionospheric results from, 187, 200, 218, 219
magnetic measurements by, 237, 240

S
Satellites, see also Geostationary satellites, Space probes
description of experiments, 42, 43, 71, 74–76, 78–82
results obtained from, 20, 21, 98, 170, 172, 173, 184, 186–189, 191, 209, 210, 212, 220, 225, 266
Satellite drag, 11, 18, 45, 46, 253, 257
Scale height
defined, 5, 6
of plasma, 143, 147, 187, 188
Scatter radar, see Incoherent scatter
Schummann–Runge continuum, 19
Scintillation of radio stars, 209, 210
Searchlight, 40, 41
Seasonal anomaly, F region, 179–181
Sen–Wyller generalization of Appleton–Hartree theory, 62
Shape preserving distribution, 153, 155
Shock front, 241, 242
Short wave fadeout (SWF), 190, 192, 262
SID, see Sudden ionospheric disturbance
Sluggishness of ionosphere, 151, 166, 195
Sodium vapor trails, 22, 43
Solar-cycle variations
absorption, 162, 163
atmospheric temperature, 12
E and F layers, 166, 167, 169, 171
ion composition, 188
total electron content, 171, 173
Solar eclipse, see Eclipse
Solar emission lines, 98–100, 102
Solar flare effect, magnetic (SFE), 191, 254
Solar flares, 76, 102, 103, 190–193, 252–254, 262
Solar radio noise, 192, 193, 232, 233, 254
Solar rotations, 233
Solar spectrum, 99, 100
Solar wind, 241, 242, 247, 253, 258
South Atlantic anomaly, 225, 244
Space probes, 77, 78
Speed of sound, 32
Spiral patterns of ionospheric phenomena, 249
Sporadic E (Es), 53, 165, 199–204, 218, 240
definition of, 199
equatorial, 200, 201, 205, 238
high latitude, 200, 249
storm, 257, 264
wind-shear theory of, 201–204
Spread F, 210, 211, 249
Sq, see Quiet-day variation
Stationary phase principle, 63
Storm current systems, 253, 258–261, 269, see also Disturbance variations
Storms, 252–270
effects in D region, 257, 261–264
in E region, 264
in F1 layer, 264
in F2 layer, 257, 264–270
electron temperature variations, 218, 219, 268
heating of neutral atmosphere, 14, 18, 43, 253, 257
magnetic variations, 229–231, 255, 258–261
magnetospheric structure variations, 188, 251
synopsis of phenomena, 252–257
Storm time variation
  F2 layer, 265, 266
  magnetic field, 229–231
Stratopause, 3
Stratosphere, 2
Stratospheric warming, 160, 163–165, 181
Sudden commencement (SC or SSC), 229, 252, 255, 258, 265
Sudden cosmic-noise absorption (SCNA), 190
Sudden enhancement of atmospherics (SEA), 191
Sudden frequency deviation (SFD), 190–192
Sudden impulse (SI), 229
Sudden ionospheric disturbance (SID), 65, 190–193, 254, 262
Sudden phase anomaly (SPA), 191, 192
Sunrise phenomena in ionosphere, 161, 170, 172, 180, 217, 219
Sunspot number, 233
Superthermal electrons, 213
Surface wave, 32
Symbols (table of), 271–274

T
Temperature variations of neutral atmosphere, 12, 17, 18, 43, 253, 257
  effects on ionosphere, 128, 156, 177, 178, 179
Thermal balance
  of electrons and ions, 213–220
  of neutral air, 11–19
Thermal conduction, see Heat conduction
Thermal conductivity, 3, 16, 33, 216, 217
Thermal diffusion, 23, 144
Thermal equilibrium, absence in ionosphere, 213
Thermal expansion and contraction
  effect on F region, 128, 156, 177, 199, 268
  on neutral gas distribution, 8–10
Thermal tide, 35
Thermal wind equation, 27
Thermopause temperature, 18
Thermosphere, 3, 10–19, 28–31, 128, 181, 253
Thomson cross section, 85
Thomson scatter, see Incoherent scatter
Three-body reactions, 105–108, 123
Tidal amplification, 36, 38, 39
Tides
  in atmosphere, 12, 25, 28, 34–40, 233
  lunar, 186, 236
  theory of, 37–40
Topside ionosphere, 146–148, 157, 186–189, 199, 211
Topside sounder, 59, 60, 81, 82, 184, 186, 188, 209, 211
Total column content of neutral gas, 7
Total electron content, 73–76, 170–173, 179, 180, 192, 193, 266–268
  solar cycle variation, 173
Transfer reactions, 119–121
Transport processes, 87, 88, 110, 117, 126–159
Transport regime, 88
Transport term in continuity equation, 126, 127
Transverse conductivity, 136
Transverse drift, 202
Transverse mobility, 134
Trapped particles, 241–244, 250, 260, see also Radiation belts
Traveling ionospheric disturbances (TID), 33, 34, 205, 211, 212
Trimethylaluminum (TMA), 43, 44
Tropopause, 2
Troposphere, 1, 2
Trough
  equatorial F region, 183–186
  high latitude F region, 183, 188, 189, 250, 251
Turbopause, 4, 13, 14, 19, 22, 42, 88
Turbulence, in atmosphere, 4, 12, 14, 21, 22, 25, 26, 31, 201, 269

U
Ultraviolet radiation
  absorbed by ozone, 2, 3
  detachment by, 112
  heating by, 3, 11, 13–16
  measurement of, 98
  nighttime, 167, 182
  production of ionization by, 48, 87, 98–103, 151
  solar flares, 76, 192, 193, 254
Universal time, phenomena related to, 182, 249
U.S. standard atmosphere, 2, 14, 22, 42

V
Valleys, in electron concentration profile, 57, 58, 60, 167, 197
Van Allen belts, see Radiation belts
Velocity divergence, 37, 38
Velocity of escape, 23
Virtual height, 51-55, 264, 268
Viscosity, 14, 26, 29, 30, 39, 139, 140
VLF emissions, 242, 249, 256
VLF radio waves, 163, 191, 254, 257, 262, 263, see also Long waves

W
Wave interaction, 61, 68-70, see also Cross modulation
Waves, see Atmospheric waves, Gravity waves
Whistlers, 189, 250
Wind direction, defined, 25
Wind shear, 27, 43, 44, 201-204
Winds, see also Drifts, Dynamo theory
diurnal and semidiurnal components, 36, 37
geostrophic, 27, 30
heat transport by, 13, 18
ionospheric effects of, 153, 156, 157, 177, 178, 181, 182, 186, 270
measured
from meteor observations, 28, 31, 36, 37, 236, 237, 239, 240
by rocket experiments, 27, 28, 41
mesospheric, 25, 27, 28, 41
movement of ions and electrons by, 128, 135, 234, 235
prevailing, 25, 28, 31, 36
thermospheric, 25, 28-31, 46
Winter anomaly, see also Seasonal anomaly
D region, 160; 163, 164
F region, 179-181

X
X-ray photographs of sun, 196
X-rays, solar
production of ionospheric layers by, 87, 98-102, 106, 109, 161, 167
relation to eclipse effects, 194, 195
solar flares, 76, 81, 102, 103, 191-193, 254