INTERNATIONAL GEOPHYSICS SERIES • VOLUME 22

# Theory of Planetary Atmospheres

## An Introduction to Their Physics and Chemistry

Joseph W. Chamberlain

# THEORY OF PLANETARY ATMOSPHERES

## An Introduction to Their Physics and Chemistry

# International Geophysics Series

EDITORS

J. VAN MIEGHEM
*Royal Belgian Meteorological Institute*
*Uccle, Belgium*

ANTON L. HALES
*Australian National University*
*Canberra, A.C.T., Australia*

*Volume* 1 BENO GUTENBERG. Physics of the Earth's Interior. 1959

*Volume* 2 JOSEPH W. CHAMBERLAIN. Physics of the Aurora and Airglow. 1961

*Volume* 3 S. K. RUNCORN (ed.). Continental Drift. 1962

*Volume* 4 C. E. JUNGE. Air Chemistry and Radioactivity. 1963

*Volume* 5 ROBERT C. FLEAGLE AND JOOST A. BUSINGER. An Introduction to Atmospheric Physics. 1963

*Volume* 6 L. DUFOUR AND R. DEFAY. Thermodynamics of Clouds. 1963

*Volume* 7 H. U. ROLL. Physics of the Marine Atmosphere. 1965

*Volume* 8 RICHARD A. CRAIG. The Upper Atmosphere: Meteorology and Physics. 1965

*Volume* 9 WILLIS L. WEBB. Structure of the Stratosphere and Mesosphere. 1966

*Volume* 10 MICHELE CAPUTO. The Gravity Field of the Earth from Classical and Modern Methods. 1967

*Volume* 11 S. MATSUSHITA AND WALLACE H. CAMPBELL (eds.). Physics of Geomagnetic Phenomena. (In two volumes.) 1967

*Volume* 12 K. YA. KONDRATYEV. Radiation in the Atmosphere. 1969

*Volume* 13 E. PALMEN AND C. W. NEWTON. Atmospheric Circulation Systems: Their Structure and Physical Interpretation. 1969

*Volume* 14 HENRY RISHBETH AND OWEN K. GARRIOTT. Introduction to Ionospheric Physics. 1969

*Volume* 15 C. S. RAMAGE. Monsoon Meteorology. 1971

*Volume* 16 JAMES R. HOLTON. An Introduction to Dynamic Meteorology. 1972

*Volume* 17 K. C. YEH AND C. H. LIU. Theory of Ionospheric Waves. 1972

*Volume* 18 M. I. BUDYKO. Climate and Life. 1974

*Volume* 19 MELVIN E. STERN. Ocean Circulation Physics. 1975

*Volume* 20 J. A. JACOBS. The Earth's Core. 1975

*Volume* 21 DAVID H. MILLER. Water at the Surface of the Earth: An Introduction to Ecosystem Hydrodynamics. 1977

*Volume* 22 JOSEPH W. CHAMBERLAIN. Theory of Planetary Atmospheres: An Introduction to Their Physics and Chemistry. 1978

# THEORY OF PLANETARY ATMOSPHERES

## An Introduction to Their Physics and Chemistry

*Joseph W. Chamberlain*

Department of Space Science and Astronomy
Rice University
Houston, Texas

*ACADEMIC PRESS New York San Francisco London 1978*
*A Subsidiary of Harcourt Brace Jovanovich, Publishers*

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

*United Kingdom Edition published by*
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Chamberlain, Joseph Wyan, Date
Theory of planetary atmospheres.

(International geophysics series)
1. Planets--Atmospheres. I. Title. II. Series.
QB603.A85C48 523.4 78-203
ISBN 0-12-167250-6

PRINTED IN THE UNITED STATES OF AMERICA

To

$O_3$, $N_2O$, $CO_2$, and $^{14}C$;
$NO_x$, $HO_x$, and $ClO_x$;
$CCl_4$, $CF_2Cl_2$, and $CFCl_3$;
$^{235}U$, Pu, D, and T;

and many others
of our effluent society
without whom this book
would not have been written.

This page intentionally left blank

# CONTENTS

This page intentionally left blank

# PREFACE

As atmospheric physics has grown and become complicated, it has (like all subjects) encouraged its practitioners to specialize. No one understands the entire atmosphere very well; the last generalist was probably Aristotle. Yet the applied problems we face today often require that the research worker cross over imaginary disciplinary lines.

Our subject is, of course, nothing more than applied physics and chemistry. But in addition to those basic sciences the student of planetary atmospheres needs an overview of atmospheric structure and physical processes as presently understood. This book is intended to help fill that need for both graduate students and research scientists. Although the approach is mainly theoretical, very little basic physics is developed here. Material that is standard fare in third- and fourth-year physics courses is simply absorbed where needed.

The subject is developed as simply as I know how, without reducing it to a purely descriptive level. I frequently make radical assumptions that simplify the mathematics in order to illustrate a point, but I have attempted to alert the reader to the complexities that would have to be coped with at the research level.

The bibliographical notes for each chapter list some of the historically important papers, a selection of books and review articles, and a few other papers containing analyses or data discussed in the text. These selections are matters of personal judgment, but the interested student should be able to start his own literature survey for a selected topic with the material provided. For those readers who desire a more descriptive introduction to the subject to supplement their physical and chemical study, a number of excellent works are listed in Appendix V.

This page intentionally left blank

# ACKNOWLEDGMENTS

Several of my colleagues have read chapters of the text in draft and made constructive suggestions. I am especially indebted to Donald E. Anderson, Ralph J. Cicerone, Von R. Eshleman, Donald M. Hunten, Andrew A. Lacis, and Darrell F. Strobel for very critical and helpful readings. The text and the drawings have been prepared by Christine Dunning; I am most grateful to her for the high quality of her work and for a constant good humor, both of which lightened my burdens enormously.

This page intentionally left blank

# *Chapter 1*
# VERTICAL STRUCTURE OF AN ATMOSPHERE

The simplest conceivable atmosphere is gravitationally bound (and therefore in hydrostatic equilibrium) and spherically symmetric. For the terrestrial planets, at least, solar radiation and the radiative properties of the atmospheric constituents fix the first-order description of the vertical thermal structure. An atmosphere controlled by sunlight can scarcely be spherically symmetric. Nevertheless, it is useful to think of a mean planetary atmosphere, with day–night and latitudinal variations occurring about the mean.

The *vertical structure* of an atmosphere is the run of pressure, temperature, density, and chemical composition with distance from the center of the planet (or with height above the surface). When these parameters are inferred in part from theory or when they are tabulated as mean or representative values, they constitute a *model atmosphere.*

Figure 1.1 shows the temperature profile for Earth's atmosphere; it serves to divide the atmosphere into different regions, where the controlling physics and chemistry differ.

The tropospheric temperature is governed by radiative and convective exchange. In the stratosphere trace amounts of $O_3$ are formed by sunlight; the remarkable ability of $O_3$ to absorb both ultraviolet and infrared radiation causes an inversion above the tropopause. The decrease in $O_3$ production and the increased rate of cooling to space by $CO_2$ reestablish a declining

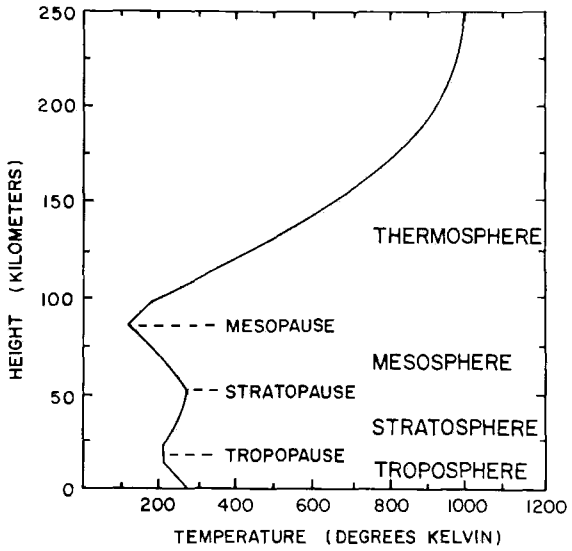



**Fig. 1.1** Schematic temperature profile for Earth's atmosphere showing the various regimes defined by the temperature gradient.

temperature in the mesosphere. Finally, heating by $O_2$ photolysis and ionization increases the thermosphere temperatures to about 1000°K. We shall examine the main processes in these regions in turn.

## 1.1 Hydrostatic Equilibrium

The vertical distribution of pressure, temperature, and density in a static, spherical atmosphere with a specified composition are governed by three relationships. First we have hydrostatic equilibrium, in which the pressure gradient is

$$\frac{dp}{dr} = -\left(\frac{G\mathcal{M}}{r^2}\right)(MN) = -g(r)\rho \tag{1.1.1}$$

Here $\mathcal{M}$ is the mass of the planet and $M$ is the mean mass of the molecules of a mixed atmosphere, $N$ is their number density, $\rho$ is the mass density, and $r$ is the distance from the center of the (spherical) planet. Over height intervals $\Delta r$ such that $\Delta r \ll r$, the local acceleration of gravity $g(r) \approx$ const.

For an equation of state, the perfect gas law is adequate:

$$p = NkT = \rho RT \tag{1.1.2}$$

where $R = k/M$ (erg/gm deg) is the gas constant appropriate to the atmospheric composition. Then hydrostatic equilibrium gives

$$\frac{dp}{p} = -\frac{G\mathcal{M}M}{kT}\frac{dr}{r^2} \approx -\frac{gM}{kT}\,dz \equiv -\frac{dz}{H} \tag{1.1.3}$$

where $z$ is height above the surface and $H$ the (pressure) *scale height*.

A third relation must fix the temperature (cf. Section 1.2). If the mean mass and temperature are constant with height, we obtain the *barometric law*:

$$\begin{aligned} p(r) &= p(r_0)\exp\left(-\frac{G\mathcal{M}M}{kTrr_0}(r - r_0)\right) \\ &\approx p(r_0)\exp\left(-\frac{r - r_0}{H}\right) \\ p(z) &= p(z_0)\exp\left(-\frac{z - z_0}{H}\right) \end{aligned} \tag{1.1.4}$$

Thus the *pressure scale height* ($H = kT/Mg$) is an $e$-folding distance. In the general case the density distribution is

$$\begin{aligned} \frac{dN}{N} &= -\frac{dT}{T} - \frac{G\mathcal{M}M}{kT}\frac{dr}{r^2} \approx -\frac{dT}{T} - \frac{dz}{H} \\ &= -\left(\frac{1}{T}\frac{dT}{dz} + \frac{Mg}{kT}\right)dz = -\frac{dz}{H^*} \end{aligned} \tag{1.1.5}$$

which defines the *density scale height* $H^*$. The *integrated density* is the number of particles in a column above a specified height. It is, from (1.1.1),

$$\begin{aligned} \mathcal{N}(r) &\equiv \int_r^\infty N(r)\,dr = \int_0^{p(r)} \frac{r^2}{G\mathcal{M}M}\,dp \\ &\approx \frac{p(r)}{g(r)M} = N(r)H \end{aligned} \tag{1.1.6}$$

The integrated density is often written in terms of the height of a column under standard temperature and pressure conditions that would contain the same number of molecules or atoms. This is the equivalent thickness in "atmo-centimeters" (alternatively, "centimeter-atmospheres" or "centimeter-amagat").

$$\xi = \frac{\mathcal{N}(z)}{N_0}\ \text{atm-cm} \tag{1.1.7}$$

where $N_0$ is Loschmidt's number ($2.687 \times 10^{19}\ \text{cm}^{-3}$).

Without perfect mixing the mean mass $M$ is a function of height, $M(z)$. At high altitudes the mixing processes are likely to be less important since the diffusion coefficient (Section 2.3.1) is large. Hence the atmosphere tends to separate out diffusively, and the composition changes as well through photochemical reactions.

## 1.2 Radiative Equilibrium

As a starting point we will regard the atmospheric temperature as governed by radiative equilibrium. Of course, it is not, but we will add convection later. In the ionosphere conduction becomes the dominant mechanism for heat transfer, and radiative equilibrium is not even a good starting approximation.

### 1.2.1 Equation of Radiative Transfer and Kirchhoff's Law

In a homogeneous medium (see Fig. 1.2) the monochromatic radiant intensity $I_\nu$ (measured in erg/cm$^2$ sec sr Hz) changes along distance $ds$ (measured in the direction of propagation of the light ray and always positive) by amount $dI_\nu$ given by

$$\frac{1}{\rho}\frac{dI_\nu}{ds} = -(\kappa_\nu + \sigma_\nu)I_\nu + j_\nu \tag{1.2.1}$$

where $\kappa_\nu$ is the mass absorption coefficient and $\sigma_\nu$ the mass scattering coefficient (both measured in cm$^2$/gm), and $j_\nu$ is the emission coefficient (erg/gm sec sr Hz). The emission $j_\nu$ may be due in part to scattering and in part to thermal excitation. The combination $\kappa_\nu + \sigma_\nu$ is the *extinction coefficient.*

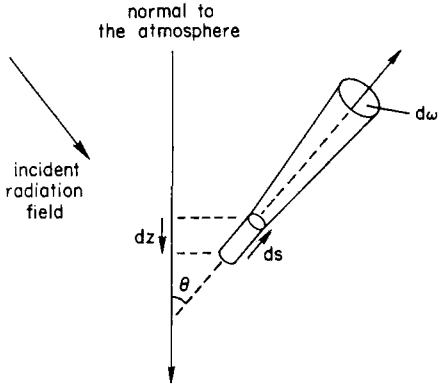



**Fig. 1.2** Geometry for the equation of radiative transfer. The element $ds$ is always positive and taken in the direction of propagation of a light ray.

Let us look at a few special cases. If emission and scattering back into the original beam are negligible, as in viewing a single star against a black sky, we have $j_\nu = 0$ and

$$I_\nu(s) = I_\nu(0)e^{-(\kappa_\nu + \sigma_\nu)\rho s} \tag{1.2.2}$$

which is *Lambert's exponential absorption law.*

If the source of emission includes scattering (as in the blue, sunlit sky), we have to specify a *scattering phase function*, $p(\cos\Theta)$, giving the angular distribution of scattered radiation. Thus, if $j_\nu$ is entirely due to scattering, the emission term is

$$j_\nu = \frac{\kappa_\nu + \sigma_\nu}{4\pi} \int I_\nu(\cos\Theta) p(\cos\Theta)\, d\Omega' \tag{1.2.3}$$

where the phase function itself is normalized so that integrated over a sphere it is

$$\frac{1}{4\pi} \int p(\cos\Theta)\, d\Omega' = \frac{\sigma_\nu}{\kappa_\nu + \sigma_\nu} \equiv \tilde{\omega}_\nu \tag{1.2.4}$$

which is called the *albedo for single scattering.*

For the scattering atmosphere, then, the equation of transfer is, for radiation in direction $\theta, \phi$,

$$\frac{dI_\nu(\theta,\phi)}{(\kappa_\nu + \sigma_\nu)\rho\, ds} = -I_\nu(\theta,\phi) + \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} I_\nu(\theta',\phi') p(\theta,\phi;\theta',\phi') \sin\theta'\, d\theta'\, d\phi' \tag{1.2.5}$$

At the other extreme from a scattering atmosphere is one in *local thermodynamic equilibrium* (LTE). It is assumed that at each point a local temperature $T$ can be defined so that the emission is given by Kirchhoff's law,

$$j_\nu = \kappa_\nu B_\nu(T) \tag{1.2.6}$$

where the Planck function is

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \tag{1.2.7}$$

We then have, since $\sigma_\nu$ is assumed zero, the monochromatic radiation in LTE given by

$$\frac{dI_\nu(\theta,\phi)}{\kappa_\nu \rho\, ds} = -I_\nu(\theta,\phi) + B_\nu(T) \tag{1.2.8}$$

The LTE approximation can never be exact, and one problem is ascertaining how inexact it is. In complete thermodynamic equilibrium, the temperature is everywhere the same; in the atmosphere the temperature has a definite gradient. Also, atmospheric emission is not Planckian at any point; the radiation field in the ultraviolet and infrared are not characteristic of the same $T$. Finally, the local kinetic temperature (given by the Maxwellian distribution law) is not the same as the effective Planckian temperature (defined by the radiation field). In a real situation we must usually treat scattering and thermal emission together. For combined isotropic scattering and thermal emission the transfer equation is

$$\frac{dI_\nu}{(\kappa_\nu + \sigma_\nu)\rho\, ds} = -I_\nu + \frac{\tilde{\omega}_\nu}{4\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} I_\nu(\theta', \phi')\sin\theta'\, d\theta'\, d\phi' + (1 - \tilde{\omega}_\nu)B_\nu \tag{1.2.9}$$

In the general case we define the source function (in the same units as $I_\nu$) as

$$\mathscr{J}_\nu = \frac{j_\nu}{\kappa_\nu + \sigma_\nu} \tag{1.2.10}$$

Then the general equation of transfer is

$$\frac{dI_\nu}{(\kappa_\nu + \sigma_\nu)\rho\, ds} = -I_\nu + \mathscr{J}_\nu \tag{1.2.11}$$

For isotropic scattering,

$$\mathscr{J}_\nu = \frac{\tilde{\omega}_\nu}{4\pi}\int I_\nu\, d\Omega \equiv \tilde{\omega}_\nu J_\nu \tag{1.2.12}$$

where $J_\nu$ is the local mean intensity. For LTE,

$$\mathscr{J}_\nu = B_\nu(T) \tag{1.2.13}$$

and the combined case has

$$\mathscr{J}_\nu = \tilde{\omega}_\nu J_\nu + (1 - \tilde{\omega}_\nu)B_\nu \tag{1.2.14}$$

Defining a *slant optical thickness* from $s$ to $s'$ as

$$\tau_\nu(s, s') = \int_s^{s'} (\kappa_\nu + \sigma_\nu)\rho\, ds \tag{1.2.15}$$

we can write down the formal solution to Eq. (1.2.11) as

$$I_\nu(s) = I_\nu(0)e^{-\tau_\nu(s,0)} + \int_0^s \mathscr{J}_\nu(s')e^{-\tau_\nu(s,s')}(\kappa_\nu + \sigma_\nu)\rho\, ds' \tag{1.2.16}$$

If the source function is known, we have the solution for the radiation field. In practice, the solution is not so simple because $\mathcal{J}_\nu(s')$ depends on $I_\nu(s')$ directly or on $B_\nu(T)$ (which in turn depends on the heating from the radiation field) or both.

### 1.2.2 Monochromatic Radiative Equilibrium

In a plane-stratified atmosphere in which height $z$ is measured upward, we measure optical depth $\tau$ downward and zenith angle of the direction of radiation flow $\theta$ from the upward vertical. Then $ds = \sec\theta\, dz$ is always positive and the vertical optical thickness is

$$d\tau_\nu = -(\kappa_\nu + \sigma_\nu)\rho\, dz \tag{1.2.17}$$

Equation (1.2.8) for LTE conditions is then

$$\mu \frac{dI_\nu(\theta,\phi)}{d\tau_\nu} = I_\nu(\theta,\phi) - B_\nu(T) \tag{1.2.18}$$

where $\mu = \cos\theta$. Integrating over a sphere we have

$$\frac{d}{d\tau_\nu}(\pi F_\nu) = 4\pi(J_\nu - B_\nu) \tag{1.2.19}$$

where the mean intensity $J_\nu$ is given by (1.2.12) and the net flux across an area parallel to the surface is

$$\pi F_\nu = 2\pi \int_{-1}^{1} I_\nu(\mu)\mu\, d\mu \quad . \tag{1.2.20}$$

We may obtain an approximate solution by the *two-stream approximation* (see Fig. 1.3). Suppose that the upward radiant intensity is $I_\nu(\mu,\tau) = I^+(\tau)$

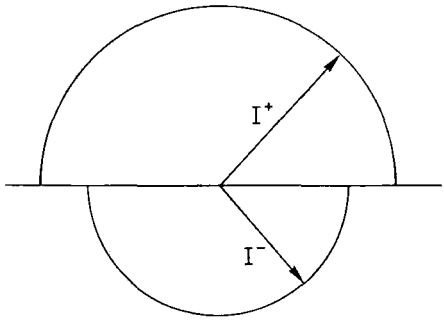



**Fig. 1.3** The two-stream approximation considers the radiation field to be composed of two simple streams, one upward and one downward.

for $0 < \mu < +1$ and the downward radiation field is $I_\nu(\mu, \tau) = I^-(\tau)$ $(-1 < \mu < 0)$. Then the mean intensity at depth $\tau$ is

$$J_\nu \equiv \tfrac{1}{2} \int_{-1}^{1} I_\nu(\mu)\, d\mu = \tfrac{1}{2}(I^+ + I^-) \tag{1.2.21}$$

and the net flux is

$$\pi F_\nu = \pi(I^+ - I^-) \tag{1.2.22}$$

To obtain a second relation between mean intensity and flux, we multiply (1.2.18) by $\mu$ and integrate over a sphere. With (1.2.20) and (1.2.21) we have

$$2\pi \frac{d}{d\tau_\nu} \int_0^1 d\mu\, \mu^2 (I^+ + I^-) = 2\pi \int_0^1 d\mu\, \mu (I^+ - I^-) \tag{1.2.23}$$

or

$$\frac{4\pi}{3} \frac{dJ_\nu}{d\tau_\nu} = \pi F_\nu \tag{1.2.24}$$

Substituting (1.2.19) for $J_\nu$ in this equation yields the flux equation

$$\frac{d^2 F_\nu}{d\tau_\nu^2} - 3F_\nu = -4 \frac{dB_\nu}{d\tau_\nu} \tag{1.2.25}$$

The concept of radiative equilibrium means that the net flux divergence is everywhere zero, with no energy lost or supplied by convection or conduction. It does not mean that the flux $\pi F_\nu$ in each separate frequency everywhere is constant. Nevertheless, it is instructive to examine this case of *monochromatic radiative equilibrium* (MRE). With $dF_\nu/d\tau_\nu = 0$ everywhere, our equation for the thermal radiation versus $\tau_\nu$ is

$$\frac{dB_\nu}{d\tau_\nu} = \tfrac{3}{4} F_\nu = \text{const} \tag{1.2.26}$$

In applying boundary conditions we have to be careful that they do not conflict with assumptions already introduced in the two-stream approximation. Let us suppose that the ground is a black body at temperature $T_g$ and a cold black sky ($T = 0$) lies above the atmosphere. With Eqs. (1.2.21) and (1.2.22) we can write

$$J_\nu = I^- + \tfrac{1}{2} F_\nu = I^+ - \tfrac{1}{2} F_\nu \tag{1.2.27}$$

Then the transfer equation (1.2.19) gives

$$\begin{aligned} \frac{dF_\nu}{d\tau_\nu} &= 4(I^+ - B_\nu) - 2F_\nu \\ &= 4(I^- - B_\nu) + 2F_\nu = 0 \end{aligned} \tag{1.2.28}$$

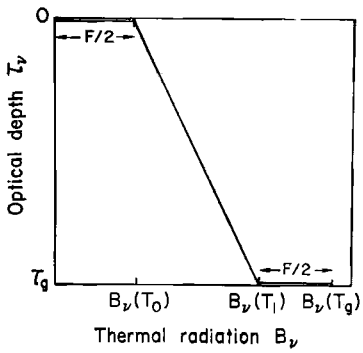



Fig. 1.4 The MRE solution for $T(\tau)$, presented as $B_\nu(T)$ vs. $\tau$. Note the discontinuity at the ground and the finite skin temperature at $\tau = 0$.

Hence the upward intensity at the ground is

$$I_g{}^+ \equiv B_\nu(T_g) = B_\nu(T_1) + \tfrac{1}{2}F_\nu \tag{1.2.29}$$

where $T_1$ is the air temperature at the ground, and the downward intensity at the top of the atmosphere is

$$I_0{}^- \equiv 0 = B_\nu(T_0) - \tfrac{1}{2}F_\nu \tag{1.2.30}$$

where $T_0$ is the air temperature at $\tau = 0$.

The solution tells us that, to fit an LTE solution with the special case of MRE, there is a discontinuity in temperature at the ground, with $T_g > T_1$, and the air at the top approaches a value, $T_0 \neq 0$. The radiant flux leaving the atmosphere is

$$\pi I_0{}^+ = \pi B_\nu(T_0) + \frac{\pi}{2} F_\nu = 2\pi B_\nu(T_0) \tag{1.2.31}$$

or twice what an opaque black body at temperature $T_0$ would emit.

The solution, then, from (1.2.26) and (1.2.30) is (see Fig. 1.4)

$$B_\nu(\tau_\nu) = B_\nu(T_0)(1 + \tfrac{3}{2}\tau_\nu) \tag{1.2.32}$$

The atmospheric radiation $\pi I_0{}^+$ is thus characteristic of the thermal emission at $\tau_\nu = \frac{2}{3}$.

### 1.2.3 Local Thermodynamic Equilibrium for a Gray Atmosphere Heated from the Ground

We are now ready to examine how the interchange of radiation governs the temperature of an atmosphere when there is no direct solar absorption by the atmosphere, when there is no conduction or convection, and when scattering can be neglected. The transfer equation (1.2.8) for LTE is

$$\frac{\mu\, dI_\nu}{\kappa_\nu \rho\, dz} = -I_\nu + B_\nu \tag{1.2.33}$$

Integrating over frequency we have

$$\frac{\mu}{\rho}\frac{d}{dz}\left(\int_0^\infty \frac{I_\nu}{\kappa_\nu}d\nu\right) = -I + B \tag{1.2.34}$$

where

$$I = \int_0^\infty I_\nu\, d\nu, \qquad B = \int_0^\infty B_\nu\, d\nu \tag{1.2.35}$$

By integrating over a sphere and setting the net flux constant, we have, analogously to Eq. (1.2.26),

$$-\frac{1}{\rho}\frac{d}{dz}\left(\int_0^\infty \frac{B_\nu}{\kappa_\nu}d\nu\right) = \tfrac{3}{4}F \tag{1.2.36}$$

where $F = \int F_\nu\, d\nu$. In the event $\kappa_\nu = \text{const}\,(=\kappa)$ the transfer equation (1.2.33) is

$$\mu\frac{dI}{d\tau} = I - B \tag{1.2.37}$$

and the thermal radiation is given by

$$\frac{dB}{d\tau} = \tfrac{3}{4}F \tag{1.2.38}$$

where

$$d\tau = -\kappa\rho\, dz \tag{1.2.39}$$

Hence the gray solution is analogous to that for MRE given above.

It would seem that, with a mean absorption coefficient properly defined, any nongray atmosphere could be treated with the gray solution. Thus writing

$$d\tau = -\langle\kappa\rangle\rho\, dz \tag{1.2.40}$$

a comparison of (1.2.36) and (1.2.38) suggests

$$\frac{1}{\langle\kappa\rangle} = \frac{1}{B}\int_0^\infty \frac{B_\nu}{\kappa_\nu}d\nu \tag{1.2.41}$$

where $B$ is given by (1.2.35). This $\langle\kappa\rangle$ is the *Rosseland mean* used widely in astrophysics. The problem with it lies in the Eddington approximation. To conserve flux, the MRE equation (1.2.19) requires $J_\nu = B_\nu$ precisely. But in LTE there is a gradual shift with depth in the frequency distribution of the radiation, since temperature varies with depth, and $J_\nu$ cannot precisely equal $B_\nu$ to conserve the flux in each frequency interval. But if there is approximate equality over the spectrum, the Rosseland mean is a good approximation.

It works well for stellar atmospheres, but is not good for most terrestrial situations. Since $T$ and $B_\nu(T)$ vary with depth, $\langle\kappa\rangle$ cannot be treated as constant with depth. Thus iterations are required to compute $B_\nu(T)$ and $\langle\kappa\rangle$ as functions of depth. An alternative and generally better procedure is to put $\kappa_\nu$ on the right-hand side of (1.2.26) and integrate over $\nu$. Comparison with (1.2.38) gives

$$\langle\kappa\rangle = \frac{1}{F}\int_0^\infty \kappa_\nu F_\nu \, d\nu \tag{1.2.42}$$

which is the *Chandrasekhar mean* coefficient. Its difficulty is that $F_\nu$ is not known in advance of solving the problem, requiring iterations, or other approximations.

We can now write the LTE gray-atmosphere solution and compare numerical results with temperatures in the Earth's atmosphere. Equation (1.2.38) gives a solution, similar to (1.2.32),

$$T^4(\tau) = T_0{}^4(1 + \tfrac{3}{2}\tau) \tag{1.2.43}$$

where $T_0$ is the temperature of the upper boundary. Here the integrated black-body intensity is

$$B(\tau) = \frac{\sigma}{\pi} T^4(\tau) \tag{1.2.44}$$

The total radiant flux from the Earth can be expressed in terms of a *mean planetary emission temperature*, obtained by integrating Eq. (1.2.31) over frequency:

$$T_e{}^4 = 2T_0{}^4 \tag{1.2.45}$$

For a rotating planet of radius $R$ at a uniform temperature over the sphere this temperature is related to the incident solar flux by

$$4\pi R^2 \sigma T_e{}^4 = (1 - \Lambda)\pi R^2(\pi \mathscr{F}_\odot) \tag{1.2.46}$$

where $\Lambda$ is the effective planetary albedo and $\pi\mathscr{F}_\odot$ is the incident solar flux. With $\Lambda = 0.4$ for Earth, we find $T_e = 246°\text{K}$ and a boundary temperature $T_0 = T_e/1.19 = 207°\text{K}$, which is close to the mid-latitude tropopause temperature. If we know the optical thickness of the atmosphere versus height, we can figure the temperature versus height—the lapse rate for radiative equilibrium. An example is given in Problem 1.3.

The discontinuity between the air and surface temperature can be expressed from (1.2.29) as

$$\begin{aligned} T_s{}^4 &= T^4(\tau_g) + \tfrac{1}{2}T_e{}^4 \\ &= T_0{}^4(2 + \tfrac{3}{2}\tau_g) \end{aligned} \tag{1.2.47}$$

**TABLE 1.1** *Approximate Percentage Deposition of Incident Solar Flux*[a]

| Reflected to space | Absorbed by ground | Absorbed in troposphere | Absorbed above troposphere | Reradiated from ground to atmosphere |
|---|---|---|---|---|
| 40 | 40 | 18 | 2 | 25 |

[a] The solar constant outside the atmosphere is 2 cal/min cm$^2$ = 1.39 × 10$^6$ erg/cm$^2$ sec; the global average heat received is one fourth this amount.

where $\tau_g$ is the optical thickness at the ground. The deposition of sunlight striking the Earth is shown in Table 1.1.

Equations (1.2.43) and (1.2.47) illustrate how high temperatures can be attained near the ground if the infrared $\tau$ is large and heating is from below. Heating from this combination of a transparent atmosphere in the visible, where the sun's energy is a maximum, and of high opacity in the infrared, where the Earth's Planckian curve peaks, is known as the *greenhouse effect*. It has long been thought that the trapping of infrared by glass is not the important thing in warming greenhouses. Rather, it is said, the glass roof merely keeps the warm air from convecting away. Purists have fought a steadily losing battle to replace "greenhouse effect" with a less picturesque term. I prefer to think of atmospheres as warming by the greenhouse effect, even if greenhouses do not.

If there is no internal heat source, the emission temperature $T_e$, computed from (1.2.46), will be equivalent to the measured *bolometric temperature* $T_b$, which is obtained by measuring the mean planetary flux from thermal emission over all frequencies and setting it equal to $\sigma T_b^4$. In the case of Jupiter and possibly other major planets, $T_b > T_e$, indicating internal generation of heat (cf. Section 1.8.3).

If an atmosphere's thermal emission is measured only in a narrow frequency interval, its intensity gives a *brightness temperature* $T_B$, defined by $I_\nu = B_\nu(T_B)$. If the atmosphere were gray, $I_\nu$ would be Planckian and $T_B$ would be the same at all frequencies and the same as $T_b$. The brightness temperature of Venus in the microwave spectrum gave the first indication of its 750°K surface (cf. Section 1.8.1).

## 1.3 Convection in the Troposphere

As we have seen, a gray atmosphere in radiative equilibrium approaches a finite "skin temperature" at high altitude. This isothermal region is stable against convective circulation. At large $\tau$, however, the radiative gradient

$dT/dz$ becomes steep (i.e., negatively large). Hence, an optically thick, gray atmosphere can be convectively unstable at low altitudes; the temperature distribution that radiative exchange tends to establish is then too steep to be hydrostatically supported.

If an element of gas moves adiabatically, the first law of thermodynamics requires that

$$C_v\,dT = -p\,dV \tag{1.3.1}$$

where $C_v$ is the specific heat at constant volume (erg/gm°K). If $V$ is the *specific volume* containing a gram of molecules, then the perfect gas law gives

$$dV = \frac{N_0 k}{p}dT - \frac{N_0 kT}{p^2}dp \tag{1.3.2}$$

where $N_0 = 1/M$ and $M$ is the molecular mass. Since $C_p = C_v + N_0 k$ (erg/gm°K), we have the alternate thermodynamic relation

$$C_p\,dT = \frac{N_0 kT}{p}dp = V\,dp = \frac{1}{\rho}dp \tag{1.3.3}$$

With hydrostatic equilibrium, (1.1.1), the first law thus gives the *dry adiabatic lapse rate*,

$$\frac{dT}{dz} = -\frac{g}{C_p} = -\frac{\gamma - 1}{\gamma}\frac{gM}{k} \tag{1.3.4}$$

where $\gamma = C_p/C_v$. For the Earth's troposphere this lapse rate is $-9.8$°K/km.

For saturated air the first law includes the latent heat released by water condensing:

$$C_v\,dT = -p\,dV - L\,dw_s, \qquad C_p\,dT = \frac{1}{\rho}dp - L\,dw_s \tag{1.3.5}$$

where $w_s$ is the mass of saturated water per mass of air and $L$ is the latent heat of vaporization. The *saturation adiabatic lapse rate* is then

$$\frac{dT}{dz} = -\frac{g/C_p}{1 + (L/C_p)(dw_s/dT)} \tag{1.3.6}$$

The wet lapse rate may be about half the dry rate or around 5°K/km. Since convection is partly moist and partly dry, the troposphere has an average value of 6.5°K/km. This value characterizes the static stability of the Earth's large-scale weather systems (see Section 2.2.4).

The temperature distribution in radiative-convective equilibrium is thus simply the adiabatic curve at low altitudes, merging into the purely radiative

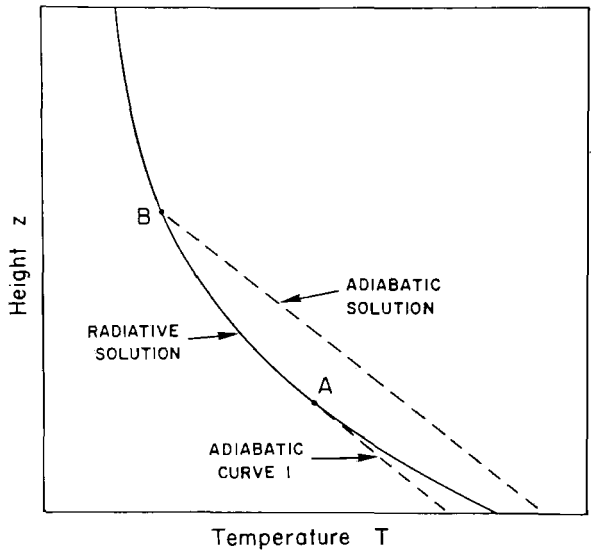



**Fig. 1.5** When convection is present the adiabatic portion of the curve not only extends to a point where the radiative solution is convectively stable; it must also supply the thermal radiation to sustain the profile above "*B*."

one at higher levels (see Fig. 1.5). Of an infinite number of parallel adiabatic atmospheres we must select the one that emits the same upward radiant flux as the radiative atmosphere itself. For example, curve 1 in Fig. 1.5 is tangent to the radiative curve and would not require a temperature discontinuity. But the adiabatic curve is everywhere below the radiative one and it clearly cannot supply the radiant flux required above point A to support the temperature there in radiative equilibrium.

A self-consistent solution is an adiabatic curve displaced to the right by an amount such that the flux boundary condition is satisfied. The temperature discontinuity at the ground must also be removed, since it is unstable convectively. The flux given by the various adiabatic distributions is readily computed from the formal solution to the transfer equation (1.2.16).

## 1.4 Latitudinal Variations of the Tropopause and Departures from Grayness

The radiative-convective model with a gray atmosphere reproduces the main features of the troposphere temperature distribution, with the radiative "skin temperature" being identified with the near-isothermal region of the tropopause. This happy state of affairs does not hold, however, when we

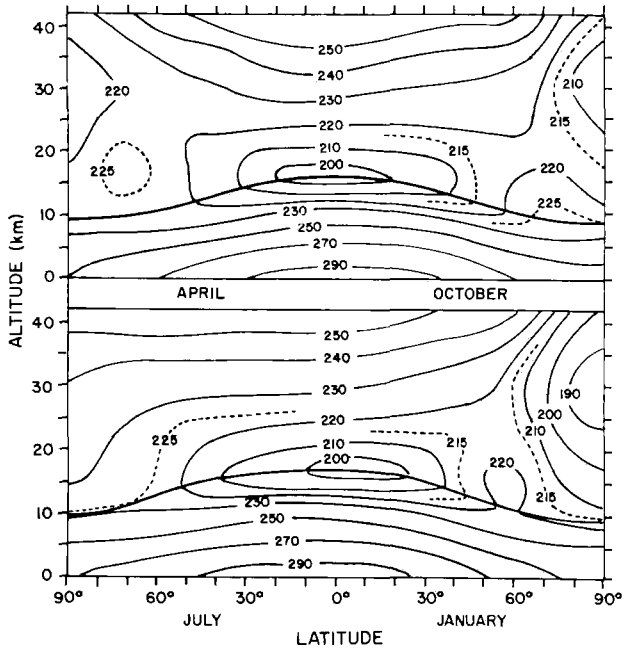



**Fig. 1.6** Distribution of the observed temperature (°K) in the northern hemisphere for different seasons. [Based on data assembled by J. LONDON; see MANABE and STRICKLER (1964).]

examine some of the finer features. Figure 1.6 shows isotherms on meridional cross sections for the various seasons. The heavy line shows the location of the tropopause. It is apparent from (1.2.45) and (1.2.46) that in the gray atmosphere, heated from below and without dynamical interchange latitudinally, the tropopause would be warmer in the tropics than in the Arctic simply because the ground temperature is higher. In fact, the tropic tropopause is around 15 km with $T \approx 195°K$, whereas over the polar cap it is as low as 8 to 10 km and $T \approx 225°K$.

Why should the tropopause be lower and warmer in the Arctic? There are several contributing factors. Quite likely the most important is the fact that the atmosphere is not only nongray but the distribution of the infrared active gases ($CO_2$, $H_2O$, and $O_3$) varies with latitude.

The departure from grayness itself diminishes the coupling between the ground and skin temperatures. With windows in the absorption spectrum, the ground can cool by radiating directly to space or to the stratosphere. The higher water vapor in the tropics inhibits this direct cooling of the ground but, more important, it increases the cooling of the upper troposphere. This dependence of absolute humidity on air temperature is probably the most important item governing tropopause height.

Another factor is the latitudinal dependence of stratospheric ozone. In a static atmosphere, ozone would be most abundant where it is most readily created—at lower latitudes. In fact (as we shall discuss in Section 3.3.2), it is more abundant at higher latitudes. The 9.6 $\mu$m band of $O_3$ is largely responsible for heating the lower stratosphere by absorbing surface infrared radiation in an otherwise clear window. In addition, we have so far neglected direct heating by the sun, but the existence of the stratosphere is primarily due to direct solar photolysis and heating. With ozone more abundant in the Arctic and distributed more toward low altitudes than in the tropics, it has a different effect on the temperature profile. The tropic tropopause is sharply defined, with an abrupt temperature inversion and rapid rise in the lower stratosphere. In the polar regions the lower stratosphere is nearly isothermal, and these differences are probably due to the different ozone distributions.

Nevertheless, detailed model calculations have not given as satisfactory agreement as one would desire, suggesting other effects are contributing. The meridional circulation of the stratosphere is known to be important (see Section 3.3.2) and it probably has an influence on the tropopause height and temperature of the lower stratosphere. Also, convection is driven more strongly in the tropics, both because solar radiation is more abundant, and because of the cooling of the upper troposphere by the increased water vapor.

The interplay of all these effects requires elaborate computational models in which one feature is varied at a time to see what its effect is. Often the effects involve such complex feedbacks that the physics at work becomes obscure.

## 1.5 The Stratosphere: Absorption of Direct Solar Radiation

For an elementary examination of the stratospheric heating, we will oversimplify the problem while retaining enough of the relevant physics to see what is happening. In Chapter 3 the ozone-related chemistry is examined more thoroughly. (Also, the basic principles of photochemistry are reviewed in Section 3.1.) For now let us suppose that the only reactions of importance involve only oxygen: the so-called "Chapman reactions."

### 1.5.1 Elementary Oxygen Chemistry

Molecular oxygen is photodissociated by absorption in the Herzberg continuum:

$$O_2 + h\nu \rightarrow O(^3P) + O(^3P) \tag{1.5.1}$$

Atomic oxygen then attaches to $O_2$ in the presence of an unspecified third body M to form ozone,

$$\begin{aligned} &O + O_2 + M \rightarrow O_3 + M \\ &\qquad k_{12} = 1.1 \times 10^{-34} e^{510/T} \text{ cm}^6/\text{sec} \end{aligned} \tag{1.5.2}$$

where $k_{12}$ is the three-body rate coefficient (see Section 3.1). Both kinds of "odd oxygen" (i.e., O and $O_3$) are destroyed by

$$\begin{aligned} &O + O_3 \rightarrow O_2 + O_2 \\ &\qquad k_{13} = 1.1 \times 10^{-11} e^{-2150/T} \text{ cm}^3/\text{sec} \end{aligned} \tag{1.5.3}$$

and $O_3$ is converted back to O by

$$O_3 + h\nu \rightarrow O + O_2 \tag{1.5.4}$$

The principal source of heating is this $O_3$ dissociation, which occurs in the strong Hartley continuum, shown in Fig. 1.7. Over this region $O_3$ is the only atmospheric substance strongly absorbing (Fig. 1.8).

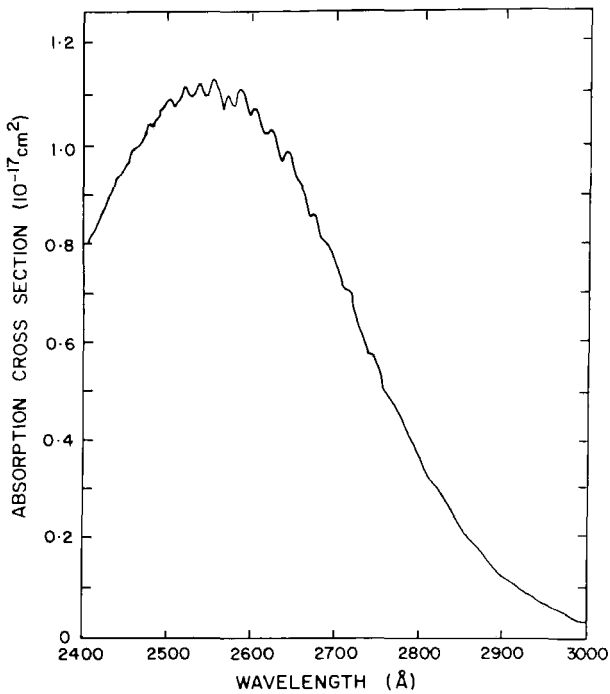



Fig. 1.7 Absorption cross sections of the Hartley bands and continuum. [Measurements by E. C. Y. Inn and Tanaka, (1953), *J. Opt. Soc. Amer.* **43**, 870.]

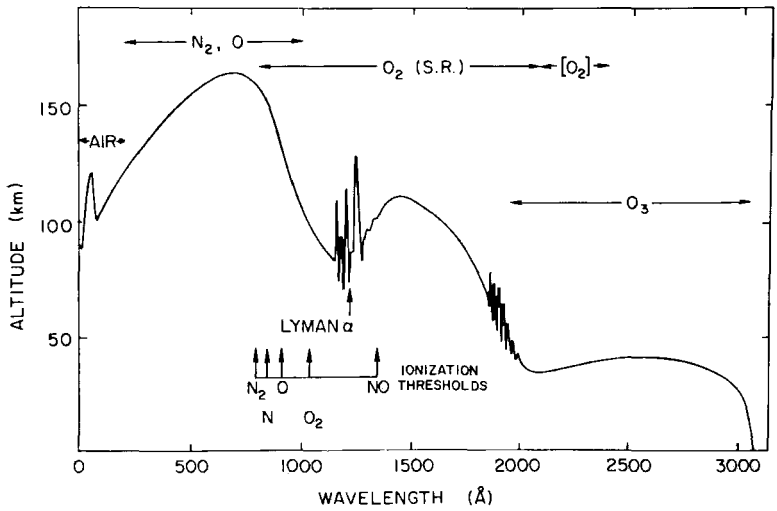



**Fig. 1.8** Altitude at which the intensity of solar radiation drops to $1/e$ of its value outside Earth's atmosphere for vertical incidence. The Schumann–Runge and forbidden Herzberg $O_2$ absorbing regions are denoted by S.R. and $[O_2]$, respectively. [Based on data assembled in 1961 by P. J. NAWROCKI, K. WATANABE, and L. G. SMITH; adapted from L. HERZBERG (1965), *in* "Physics of the Earth's Upper Atmosphere," (C. O. Hines *et al.*, eds.) p. 40, Prentice-Hall, Englewood Cliffs, New Jersey.]

Thus not only is this absorption responsible for heating the stratosphere, but the $O_3$ there prevents sunlight in the 0.2–0.3 $\mu$m region from reaching the ground. Since many forms of life (including the DNA molecule) cannot tolerate the ultraviolet, the 0.3 atm-cm or so in the stratosphere seems a very thin shield indeed. Only because $O_3$ is a strong absorber in the Hartley region (cross section of $10^{-18}$–$10^{-17}$ cm$^2$) can so few molecules accomplish so much.

The $O_2$ absorption (1.5.1), on the other hand, is intrinsically very weak. It is important to keep in mind that there are two photodissociation continua for $O_2$ important in atmospheric physics (cf. Fig. 1.9). The stronger one, the *Schumann–Runge continuum*, lies at wavelengths shortward of 1750 Å (7.1 eV). It involves the electronic transition $B\,^3\Sigma_u^- \leftarrow X\,^3\Sigma_g^-$ and hence is permitted as an electric-dipole transition. The measured cross section is in the range $10^{-18}$–$10^{-17}$ cm$^2$ and $O_2$ dissociates into $O(^3P) + O(^1D)$, that is, one atom in the ground term and one in the metastable $^1D$ term at 2 eV. This dissociation occurs in the 100 km region and sunlight at $\lambda < 1750$ Å is totally extinguished above the stratosphere.

The weaker absorption occurs at $\lambda < 2420$ (5.1 eV). It involves the forbidden Herzberg continuum $A\,^3\Sigma_u^+ \leftarrow X\,^3\Sigma_g^-$ whose cross section is $10^{-24}$–$10^{-23}$ cm$^2$, and the dissociation products are both in the ground term $O(^3P)$.

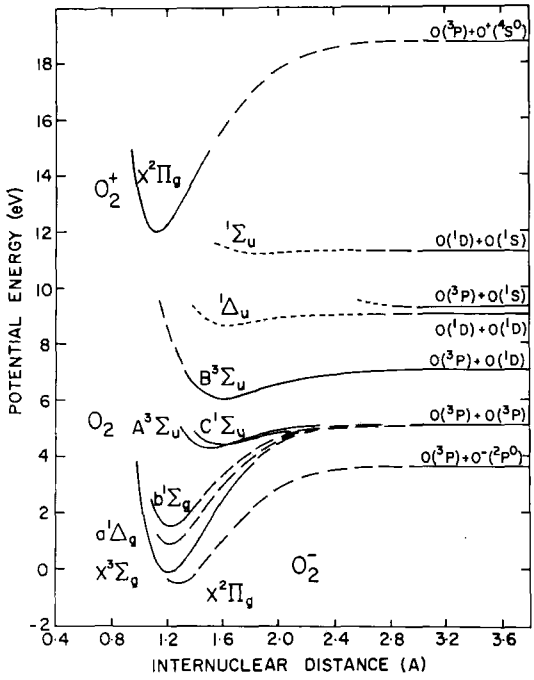



**Fig. 1.9** Simplified potential energy diagram for oxygen. Along the right side are listed the dissociation products; an energy level diagram of O appears as Fig. 6.3.

When the source function (representing local scattering or emission) is negligible, the transfer equation (1.2.11) gives the absorption per unit volume as

$$\mu \frac{dI_\nu(z)}{dz} = -\kappa_\nu \rho I(z) = -\kappa_\nu \rho I_\nu(\infty) e^{-\tau_\nu/\mu} \tag{1.5.5}$$

We will pursue this equation further in discussing the formation of the ionosphere (Section 5.1.1), but for the moment we will simply note that, in a barometric atmosphere (1.4), the maximum absorption occurs at (see Problem 1.4).

$$\tau_\nu/\mu = \kappa_\nu \rho H/\mu = \alpha_\nu N(z) H/\mu = 1 \tag{1.5.6}$$

where $\alpha_\nu$ (cm$^2$) is the molecular cross section. Hence the Schumann–Runge dissociation peaks near 100 km; the Herzberg dissociation would peak in the 20–25 km region, except that $O_3$ absorption between 1800 and 2400 Å raises the $\tau_\nu = 1$ level to about 40 km. The actual absorption coefficients are not gray and the destruction of $O_3$, involving reactions (1.5.2), (1.5.3), and (1.5.4), places the maximum ozone abundance somewhat lower, around 25–30 km altitude.

We can make an approximate assessment of the ozone distribution with the above four equations in daytime photochemical equilibrium. First, O is created by dissociation (1.5.1) at a rate $2J_2(z)$ ($sec^{-1}$ per $O_2$ molecule), which depends on the photon flux ($cm^{-2}\ sec^{-1}$) of dissociating radiation at height $z$ and the $O_2$ absorption cross section ($cm^2$) in the Herzberg continuum. Similarly (1.5.4) proceeds at rate $J_3$ ($sec^{-1}$ per $O_3$ molecule). The destruction of O by the two-body reaction (1.5.3) is given by the rate coefficient $k_{13}$ ($cm^3$/sec), which is the mean of the O—$O_3$ association cross section times the velocity of collision. Cross sections for chemical reactions may have a strong velocity dependence, making the rate coefficients strongly temperature dependent. Finally, the three-body reaction (1.5.2) has a rate coefficient $k_{12}$($cm^6$/sec). Rate coefficients are discussed more thoroughly in Section 3.1.

Thus the net change in O abundance is

$$\frac{d[O]}{dt} = 2J_2(z)[O_2] + J_3(z)[O_3] - k_{13}[O][O_3] - k_{12}[O][O_2][M] \quad (1.5.8)$$

where brackets denote concentrations (molecule/$cm^3$). Similarly, the $O_3$ rate of change is

$$\frac{d[O_3]}{dt} = k_{12}[O][O_2][M] - k_{13}[O][O_3] - J_3[O_3] \quad (1.5.9)$$

In chemical equilibrium these two equations lead to

$$[O] = \frac{J_2[O_2]}{k_{13}[O_3]} \quad (1.5.10)$$

and

$$[O_3] = \frac{k_{12}[O][O_2][M]}{k_{13}[O] + J_3} \quad (1.5.11)$$

Below 60 km the term $k_{13}[O] \ll J_3$ because of the low O abundance. Then we have

$$\frac{[O_3]_{day}}{[O_2]} \approx \left(\frac{k_{12}J_2[M]}{k_{13}J_3}\right)^{1/2} \quad (1.5.12)$$

and

$$[O]_{day} \approx \left(\frac{J_2J_3}{k_{12}k_{13}[M]}\right)^{1/2} \quad (1.5.13)$$

The daytime distribution of oxygen allotropes is shown in Fig. 1.10. At nighttime $O_3$ grows at the expense of O through (1.5.2), which proceeds rapidly because $O_2$ is large. Nevertheless, since $[O] \ll [O_3]$ in the strato-

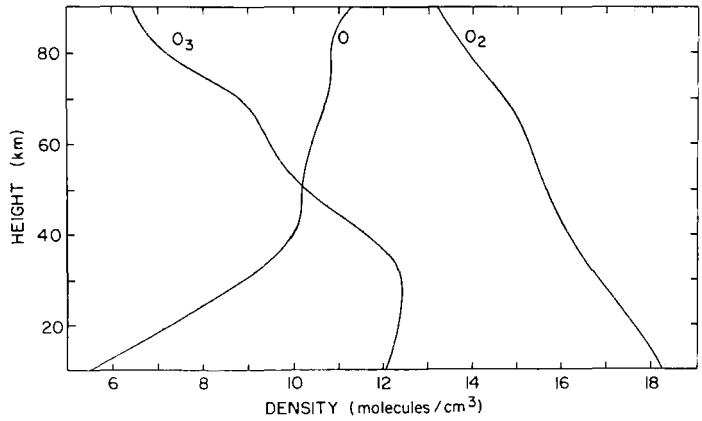



**Fig. 1.10** Daytime equilibrium of oxygen allotropes according to the Chapman theory. The abscissa is $\log_{10} N$. [Calculations by D. R. BATES (1954) *in* "The Earth as a Planet," (G. P. Kuiper, ed.), p. 581, Univ. Chicago Press, Chicago.]

sphere, the resulting increase in $O_3$ is relatively small. Were there a plentiful supply of O, the nighttime equilibrium of $O_3$ would grow to

$$\frac{[O_3]_{\text{night}}}{[O_2]} \to \frac{k_{12}[M]}{k_{13}} \tag{1.5.14}$$

but in the stratosphere the O rapidly disappears at night and the $O_3$ becomes essentially inert.

### 1.5.2 Stratospheric Heating

A calculation of the stratospheric temperature profile is not a simple task even with simplified photochemistry, the absence of vertical mixing, and ignored meridional dynamics. While the solar heating is not too complicated (although the solar-energy input at each level must be averaged over the day for any latitude and time of year) the infrared net cooling (or heating) is less straightforward. For example, weak rotational lines in the 15 $\mu$m fundamental band of $CO_2$ and lines in isotopic bands and weak *hot bands* (which arise from an excited lower vibrational level) are important in radiating directly to space, whereas stronger transitions in $CO_2$, $H_2O$, and $O_3$ provide a complex interchange of radiation within the atmosphere. The problem can be handled only by intricate computer codes.

With it understood that these complexities are being greatly oversimplified, we can at least examine the fundamental physics of solar heating. The Herzberg $O_2$ continuum is screened by not only the overhead $O_2$ but by $O_3$.

Hence we have

$$J_2(z) = J_2^{(0)} \exp\left( -\frac{\tau_2(O_2, z)}{\mu} - \frac{\tau_2(O_3, z)}{\mu} \right) \tag{1.5.15}$$

where $J_2^{(0)}$ is the value outside the atmosphere. The extinction due to $O_2$ is

$$\tau_2(O_2, z) = \alpha_2(O_2) \int_z^\infty [O_2(z)]\, dz \tag{1.5.16}$$

and that due to $O_3$ is

$$\tau_2(O_3, z) = \alpha_2(O_3) \int_z^\infty [O_3(z)]\, dz \tag{1.5.17}$$

The extinction of $J_3(z)$ is similarly defined by

$$\tau_3(O_3, z) = \alpha_3(O_3) \int_z^\infty [O_3(z)]\, dz \tag{1.5.18}$$

where $\alpha_3$ is the average absorption cross section in the Hartley continuum.

The evaluation of (1.5.12) for $[O_3]$ thus involves $[O_3]$ on the right-hand side, but we may proceed by successive approximations. Let us begin by making the obviously crude assumption that all constituents follow the same atmospheric scale height $H$. Then (1.5.12) gives for a first approximation

$$\frac{[O_3(z)]_{\text{day}}}{[O_2(z)]} \approx \left( \frac{k_{12}[M(z_0)]J_2^{(0)}}{k_{13}J_3^{(0)}} \right)^{1/2} e^{-(z-z_0)/2H} \exp\left( -\frac{\tau_0}{2\mu} e^{-(z-z_0)/H} \right) \tag{1.5.19}$$

where

$$\tau_0 = \tau_2(O_2, z_0) + \tau_2(O_3, z_0) - \tau_3(O_3, z_0) \tag{1.5.20}$$

Equation (1.5.19) has the form of a so-called "Chapman layer," which we will encounter again in ionosphere theory (see Section 5.1.1). The simple exponential represents diminished absorption with increasing altitude simply because of the decreasing abundance of the absorber. The compound exponential is due to attenuation of sunlight, which causes the "layer" to diminish deep in the atmosphere. The relative abundance of $[O_3]$ would, in this approximation, peak at $\tau_0/\mu = 1$; that is, at the $\tau_0$ "absorption" peak [see Eq. (1.5.6)]. The absolute $[O_3]$ peak, however, is at $\tau_0/\mu = 3$, or well below the $\tau_0$ "absorption" peak. The reader should be aware that (1.5.19) is a poor approximation for a real ozone layer since $\tau_0$ itself depends on $[O_3]$ and can even be negative.

The instantaneous heating rate from dissociation is then

$$\left( \frac{dT}{dt} \right)_{\text{heat}} = \frac{[O_2(z)]}{\rho(z)C_p} \left( J_2(z)h\nu_2 + J_3(z)h\nu_3 \frac{[O_3(z)]_{\text{day}}}{[O_2(z)]} \right) \tag{1.5.21}$$

where $\rho(z)$ is the local mass density, $C_p$ is the specific heat of air (erg/gm°K), and $\nu_2$ and $\nu_3$ are average frequencies of the $O_2$ Herzberg and $O_3$ Hartley continua, respectively. It is a good approximation to assume that all the solar

energy absorbed in the stratosphere goes into heating. Part of it appears as kinetic energy at the moment of dissociation; part goes into metastable states, such as $O(^1D)$, and is collisionally deactivated (i.e., transformed into heat); and part (the dissociation energy itself) appears as heat when the dissociation products react chemically to re-associate by (1.5.2) and (1.5.3).

To find the equilibrium temperature we equate the heat absorbed to that emitted. At high altitudes the infrared region of the atmosphere becomes optically thin and radiates directly to space. This assumption is occasionally used to examine small fluctuations in heating and temperature and is called the *transparent approximation.* Assuming grayness for the emitting gases, we have [cf. (1.2.44)]

$$4\pi\kappa B(z) \equiv 4\kappa\sigma T^4 = C_p\left(\frac{dT}{dt}\right)_{\text{cool}} \tag{1.5.22}$$

where $\kappa$ ($\text{cm}^2/\text{gm}$) is the absorption coefficient. The actual atmosphere is far from gray and a *linear* or *Newtonian* law of cooling is often adopted:

$$C_p\left(\frac{dT}{dt}\right)_{\text{cool}} = aT \tag{1.5.23}$$

where $a$ is a constant ($\sim 4\kappa\sigma {T_0}^3$). An example of the use of these simplified equations is given in Problem 1.6. The average overhead absorption for atmospheric gases in the infrared is shown in Fig. 1.11.

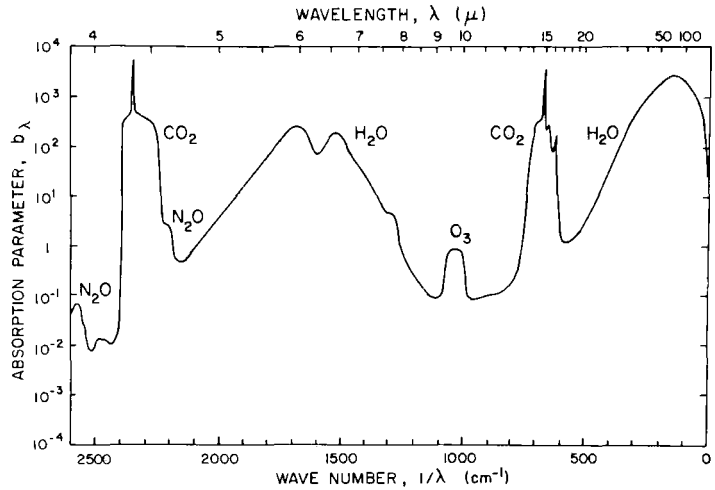



**Fig. 1.11** An indication of the absorbing properties of the lower atmosphere is $b_\lambda$, where $1/b_\lambda$ is the thickness in atmospheres that would give 50 percent absorption and 50 percent transmission. In this limited sense, $b_\lambda$ is a kind of optical thickness, even though over finite wavelength intervals the atmosphere does not absorb according to the Lambert exponential-attenuation law (1.2.2). [Adapted from C. W. ALLEN (1973), "Astrophysical Quantities," 3rd ed., p. 130, Athlone Press, London.]

### 1.5.3 Water in the Stratosphere

The predominant carrier of radiative energy in the troposphere is $H_2O$, although $CO_2$ is also important, especially at higher altitudes. In the stratosphere and above, $H_2O$ is much less important owing to its small mixing ratio. The mixing ratio of stratospheric $H_2O$ is about $5 \times 10^{-6}$ by volume and this amount would produce a saturated vapor pressure at the low temperature of the tropical tropopause ($\sim 195°K$ at 15 km). Hence it is commonly thought that the tropopause acts as a kind of cold trap. Any higher fraction of water would eventually find condensation nuclei, condense as ice, and fall out.

## 1.6 The Mesopause: Vibrational Relaxation of $CO_2$

Above the statosphere the direct absorption of solar ultraviolet diminishes sharply because the ozone mixing ratio is down. The temperature thus monotonically decreases at an average gradient of about $-5°K/km$, quite stable in comparison with the dry adiabatic gradient (1.3.4) of $-9.8°K/km$.

The direct absorption of sunlight is thus small, and radiation transport, primarily in the 15 $\mu$m band of $CO_2$, becomes dominant in governing the temperature distribution.

### 1.6.1 The $CO_2$ $\nu_2$ Fundamental Transition

The $CO_2$ molecule is linear (see Fig. 1.12). Consequently, the $\nu_1$ mode, symmetric vibration, has no dipole moment for $^{16}O^{12}C^{16}O$, and the vibration-rotation spectrum is absent, just as it is for homonuclear diatomic molecules. The $\nu_2$ or bending mode has the lowest fundamental, occurring at 667.3 $cm^{-1}$ or 15 $\mu$m. The $\nu_2$ degeneracy (meaning the energy is equivalent for a vibration in the plane of the page and one perpendicular to it) is removed with rotation. Hence the transition is formally written $(0\,1^1\,0) \leftrightarrows (0\,0^0\,0)$,

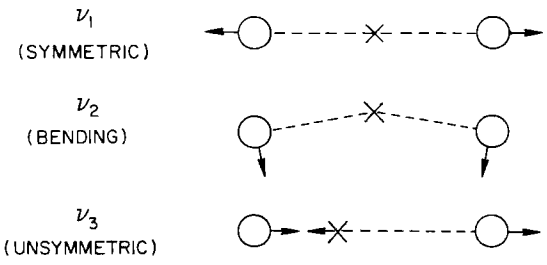



**Fig. 1.12** Vibrational modes for the linear OCO molecule.

where the principal numbers are the three vibrational quantum numbers $v_i$, and where the superscript is $l$, the quantum number specifying the *angular momentum of vibration* ($0 \leq l \leq v_2$). The $v_3$ mode, unsymmetric vibration, is $(0\,0^0\,1) \leftrightarrows (0\,0^0\,0)$, which occurs at 2350 $cm^{-1}$ or 4.3 $\mu$m and is an order of magnitude stronger in absorption. But because the 15 $\mu$m band is more readily excited at the low mesospheric temperatures, is strong enough to be optically thick, and remains in local thermodynamic equilibrium (LTE) with the atmosphere to higher altitudes, it is the more important. The latter point requires some discussion, for the breakdown of LTE occurs through the phenomenon of *vibrational relaxation.*

### 1.6.2. Kirchhoff's Law and Vibrational Relaxation

In treating radiation in the troposphere (see Section 1.2.3), where many substances (e.g., $H_2O$, $CO_2$) are radiatively active, we adopted a gray-atmosphere approximation for the infrared. Now we are concerned with a single vibrational transition for transporting heat, and it might seem that the use of Kirchhoff's law is no longer appropriate. We will demonstrate here that the approaches of thermodynamics and molecular physics are equivalent.

First, assume there is no scattering. The emitted radiation (erg/$cm^3$ sec sr Hz) from unit volume is given by the emission coefficient of (1.2.1),

$$dI_\nu/ds = \rho j_\nu \tag{1.6.1}$$

where by (1.2.6), $j_\nu = \kappa_\nu B_\nu$. Integrating over the band profile (which we cannot now regard as gray but must represent with a realistic $\kappa_\nu$ variation), we have an omnidirectional emission of energy density $E$ in the entire $v' = 1 \rightarrow v = 0$ band of

$$\frac{dE(1,0)}{dt} \equiv \int_{\text{sphere}} \int_{\text{band}} \frac{dI_\nu}{ds}\, d\nu\, d\Omega = 4\pi\rho B_\nu \int_\nu \kappa_\nu\, d\nu \tag{1.6.2}$$

where $B_\nu$ comes out of the integrals since we are concerned with isotropic emission over a small frequency interval wherein $B_\nu \approx$ const. Making use of the identity

$$\rho\kappa_\nu = N(v)\alpha_\nu(v, v') \tag{1.6.3}$$

to write the absorption coefficient in terms of the molecular cross section $\alpha_\nu(v, v')$, and the $CO_2$ population in the lower vibrational level $N(v)$, we have

$$\frac{dE(1,0)}{dt} = \frac{4\pi N(0) 2h\nu^3}{c^2(e^{h\nu/kT} - 1)} \int \alpha_\nu(0,1)\, d\nu \tag{1.6.4}$$

The integrated absorption cross section is expressed commonly in a variety of equivalent ways, including the *band strength* $S$, the *oscillator strength* $f$, the Einstein *absorption coefficient* $B$, or (since the three Einstein coefficients are uniquely related) the emission *transition probability* $A$. Thus we may choose from

$$\begin{aligned}\int \alpha_\nu(0,1)\,d\nu &= S(0,1)\\ &= \frac{\pi e^2}{mc} f(0,1)\\ &= \frac{h\nu}{4\pi} B(0,1) = \frac{c^2}{8\pi\nu^2}\frac{\tilde{\omega}(1)}{\tilde{\omega}(0)} A(1,0)\end{aligned} \tag{1.6.5}$$

where $\tilde{\omega}$ represents the *statistical weight* of a level. Opting for the transition probability, we write

$$\frac{dE(1,0)}{dt} = \frac{N(1)A(1,0)h\nu}{1 - e^{-h\nu/kT}} \text{ erg/cm}^3\,\text{sec} \tag{1.6.6}$$

since in thermodynamic equilibrium the relative populations of levels $v$ and $v'$ are given by Boltzmann's excitation equation,

$$\frac{N(v')}{N(v)} = \frac{\tilde{\omega}(v')}{\tilde{\omega}(v)} \exp\left(-\frac{h\nu(v'v)}{kT}\right) \tag{1.6.7}$$

For a diatomic molecule we can regard the vibrational levels (summed over all rotational levels) as having the same statistical weights, $\tilde{\omega}(v') = \tilde{\omega}(v)$.

From the approach of molecular physics, the rate of energy emission from $v' = 1 \to v = 0$ is

$$\frac{dE(1,0)}{dt} = N(1)[A(1,0) + B_\nu(T)B(1,0)]h\nu \tag{1.6.8}$$

where $B(1,0)$ is the *Einstein coefficient for stimulated emission*, and we take the incident intensity as Planckian and characteristic of the local temperature. The Einstein relationships are

$$\tilde{\omega}(1)B(1,0) = \tilde{\omega}(0)B(0,1) = \frac{\tilde{\omega}(1)c^2}{2h\nu^3} A(1,0) \tag{1.6.9}$$

With $B_\nu$ given by (1.2.7), Eq. (1.6.8) reduces to (1.6.6).

Of course, it is scarcely surprising that the two approaches are equivalent when one considers that we have assumed Planckian radiation and a Boltzmann distribution of level populations. These conditions are, after all, used to establish the relationships among the Einstein $A$ and $B$'s. Conversely, if the radiation field is far from isotropic or not characteristic of the local

temperature or if collisions are not frequent enough to maintain the Boltzmann distribution (1.6.7), the Kirchhoff relationship does not apply and we must look at the molecular physics to ascertain the emission.

Let us examine the collision and radiation physics to see when thermodynamic equilibrium becomes a poor approximation. Radiation is transferred to heat by absorption,

$$CO_2(v = 0) + h\nu \rightarrow CO_2(v' = 1) \tag{1.6.10}$$

followed by collisional deactivation,

$$CO_2(v' = 1) + M \rightarrow CO_2(v = 0) + M + \text{K.E.} \tag{1.6.11}$$

where K.E. ($= h\nu$) is the kinetic energy supplied to the gas. The rate coefficient for deactivation is $\eta(1,0)$ ($cm^3$/sec) so that

$$\frac{d[CO_2(1)]}{dt} = -[CO_2(1)][M]\eta(1,0)\ \text{cm}^{-3}\ \text{sec}^{-1} \tag{1.6.12}$$

On the other hand, absorption by (1.6.10) may be followed by spontaneous emission,

$$CO_2(v' = 1) \rightarrow CO_2(v = 0) + h\nu \tag{1.6.13}$$

in which case there is no radiation transferred into kinetic energy and the combination (1.6.10) plus (1.6.13) amounts to scattering. Neglecting the small stimulated-emission term in the denominator of (1.6.6), the loss of $CO_2(v' = 1)$ by spontaneous emission is

$$\frac{d[CO_2(1)]}{dt} = -[CO_2(1)]A(1,0) \tag{1.6.14}$$

which gives an albedo for single scattering of

$$\tilde{\omega}^* = \frac{A(1,0)}{A(1,0) + \eta(1,0)[M]} \tag{1.6.15}$$

which is independent of frequency within the 1-0 band.

The equation of transfer (1.2.11), with (1.2.14) and (1.2.17), is then

$$\mu \frac{dI_\nu}{d\tau_\nu} = I_\nu - \tilde{\omega}^* J_\nu - (1 - \tilde{\omega}^*)B_\nu \tag{1.6.16}$$

When $\tilde{\omega}^* \approx 0$ the coupling of the gas to the radiation field is strong; this is the state of LTE. But when $\tilde{\omega}^* \approx 1$, the upper state scatters the incident photons and the kinetic temperature of the gas is not strongly affected by the incident radiation.

The shift from $\tilde{\omega}^* \approx 0$ to $\tilde{\omega}^* \approx 1$ is called *vibrational relaxation* [meaning that vibrational populations are no longer governed by the Boltzmann

equation (1.6.7)], and it occurs at

$$[\mathrm{M}] \lesssim \frac{A(1,0)}{\eta(1,0)} \tag{1.6.17}$$

For the $\nu_2$ band of $CO_2$, the band strength is $S(0,1) = 2.37 \times 10^{-7}$ cm$^2$/sec, giving $A(0,1) = 1.35$ sec$^{-1}$. Adopting $\eta(1,0) = 2.5 \times 10^{-15}$ cm$^3$/sec, we find relaxation occurs for densities $[\mathrm{M}] \lesssim 10^{15}$ cm$^{-3}$, which occurs around 70–75 km.

Above this altitude radiative absorption is negligible compared with collisional excitation, although the gas still radiatively cools by emission of energy at the rate

$$\frac{dE(1,0)}{dt} = \eta(0,1)\tilde{\omega}^*[\mathrm{M}][\mathrm{CO}_2]h\nu \tag{1.6.18}$$

where the excitation rate coefficient is related to the deactivation coefficient (through thermodynamic considerations of detailed balancing) by

$$\tilde{\omega}(0)\eta(0,1) = \tilde{\omega}(1)\eta(1,0)e^{-h\nu/kT} \tag{1.6.19}$$

When the relative populations are not in thermodynamic equilibrium, they must in general be determined by the relative importance of collisions and the radiation field. If the local kinetic temperature is $T$ and radiation characteristic of some other temperature $T_0$ is incident from below, the excitation is

$$\frac{N(1)}{N(0)} = \frac{\tilde{\omega}(1)}{\tilde{\omega}(0)}\left((1-\tilde{\omega}^*)e^{-h\nu/kT} + \frac{\tilde{\omega}^*}{2}e^{-h\nu/kT_0}\right) \tag{1.6.20}$$

in lieu of the Boltzmann equation (1.6.7). The $\frac{1}{2}$ factor enters because the gas is irradiated only from below; the equation is based on negligible induced emission (see Problem 1.9).

The temperature of the upper mesosphere will thus decrease upwards until heating from the thermosphere offsets the radiant losses. To estimate where this occurs we must first look at the thermospheric processes.

## 1.7 Ionization, Dissociation, and Heat Transfer in the Thermosphere

Above 80 km in the Earth's atmosphere the direct absorption of solar energy again becomes important for two reasons: Photoionization, at the lower altitudes mainly by X-rays, and photodissociation in the Schumann–Runge continuum of $O_2$. (The principles of photochemistry are reviewed in Section 3.1.)

### 1.7.1 Absorption of Sunlight

The heating that occurs from photoionization involves a number of processes. First is ejection of the electron itself. For an example, let us take the $O_2$ molecule, although $N_2$, O, and minor constituents are also ionized by X-rays and NO is ionized by the strong solar Ly $\alpha$ emission of hydrogen at 1215 Å. Thus photoionization,

$$O_2 + h\nu \rightarrow O_2^+ + e \tag{1.7.1}$$

where the ionization potential is about 12 eV and the electron carries away additional kinetic energy. This energy is fed into the neutral gas by collisions that excite electronic states or high vibrational levels. A small amount of the energy is lost to space by molecular radiation, but $N_2$ and $O_2$ have no vibrational-rotational spectra, since they are *homonuclear* and have no dipole moment. Hence the excitation energy is converted to kinetic energy by energy exchange,

$$O_2(v') + M \rightarrow O_2(v < v') + M \tag{1.7.2}$$

or, through atom-atom interchange,

$$O_2(v') + O \rightarrow O + O_2(v < v') \tag{1.7.3}$$

How the energy of ionization is converted to heat depends somewhat on the altitude. Below the 90 km region recombination occurs through (a) *electron attachment* to a neutral molecule followed by (b) *mutual neutralization*, a collision between a negative and a positive ion that leaves neutral products. Most of the ionization potential is thus released in the same region where ionization occurs.

At higher altitudes negative ions do not form readily and the positive charge is predominantly on $O^+$, $O_2^+$, or $NO^+$. However, $O^+$ is very slow to recombine and sooner or later loses its charge to a molecule. Recombination is rapid for charged molecules, proceeding by, for example,

$$O_2^+ + e \rightarrow O + O \tag{1.7.4}$$

*dissociative recombination.*

Thus the difference between the ionization energy ($\sim$12 eV) and the dissociation energy ($\sim$5 eV) is released in the ionization region, but the recombination energy is released lower, in the 100 km region, where oxygen associates by

$$O + O + M \rightarrow O_2 + M \qquad k_{11} = 2.76 \times 10^{-34} e^{710/T} \tag{1.7.5}$$

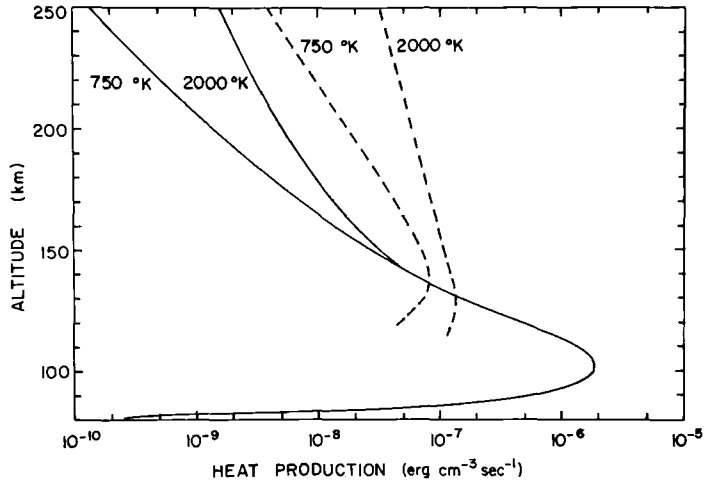



**Fig. 1.13** Deposition of heat in the upper atmosphere for an overhead sun ($\mu = 1$) and heating efficiency $\varepsilon = 0.3$, computed for two models having temperatures in the high thermosphere and exosphere of 750 and 2000° K. The curves give the second term on the right of (1.7.19). Solid curves—Schumann–Runge continuum; dashed curves—ionization by $\lambda < 1026$ Å. [Calculations by P. M. BANKS and G. KOCKARTS (1973), "Aeronomy," **B**, p. 19, Academic Press, New York.]

In the same way $O_2$ is dissociated in the Schumann–Runge continuum and somewhat higher in altitude than the main level of re-association. The vertical interchange of matter, either by molecular diffusion, large-scale wind systems, or eddy mixing, thus plays a significant role in the heat deposition. Figure 1.13 shows heating rates computed for two assumed temperature and density distributions. The Schumann–Runge region is important in the lower thermosphere, and the 80–1026 Å interval that produces ionization dominates in the middle and upper thermosphere. The heating efficiency was taken as 30 percent. The remainder of the incident energy is either radiated away or transported downward in the form of chemical energy to the region where O atoms readily associate.

### 1.7.2 Oxygen Chemistry and Distribution in the Thermosphere

The strong Schumann–Runge continuum (see Section 1.5.1) at $\lambda < 1750$ Å produces

$$O_2 + h\nu = O(^3P) + O(^1D) \tag{1.7.6}$$

with $J_2(\infty) \sim 4 \times 10^{-6}$ sec$^{-1}$. The $O(^1D)$ term lies 1.96 eV above ground. It is the upper term of the red 6300 and 6364 Å forbidden lines that appear in

both the airglow and aurora. Association occurs in the presence of a third body by (1.7.5) and by *radiative association*,

$$O + O \rightarrow O_2 + h\nu \tag{1.7.7}$$

although the latter is unimportant compared with the three-body process except well above the main dissociation region. Neglecting (1.7.7) we have an equilibrium distribution of

$$\frac{[O]}{[O_2]} = \left(\frac{J_2}{k_{11}[O_2][M]}\right)^{1/2} \tag{1.7.8}$$

However, the diffusive separation of gases becomes important in the same region (around 100–110 km) where dissociation peaks. The theoretical time for uniformly mixed $O_2$ and O to approach diffusive equilibrium at 115 km is about 10 days, which is also the lifetime of an $O_2$ molecule against photodissociation. But in the 100–115 km region, the ratio $[O_2]/[O] \approx \frac{1}{3}$ and is nearly constant. Hence the characteristic time for the lower thermosphere to become thoroughly mixed, from large scale circulation and turbulence, may also be the order of 10 days, with diffusion, mixing, and $O_2$ photolysis having comparable time scales. In the next chapter we discuss diffusion and mixing more thoroughly. For now we note that somewhere between 115 and 150 km the diffusion time (which varies as the density) becomes sufficiently short that mixing is overcome and $O_2$, O, and $N_2$ are all in diffusive equilibrium, meaning that each gas follows hydrostatic equilibrium, (1.1.5), independently for its own mass $M_j$. The atmospheric scale height is then

$$H = \left(\sum_j \frac{N_j/N}{H_j}\right)^{-1} \tag{1.7.9}$$

where $H_j = kT/M_j g$.

There is no $N_2$ continuum where solar ultraviolet radiation is strong enough to dissociate it, but N atoms are formed through photoionization,

$$N_2 + h\nu \rightarrow N_2^+ + e \tag{1.7.10}$$

which occurs for $\lambda < 796$ Å, followed by dissociative recombination,

$$N_2^+ + e \rightarrow N + N \tag{1.7.11}$$

Once formed, N can re-associate by the three-body process, but it also reacts with oxygen, forming nitric oxide by

$$N + O + M \rightarrow NO + M \tag{1.7.12}$$

and

$$N + O_2 \rightarrow NO + O \tag{1.7.13}$$

However, ion-neutral reactions are the dominant source of N atoms,

$$N_2^+ + O \rightarrow NO^+ + N \tag{1.7.14}$$

$$O^+ + N_2 \rightarrow NO^+ + N \tag{1.7.15}$$

and

$$O_2^+ + N_2 \rightarrow NO^+ + NO \tag{1.7.16}$$

although the latter reaction is probably very slow. Dissociative recombination of $NO^+$ yields the N atom preferentially in the $^2D$ state.

A different type of production mechanism involves $N(^2D)$ formed by recombination, (1.7.11). The $^2D$ term is metastable, at 2.37 eV above ground, $N(^4S)$. Once formed it can react with $O_2$,

$$N(^2D) + O_2 \rightarrow NO + O(^3P) \tag{1.7.17}$$

Nitric oxide formed in the thermosphere is important to the formation of the lower ionosphere, and we shall return to it in Chapter 5.

### 1.7.3 Radiative Exchange and Loss

Some of the processes mentioned above emit radiation as they occur. Electron attachment to form a negative ion may proceed either by emitting a stabilizing photon or in the presence of a third body. *Three-body association*, (1.7.5), may leave $O_2$ in the ground state, $X\,^3\Sigma_g^-$, but it can just as well associate into one of the other (excited) states that correspond to two ground-term $O(^3P)$ atoms (see Fig. 1.9). In that case the excited molecule might emit radiation in one of the $O_2$ band systems observed in the airglow (see Section 6.1.2), and this loss represents energy not available for thermospheric heat.

The net heating efficiency must also include the loss of infrared radiation due to thermal collisions, such as the $CO_2$ loss above its level of vibrational relaxation. But the abundant molecules $N_2$ and $O_2$ are forbidden to radiate in the infrared by electric-dipole transitions and the minor constituents that can radiate (NO, CO, etc.) are too rare in the Earth's atmosphere to account for much loss, although CO rotational spectra are probably important in cooling the high atmospheres of Mars and Venus.

The principal radiative loss for the upper thermosphere of Earth is a forbidden (by electric-dipole radiation) transition between two of the fine-structure levels, $J = 1 \rightarrow J = 2$, of the ground term of $O(^3P_{2,1,0})$ (cf. Fig. 6.3). The transition occurs at 63 $\mu$m. The level separation (0.02 eV) is comparable to atmospheric thermal energies (0.02 eV = 232°K). Because the radiative lifetime is so long ($1/A = 3.1$ hr) compared with the time between thermal

collisions, the levels of the O ground term are in a Boltzmann distribution. Hence the emission rate at 63 $\mu$m is

$$\frac{dE(1,2)}{dt} = [\mathrm{O}] \frac{\tilde{\omega}_1 \exp(-E_1/kT)}{\sum_{J=0}^{2} \tilde{\omega}_J \exp(-E_J/kT)} A(1,2)E_1 \tag{1.7.18}$$

where $E_J$ is the energy of the $J$th level above ground ($J = 2$). The cooling effect of this radiation is important, however, only above the level where the transition ceases to the optically thick and this consideration limits the cooling to the $F$ region (see Problem 1.7). The $J = 0 \rightarrow J = 1$ transition at 147 $\mu$m also occurs, but is less important.

### 1.7.4 Conductive Flow of Heat and Location of the Mesopause

With a positive temperature slope, preventing convection, and a not very effective exchange of heat by radiation, the primary energy transfer in the lower thermosphere is by conduction. The instantaneous rate of heating is

$$\rho \frac{\partial}{\partial t}(C_p T) = \frac{\partial}{\partial z}\left(K \frac{\partial T}{\partial z}\right) + \sum_i \pi \mathscr{F}_i \varepsilon_i \alpha_i N(z) e^{-\tau_i(z)/\mu} - \mathscr{R}(z) \tag{1.7.19}$$

We here divide the solar flux into a number of spectral intervals $\pi\mathscr{F}_i$, chosen so that the absorption coefficient has little variation within each one. Thus $\pi\mathscr{F}_i$ is the solar energy flux (erg/cm$^2$ sec) within each interval, $\varepsilon_i$ is the fraction of absorbed solar energy that appears locally as heat, and $\tau_i(z)$ is the vertical optical thickness in the ultraviolet intervals. With $\alpha_i$ the mean absorption cross section (cm$^2$/molecule), we have $d\tau_i = -N(z)\alpha_i dz$. Also $K(T)$ is the thermal conductivity (erg/cm sec°K) and $\mathscr{R}(z)$ is the radiant heat loss (erg/cm$^3$ sec).

In a steady state, averaged over a day, this equation integrates to

$$K \frac{dT}{dz} = \tfrac{1}{2} \sum_i \pi \mathscr{F}_i \varepsilon_i \langle \mu \rangle \left[1 - \exp\left(\frac{-\tau_i}{\langle \mu \rangle}\right)\right] - \int_z^\infty \mathscr{R}(z)\, dz \tag{1.7.20}$$

which is equivalent to the downward conductive heat flux. The use of an average solar zenith angle, $\cos^{-1}\langle \mu \rangle$, is an approximation (as is the factor $\frac{1}{2}$ for nighttime). The flux should actually be integrated through the day with $\mu = \mu(t)$.

It is clear that at high altitude the right side vanishes and $T \rightarrow T_{ex}$, the exospheric temperature. At other heights $z$, the gradient must be large enough to sustain a downward flow of heat equivalent to the total ultraviolet energy absorbed above $z$ less the infrared loss.

Figure 1.13 shows where the energy is absorbed in the two intervals 80–1026 Å and the Schumann–Runge region, $\lambda < 1750$ Å. The ionization energy varies with solar activity, but is the order of 2 erg/cm$^2$ sec. The Schumann–Runge flux is the order of 15 erg/cm$^2$ sec. The radiant loss from 63 $\mu$m radiation of O is probably only a few tenths of an erg/cm$^2$ sec, although that could be important above 150 km. The high altitudes are well insulated from the mesosphere by the thick thermosphere through which heat must flow to escape. Consequently, small changes in the heat absorbed or emitted can have a big impact on the local temperature.

A good approximation for $K(T)$ at atmospheric temperatures is

$$K(T) = 36 \times 10^{-5} T^{3/4} \text{ erg km/cm}^2 \text{ sec°K} \tag{1.7.21}$$

for $O_2$ and $N_2$ and

$$K(T) = 54 \times 10^{-5} T^{3/4} \text{ erg km/cm}^2 \text{ sec°K} \tag{1.7.22}$$

for O. The net heat flux deposited above 120 km, where $T \approx 325$°K, is therefore

$$K(T) \frac{dT}{dz} = \tfrac{1}{2}\pi \mathscr{F} \varepsilon \langle \mu \rangle \gtrsim 0.4 \text{ erg/cm}^2 \text{ sec} \tag{1.7.23}$$

in order to support a gradient of at least 15°K/km, higher gradients occurring with high solar activity. Thus the ionization energy alone deposited above 120 km can probably account for the gradient with, say, $\langle \mu \rangle = \frac{1}{2}$ and $\varepsilon = 0.8$.

The gradient at 90 km is only some 10°K/km, but a glance at Fig. 1.13 shows that the overhead heat input has increased manyfold. Even with some differences in efficiency, the conclusion is clear that radiative loss is important in the 100 km region. The location of the mesopause is given by the altitude at which the entire thermospheric flux has been radiated away by $CO_2$, at 15 $\mu$m.

With $CO_2$ in vibrational relaxation, Eqs. (1.6.18) and (1.6.19) give the mesopause condition as

$$f(CO_2)\eta(1,0)h\nu \int_{z_0}^{\infty} N^2(z) e^{-h\nu/kT}\, dz = \tfrac{1}{2}\pi \mathscr{F} \varepsilon \langle \mu \rangle \tag{1.7.24}$$

where $f(CO_2)$ is the mixing ratio of $CO_2$ ($3 \times 10^{-4}$) and $\eta(1,0) = 2.5 \times 10^{-15}$ cm$^3$/sec. The conductive flux on the right is about 3.4 erg/cm$^2$ sec. Evaluating the integral in the isothermal approximation (1.1.4) at $T_0 = 175$°K (the solution is not sensitive to $T_0$) we find

$$N(z_0) = 1.3 \times 10^{14} \text{ cm}^{-3} \tag{1.7.25}$$

which occurs around 85 km.

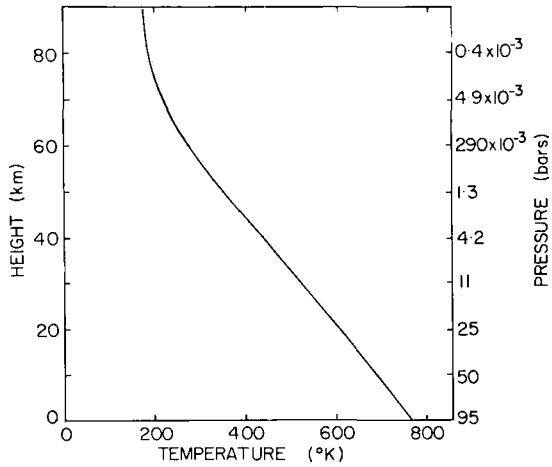



**Fig. 1.14** Temperature profile of Venus' atmosphere. [This model is the NASA 1972 standard atmosphere, NASA Report S.P. 8001.]

## 1.8 Atmospheric Structure of Venus, Mars, and Jupiter

### 1.8.1 Venus

The high surface temperature (Fig. 1.14) of Venus (750°K) was first detected from its emission in the microwave radio spectrum, to which the atmosphere is nearly transparent. The thick atmosphere contains nearly 100 atmospheres of $CO_2$, and the high temperature is evidently due to a greenhouse effect, although only a small percent of the solar flux is deposited at the ground.

The temperature profile was measured by the radio occultation experiments (cf. Section 5.2.2) aboard Mariner 5 and by *in situ* sensing on Venera 8 probes. The profile is essentially adiabatic from the ground to 50 km. Horizontal temperature contrasts are quite small, both in the high and deep atmosphere, indicating dynamical activity (see Section 2.4).

A layer of clouds exists around 50 to 70 km. Polarization studies show their composition to be mainly $H_2SO_4$, sulfuric acid (see Section 4.3.4). The high reflectivity of the clouds accounts for the approximate 80 percent albedo of the planet. The clouds are moderately thick extending probably over some 20 or 30 kilometers and in the visual have a total optical thickness $\tau \gg 1$.

Above the clouds a stratosphere has not been found nor is one especially expected. The thermosphere (Fig. 1.15) rises to temperatures of around 300°K (Section 7.3.3).

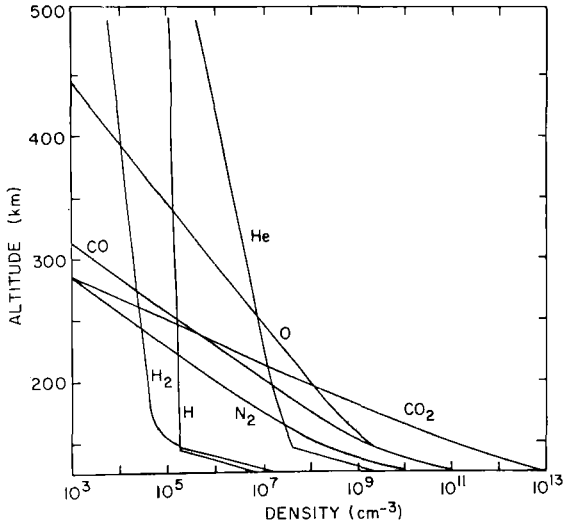



**Fig. 1.15** Model for the upper atmosphere of Venus with an exospheric $T = 350°K$. Fractional abundances are $CO_2(0.98)$, $CO(0.01)$, $O(0.01)$, $N_2(10^{-3})$, $H_2(2 \times 10^{-6})$, and $He(2 \times 10^{-4})$. Total density is $N = 1.0 \times 10^{13}$ cm$^{-3}$ at $z = 125$ km; eddy diffusion coefficient is $K = 10^8$ cm$^2$/sec (see Section 2.3.2) below 145 km. [Based on S. KUMAR and D. M. HUNTEN (1974), *J. Geophys. Res.* **79**, 2529.]

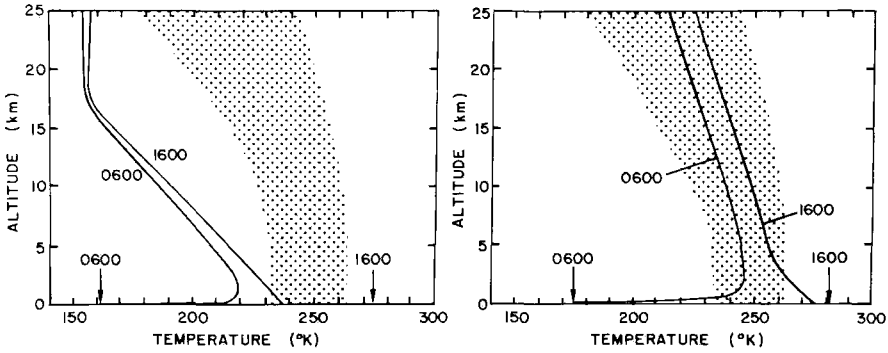



**Fig. 1.16** Temperature profiles for Mars showing the effect of direct solar heating due to dust absorption. The shaded areas show the region of Mariner 6 and 7 observations. The curves at the left are calculated for a clear $CO_2$ atmosphere and at the right for one containing dust. [Calculations are due to GIERASCH and GOODY (1972).]

### 1.8.2 Mars

Figure 1.16 shows a shaded area that includes the temperature measurements with height at various positions and times of day for dusty conditions. The temperature gradient is not even close to the predicted adiabatic lapse rate of $-5°K/km$ for a clear $CO_2$ atmosphere in radiative-convective equilibrium. The actual temperature profile is due mainly to absorption of direct solar radiation by dust in the atmosphere. The effect is important even when the atmosphere appears to be clear; although the lapse rate is more nearly adiabatic for clear conditions, it is seldom steeper than $-3°K/km$. Models of the large-scale circulation (Section 2.5) also give a smaller (numerical) lapse rate than does convection.

Because the surface pressure is so low ($\lesssim 8$ mbar depending on geography and season) the greenhouse effect is weak, and the surface temperatures undergo large diurnal variations during the Martian day of 24.660 hr (Fig. 1.17).

The amount of water vapor in the atmosphere appears to be established by the saturation vapor pressure at night. For the relatively clear conditions of Fig. 1.17, $T_{min} \leq 190°K$ corresponding to 10 $\mu$m of precipitable water. As much as 30 $\mu$m has been observed over the north polar cap during northern hemisphere spring when $H_2O$ was probably being released by thawing ice.

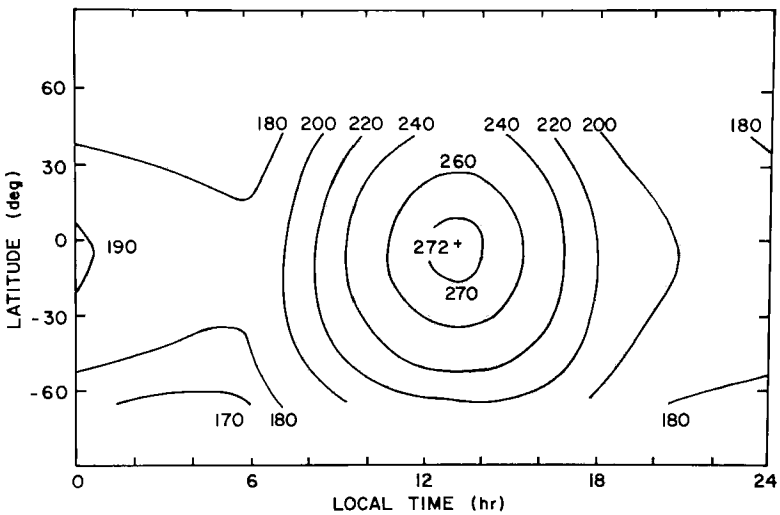



**Fig. 1.17** Surface temperatures on Mars, calculated for a clear atmosphere based on Mariner 9 infrared radiometer observations. [Adapted from KIEFFER *et al.* (1973).]

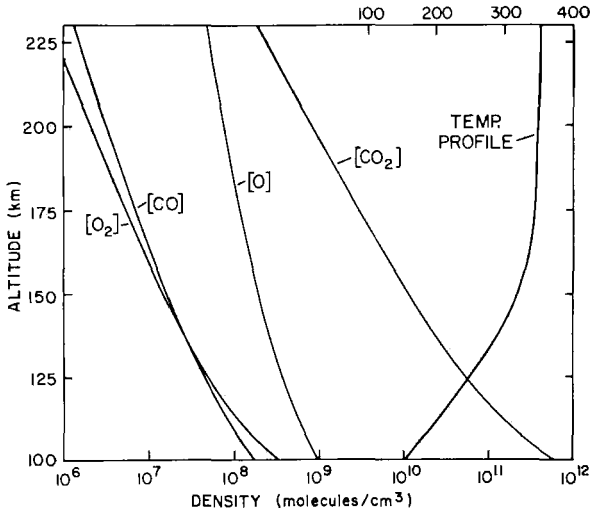



**Fig. 1.18** Model for the upper atmosphere of Mars, with an exospheric temperature of 350°K. The model is based on radio-occultation and ultraviolet airglow observations with Mariner 9. [After C. A. BARTH *et al.* (1972), *Icarus* **17**, 457.]

Figure 1.18 shows a model upper atmosphere based on Mariner 9 observations. The exospheric temperature is about 350°K. (Both the Viking spacecraft found much lower temperatures; there seems to be a real variability, but it is not understood.)

### 1.8.3 Jupiter

Figure 1.19 gives a schematic picture of the temperature profile on Jupiter. The adiabatic lapse rate (see Section 2.6) below the solid $NH_3$ (ammonia) crystals that form the highest clouds is $-3$°K/km. The region of radiation equilibrium in the troposphere involves radiation exchange in the weak bands of $CH_4$ and solar absorption in the ultraviolet, probably by a "dust"—possibly condensed $N_2H_4$ (hydrazine), a by-product of ammonia photochemistry.

The mesosphere arises from solar heating in the 3.3 $\mu$m band ($\nu_3$ fundamental) of $CH_4$ (methane) with subsequent re-radiation in the 7.7 $\mu$m ($\nu_4$ fundamental) $CH_4$ band, which has a brightness in the Jovian spectrum corresponding to 145°K or more. The emission intensity of the 7.7 $\mu$m band and limb darkening (Section 4.3.3) in the infrared continuum are the critical data that radiative-equilibrium models must satisfy. The effective global temperature of Jupiter is 134°K although the equilibrium temperature of

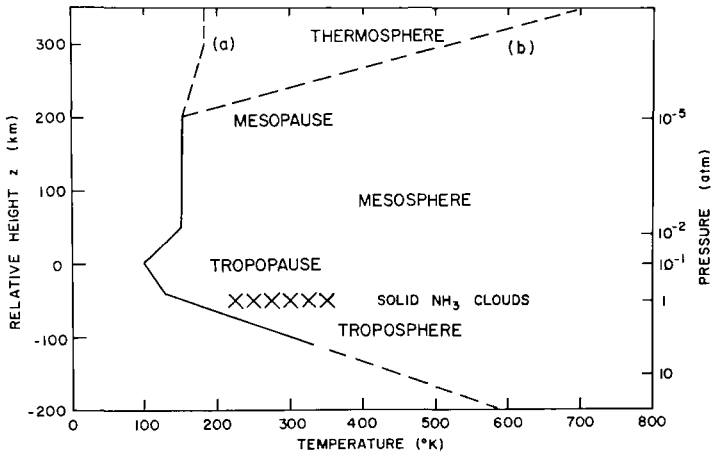



**Fig. 1.19** Schematic structure of the Jovian atmosphere. Even the adiabatic curve for the troposphere is uncertain, since it depends on the $H_2$/He mixture and the role of moist convection; also at low $T$, $\gamma$ for $H_2$ is sensitive to $T$. Here we have taken a lapse rate of 2.9°K/km corresponding to $\gamma \approx 1.6$. The tropopause occurs at a minimum $T$ of 100–120°K. A stratosphere is not shown because the profile is not known accurately enough for one to be sure that there is a temperature decrease above the tropopause. In the region just above the tropopause, $T$ probably increases because of direct solar absorption, with a low level of infrared activity providing radiative cooling. At the mesopause the main vibrational relaxation occurs in $C_2H_2$ and $CH_4$. In the thermosphere, curve (a) is the expected profile for heating by solar extreme ultraviolet and curve (b) is a profile inferred from the Pioneer 10 radio occultation. The discrepancy presumably indicates an additional heat source for the atmosphere (see Section 5.3.3).

Eq. (1.2.46) is only 106°K, the difference evidently being due to Jupiter's self-radiation as it continues its *Helmholtz contraction*, the early gravitational collapse experienced by a protostar.

Of the infrared active molecules in the upper atmosphere, $C_2H_2$ (acetylene) is third in abundance, and since it has a low lying vibrational level ($\nu_5$ fundamental) that radiates at 13.7 $\mu$m, its emission plays an analogous role to the 15 $\mu$m band of $CO_2$ in the Earth's mesosphere. The Jovian mesosphere is thought to be nearly isothermal ($\sim$150°K) possibly decreasing by 10°K from a stratopause to the mesopause.

The thermosphere was formerly thought on theoretical grounds to increase only about 15°K from the mesopause up to the exosphere. Compared with the radiation governing the Earth's thermosphere the solar ionizing flux is down a factor of 27 at Jupiter, the thermal conductivity of the $H_2$ thermosphere is higher, and the scale heights are much less (decreasing the thermal insulation that allows high temperatures to be achieved). The latter point is really a feedback factor: If the temperature does not build up in the

lower thermosphere, the scale height cannot increase to a value that will insulate the upper thermosphere. However, a surprise lay in store.

The radio occultation experiments that were so useful in obtaining atmospheric pressure-temperature curves for Venus and Mars did not yield sound results for the neutral atmosphere when flown past Jupiter on Pioneers 10 and 11. (Subsequent analysis indicated that errors were magnified greatly by problems involving the planet's oblateness and technical aspects of the experiment's equipment.) Nevertheless, the ionospheric results should be reasonably accurate, and they show the thermosphere rising steeply to about 750°K. Clearly, if this result holds up, there is an unexpected heat source for the high atmosphere. We will defer further discussion to Section 5.3.3.

## BIBLIOGRAPHICAL NOTES

**Section 1.2 Radiative Equilibrium**

Atmospheric radiation is the subject of three major books:

GOODY, R. M. (1964), "Atmospheric Radiation," Clarendon Press, Oxford,

is mainly a theoretical account of absorption and heat transfer by atmospheric gases.

KONDRATYEV, K. YA. (1969), "Radiation in the Atmosphere," Academic Press, New York,

deals more with measurements, radiation scattering, global heat balance, and so forth.

ZUEV, V. E. (1974), "Propagation of Visible and Infrared Radiation in the Atmosphere," Wiley, New York,

treats absorption and scattering as well, but the author devotes the final third of his book to the propagation of laser beams.

Radiative transfer is treated extensively in all the above books, but the standard reference is

CHANDRASEKHAR, S. (1950), "Radiative Transfer," Clarendon Press Oxford, (reprinted by Dover, New York, 1960).

The common belief that the heating of a greenhouse is due to the inhibition of convection by the roof can be traced to experiments described in 1909 by R. W. WOOD. This view has been challenged by

SILVERSTEIN, S. D. (1976), Effect of infrared transparency on the heat transfer through windows: A clarification of the greenhouse effect, Science **193**, 229–231,

who finds that the infrared radiative blanketing is not trivial and the physical analogy with the atmosphere is fairly appropriate after all.

Radiative–convective equilibrium has been the subject of many computer modeling studies. Notable as one of the first thorough attempts along these lines is

MANABE, S. and STRICKLER, R. F. (1964), Thermal equilibrium of the atmosphere with a convective adjustment, *J. Atmos. Sci.* **21**, 361-385.

A thorough review of modeling techniques and results is

RAMANATHAN, V. and COAKLEY, J. A., JR. (1978), Climate modeling through radiative-convective models, *Rev. Geophys. Space Phys.*, **16**, No. 4 (Nov).

Some of the modelling has been concerned with climatic variations due to compositional changes (see Section 7.5).

**Section 1.3 Convection in the Troposphere**

The breakdown of radiative equilibrium by convective activity was first noted by R. EMDEN in 1913. The surge of theoretical activity following discovery of the stratosphere in 1900 is documented by GOODY (op. cit. Chapter 8).

**Section 1.4 Latitude Variations of the Tropopause**

The problem of the latitude variation of the tropopause and its dependence on departures from grayness was treated by

GOODY, R. M. (1949), The thermal equilibrium at the tropopause and the temperature of the lower stratosphere, *Proc. Roy. Soc.* **A 197**, 487–505.

**Section 1.5 The Stratosphere**

The photochemical theory for oxygen in the stratosphere is due to

CHAPMAN, S. (1930a), A theory of upper-atmospheric ozone, *Mem. R. Meteorol. Soc.* **3**, 103–125.

CHAPMAN, S. (1930b), On ozone and atomic oxygen in the upper atmosphere, *Phil. Mag.* **10**, 369–383.

CHAPMAN, S. (1931a), Some phenomena of the upper atmosphere (Bakerian Lecture), *Proc. Roy. Soc.* **A 132**, 353–374.

and to

WULF, O. R. and DEMING, L. S. (1936a), The theoretical calculation of the distribution of photochemically-formed ozone in the atmosphere, *Terr. Magn.* **41**, 299–310.

WULF, O. R. and DEMING, L. S. (1936b), The effect of visible solar radiation on the calculated distribution of atmospheric ozone, *ibid.* pp. 375–378.

WULF, O. R. and DEMING, L. S. (1937), The distribution of atmospheric ozone in equilibrium with solar radiation and the rate of maintenance of the distribution, *ibid.* **42**, 195–202.

The concept that ionizing or dissociating radiation will create "layers" in the atmosphere was first stated quantitatively in

LENARD, P. (1911), Über die Absorption der Nordlichtstrahlen in der Erdatmosphäre, *SitzungsBer. d. Heidelberger Akad. d. Wissenschaften* **12**, 1–9,

and developed more fully for electromagnetic radiation in the first full-length book on radio propagation,

PEDERSON, P. O. (1927), "The Propagation of Radio Waves," Copenhagen, Danmarks Naturvidenskabelige Samfund **A**, Nr. 15a, 15b.

However, it was CHAPMAN who developed the theory to an incredible degree of completeness, including the numerical integration of the equations for a rotating Earth. The "Chapman-layer" papers are

CHAPMAN, S. (1931b), The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, *Proc. Phys. Soc.* **43**, 26–45.

CHAPMAN, S. (1931c), The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, *Proc. Phys. Soc.* **43**, 483–501.

It was first proposed that the water vapor at the tropopause is saturated and that this cold trap limits the abundance of $H_2O$ at higher altitudes by E. GOLD in 1909 in a report to the British Association in Winnipeg. The concept is developed in

GOLD, E. (1913), The international kite and balloon ascents, *Geophys. Mem.* **1**, No. 5, M.O. 210e (cf. p. 129ff.)

From a series of 16 ascents into the stratosphere between 1943 and 1945 during which humidity measurements were made, BREWER concluded that the low temperature of the tropical tropopause provides the cold trap and that meridional circulation accounts for the low humidity and high ozone content at higher latitudes:

BREWER, A. W. (1949), Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere, *Quart. J. Roy. Meteorol. Soc.* **75**, 351–363.

**Section 1.6 The Mesopause**

The concept of the cold mesopause being strongly influenced by the onset of vibrational relaxation was developed by

CURTIS, A. R. and GOODY, R. M. (1956), Thermal radiation in the upper atmosphere, *Proc. Roy. Soc.* **A 236**, 193–206.

**Section 1.7 The Thermosphere**

The importance of conduction in maintaining the high temperatures in the thermosphere (which at that time could be only indirectly inferred) was first demonstrated in a classic paper by

SPITZER, L. (1949), The terrestrial atmosphere above 300 km, in "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), pp. 211–247, Univ. Chicago Press, Chicago, [2nd ed., 1952].

Thermospheric theory has since been developed extensively, most notably by

BATES, D. R. (1951), The temperature of the upper atmosphere, *Proc. Phys. Soc.* **B 64**, 805–821.

BATES, D. R. (1959), Some problems concerning the terrestrial atmosphere above about the 100 km level, *Proc. Roy. Soc.* **A 253**, 451–462,

NICOLET, M. (1960), The properties and constitution of the upper atmosphere, *in* "Physics of the Upper Atmosphere," (J. A. Ratcliffe, ed.), pp. 17–71, Academic Press, New York.

The heating of an upper atmosphere by the dissipation of mechanical energy is discussed and compared with radiation sources and sinks by

LEOVY, C. B. (1975), Some energy sources and sinks in the upper atmosphere, *in* "Atmospheres of Earth and the Planets," (B. M. McCormac, ed.), pp. 73–86, D. Reidel Publ. Co., Dordrecht, The Netherlands.

The treatment here of the mesopause condition, defining the mesopause as the height where vibrationally relaxed $CO_2$ has emitted the entire conductive flux drained from the thermosphere, follows

CHAMBERLAIN, J. W. (1962), Upper atmospheres of the planets, *Astrophys. J.* **136**, 582–593.

**Section 1.8.1 Venus**

Radio emission from Venus and its interpretation in terms of the temperature and pressure at the planet's surface are discussed by

BARRETT, A. H. and STAELIN, D. H. (1964), Radio observations of Venus and the interpretations. *Space Sci. Rev.* **3**, 109–135,

WOOD, A. T., JR.; WATTSON, R. B.; and POLLACK, J. B. (1968), Venus: Estimates of the surface temperature and pressure from radio and radar measurements, *Science* **162**, 114–116.

Result of the Soviet Venera experiments have been summarized in

AVDUEVSKY, V.; MAROV, M. YA.; and ROZHDESTVENSKY, M. (1970), A tentative model of the atmosphere of planet Venus based on the results of measurements of spaceprobes Venera 5 and Venera 6, *J. Atmos. Sci.* **27**, 561–568.

MAROV, M. YA.; AVDUEVSKY, V.; KERZHANOVICH, V.; ROZHDESTVENSKY, M.; BORODIN, N.; and RYABOV, O. (1973), Venera 8: Measurements of temperature, pressure, and wind velocity on the illuminated side of Venus, *J. Atmos. Sci.* **30**, 1210–1214,
AVDUEVSKY, V.; MAROV, M. YA.; MOSHKIN, B. E.; and RONOMOV, A. E. (1973), Venera 8: Measurements of solar illumination through the atmosphere of Venus, *J. Atmos. Sci.* **30**, 1215–1218.

The Mariner observations of Venus' temperature profile are reported in

FJELDBO, G.; KLIORE, A.; and ESHLEMAN, V. (1971), The neutral atmosphere of Venus as studied with the Mariner 5 radio occultation experiments, *Astron. J.* **76**, 123–140,
HOWARD, H. T., *et al.* (1974), Venus: Mass, gravity field, atmosphere, and ionosphere as measured by the Mariner 10 dual-frequency radio system, *Science* **183**, 1297–1301.

The temperature of the upper atmosphere of Venus is examined theoretically by

DICKINSON, R. E. (1976), Venus mesosphere and thermosphere temperature structure, Icarus **27**, 479–493.

**Section 1.8.2 Mars**

Infrared radiometer and radio occultation experiments on the Martian surface and atmosphere by Mariner 4 (1965), Mariners 6 and 7 (1969), Mariner 9 (1971–1972), and the Soviet Mars 2 and 3 (1971–1972) have been reviewed in

BARTH, C. A. (1974), The atmosphere of Mars, *Ann. Rev. Earth Planet. Sci.* **2**, 333–367.

The analysis of the low-atmosphere temperatures in terms of heating by dust absorption is due to

GIERASCH, P. J., and GOODY, R. M. (1972), The effect of dust on the temperature of Martian atmosphere, *J. Atmos. Sci.* **29**, 400–402.

Sirface temperatures measured by Mariner 9 have been analyzed in

KIEFFER, H. H.; CHASE, S. C.; MINER, E.; MÜNCH, G.; and NEUGEBAUER, G. (1973), Preliminary report on infrared radiometric measurements from the Mariner 9 spacecraft, *J. Geophys. Res.* **78**, 4291–4312.

The upper atmosphere model given here is from

BARTH, C. A.; STEWART, A. I.; HORD, C. W.; and LANE, A. L. (1972), Mariner 9 ultraviolet spectrometer experiment: Mars airglow spectroscopy and variations in Lyman alpha, *Icarus* **17**, 456–468.

**Section 1.8.3 Jupiter**

A radiative-transfer analysis for deriving the structure of the troposphere and mesosphere of Jupiter is given in

WALLACE, L.; PRATHER, M.; and BELTON, M. J. S. (1974), The thermal structure of the atmosphere of Jupiter, *Astrophys. J.* **193**, 481–493.

The mesosphere and thermosphere results quoted here are from

STROBEL, D. F., and SMITH, G. R. (1973), On the temperature of the Jovian atmosphere, *J. Atmos. Sci.* **30**, 718–725.

The problems associated with the use of radio occultation experiments to obtain temperature profiles on Jupiter are discussed in

HUBBARD, W. B.; HUNTEN, D. M.; and KLIORE, A. (1975), Effect of the Jovian oblateness on Pioneer 10/11 radio occultations, *Geophys. Res. Lttrs.*, **2**, 265–268,
HUNTEN, D. M. and VEVERKA, J. (1976), Stellar and spacecraft occultations by Jupiter: A critical review of derived temperature profiles, *in* "Jupiter," ed. T. Gehrels, pp. 247–283, Univ Arizona Press, Tucson,

ESHLEMAN, V. R. (1975), Jupiter's atmosphere: Problems and potential of radio occultation, *Science* **189**, 876–878.

## PROBLEMS

**1.1** *Diffusive equilibrium.* Show that for an atmosphere in diffusive equilibrium the atmospheric scale height is

$$H = \left(\sum_j \frac{N_j/N}{H_j}\right)^{-1}$$

where $H_j = kT/M_j g$.

**1.2** *Linear variation of scale height.* Over a region where the temperature changes linearly with height and where $g(r) \approx$ const, show that

$$\frac{p}{p_0} = \left(\frac{H}{H_0}\right)^{-1/\beta} \quad \text{and} \quad \frac{N}{N_0} = \left(\frac{H}{H_0}\right)^{-(1+\beta)/\beta}$$

where $\beta = dH/dr$ and $p_0$, $N_0$, and $H_0$ denote values at a starting distance $r_0$.

**1.3** *Radiative equilibrium.* Assume that solar radiation is absorbed only at the Earth's surface where the albedo is 40%. The re-radiated energy is absorbed mainly by water vapor, which we approximate as a gray absorber with a density scale height of 2 km and a total $\tau_g = 2$. Plot the temperature distribution with height for radiative equilibrium. What is the temperature discontinuity at the ground? What is the gradient in the air temperature near the ground?

**1.4** *Chapman Layer.* Show that the maximum absorption in a barometric atmosphere occurs at $\tau_\nu/\mu = 1$ [Eq. (1.5.6)].

**1.5** *Chapman stratospheric chemistry.* With the Chapman reactions, what is the daytime rate of change of $O_2$? Show that this equation adds no new information, beyond what is given by the analogous equations for O and $O_3$, but merely confirms the conservation of matter.

**1.6** *Stratospheric heating.* (a) With the data given below and model atmosphere in Appendix III, find $J_2(z)$ and $J_3(z)$. (b) Use the $J$'s to recompute $[O_3]$ with the "gray u.v." approximation and Chapman chemistry and compare the results for self-consistency with $[O_3]$ in the assumed model. (c) What are the heating rates in degrees per day? What portion arises from $O_2$ dissociation? (d) Assuming a transparent "gray i.r." atmosphere, compute the temperatures and compare with the model. Show that the $\kappa_\nu$ assumed is consistent with the assumption of optical thinness. Data provided:

$$\alpha_2(O_3) = 1 \times 10^{-18} \text{ cm}^2 \quad (1800\text{–}2400 \text{ Å})$$
$$\alpha_2(O_2) = 5 \times 10^{-24} \text{ cm}^2$$
$$\alpha_3(O_3) = 5 \times 10^{-18} \text{ cm}^2 \quad (2200\text{–}2900 \text{ Å})$$
$$J_2^{(0)} = 6 \times 10^{-10} \text{ sec}^{-1}$$
$$J_3^{(0)} = 4 \times 10^{-3} \text{ sec}^{-1}$$
$$k_{12} = 1.1 \times 10^{-34} e^{510/T} \text{ cm}^6/\text{sec}$$
$$k_{13} = 1.1 \times 10^{-11} e^{-2150/T} \text{ cm}^3/\text{sec}$$
$$\langle \mu \rangle = \tfrac{1}{2}$$
$$\kappa_\nu(\text{i.r.}) = 1 \times 10^{-3} \text{ cm}^2/\text{gm}$$

**1.7** *Thermospheric forbidden oxygen emission.* The triplet ground term, $^3P_{2,1,0}$, of O has the $J = 1$ level at 0.0197 eV and $J = 0$ at 0.0281 eV. The transition probabilities are $A(1,2) = 8.95 \times 10^{-5}$ sec$^{-1}$ and $A(0,1) = 1.70 \times 10^{-5}$ sec$^{-1}$. (a) What is the rate of emission in the two transitions at 200 km? Take $T = 915$°K and $[O] = 4 \times 10^9$ cm$^{-3}$. (b) What is the overhead optical thickness in the centers of each of the two lines? Assume pure Doppler broadening.

**1.8** *Mesopause condition.* Compute and make a log plot of the 15$\mu$m $CO_2$ thermal emission rate in the 70–100 km region with the model in Appendix III. Estimate a slope for the curve around 85–90 km and use it to integrate the emission. Find a value of $z_0$ that satisfies the mesopause condition and compare with the solution (1.7.25).

**1.9** *Vibrational relaxation.* Write the equation for statistical equilibrium for the excited vibrational level of the $\nu_2$ fundamental of $CO_2$ at temperature $T$. Assume that it is illuminated from below by black-body radiation at temperature $T_0$ and allow for induced emission as well as absorption. Show that in the limit of low $T_0$ and low local densities, the population is given by (1.6.20).

**1.10** *Analytic model thermosphere* [after Bates (1959)]. (a) The *geopotential distance* above height $z_0$ is defined as

$$\zeta = \int_{z_0}^{z} \frac{g(z)}{g(z_0)} dz$$

Show that the height may be written as

$$z - z_0 = \frac{\zeta}{1 - \dfrac{\zeta}{r_0 + z_0}}$$

where $r_0$ is the radius of the planet. (b) Let the temperature be represented by the formula

$$T(z) = T(\infty)[1 - a\exp(-\tau\zeta)]$$

where $a$ and $\tau$ are specified constants:

$$a \equiv 1 - \frac{T(z_0)}{T(\infty)}$$

$$\tau \equiv \frac{(dT/dz)z_0}{T(\infty) - T(z_0)}$$

Show by integrating (1.1.5) that the number density of a constituent in diffusive equilibrium is

$$N(z) = N(z_0)\left(\frac{1-a}{\exp(\tau\zeta) - a}\right)^{1+\gamma} \exp(\tau\zeta)$$

where

$$\gamma = \frac{1-a}{\tau H_0}$$

and $H_0$ is the pressure scale height of the constituent at $z_0$. (c) Take $T(\infty) = 1000°K$ and at $z_0 = 120$ km take the O density as $N_0 = 9.3 \times 10^{10}$ cm$^{-3}$, $T(z_0) = 360$ and $(dT/dz)_{z_0} = 12$°K/km. Compute the temperature and O density at $\zeta = 79$ and 359 km and compare with the model in Appendix III.

# *Chapter 2*
# HYDRODYNAMICS OF ATMOSPHERES

## 2.1 Basic Equations

### 2.1.1 Equations in an Inertial Frame

The basic equations relating mass density $\rho$, pressure $p$, temperature $T$, and the three velocity components of $\mathbf{v}$ $(u, v, w)$ are the equations of continuity, motion, state, and energy conservation. The continuity equation expresses the conservation of mass,

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}) \tag{2.1.1}$$

which states that the local rate of increase in mass density is equivalent to the fluid's "convergence." Alternatively, it is common to write continuity in terms of the *substantial derivative*, taken along the path followed by an element of fluid. Thus

$$\frac{d\rho}{dt} \equiv \frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho = -\nabla \cdot (\rho \mathbf{v}) + \mathbf{v} \cdot \nabla \rho \tag{2.1.2}$$

Expanding the divergence we have

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v} \tag{2.1.3}$$

In the special case of an incompressible fluid,

$$\nabla \cdot \mathbf{v} = 0 \tag{2.1.4}$$

*Euler's equation* of motion for an ideal gas (i.e., one with no viscosity) is established by equating the force (per unit volume) on a fluid to the external force ($\rho\mathbf{g}$) plus the pressure gradient:

$$\rho \frac{d\mathbf{v}}{dt} \equiv \rho\left(\frac{\partial \mathbf{v}}{dt} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right)$$
$$= -\nabla p + \rho\mathbf{g} \tag{2.1.5}$$

If the fluid is not frictionless, we must add a term accounting for the viscous transfer of momentum from one element of fluid to another. The viscosity stress tensor (which enters the equation in a manner analogous to the role of pressure) must have terms proportional to

$$\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \tag{2.1.6}$$

since internal friction arises only when there is a relative motion between different parts of the fluid. For an incompressible fluid, terms with $i = j$ vanish because of Eq. (2.1.4) and the equation of motion becomes the *Navier–Stokes equation*,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \frac{\eta}{\rho}\nabla^2\mathbf{v} \tag{2.1.7}$$

The proportionality constant $\eta$ is the *dynamic viscosity* (gm/cm sec) while the ratio

$$\nu = \eta/\rho \tag{2.1.8}$$

is the *kinematic viscosity* ($\text{cm}^2$/sec).

When the heat remains constant within all elements of the system and there is no interchange of heat by conduction and radiation, we have the special situation that the flow of gas is isentropic and adiabatic. An atmosphere scarcely conforms to these conditions, but in many problems the departures are not serious and the approximation is useful. Thus temperature is given by (1.3.4) (or its equivalent for moist air), and our set of equations is completed with the perfect gas law, (1.1.2). Integrating (1.3.3) we obtain the adiabatic condition for dry gas,

$$Tp^{-(\gamma-1)/\gamma} = \text{const} \quad \text{or} \quad p\rho^{-\gamma} = \text{const.} \tag{2.1.9}$$

There are other problems, such as those dealing with instabilities and their driving forces, where one cannot take an adiabatic law but can regard the gas as "almost incompressible." That is, one takes $\rho$ as a constant except in the buoyancy term of (2.1.7). This approach gives the *Boussinesq approximation.* Continuity is given by Eq. (2.1.4) and the Navier–Stokes equation

becomes

$$\frac{\partial \mathbf{v}}{dt} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho_0}\nabla p + \left(1 + \frac{\delta\rho}{\rho_0}\right)\mathbf{g} + \nu\nabla^2\mathbf{v} \tag{2.1.10}$$

where $\rho = \rho_0 + \delta\rho$. Energy conservation in a moving element with energy inflow gives

$$\rho_0 C_p\left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T\right) = \nabla \cdot (K\nabla T) + \rho_0 \mathbf{v} \cdot \mathbf{g} + Q \tag{2.1.11}$$

where $K$ is the coefficient of thermal conductivity and where $Q$ represents the source terms arising from the heat of viscous dissipation and radiation. For the troposphere and stratosphere, conduction is unimportant and radiation is the dominant term on the right.

### 2.1.2 Motion in a Rotating Atmosphere

As seen in a coordinate system $\Sigma'$ rotating with angular velocity $\Omega$ in an inertial frame $\Sigma$, a particle has a velocity $v'$ such that

$$\mathbf{v} = \mathbf{v}' + \boldsymbol{\Omega} \times \mathbf{r} \tag{2.1.12}$$

Thus the absolute velocity is the velocity of a particle relative to Earth plus the Earth's rotational velocity. Differentiating a vector $\mathbf{A}$ in the rotating system $d'\mathbf{A}/dt$ also requires an adjustment for the rotation:

$$\frac{d\mathbf{A}}{dt} = \frac{d'\mathbf{A}}{dt} + \boldsymbol{\Omega} \times \mathbf{A} \tag{2.1.13}$$

in general, or

$$\frac{d\mathbf{v}}{dt} = \frac{d'\mathbf{v}}{dt} + \boldsymbol{\Omega} \times \mathbf{v} \tag{2.1.14}$$

in particular. Then with (2.1.12) we have

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= \frac{d'}{dt}(\mathbf{v}' + \boldsymbol{\Omega} \times \mathbf{r}) + \boldsymbol{\Omega} \times (\mathbf{v}' + \boldsymbol{\Omega} \times \mathbf{r}) \\ &= \frac{d'\mathbf{v}'}{dt} + 2\boldsymbol{\Omega} \times \mathbf{v}' - \Omega^2\mathbf{R}\end{aligned} \tag{2.1.15}$$

since $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) = -\Omega^2\mathbf{R}$, where $\mathbf{R}$ is the distance from the axis of rotation. Equation (2.1.15) gives the acceleration of a particle in an inertial system as composed of the acceleration measured in the rotating system plus the Coriolis and centrifugal terms.

The Navier–Stokes equation for a rotating planet is thus given by writing (2.1.7) for $d\mathbf{v}/dt$ in (2.1.15). We find

$$\frac{d'\mathbf{v}'}{dt} = -2\boldsymbol{\Omega} \times \mathbf{v}' - \frac{1}{\rho}\nabla p + \mathbf{g}_e + \nu\nabla^2\mathbf{v}' \tag{2.1.16}$$

where the centrifugal force has been absorbed into the gravitational term by writing an effective gravity,

$$\mathbf{g}_e = \mathbf{g} + \Omega^2\mathbf{R} \tag{2.1.17}$$

In planetary problems it is conventional to take rotating coordinates $x$, $y$, and $z$ (components of $\mathbf{r}'$) as aligned eastward, northward, and upward. Then the velocities are correspondingly $u = d'x/dt$, $v = d'y/dt$, and $w = d'z/dt$. These velocities are not Cartesian, however, and the acceleration components in spherical coordinates are (for $\mathbf{v}' = \hat{\mathbf{i}}u + \hat{\mathbf{j}}v + \hat{\mathbf{k}}w$)

$$\begin{aligned}\frac{d'\mathbf{v}'}{dt} = &\left(\frac{d'u}{dt} - \frac{uv\tan\phi}{a} + \frac{uw}{a}\right)\hat{\mathbf{i}} \\ &+ \left(\frac{d'v}{dt} + \frac{u^2\tan\phi}{a} + \frac{wv}{a}\right)\hat{\mathbf{j}} \\ &+ \left(\frac{d'w}{dt} - \frac{u^2+v^2}{a}\right)\hat{\mathbf{k}}\end{aligned} \tag{2.1.18}$$

where $a$ is the planetary radius and $\phi$ is latitude.

We now expand the various force terms. Thus we have

$$\boldsymbol{\Omega} \times \mathbf{v}' = \Omega(w\cos\phi - v\sin\phi)\hat{\mathbf{i}} + u\sin\phi\hat{\mathbf{j}} - u\cos\phi\hat{\mathbf{k}} \tag{2.1.19}$$

and we can usually write

$$\mathbf{g}_e = -g\hat{\mathbf{k}} \tag{2.1.20}$$

(where $g \approx 980\ \text{cm}^2/\text{sec}$ at Earth's surface). Finally we have the equation for the eastward, northward, and upward components of the momentum equation. Omitting the primes for derivatives in the rotating system, we have

$$\frac{du}{dt} - \frac{uv\tan\phi}{a} + \frac{uw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + 2\Omega v\sin\phi - 2\Omega w\cos\phi + \nu\nabla^2 u \tag{2.1.21}$$

$$\frac{dv}{dt} + \frac{u^2\tan\phi}{a} + \frac{vw}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - 2\Omega u\sin\phi + \nu\nabla^2 v \tag{2.1.22}$$

and

$$\frac{dw}{dt} - \frac{u^2+v^2}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g + 2\Omega u\cos\phi + \nu\nabla^2 w \tag{2.1.23}$$

The curvature terms (those in $a^{-1}$) are quadratic and awkward to handle, but can often be ignored. The acceleration term is also nonlinear, since (for example),

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} \tag{2.1.24}$$

These velocity interaction terms are called "inertial" in turbulence theory and "advective" in atmospheric dynamics.

## 2.2 Horizontal Circulation of the Troposphere

### 2.2.1 Geostrophic Winds and the Gradient Flow

To simplify the horizontal equations of motion (2.1.21) and (2.1.22) we first examine the magnitude of the various terms. On a laboratory scale when the flow is not turbulent, energy dissipation occurs by the conversion of ordered, laminar flow directly into random molecular motions (i.e., heat) through the ordinary molecular viscosity. For air, $\eta \sim 2 \times 10^{-4}$ gm/cm sec and at STP, $\nu \sim 0.15$ cm$^2$/sec, which makes the friction term negligible. In a turbulent atmosphere, however, there is an analogous, large-scale dissipation of ordered flow (when there are wind shears) into turbulent eddies of various sizes. The kinetic energy in this spectrum of eddies migrates from larger eddies toward smaller ones (from the more ordered flow into the more disordered), ultimately ending up as heat through the molecular viscosity. But from the point of view of atmospheric winds, the kinetic energy is lost at the rate it is convected into turbulence of any size. Without going into the mechanism of inertial transfer of energy, we can describe it phenomenologically by an *eddy viscosity* $\nu_E$. This viscosity operates on a scale, $l \sim 3 \times 10^3$ cm in the troposphere, called the *mixing length*, which is analogous to the mean free path in molecular viscosity. Eddy viscosity is important in the lowest kilometer of the atmosphere; since wind velocities must vanish at the surface, wind shears are large in this *boundary layer* of the atmosphere. In this region $\nu_E$ is found empirically to be the order of $5 \times 10^4$ cm$^2$/sec and with wind gradients $\Delta u/\Delta z$ and $\Delta v/\Delta z \sim 5$ m/sec km, $\nabla^2 u \sim \nabla^2 v \sim 5 \times 10^{-8}$ cm$^{-1}$ sec$^{-1}$. The friction term is of order $3 \times 10^{-3}$ cm/sec$^2$ and has to be considered. Higher in the troposphere it is usually negligible and we shall not consider the dissipation term further in this section. Incidentally, turbulent eddies are responsible for transporting mass and energy, as well as momentum, and in those connections we shall later have need for the *eddy diffusion coefficient* (see Section 2.3.2) and the *coefficient of thermal diffusivity* (see Section 2.6).

On a planetary scale ($L \sim 10^8$ cm) typical horizontal velocities $u$ and $v$ are $10^3$ cm/sec ($\approx 20$ knots), but mean vertical motions $w$ are much smaller. Thus omitting the curvature terms and the small Coriolis term in $\Omega w$, we obtain the simplified equations

$$\frac{du}{dt} = 2\Omega v \sin\phi - \frac{1}{\rho}\frac{\partial p}{\partial x} \tag{2.2.1}$$

and

$$\frac{dv}{dt} = -2\Omega u \sin\phi - \frac{1}{\rho}\frac{\partial p}{dy} \tag{2.2.2}$$

With an advective time scale the order of $L/V \sim 10^5$ sec ($\sim 1$ day), $\Omega \sim 10^{-4}$ sec$^{-1}$, and pressure differences $\Delta p \sim 10^4$ dynes/cm$^2$ ($\sim 10^{-2}$ atm), the acceleration terms are an order of magnitude smaller than the force terms. Thus these typical synoptic characteristics of terrestrial weather indicate that the pressure gradients are nearly balanced by so-called *geostrophic winds* arising from the Coriolis force. Knowing the pressure distribution allows one to derive, in a purely diagnostic fashion, the zonal ($u$) and meridional ($v$) winds.

The relative magnitude of the acceleration term ($V/t \sim V^2/L$) and the Coriolis term at mid-latitudes ($\Omega V$) is the Rossby number

$$R_0 = V/L\Omega \tag{2.2.3}$$

A value of $R_0 \ll 1$ indicates that the geostrophic approximation is good.

The analogous approximation for the vertical pressure gradient is simply hydrostatic equilibrium because of the dominance of $g$ in (2.1.23). To a good approximation, therefore,

$$\frac{1}{\rho}\frac{\partial p}{\partial z} = -g \tag{2.2.4}$$

The above simplified equations of horizontal motion may be rewritten vectorially as

$$\frac{d\mathbf{V}}{dt} = f\,\mathbf{V} \times \hat{\mathbf{k}} - \frac{1}{\rho}\nabla p \tag{2.2.5}$$

where $\mathbf{V} = \hat{\mathbf{i}}u + \hat{\mathbf{j}}v$, $\nabla = \hat{\mathbf{i}}\partial/\partial x + \hat{\mathbf{j}}\partial/\partial y$, and the Coriolis parameter is

$$f \equiv 2\Omega \sin\phi \tag{2.2.6}$$

A set of *natural coordinates* simplifies the discussion (see Fig. 2.1). The unit vector $\hat{\mathbf{t}}$ is tangent to the flow, $\hat{\mathbf{n}}$ is positive to the left of the flow, and $\hat{\mathbf{k}}$ is

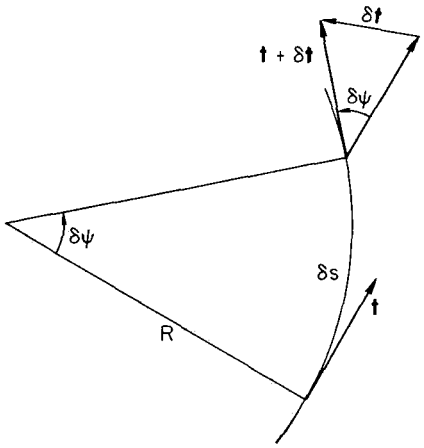



**Fig. 2.1** Curvilinear coordinates, showing the differential change in the unit tangent vector $\hat{\mathbf{t}}$. [Adapted from HOLTON (1972).]

vertically upward (normal to the page). Then $\mathbf{V} = V\hat{t}$ and, as in polar coordinates, $\delta\hat{t} = \hat{n}\delta\psi = \hat{n}\delta s/R$. Hence

$$\frac{d\mathbf{V}}{dt} = \hat{\mathbf{t}}\frac{dV}{dt} + V\frac{\hat{\mathbf{n}}}{R}\frac{ds}{dt}$$

$$= \hat{\mathbf{t}}\frac{dV}{dt} + \hat{\mathbf{n}}\frac{V^2}{R} \tag{2.2.7}$$

The Coriolis force is $f\mathbf{V} \times \hat{\mathbf{k}} = -fV\hat{\mathbf{n}}$, so that (2.2.5) can be written as the scalar equations,

$$\frac{dV}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial s} \tag{2.2.8}$$

and

$$\frac{V^2}{R} = -fV - \frac{1}{\rho}\frac{\partial p}{\partial n} \tag{2.2.9}$$

The equations of motion are now given in terms of forces parallel and normal to the flow, respectively.

The geostrophic approximation, in which $d\mathbf{V}/dt = 0$, thus gives $p = \text{const.}$ along the direction of flow. The magnitude of the velocity is

$$V = -\frac{1}{\rho f}\frac{\partial p}{\partial n} \tag{2.2.10}$$

the flow governed by the Coriolis force being perpendicular to the pressure gradient.

The special case of a vanishing horizontal pressure gradient is considered in Problem 2.1. If, on the other hand, the horizontal scale of a disturbance is very small, the Coriolis force is negligible. Then Eq. (2.2.9) gives

$$V = \left( -\frac{R}{\rho} \frac{\partial p}{\partial n} \right)^{1/2} \tag{2.2.11}$$

and the motion is also perpendicular to the pressure gradient with the pressure decreasing toward the center of rotation. Thus the force of the pressure gradient inward balances the outward centrifugal force. This *cyclostrophic flow* can be in either direction (cyclonic or anticyclonic) and is descriptive of tornados and dust devils.

The *gradient wind* is given by the full solution of (2.2.9), which describes the balance of centrifugal force with Coriolis force and the normal component of the pressure gradient:

$$V = -\frac{fR}{2} \pm \left( \frac{f^2 R^2}{4} - \frac{R}{\rho} \frac{\partial p}{\partial n} \right)^{1/2} \tag{2.2.12}$$

A physical solution must have $V \geq 0$. In Fig. 2.1 we have chosen the convention of taking $R$ as positive when it is antiparallel to $\mathbf{n}$—that is, for counterclockwise motion. In this case ($R > 0$) in the northern hemisphere ($f > 0$), we have $V > 0$ only if $\partial p/\partial n < 0$, which gives a low pressure (the *regular low*) at the center of curvature (see Fig. 2.2). Only the positive root is allowed.

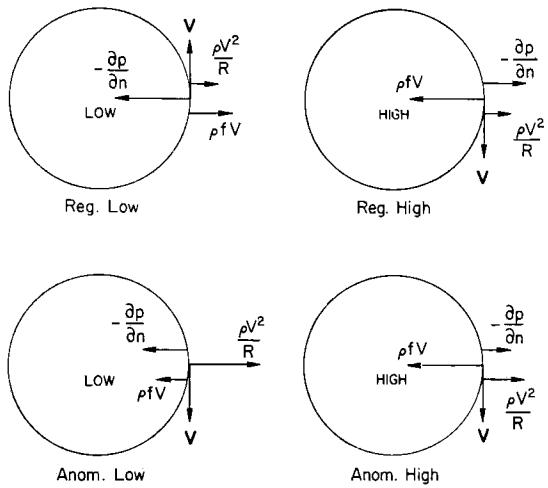



**Fig. 2.2** The four classes of gradient flow, showing how the pressure gradient is balanced by the Coriolis and centrifugal forces normal to the direction of flow. See Eq. (2.2.9), [Adapted from HOLTON (1972).]

On the other hand, if the curvature is clockwise ($R < 0$), there are three possible solutions: (a) $\partial p/\partial n > 0$, which also (with the positive root) gives a low pressure, the *anomalous low*, at the center; (b) $\partial p/\partial n < 0$, which gives high-pressure solutions for either (b.1) the positive root or *anomalous high* or (b.2) the negative root or *regular high.*

For cases (b), either high-pressure solution, the pressure gradient is restrained by the condition that the radical must be real, or

$$\frac{\partial p}{\partial n} < \frac{\rho R f^2}{4} \tag{2.2.13}$$

Thus the pressure in a high tends to be nearly constant at the center with little concomitant wind.

### 2.2.2 Hadley Circulation

In 1735 Hadley proposed that the trade winds were driven by upward and poleward motion of air heated hear the surface in the tropics. This air is diverted eastward by the Coriolis force and is replaced by air moving southwestward (in the northern hemisphere) near the surface (see Fig. 2.3). These *Hadley cells*, in tending to conserve their angular momentum about the rotational axis, thus have a zonal as well as meridional component. These cells were known not to extend to the polar regions because there is predominantly descending air around $\pm 30°$ latitude. However, cold surface air in the polar regions moves toward lower latitudes causing air motions towards the poles in the high troposphere.

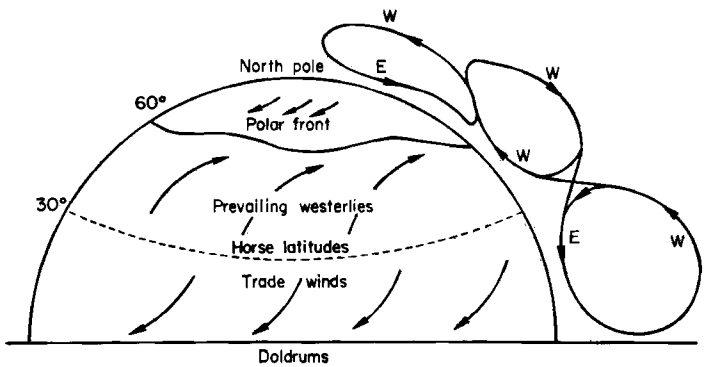



**Fig. 2.3** The three-cell circulation scheme, illustrating tropical convection with subsidence in the "horse latitudes," ascent in the zone of polar fronts, and subsidence over the polar cap. The notations W and E refer to the directions from which the prevailing winds arise. [Adapted from C.-G. ROSSBY (1941), *in* "Yearbook of Agriculture, Climate, and Man," G. Hambridge, ed.), pp. 599–655, U.S. Gov. Printing Off., Washington, D.C.]

These two sets of convection cells are driven by solar energy, with heat being converted to kinetic energy. An intermediate cell develops between the Hadley and polar cells with low-level air drifting northward and eastward. This mid-latitude or *Ferrel cell* must be supplied kinetic energy for its maintenance. Mid-latitude circulation is dominated, however, by other dynamical forces than cellular overturning. Thus a simple conservation of angular momentum within the mid-latitude cell would require the surface westerlies to diminish in strength and even change to easterlies at high altitude, whereas in fact the westerlies increase in velocity with height. The reason for this increase is described below.

### 2.2.3 Thermal Wind: Jet Streams

As a practical matter it is convenient to examine horizontal circulations relative to surfaces of constant pressure (*isobars*) rather than constant altitude. Hence to use *isobaric* coordinates we need to make the transformations (see Fig. 2.4)

$$\left(\frac{\partial p}{\partial x}\right)_z = -\left(\frac{\partial p}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_p \tag{2.2.14}$$

With the hydrostatic approximation (2.2.4), we have

$$\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right)_z = g\left(\frac{\partial z}{\partial x}\right)_p \tag{2.2.15}$$

Thus the geostrophic wind (2.2.10) becomes

$$u = -\frac{g}{f}\left(\frac{\partial z}{\partial y}\right)_p \quad \text{and} \quad v = \frac{g}{f}\left(\frac{\partial z}{\partial x}\right)_p \tag{2.2.16}$$

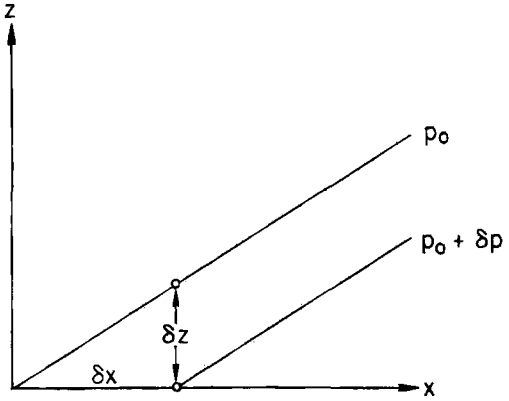



**Fig. 2.4** Isobaric coordinates, showing the relation of constant-pressure surfaces to $x$, $z$, coordinates. See Eq. (2.2.14). [After HOLTON (1972).]

The *geopotential* is the work required to raise unit mass from the surface to height $z$, namely,

$$\Phi \equiv \int_0^z g\,dz = -\int_{p_0}^{p} \frac{dp}{\rho} \tag{2.2.17}$$

Thus in isobaric coordinates the geostrophic wind is

$$u = -\frac{1}{f}\left(\frac{\partial \Phi}{\partial y}\right)_p \quad \text{and} \quad v = \frac{1}{f}\left(\frac{\partial \Phi}{\partial x}\right)_p \tag{2.2.18}$$

Since

$$\frac{\partial \Phi}{\partial p} = -\frac{RT}{p} \tag{2.2.19}$$

differentiation of (2.2.18) yields

$$p\frac{\partial u}{\partial p} \equiv \frac{\partial u}{\partial \ln p} = -\frac{p}{f}\frac{\partial}{\partial p}\left(\frac{\partial \Phi}{\partial y}\right)_p = \frac{p}{f}\left[\frac{\partial}{\partial y}\left(\frac{RT}{p}\right)\right]_p$$
$$= \frac{R}{f}\left(\frac{\partial T}{\partial y}\right)_p \tag{2.2.20}$$

and

$$\frac{\partial v}{\partial \ln p} = -\frac{R}{f}\left(\frac{\partial T}{\partial x}\right)_p \tag{2.2.21}$$

These *thermal wind equations* give the vertical wind shear due to nonuniform horizontal heating. Integrating these equations vertically over a slab having a mean temperature $T$ and extending from pressure $p_1$ up to $p_2$ ($<p_1$), we obtain a zonal velocity increment of

$$u(z_2) - u(z_1) = -\frac{R}{f}\left(\frac{\partial T}{\partial y}\right)_p \ln\frac{p_1}{p_2} \tag{2.2.22}$$

with a similar expression for meridional flow.

In a so-called *barotropic atmosphere* the density and pressure are assumed to be uniquely related, or $\rho \equiv \rho(p)$. In this situation an isobaric surface has a uniform density and therefore temperature. Consequently, the right side of (2.2.22) vanishes, geostrophic flow is uniform with height, and there is no thermal wind (i.e., vertical shear). In the more realistic *baroclinic atmosphere*, density depends on temperature as well as pressure, or $\rho = \rho(p, T)$, and thermal winds can exist.

Figure 2.4 shows mean meridional cross sections of temperature and wind velocity. The pole-to-equator temperature gradient $(\partial T/\partial y)_p$ is much

larger in middle latitudes in January than in July and gives rise to a strong jet stream around 12 km and latitude 30°. In summer the jet stream weakens and moves northward and downward (see Problem 2.4). In Fig. 2.5 the discontinuity or "gap" in the tropopause, near the core of the jet stream, is a characteristic feature.

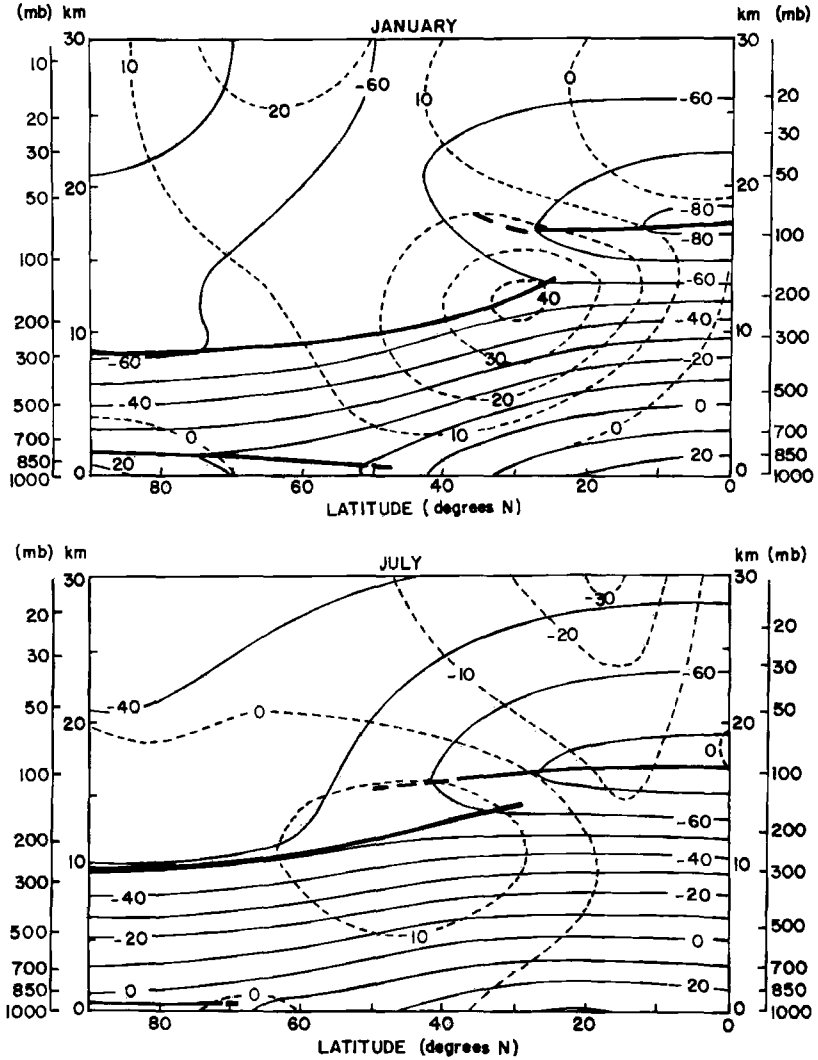



Fig. 2.5 Meridional winds (dashed lines in meters per sec) and isotherms (solid lines in degrees Centigrade). Heavy solid lines representing the tropopause show mid-latitude discontinuities or "gaps." Note the seasonal change in the peak of the thermal wind (jet stream). [Adapted from "Arctic Forecast Guide," Navy Weather Research Facility, April 1962.]

### 2.2.4 Propagation of Disturbances on the Planetary Scale

We return now to the full set of hydrodynamic equations and write them in *isobaric coordinates*—that is, relative to surfaces of constant pressure rather than constant height. Restoring the acceleration term to the velocity equations (2.2.18) recovers the simplified equations of motion, (2.2.1) and (2.2.2):

$$\frac{du}{dt} - fv = -\left(\frac{\partial \Phi}{\partial x}\right)_p \tag{2.2.23}$$

and

$$\frac{dv}{dt} + fu = -\left(\frac{\partial \Phi}{\partial y}\right)_p \tag{2.2.24}$$

By (2.2.17), the hydrostatic approximation is

$$\frac{\partial \Phi}{\partial p} = -\frac{1}{\rho} = -\frac{RT}{p} \tag{2.2.25}$$

Also, with the vertical velocity being negligible in this approximation, we can simplify the equation of continuity, (2.1.3), by writing $-(1/g)\,\partial p/\partial z$ for $\rho$ and reversing the order of differentiation. Thus continuity becomes

$$\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_p + \frac{\partial \omega}{\partial p} = 0 \tag{2.2.26}$$

where $\omega \equiv dp/dt$ is the pressure change along the trajectory of a parcel of air. If the atmosphere is nearly adiabatic, the thermodynamic relation (1.3.3) is

$$C_p \frac{d \ln T}{dt} \approx R \frac{d \ln p}{dt} \tag{2.2.27}$$

and the set of equations is completed with the perfect gas law, (1.1.2),

$$T = p/\rho R \tag{2.2.28}$$

In these equations the substantial derivative, taken along the trajectory, is

$$\frac{\partial}{dt} \equiv \left(\frac{\partial}{\partial t}\right)_p + u\left(\frac{\partial}{\partial x}\right)_p + v\left(\frac{\partial}{\partial y}\right)_p + \omega \frac{\partial}{\partial p} \tag{2.2.29}$$

in lieu of (2.1.24), since $p$ and $\omega$ now play the role formerly held by $z$ and $w$. In the equations of motion and continuity, the density does not enter, and there is no time derivative in the continuity equation, which leads to considerable simplification. It should be emphasized that all velocities and spatial gradients are measured with respect to isobaric surfaces, not horizontal surfaces.

The independent variables are now $x$, $y$, $p$, and $t$ whereas the dependent variables are $u$, $v$, $\omega$, $\rho$, $T$, and $\Phi$ (where $\Phi$, being essentially $z$, has interchanged roles with $p$).

It is sometimes convenient to replace temperature $T$ with the concept of *potential temperature* $\theta$, which is the temperature a volume of dry air at $p$, $T$ would have if compressed to pressure $p_0 = 1$ atm. The substitution of $\theta$ for $T$ leads us to alternate forms of the fifth and sixth equations of this section. Thus from (2.2.27) or its integrated form (2.1.9), we write

$$\theta \equiv T\left(\frac{p_0}{p}\right)^{(\gamma-1)/\gamma} \approx \text{const.} \tag{2.2.30}$$

where $(\gamma - 1)/\gamma \equiv R/C_p$. Then (2.2.28) substituted into (2.2.30) relates $\theta$ and $\rho$ by

$$\theta = \frac{p}{\rho R}\left(\frac{p_0}{p}\right)^{R/C_p} \tag{2.2.31}$$

In developing the concept of geostrophic wind, we noted that some terms in the equations are found empirically to be small. The remaining terms then tell how the atmosphere generally behaves—that is, to the first order, winds are produced by pressure gradients and the Earth's rotation. If we had a synoptic picture of the pressure the world over, we could construct a corresponding snapshot of the flow. The analysis would be diagnostic but not predictive. It would tell us nothing about the pressure and wind fields the next day.

We will now extract from the six basic equations two simplified expressions that allow prediction. Of course, the actual solution of such equations requires an enormous set of starting data, and a large computer is required to obtain even fragmentary solutions. In the following we will concentrate on predicting the two dependent variables, $\Phi$, the geopotential, and $\omega = dp/dt$, the vertical motion. We start with the thermodynamic equation (2.2.30) which we will write in terms of $\Phi$ by using the hydrostatic relation (2.2.25).

Differentiating the logarithm of (2.2.31) yields, for example,

$$\left(\frac{\partial \ln \theta}{\partial x}\right)_p = -\left(\frac{\partial \ln \rho}{\partial x}\right)_p \tag{2.2.32}$$

The thermodynamic equation states simply that the external heat input is small, or $d \ln \theta/dt \approx 0$. Expanding this derivative with (2.2.29) and substituting with (2.2.32) and analogous derivatives, we obtain

$$\left(\frac{\partial}{\partial t}\ln \rho\right)_p + u\left(\frac{\partial \ln \rho}{\partial x}\right)_p + v\left(\frac{\partial \ln \rho}{\partial y}\right)_p \approx \omega \frac{\partial \ln \theta}{\partial p} \tag{2.2.33}$$

The hydrostatic equation (2.2.25) eliminates $\rho$ for $\Phi$, giving after some simplification

$$\frac{\partial}{\partial t}\left(-\frac{\partial \Phi}{\partial p}\right)+u\frac{\partial}{\partial x}\left(-\frac{\partial \Phi}{\partial p}\right)+v\frac{\partial}{\partial y}\left(-\frac{\partial \Phi}{\partial p}\right)\approx \sigma\omega \tag{2.2.34}$$

where

$$\sigma=-\frac{\partial\theta/\partial p}{\rho\theta} \tag{2.2.35}$$

is the *static stability parameter*. When $\partial\theta/\partial p<0$ (so that $\sigma>0$) a volume of air in adiabatic motion is stable to a small vertical displacement. Adopting the geostrophic velocity (2.2.18) for $u$, $v$, we have

$$\frac{\partial}{\partial t}\left(-\frac{\partial \Phi}{\partial p}\right)\approx -\frac{1}{f}(\hat{\mathbf{k}}\times\nabla_p\Phi)\cdot\nabla_p\left(-\frac{\partial \Phi}{\partial p}\right)+\sigma\omega \tag{2.2.36}$$

where $\nabla_p$ signifies a gradient along an isobar.

Since $\sigma$ is a function of $\Phi$ (see Problem 2.5), this equation involves only the two dependent field variables $\Phi$ and $\omega$. By Eq. (2.2.25) $\partial\Phi/\partial p$ is proportional to temperature. Hence (2.2.36) essentially gives the local change of temperature due to the *advection* of temperature by the geostrophic wind on the same isobaric surface plus the adiabatic heating (or cooling) from descending (or rising) air. Had we explicitly carried an external heating (or cooling) term in the thermodynamic equation (2.2.27) or (2.2.30), in lieu of writing an approximate equality, a corresponding additional term would appear in (2.2.36).

Just as the thermodynamic equation is the basis for one $\Phi$ vs. $\omega$ relationship, (2.2.36), the equations of motion will provide another. Using (2.2.23) and (2.2.24), we find an expression for the local rate of change in the vertical component of the *vorticity vector*, $\zeta=\hat{\mathbf{k}}\cdot\nabla\times\mathbf{V}$:

$$\begin{aligned}\frac{\partial\zeta}{\partial t}&\equiv\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)\\&=-\mathbf{V}\cdot\nabla_p(\zeta+f)-\omega\frac{\partial\zeta}{\partial p}-(\zeta+f)\nabla_p\cdot\mathbf{V}\\&\quad+\left[\frac{\partial u}{\partial p}\left(\frac{\partial\omega}{\partial y}\right)_p-\frac{\partial v}{\partial p}\left(\frac{\partial\omega}{\partial x}\right)_p\right]\end{aligned} \tag{2.2.37}$$

Since $\zeta$ is measured in the rotating coordinate system, it is a *relative vorticity* and $\zeta+f$, including the Coriolis motion, is the *absolute vorticity*. Thus the first term on the right is the horizontal advection term for vorticity and

the next term represents vertical advection of vorticity. The final two terms are called the divergence and twisting terms.

The vertical advection and twisting terms are small, and normally $\zeta \ll f$. Hence a simplified version of the vorticity equation is

$$\frac{\partial \zeta}{\partial t} \approx -\mathbf{V} \cdot \nabla_p(\zeta + f) + f\frac{\partial \omega}{\partial p} \tag{2.2.38}$$

where the velocity divergence has been eliminated by the equation of continuity, (2.2.26).

Just as we obtained a geostrophic velocity, $\mathbf{V}_g = (1/f)\mathbf{k} \times \nabla_p\Phi$, by setting the acceleration term equal to zero, so can we obtain a geostrophic vorticity, $\zeta_g = (1/f)\nabla_p{}^2\Phi$, from (2.2.23) and (2.2.24). Putting these approximate values in the simplified vorticity equation gives the *quasi-geostrophic vorticity equations*,

$$\frac{\partial}{\partial t}(\nabla_p{}^2\Phi) \approx -(\hat{\mathbf{k}} \times \nabla_p\Phi) \cdot \nabla_p\left(\frac{1}{f}\nabla_p{}^2\Phi + f\right) + f^2\frac{\partial \omega}{\partial p} \tag{2.2.39}$$

Equations (2.2.36) and (2.2.39) govern the time changes in $\Phi$ and $\omega$. It is possible to proceed further and eliminate one or the other dependent variable to obtain separate equations for $\partial\Phi/\partial t$, the *geopotential tendency*, and $\omega \equiv dp/dt$, the vertical motion. Thus by differentiating (2.2.36) with respect to $p$ and for constant $\sigma$, we can eliminate $\omega$, obtaining an equation for $\partial\Phi/\partial t$,

$$\left(\nabla_p{}^2 + \frac{f^2}{\sigma}\frac{\partial^2}{\partial p^2}\right)\frac{\partial \Phi}{\partial t} = -f\mathbf{V} \cdot \nabla_p\left(\frac{1}{f}\nabla_p{}^2\Phi + f\right) - \frac{f^2}{\sigma}\frac{\partial}{\partial p}\left(\mathbf{V} \cdot \nabla_p\frac{\partial \Phi}{\partial p}\right) \tag{2.2.40}$$

Taking the spatial and pressure dependencies of the geopotential tendency to vary sinusoidally, we write

$$\frac{\partial \Phi}{\partial t} = (\text{const.}) \sin kx \sin ly \sin\frac{p\pi}{p_0} \tag{2.2.41}$$

where the $p$ variation is an approximation based on the supposition that the vertical scale of disturbances is one or two scale heights (i.e., about the tropospheric height). (The wavenumbers $k$ and $l$ are defined by $k = 2\pi/L_x$ and $l = 2\pi/L_y$, where the $L$'s are wavelengths.) Hence the left side of (2.2.40) is proportional to $-\partial\Phi/\partial t$.

On the right side of (2.2.40), the first term represents advection of vorticity by the geostrophic wind (see Fig. 2.6). Thus barometric highs and lows are carried either upstream or downstream, depending on the relative importance of the relative geostrophic vorticity $\zeta = \nabla_p{}^2\Phi/f$ and planetary vorticity $f$ and on the wavelength of the disturbance.

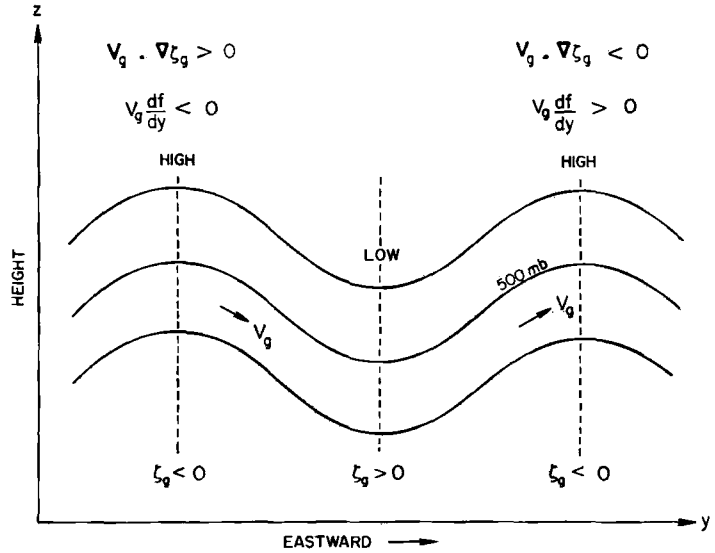



**Fig. 2.6** Schematic isobars near the 500 mb level in the $x,z$ plane, showing regions of positive and negative advections of relative and planetary vorticity. [After HOLTON (1972).]

The second term on the right side gives the growth (or decay) of a high- or low-pressure system. By (2.2.25), $\partial\Phi/\partial p$ is essentially the temperature, so that this second term gives the height (i.e., pressure) gradient of temperature advection. For example, if cold air is advected into the lower regions of a low-pressure system, the thickness of the air column is reduced and the height of a fixed pressure is lowered—the low-pressure system thus grows.

In a similar manner we could solve (2.2.36) and (2.2.39) for $\omega$ as a function of $\Phi$. That is, the vertical motion depends exclusively on the geopotential field. It may also be shown that the resulting $\omega$ ensures that the vorticity remains geostrophic and the temperature distribution with height remains hydrostatic. For further discussion and application of these diagnostic equations to the development of *baroclinic disturbances* the reader may consult the bibliographical notes.

Disturbances in the pressure field (other than compressional or sound waves) propagate in various forms of waves. *Gravity waves* are the familiar waves that propagate on the surface of water, but they also affect the interior of a fluid. In a stratified fluid such as the atmosphere, these waves may propagate nearly vertically as well as horizontally. A disturbance set up, say, by air flowing over a mountain range will propagate upward and be refracted toward the horizontal with increasing height. At high altitudes dissipation of the wave may be an important source of local energy.

A second type of wave responsible for free oscillations of the atmosphere is the *Rossby* or *planetary wave*. The driving mechanism is the latitudinal

variation of the Coriolis force, which causes the Rossby wave to propagate slowly (a few meters per second) westward (relative to the mean zonal flow).

A similar kind of traveling disturbance, because it also has wavelengths the order of the planetary diameter, is the atmosphere tide. Tidal waves are almost imperceptible in the troposphere, except in the tropics, but they are important on Mars and in the Earth's upper atmosphere. Tides are driven by solar heating as well as by the gravitational forces of the sun and moon.

## 2.3 Vertical Transport

### 2.3.1 Molecular Diffusion

Assume that a minor constituent has a density distribution $N_1(z)$ in an atmosphere where its density distribution in diffusive equilibrium would be $N_{1E}(z)$. It will have an upward flux $\Phi$, which can be written in terms of an upward diffusion velocity $w$, given by the diffusion equation

$$\Phi_1 \equiv N_1 w_1 = -D N_{1E} \frac{\partial (N_1/N_{1E})}{\partial z} \tag{2.3.1}$$

Here $D$ (cm$^2$/sec) is the coefficient of diffusion, which varies inversely with the total ("background") density $N(z)$. For an atmosphere in hydrostatic equilibrium the distribution $N_{1E}(z)$ is given by (1.1.5) and

$$\Phi_1 = -N_1 D \left( \frac{1}{N_1} \frac{\partial N_1}{\partial z} + \frac{M_1 g}{kT} + \frac{1}{T} \frac{dT}{dz} \right) \tag{2.3.2}$$

where $M_1$ is the mass of the diffusing gas. The second and third terms on the right compose the reciprocal of the *density scale height* $H_{1E}^*$ of an equilibrium distribution [different from the usual pressure scale height $H$ unless the atmosphere is isothermal—cf. (1.1.5)]. Hence a convenient form of (2.3.2) is

$$\Phi_1 \equiv N_1 w_1 = N_1 D \left( \frac{1}{H_1^*} - \frac{1}{H_{1E}^*} \right) \tag{2.3.3}$$

It is often desirable to compare diffusion with other physical processes, such as dissociation or recombination. For these purposes we may use the concept of *time of diffusion*, $\tau_{\text{dif}} \sim H_1^*/w_1$. Molecular diffusion becomes important above 90 km in the Earth's atmosphere.

The discussion so far has dealt with diffusion due to concentration gradients relative to the equilibrium gradient. Diffusion may also occur due to temperature gradients. To allow for thermal diffusion we merely add a term $(\alpha_T/T)\, dT/dz$ within the parentheses of (2.3.2) and onto the definition of $1/H_{1E}^*$. The coefficient $\alpha_T$ is the *thermal diffusion factor* [cf. Eq. (7.3.3)].

### 2.3.2 Eddy Diffusion

The equations of motion are not very useful for describing turbulence or mixing of the atmosphere because of the complexity that arises from the inertial interaction of different mass elements. Hence it is useful to simplify mixing by assigning it the kind of averaged macroscopic properties that molecular diffusion possesses. Substituting an *eddy diffusion coefficient*, $K$ (cm$^2$/sec) everywhere for $D$ and the total density $N(z)$ for $N_{1E}(z)$ in (2.3.1) will, in form at least, accomplish this objective. Then in (2.3.2) we replace $K$ for $D$ and $M \equiv \langle M \rangle$ for $M_1$.

Let us first examine a very simple case where the atmosphere is isothermal and $K$ is constant (over a limited height range). Also there are no sources or sinks of the minor constituent in the region in question. Then (2.3.1) (modified for eddy diffusion) integrates to

$$N_1(z) = AN(z_0)e^{-(z-z_0)/H} - \frac{\Phi_1}{K} \tag{2.3.4}$$

where $H$ is the scale height of the background gas and $A$ is an integration constant to be fixed by $N_1(z_0)$ and $\Phi_1$. The solution has a component proportional to a constant mixing ratio $A$ (which can be negative), and a component dependent on flux. For upward diffusion the first term must be positive and dominate numerically; the second term gives a correction to perfect mixing. For downward diffusion ($\Phi_1$ = negative), the second term is positive and can dominate, giving a density nearly independent of height. Other height variations of $K$ are considered in Problems 2.9 and 3.4. For the application of these equations to the escape of planetary atmospheres see Section 7.3.1.

## 2.4 Circulation of the Venus Atmosphere

There are several indications that Venus' atmosphere is dynamically active: (1) The temperature profile (see Section 1.8.1) is nearly adiabatic up to 50 km. Only 1 percent or so of the incident solar flux reaches the ground, most of the remainder being reflected to space by Mie scattering of the cloud droplets and Rayleigh scattering by the nearly 100 atmospheres of $CO_2$. (2) Infrared thermal emission from the upper atmosphere and temperature measurements in deep atmosphere indicate small horizontal gradients, even though Venus' solar day is 117 days. (3) The very existence of clouds indicates large scale rising motions or at least small scale turbulence or both. (4) Fluctuations in radio signals suggest turbulent layers around 45 and 60 km altitude, but not in the deep atmosphere.

To understand the main physical processes at work we first examine the pertinent time scales. The thermal capacity above a level at pressure $p$ for a hydrostatic atmosphere may be regarded [cf. (1.3.3) and (1.1.6)] as $(C_p/R)pH$. If the atmosphere radiates energy to space at a rate $-\sigma T_e^4$ characteristic of some effective temperature $T_e$, given by (1.2.46), the cooling time scale (after solar heating stops) is

$$\tau_{\text{rad}} = \frac{\gamma pH}{(\gamma - 1)\sigma T_e^4} \approx 6 \times 10^4 Tp \,(\text{atm}) \tag{2.4.1}$$

since $C_p/R = \gamma/(\gamma - 1)$. At the surface $\tau_{\text{rad}} > 10^2$ yr.

The dynamical time scale, as noted after (2.2.2), is

$$\tau_{\text{dyn}} = L/V \sim 10^5 \text{ sec} \tag{2.4.2}$$

for speeds of several meters per second.

The length of the Venus day is

$$\tau_{\text{day}} = 1.01 \times 10^7 \text{ sec} \tag{2.4.3}$$

Hence at height $z \lesssim 45$ km, $\tau_{\text{rad}} > \tau_{\text{day}}$ and it is not surprising that diurnal temperature changes are small. At the cloud level, $z \approx 60$ km, $\tau_{\text{rad}} \sim 0.5 \times 10^7$ sec and the small variation in thermal emission from that region indicates the presence of planetary scale winds. Indeed, Venera 9 and 10 measured zonal winds of 60 m/sec between 40 and 60 km.

The sidereal rotation period (244 days) is

$$2\pi/\Omega = \tau_{\text{rot}} = 2.11 \times 10^7 \text{ sec} \tag{2.4.4}$$

giving a Rossby number (2.2.3)

$$R_0 = \tau_{\text{rot}}/2\pi\tau_{\text{dyn}} \sim 3 \times 10^1 \tag{2.4.5}$$

which means that Coriolis forces are very small on Venus. Hence we conclude that the important parameter governing the circulation of the lower atmosphere is the *Golitsyn number*, the controlling parameter for a nonrotating planet,

$$G_0 = \tau_{\text{dyn}}/\tau_{\text{rad}} \tag{2.4.6}$$

If $G_0 \gg 1$, implying that the global circulation requires a long period compared with the time to transport heat radiatively, then radiative equilibrium would govern the horizontal temperature variation (for a nonrotating planet). For $G_0 \lesssim 1$ the motions will be important in transporting heat. On Venus $G_0 \ll 1$ everywhere below the clouds. Thus in the lower atmosphere the time scales are ordered by

$$\tau_{\text{dyn}} \ll \tau_{\text{day}} \sim \tau_{\text{rot}} < \tau_{\text{rad}} \tag{2.4.7}$$

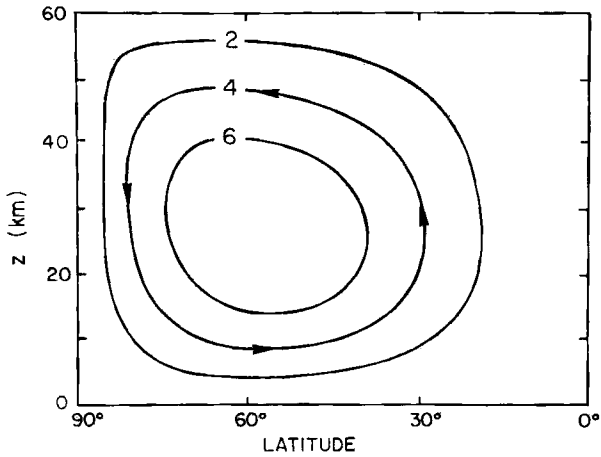



**Fig. 2.7** Streamlines in the Venus Hadley cell. [Based on calculations of E. KALNAY DE RIVAS (1973), *J. Atmos. Sci* 30, 763.]

This scaling analysis indicates that the general circulation of Venus is, to a first approximation, a Hadley cell (Fig. 2.7), with warm air rising in the equatorial region and sinking at the poles. Although the rotation is slow enough to make Coriolis forces relatively unimportant, it is fast enough to keep the equator uniformly heated, so that meridional circulation must be almost axially symmetric. To transport heat poleward, the lapse rate must be slightly *subadiabatic*—that is, on the stable side of the adiabatic lapse rate. A different way of saying the same thing is that the potential temperature $\theta$ increases along the rising, poleward branch of the cell, or the static stability (2.2.35) is positive. Thus on Venus the near-adiabatic lapse rate arises from large-scale Hadley circulation, rather than from small-scale convection.

A number of computer solutions of the six basic equations with the Boussinesq approximation have indicated similar results, namely, Hadley-type circulation. The major problem with solutions that seek an equilibrium configuration with time-marching techniques has been the extremely large value of $\tau_{rad}$ at the surface.

In the 60 km region $\tau_{rad} \sim 0.5 \times 10^7$ sec and

$$\tau_{dyn} \ll \tau_{rad} \lesssim \tau_{day} \sim \tau_{rot} \tag{2.4.8}$$

replaces (2.4.7). Thus diurnal heating is more likely to be faster than in the deep atmosphere. Before it was clearly realized that the clouds are in a different dynamical regime from the lower atmosphere, it was thought that the general circulation should lead to an absence (or at least a partial clearing) of clouds near the descending branch of the Hadley cell. The clouds may in fact be less dense, or lie lower, over the polar caps; but the matter now seems less important.

Mariner 10 observations disclosed a distinct pattern in the ultraviolet images of clouds centered on the sub-solar region that is probably due to convection. Also the near uniformity of thermal emission indicates a controlling influence of dynamics. The question arises as to what causes the ultraviolet contrast patterns, which do not appear in the visual or infrared. Probably there is an unidentified ultraviolet absorber; but perhaps the ultraviolet penetrates only into the uppermost thin haze. A highly irregular upper surface, with peaks and troughs due to convection, could produce highlights and shadows (planetary photometry is treated in Section 4.3.3).

The predominant feature of the upper-atmosphere dynamics is an approximate *four-day retrograde rotation.* Not only is such a motion well documented from telescopic photographs from Earth of cloud contrasts in the ultraviolet, but spectroscopic Doppler shifts and the zonal drifts of descending Venera spacecraft indicate real but highly variable velocities of the order of 100 m/sec. Also, see Fig. 2.8.

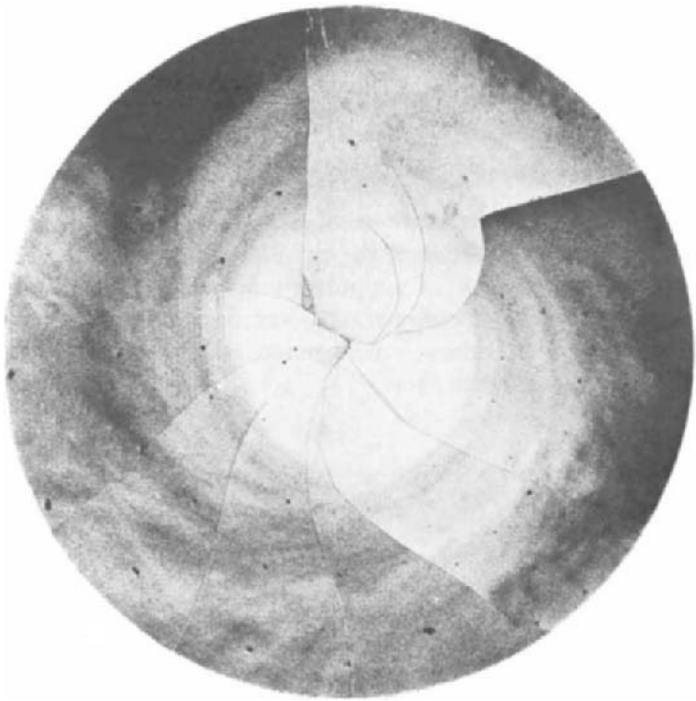


**Fig. 2.8** Space–time composite of ultraviolet images of the southern hemisphere of Venus as observed from Mariner 10. Note the counterclockwise (westward) flowing vortex centered on the south pole. Spiral streaks, resembling dense terrestrial stratus, converge into the vortex from low latitudes, akin to the spiral bands of a hurricane. The outer edge of the photograph is the equator. [Courtesy VERNER E. SUOMI and SANJAY S. LIMAYE; also see *J. Atmos. Sci.* **34**, 205 (1977).]

Most of the speculation on the driving mechanism for the four-day rotation involves the sun's apparent motion, which is direct (west to east), as seem from Venus. The "moving flame" mechanism has solar energy deposited within the clouds and convected upward, but with an appreciable time lag. As a convecting volume expands and rises, it reacts on those convection cells preceding it. Thus the stream lines of upward convection following the sun's trace across the planet will be tilted away from the sun, or in the retrograde sense. It is still not established whether this or some other mechanism is responsible for the four-day effect.

## 2.5 Diurnal Winds on Mars

The atmosphere of Mars, while composed primarily of $CO_2$, is otherwise vastly different in character from Venus' atmosphere. With a surface pressure around 7.5 mb, it radiates heat rapidly. The temperature at the Viking I landing site (22.5° N) in early summer varies from 187°K at 05 00 local solar time to 242°K at 15 00 hours. The phase lag for the maximum temperature after local noon is not unlike the response over terrestrial deserts, but the 55°K amplitude suggests a cooling time of

$$\tau_{rad} \sim \tau_{rot} = 8.88 \times 10^4 \text{ sec} \tag{2.5.1}$$

On the other hand, Coriolis forces are important, as on Earth.

Well above the high mountains, the circulation may be relatively simple. Figure 2.9 shows, from results of computer modeling, that geostrophic winds, mainly due to latitudinal pressure gradients arising from uneven solar heating, yield weak easterlies in the summer hemisphere and stronger but irregular westerlies in the winter hemisphere.

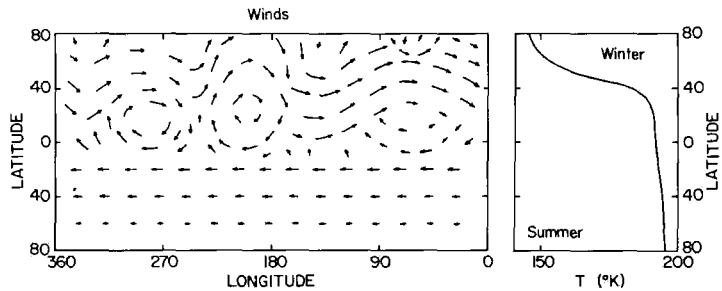



**Fig. 2.9** Schematic air flow on Mars well above the ground, computed for the southern summer solstice. The strong west winds in the middle and high latitudes of the winter hemisphere are produced by the net eastward Coriolis torque that accompanies the poleward transfer of air mass, caused in turn by condensation of $CO_2$ over the winter polar cap. [Based on numerical modeling by LEOVY and MINTZ (1969).]

Computer modeling of the Martian surface winds has also shown them to be critically dependent on the topography. In addition, both the numerical models and the pressure measurements show that the atmosphere is driven by thermal tides. Not only is there a diurnal (wavenumber 1 or $l = 1/a \sin \phi$) component present, but there is an equally important semidiurnal (wavenumber 2) component. The Earth also shows the semidiurnal component, which is driven westward by solar heating but has a strong amplitude because of an accidental resonance with the horizontal channel formed by the ground and tropopause. The Martian semidiurnal mode is similar in character to Earth's but had not been predicted by the modeling.

The Earth's diurnal mode is also westward driven by the sun, but the Martian diurnal mode shows peculiarities in phase and in the direction of associated winds. Presumably it is either modified severely by Martian topography or it is in an eastward traveling mode—the so-called *Kelvin resonance mode.*

Figure 2.10 shows the predominant southwesterlies composing the winds through the Martian day. Observations over a short period at one station do not by themselves tell much of the general surface circulation on the planet. Nevertheless, they give the required data base for a comparison with numerical models that can then lead to a more confident evaluation of the planetary circulation.

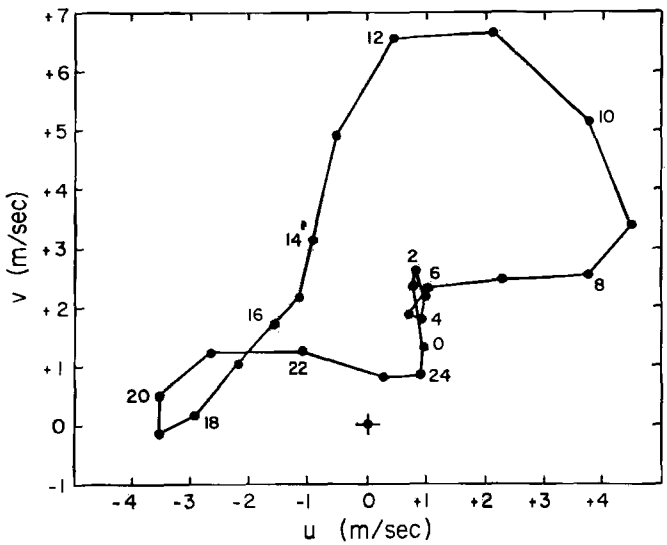



**Fig. 2.10** Hodograph of the horizontal winds measured by Viking I. The values next to the points are times in hours from local midnight. The $u$ (eastward) and $v$ (northward) velocity components are found by drawing a vector from the cross to a point on the curve. [After HESS *et al.* (1976).]

The Viking landers also showed a slow but steady decrease of total pressure at their northern-hemisphere locations. The effect was expected, as the southern ice cap was growing by the condensation of $CO_2$ from the atmosphere.

## 2.6 Convection in the Jovian Atmosphere

Jupiter, as the prototype for the major planets, offers many characteristics that are vastly different from those of terrestrial planets. It has a diameter 11.2 times the Earth's and a rotation period of less than 10 hr or

$$\tau_{\text{rot}} = 3.54 \times 10^4 \text{ sec} \tag{2.6.1}$$

Zonal speeds at the boundaries of the bright *zones* and dark *belts* may reach 100 m/sec, so that (for zonal motions, at least)

$$\tau_{\text{dyn}} \sim 10^6 \text{ sec} \tag{2.6.2}$$

and the Rossby number (2.2.3) is

$$R_0 \sim 10^{-2} \tag{2.6.3}$$

meaning the geostrophic velocity (2.2.10) should be a good approximation.

The Bond (omnidirectional) albedo has been accurately determined with the aid of the phase integral (see Section 4.3.3) measured by Pioneer 10 and is $\Lambda_B = 0.45$. Hence the equilibrium temperature (1.2.46) is $T_e = 106°\text{K}$. However, photometry in the infrared from aircraft above the Earth's water vapor gives a radiation temperature of $T = 134°\text{K}$, indicating that most of Jupiter's atmospheric heat energy comes from the interior, not from the sun (see Section 1.8.3).

It would appear also that convection is strong for two reasons: First, the bright zones are cooler and higher than the dark belts, their high albedo arising from solid $NH_3$ crystals. Thus the zones are regions of upward motion and high pressure (see Fig. 2.11). Second, the strong zonal flow is evidently due to Coriolis interaction with meridional flow between zones and belts. The highest wind velocities are found along the transitions between zones and belts, with eastward velocities between a zone and a poleward belt, westward velocities between a zone and an adjacent equatorial belt.

On the other hand, if Jupiter were not strongly heated from the interior, one would not expect to find such strong convection. To understand how rotation inhibits convection, we refer to the vorticity equation (2.2.37), which came from the equations of motion (2.1.21–2.1.23) with the rotation terms (in $a^{-1}$) and viscosity terms (in $\nu$) discarded. As in the derivation of (2.2.38), let us discard terms in the square of the velocity (remembering that

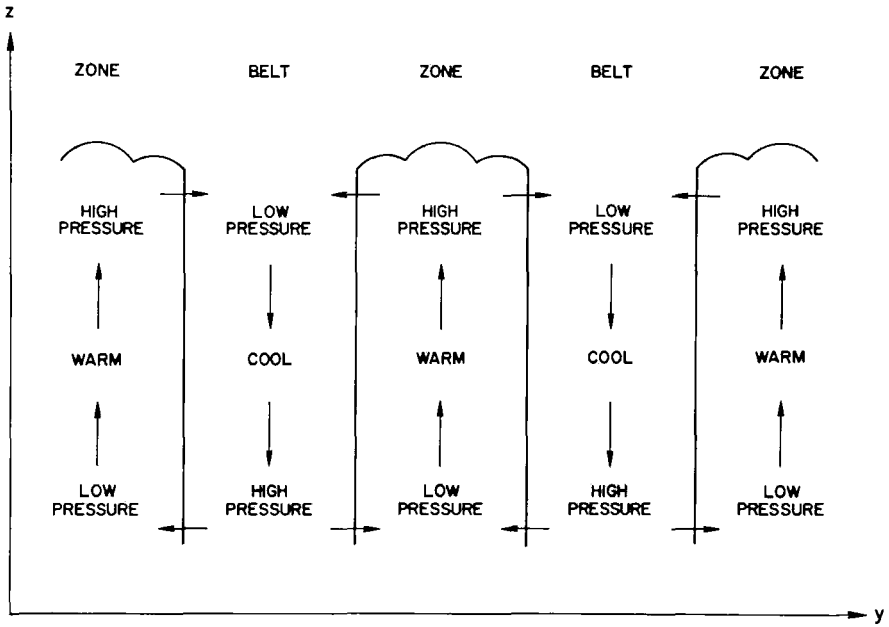



Fig. 2.11 Schematic diagram of Jupiter's meridional circulation. [After STONE (1976).]

$\omega = dp/dt$ is essentially a velocity in isobaric coordinates). Then for steady ($\partial\zeta/\partial t = 0$), slow motions we have from (2.2.37)

$$\frac{\partial(\zeta + f)}{\partial t} \approx -\nabla_p \cdot [(\zeta + f)\mathbf{V}] \approx -f\nabla_p \cdot \mathbf{V} = 0 \tag{2.6.4}$$

(since $\zeta\mathbf{V}$ is of second order in velocities), where $\mathbf{V} = \hat{\imath}u + \hat{\jmath}v$. Then from the isobaric equation of continuity (2.2.26) we have

$$\partial\omega/\partial p = 0 \tag{2.6.5}$$

which states that an element of fluid must move along a line of constantly changing pressure, $\omega \equiv dp/dt = \text{const}$. Clearly the constant can only be zero, so that to a first approximation *all steady motions in a rotating atmosphere with zero viscosity must be barotropic* (*i.e., two dimensional*). This statement amounts to the *Taylor–Proudman theorem* in hydrodynamic stability theory.

In the atmospheres of Venus, Mars, and Earth (at least in middle latitudes) the barotropic atmosphere gives way to a *baroclinic* one in which temperature varies systematically with latitude at a given pressure level. This latitudinal variation $(\partial T/\partial\phi)_p$ carries profound results for Earth and probably Mars. It gives rise to thermal winds or large velocity shears (see Section 2.2.3). In

turn, small perturbations in velocity (or pressure or temperature) tend to become greatly amplified in regions where large shears already exist. These *baroclinic instabilities* set up strong disturbances consisting of cyclonic winds and turbulent mixing. Thus a baroclinic atmosphere, which violates the Taylor–Proudman theorem, can be established because convection is driven by horizontal gradients, the motion is not steady, and the eddy viscosity $\nu_E$ (see Section 2.2.1) cannot be ignored. The important point is that, in a baroclinic atmosphere, convection is driven as a by-product of the horizontal temperature gradients.

Were Jupiter's atmosphere driven by its baroclinicity, it is not clear why the zones and belts should exist as they do and especially why the equatorial zone should have a strong eastward flowing jet. Alternatively, if Jupiter is largely heated from below, the atmosphere could be convectively unstable (with a superadiabatic gradient) so that the static stability (2.2.35) is negative and convection is the primary driving force. The question is whether the rotation of Jupiter would prevent a primary convection from being established.

The circulation problem thus posed has some similarity to the *Bénard problem*, a classical problem in hydrodynamic instability. In its simplest form the Benard problem deals with a horizontal layer of fluid heated from below. Because of thermal expansion the fluid at lower levels is lighter than at the top. But the convection tendency, which the fluid develops in order to correct this abnormal state of affairs, is inhibited by viscosity. Thus there will be some critical temperature gradient at which convection will begin. Smaller abnormal gradients of temperature will be stable, and the lighter, lower levels will support the heavier mass above.

The hydrodynamic equations for this problem may be solved in the Boussinesq approximation. The solutions are eigenvalue in nature and indicate that a fixed pattern of convection cells is set up when the unstable temperature gradient is achieved. Theory gives a critical value of the *Rayleigh number* at which convection will begin. The Rayleigh number is a combination of parameters in the problem,

$$R \equiv \frac{g\alpha\beta d^4}{\nu\chi} \tag{2.6.6}$$

where $\alpha$ and $\chi$ are coefficients of volume expansion and thermometric conductivity, $\beta$ is the abnormal temperature gradient, and $d$ the depth of the fluid. Without rotation, the critical value is $R_c = 1100$. The effect of rotation about a vertical axis is specified by another dimensionless parameter, the Taylor number,

$$Ta \equiv \frac{4\Omega^2 d^4}{\nu^2} = \left(\frac{fd^2}{\nu}\right)^2 \tag{2.6.7}$$

where $\Omega$ becomes $\Omega \sin \phi$ in the spherical problem. For example, when $Ta = 2 \times 10^3$, convection does not set in until $R_c = 2376$. As an example of the Taylor number, the Earth's boundary layer (see Section 2.2.1) has $d \sim 10^5$ cm and $\nu_E \sim 5 \times 10^4$ cm$^2$/sec; with $f \sim 10^{-4}$/sec, we have $Ta \sim 4 \times 10^2$.

Computer modeling of the Bénard problem for a rotating spherical planet with nonisotropic *eddy viscosity* $\nu_E$ and *thermal diffusivity* $\chi_E$ indicates that turbulent convection will create the zones and belts, with the proper velocities. At high latitudes, the banded structure disappears, presumably because the Taylor number's dependence on latitude leads to a suppression of convection. Figure 2.12 shows the results of a computer simulation of circulation in the Jovian atmosphere.

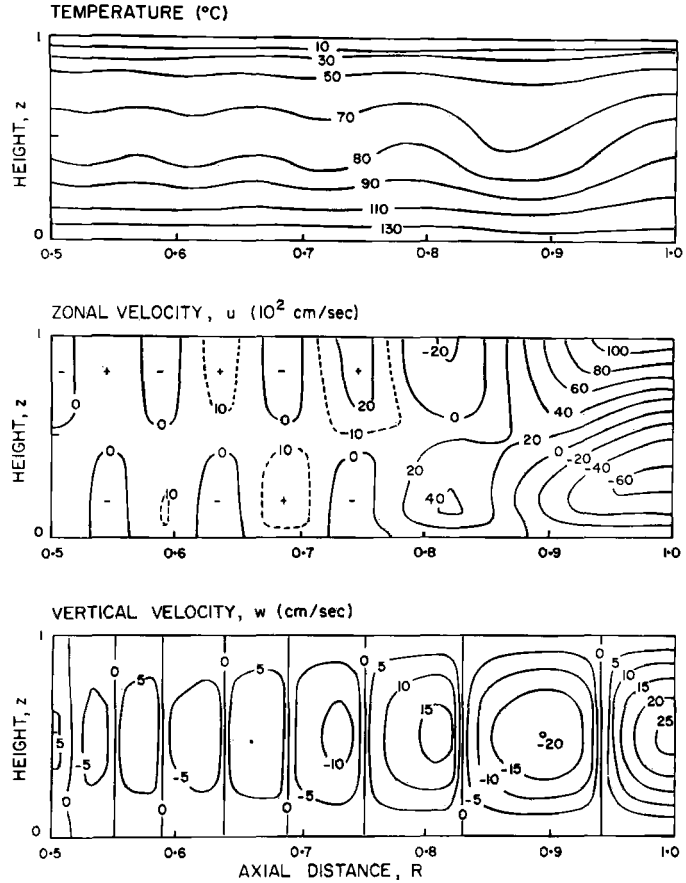



**Fig. 2.12** Contours of temperature, zonal winds, and vertical motions in a numerical simulation of Jupiter's atmosphere for a large-scale convection driven by internal heating. [Adapted from calculations by WILLIAMS and ROBINSON (1973).]

Near the equator there is an eastward jet of 100 m/sec at high levels with a retrograde motion of 60 m/sec below it. Between these jets and the inactive polar regions (beginning at $\phi = \pm 45°$) are alternating positive and negative zonal flow correlated with the banded structure. These upper level jets are countered by opposite flows below.

Attractive as it is, this model of free convection, driven by heating from below, has problems. The model calculations use the observable quantities to constrain the adopted parameters, but there is one observable quantity left over—the energy flux through the atmosphere. The main source of flux in the model is the convective flow and the parameters that produce the right scales of the motions also produce much too much outward energy flow.

Other models of the general circulation of Jupiter involve a more stable dynamical situation than free, dry convection. Below the visible clouds (near the 300°K level; see Fig. 1.19) condensation of water is important. With moist convection much of the heat flux from below is latent rather than sensible heat. Very likely the banded structure of Jupiter is due to a large-scale convection pattern driven by the internal heat source but with the phase change of water playing a controlling thermodynamic role.

Baroclinic instabilities may become important at middle latitudes, where they would have the effect of over-riding (i.e., stabilizing) the convection overturning. However, as noted above, the horizontal temperature gradients are smaller than on Earth and Mars because of the larger scale of Jupiter and its internal heat source, and consequently baroclinic instabilities must have little effect in equatorial latitudes.

## BIBLIOGRAPHICAL NOTES

**Section 2.1 Basic Equations**

An advanced but concise and clear text on hydrodynamics is

LANDAU, L. D., and LIFSHITZ, E. M. (1959), "Fluid Mechanics," Pergamon Press, New York.

**Section 2.2 Horizontal Circulation of the Troposphere**

The pioneering paper on the general circulation is

HADLEY, G. (1735), Concerning the cause of the general trade-winds, *Phil. Trans. Roy. Soc. London* **39**, 58–60 [reprinted in *Smithsonian Inst. Misc. Collections* **51**, 5–7, 1910].

Modern meteorology was founded around the close of World War I by the Norwegian school at Bergen led by VILHELM BJERKNES. One of the principal papers of that period, by his son JACOB, identifies warm and cold fronts for the first time:

BJERKNES, J. (1919), On the structure of moving cyclones, *Geofys. Publ.* 1, No. 2, 8 pp.

Much of the present thinking on planetary flow patterns is due to ROSSBY (another product of Bergen) and his collaborators. Much of their work is reviewed in

ROSSBY, C.-G. (1949), On the nature of the general circulation of the lower atmosphere, *in* "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), pp. 16–48, Univ. Chicago Press, Chicago, [2nd ed., 1952].

The theory of planetary circulation is developed in

LORENZ, E. N. (1967), "The Nature and Theory of the General Circulation of the Atmosphere," World Meteorological Organization, Geneva.

Most of the discussion in Section 2.2 has been based on Chapters 2 through 7 of the excellent text

HOLTON, J. R. (1972), "An Introduction to Dynamic Meteorology," Academic Press, New York.

A more descriptive text, which in many ways is complementary to Holton, is

PALMÉN, E. and NEWTON, C. W. (1969), "Atmospheric Circulations Systems," Academic Press, New York.

That dynamical instabilities associated with mid-latitude westerlies were crucial to the origin and development of cyclonic disturbances was demonstrated in

CHARNEY, J. G. (1947), The dynamics of long waves in a baroclinic westerly current, *J. Meteorology* **4**, 135–162.

A number of classic papers in dynamic meteorology (including CHARNEY, 1947) are reprinted in

SALTZMAN, B. (ed.) (1962), "Theory of Thermal Convection," Dover, New York.

A theoretical comparative study of circulation in atmospheres of the solar system has been developed by

GOLITSYN, G. S. (1970), A similarity approach to the general circulation of planetary atmospheres, *Icarus* **13**, 1–24.

The role of planetary waves and tides at high altitudes is discussed in

VOLLAND, H. and MAYR, H. G. (1977), Theoretical aspects of tidal and planetary wave propagation at thermospheric heights, *Rev. Geophys. Space Phys.* **15**, 203–226.

**Section 2.3 Vertical Transport**

The importance of molecular diffusion above 100 km in Earth's atmosphere was first emphasized by

NICOLET, M. and MANGE, P. (1954), The dissociation of oxygen in the high atmosphere, *J. Geophys. Res.* **59**, 15–45.

NICOLET, M. (1954), Dynamic effects in the high atmosphere, *in* "The Earth as a Planet," (G. P. Kuiper, ed.), Univ. Chicago Press, Chicago.

The standard theoretical work on molecular diffusion, viscosity, and conductivity is

CHAPMAN, S. and COWLING, T. G. (1970), "The Mathematical Theory of Non-Uniform Gases," 3d. ed., Cambridge University Press, Cambridge.

The importance of eddy diffusion to atmospheric structure and composition was established by

COLEGROVE, F. D.; HANSON, W. B.; and JOHNSON, F. S. (1965), Eddy diffusion and oxygen transport in the lower thermosphere, *J. Geophys. Res.* **70**, 4931–4941.

COLEGROVE, F. D.; HANSON, W. B.; and JOHNSON, F. S. (1966), Atmospheric composition in the lower thermosphere, *ibid.* **71**, 2227–2236.

An excellent discussion of eddy diffusion is given in

HUNTEN, D. M. (1975), Vertical transport in atmospheres, *in* "Atmospheres of Earth and the Planets," (B. M. McCormac, ed.), pp. 59–72, D. Reidel Publ. Co., Dordrecht, The Netherlands.

**Section 2.4 Circulation of the Venus Atmosphere**

The discussion here of the lower Venus atmosphere is based on

STONE, P. H. (1975), The dynamics of the atmosphere of Venus, *J. Atmos. Sci.* **32**, 1005–1016.

Computerized modeling of the Venus circulation has been conducted in two laboratories with results that are not entirely consistent with one another:

KÁLNAY DE RIVAS, E. (1975), Further numerical calculations of the circulation of the atmosphere of Venus, *J. Atmos. Sci.* **32**, 1017–1024.
POLLACK, J. B. and YOUNG, R. (1975), Calculations of the radiative and dynamical state of the Venus atmosphere, *J. Atmos. Sci.* **32**, 1025–1037,
YOUNG, R. E. and POLLACK, J. B. (1977), A three-dimensional model of dynamical processes in the Venus atmosphere, *J. Atmos. Sci.* **34**, 1315–1351.

The four-day rotation of the upper atmosphere of Venus was established by

BOYER, C. and CAMICHEL, H. (1961), Observations photographiques de la planète Vénus, *Ann. Astrophys.* **24**, 531–535.

The "moving flame" mechanism was developed by

SCHUBERT, G. and WHITEHEAD, J. (1969), Moving flame experiment with liquid mercury: Possible implications for the Venus atmosphere, *Science* **163**, 71–72,
YOUNG, R. and SCHUBERT, G. (1973), Dynamical aspects of the Venus 4-day circulation, *Planet. Space Sci.* **21**, 1563–1580.

**Section 2.5 Diurnal Winds on Mars**

A numerical model for the circulation of the Martian atmosphere was first developed by

LEOVY, C. B. and MINTZ, Y. (1969), Numerical simulation of the atmospheric circulation and climate of Mars, *J. Atmos. Sci.* **26**, 1167–1190.

Viking lander measurements of temperatures, winds, and pressures have been reported in

HESS, S., *et al.* (1976), Mars climatology from Viking I after 20 sols, *Science* **194**, 78–81.

The theory of tides in atmospheres is developed in

CHAPMAN, S. and LINDZEN, R. S. (1970), "Atmospheric Tides," Gordon and Breach, New York.

A discussion of tides expected in the Martian atmosphere is given by

ZUREK, R. W. (1976), Diurnal tides in the Martian atmosphere, *J. Atmos. Sci.* **33**, 321–337.

Growth of the ice caps is treated theoretically in

BRIGGS, G. A. (1974), The nature of the residual Martian polar caps, *Icarus* **23**, 167–191

**Section 2.6 Convection in the Jovian Atmosphere**

The nature of Jupiter's circulation has been investigated with the collective empirical evidence assembled by

CHAPMAN, C. R. (1969), Jupiter's zonal winds: Variation with latitude, *J. Atmos. Sci.* **26**, 986–990.

The first dynamical discussion of the banded structure along the lines illustrated in Fig. 2.10 is due to

HESS, S. L. and PANOFSKY, H. A. (1951), The atmospheres of the other planets, *in* "Compendium of Meteorology," pp. 391–400, Amer. Meteor. Soc., Boston.

Interpretation of the zonal flow as a baroclinic mechanism has been discussed by

GIERASCH, P. J., and STONE, P. H. (1968), A mechanism for Jupiter's equatorial acceleration, *J. Atmos. Sci.* **25**, 1169–1170.
STONE, P. H. (1972a), A simplified radiative-dynamical model for the static stability of rotating atmospheres, *J. Atmos. Sci.* **29**, 405–418.
STONE, P. H. (1972b) A non-geostrophic baroclinic stability: part III. The momentum and heat transports, *ibid.* **29**, 419–426.

The Bénard problem with rotation is treated in

CHANDRASEKHAR, S. (1961), "Hydrodynamic and Hydromagnetic Stability," Clarendon Press, Oxford.

The convective model discussed here is due to

WILLIAMS, G. P. and ROBINSON, J. B. (1973), Dynamics of a convectively unstable atmosphere: Jupiter? *J. Atmos. Sci.* **30**, 684–717.

A qualitatively similar model is suggested in

BUSSE, F. H. (1976), A simple model of convection in the Jovian atmosphere, *Icarus* **29**, 255–260.

Since the publication of WILLIAMS and ROBINSON (1973), Williams has run computer models based on conventional (i.e., terrestrial) meteorological equations scaled to Jupiter. He found that the banded structure and oval-shaped disturbances (including the Great Red Spot) may be characteristic of turbulent barotropic vorticity exchanges in a rapidly rotating atmosphere. As on Earth, baroclinic instability would be the primary energy conversion process, and strong convection would not be required. A summary is given in

WILLIAMS, G. P. (1975), Jupiter's atmospheric circulation, *Nature* **257**, 778 (only).

The controlling influence of the latent heat of water on the Jovian meteorology was proposed by

BARCILON, A. and GIERASCH, P. J. (1970), A moist Hadley cell model for Jupiter's cloud bands, *J. Atmos. Sci.* **27**, 550–560.

The role of water in a large-scale convective circulation has been examined by

GIERASCH, P. J. (1976), Jovian meteorology: Large-scale moist convection, *Icarus* **29**, 445–454.

An excellent review of the dynamics of Jupiter's atmosphere is

STONE, P. H. (1976), The meteorology of the Jovian atmosphere, *in* "Jupiter," (T. Gehrels, ed.), pp. 586–617, Univ. Arizona Press, Tucson.

## PROBLEMS

**2.1** *Inertial oscillation.* Solve the simplified equation of horizontal motion for the case of uniform horizontal pressure and show that air parcels will move in anticyclonic (clockwise) circles with periods of one-half the rotation period of a Foucault pendulum.

**2.2** *Tornado pressure.* A tornado rotates with constant angular velocity and has a uniform temperature. Find an expression for the outward pressure distribution in terms of the central pressure $p_0$. Take $T = 300°\text{K}$ and, at 0.1 km from the center, take $p = 1$ atm and $V = 10^4$ cm/sec. What is $p_0$?

**2.3** *Gradient flow.* (a) What is the geostrophic wind speed at the surface at $\phi = 30°$ when the pressure gradient is 1 mb/100 km? (b) What are the cyclonic and anticyclonic gradient wind speeds for the same conditions and a radius of curvature of 600 km?

**2.4** *Jet stream.* What is the jet stream velocity (a) in summer at 250 mb (10 km) and $\phi = 45°$ if the mean $dT/d\phi = -0.5°\text{K/deg}$, and (b) in winter at 200 mb (12 km) and $\phi = 30°$, with $dT/d\phi = -1°\text{K/deg}$. In both cases take the zonal wind to be a westerly of 10 m/sec at the 600 mb level. (c) Estimate the difference in fuel costs for a jet transport crossing the U.S. traveling eastward compared with westward in the winter jet stream. (Figure fuel consumption at 10,000 lb/hr; jet fuel costs about half the price of gasoline.)

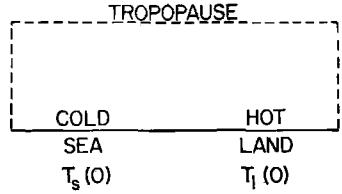
**2.5** *Static stability parameter (I).* Show that $\sigma(\Phi)$ is

$$\sigma = \frac{1}{p^2}\left(\frac{\partial}{\partial \ln p} - \frac{R}{C_p}\right)\frac{\partial \Phi}{\partial \ln p}$$

**2.6** *Static stability parameter (II).* Use the expression in Problem 2.5 to find $\sigma(p)$ when the atmosphere is (a) adiabatic and (b) isothermal.

**2.7** *Vorticity equation.* Show that the isobaric equations of motion give the vorticity equation (2.2.37).

**2.8** *Sea breeze.* Idealize a cross section of a coastal region as being bounded on either side and the top by impenetrable walls. Over the sea the surface temperature is $T_s(0) = 12°C$. Over land, $T_l(0) = 27°C$. The surface pressure over both regions is nearly equal, $p(0) \approx 1000$ mb, and $T$ decreases with height over both regions at a rate $dT/dz = -6.5°C/km$. (a) Find the pressures at 10 km over each region. (b) From considerations of continuity, what is the direction of circulation within the closed cell? (c) What will happen at night if the land air cools rapidly enough to become colder than the sea air?




**2.9** *Eddy diffusion.* What is the density distribution for a minor constituent in an isothermal atmosphere when the flux is constant and the eddy diffusion coefficient varies as $\exp[(z - z_0)/L]$ when (a) $L = \frac{1}{2}H$ and (b) $L = H$?

**2.10** *Bénard cells.* Suppose that in the horizontal plane at the onset of convection the Bénard cells consist of a network of uniform, regular, and contiguous polygons. (Within each cell the convection flow is upward near the center and downward near the perimeter.) Show that these polygons can only be equilateral triangles, squares, or regular hexagons.

# *Chapter 3*
# CHEMISTRY AND DYNAMICS OF EARTH'S STRATOSPHERE

## 3.1 Principles of Photochemistry

Photochemistry is concerned with chemical reactions that are initiated by the absorption of radiation. Thus schematically we write

$$A + h\nu \rightarrow A^* \tag{3.1.1}$$

meaning that molecule A absorbs a quantum of radiation and makes an energy transition to the excited molecular state designated A*. The excited state could be an excited rotational or vibrational level, an excited electronic state, or it could be a dissociated or an ionized molecule. The *quantum yield* of the process (3.1.1) is the number of A* molecules produced for each absorbed photon that has adequate energy to produce state A*.

Once excited molecules are produced, they enter into secondary chemical reactions, forming species that would not exist (in the same proportions, at least) in a state of thermodynamic equilibrium at the local kinetic temperature. Hence photochemical processes are non-LTE mechanisms (see Section 1.2.1), and they have to be handled quantitatively by examining the rate at which each reaction occurs. This fact constitutes the basic complexity of photochemical theory.

Suppose reaction (3.1.1) has an absorption cross section $\alpha(\nu)$ ($cm^2$). If the total cross section of A at frequency $\nu$ is $\alpha_T(\nu)$, the monochromatic quantum efficiency is $\alpha(\nu)/\alpha_T(\nu)$. The instantaneous rate of production of A* (square brackets denote concentration in molecules/$cm^3$) is

$$\frac{d[A^*]}{dt} = [A] \int_{\nu_0}^{\infty} \pi \mathscr{F}_\nu e^{-\tau_\nu/\mu} \alpha(\nu)\, d\nu$$
$$\equiv [A] J \tag{3.1.2}$$

Here $hv_0$ is the threshold energy for reaction (3.1.1), $\pi\mathscr{F}_\nu$ (photon/cm$^2$ sec Hz) is the photon flux per unit frequency interval outside the Earth's atmosphere. The vertical optical thickness is

$$\tau_\nu(z) = \sum_i \alpha_i(\nu) \int_z^\infty N_i(z')\, dz' \tag{3.1.3}$$

where the summation is taken over every constituent with abundance $N_i$ that absorbs at frequency $\nu$. The factor $1/\mu$ is the slant air-mass factor for the zenith angle $\chi$ of the sun. Unless one is especially concerned with sunrise or sunset effects, $\mu = \cos\chi$ is an adequate approximation. The second equality in (3.1.2) defines $J$ (sec$^{-1}$), the production rate of A* per molecule of A. If A* represents a partial or total dissociation of molecule A, then J is commonly called the *photolysis rate.*

A two-body reaction,

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{C} \tag{3.1.4}$$

proceeds at the rate

$$\frac{d[\mathrm{C}]}{dt} = [\mathrm{A}][\mathrm{B}] \iint Q_{\mathrm{AB}}(g) f_{\mathrm{A}}(v_{\mathrm{A}}) f_{\mathrm{B}}(v_{\mathrm{B}}) g\, d^3v_{\mathrm{A}}\, d^3v_{\mathrm{B}}$$

$$\equiv [\mathrm{A}][\mathrm{B}] k_{\mathrm{AB}} \tag{3.1.5}$$

The $f$'s represent normalized distribution functions for the molecular velocities; $g$ is the relative scalar velocity of collision, so that $g = |\mathbf{v}_{\mathrm{A}} - \mathbf{v}_{\mathrm{B}}|$; and $Q_{\mathrm{AB}}(g)$ is the collision cross section. This cross section is zero at velocities

$$g < g_0 \equiv \left(\frac{2\varepsilon_0}{M_{\mathrm{AB}}}\right)^{1/2} \tag{3.1.6}$$

where $M_{\mathrm{AB}}$ is the reduced mass of A and B, and $\varepsilon_0$ is the *activation energy* that must be supplied to overcome any potential barrier that (3.1.4) might have (see Problem 3.1).

The *rate coefficient* $k_{\mathrm{AB}}$ (cm$^3$/sec) defined by (3.1.5) is conventionally written to show a temperature dependence as

$$k_{\mathrm{AB}} = a(T/300)^b e^{-c/T} \tag{3.1.7}$$

where $T$ is expressed in degrees Kelvin. The quantity $c$ is the activation energy in °K (i.e., $c = \varepsilon_0/k$) and the exponential is essentially a Boltzmann factor giving the fraction of a population that possesses the necessary kinetic energy to make a reaction occur.

In atmospheric physics it is occasionally desirable to refer a rate coefficient to the gas-kinetic rate for elastic collisions. Most atmospheric molecules have diameters several times that of the first Bohr orbit or cross sections of $Q \approx 5 \times 10^{-15}$ cm$^2$. At 300°K the mean speed of mass 28($N_2$) is $\langle v \rangle \sim 4 \times$

$10^4$ cm/sec. Hence the gas kinetic rate is about

$$k_{\text{g.k.}} = Q\langle v\rangle \approx 2 \times 10^{-10}\,(T/300)^{1/2}\ \text{cm}^3/\text{sec} \tag{3.1.8}$$

Unless there are long-range (e.g., Coulomb) forces at work, this value is an upper limit for chemical rates.

By analogy with (3.1.5) we can define a rate coefficient for the three-body collision

$$\text{A} + \text{B} + \text{M} \rightarrow \text{C} + \text{M} \tag{3.1.9}$$

in which M serves to stabilize intermediate products of the reaction and to help balance the kinetic energy and momentum. This reaction proceeds at the rate

$$d[C]/dt = [\text{A}][\text{B}][\text{M}]\text{k}_{\text{AABM}} \tag{3.1.10}$$

Proceeding with our discussion of gas-kinetic collisions, we make an estimate for an upper limit to $k_{\text{ABM}}$. The two-body collision frequency for molecule A is

$$\nu_c(2) = \frac{1}{[\text{A}]}\frac{d[\text{A}]}{dt} = [\text{B}]k_{\text{AB}} \tag{3.1.11}$$

The mean time between collisions is $\nu_c^{-1}$, but the duration of a collision is of order $\tau_{\text{coll}} = d/\langle v\rangle$, where $d$ ($=4 \times 10^{-8}$ cm) is the molecular diameter. Hence the probability that, during a two-body collision, a third body M will also be in the immediate neighborhood is $\tau_{\text{coll}}/\nu_c^{-1}$. The three-body collision frequency is thus $\nu_c(2)$ times the number of collisions occurring during the time of the two-body collision, or

$$\begin{aligned}\frac{1}{[\text{A}]}\frac{d[\text{A}]}{dt} = \nu_c(3) &= \nu_c(2)[\nu_c(2)\tau_{\text{coll}}] \\ &= [\text{B}]k_{\text{AB}} \times [\text{M}]k_{\text{AB}} \times 1 \times 10^{-12}\end{aligned} \tag{3.1.12}$$

Comparison of this expression with (3.1.10) gives

$$k_{\text{ABM}} = 1 \times 10^{-12}k_{\text{AB}}^2 = 4 \times 10^{-32}\,(T/300)\ \text{cm}^6/\text{sec} \tag{3.1.13}$$

where we use the gas-kinetic value for $k_{\text{AB}}$.

For gas-kinetic collisions, Eq. (3.1.5) and (3.1.10) give equal rates for

$$[\text{M}] = k_{\text{AB}}/k_{\text{ABM}} = 5 \times 10^{21}\ \text{cm}^{-3} \tag{3.1.14}$$

Consequently, if all two-body reactions occurred at least once every $10^2$ (gas-kinetic) collisions, three-body reactions would be negligible everywhere, even at the ground. But in some cases $k_{\text{AB}}$ is extremely slow (for example, $\text{O} + \text{O} \rightarrow \text{O}_2 + h\nu$ has $k < 10^{-20}$ cm$^3$/sec), and the three-body process is the dominant mechanism.

Table 3.1 lists dissociation energies for the major, and a number of minor, constituents of Earth's atmosphere. Table 3.2 gives sample photolysis rates

TABLE 3.1 *Dissociation Energies of Neutral Molecules*[a]

| Species | Dissociation energy | | |
|---|---|---|---|
| | eV/molecule | kcal/mole | Å (vac) |
| $H_2$ | 4.479 | 103.266 | 2768 |
| CO | 11.111 | 256.163 | 1116 |
| $N_2$ | 9.762 | 225.061 | 1270 |
| NO | 6.509 | 150.055 | 1905 |
| $O_2$ | 5.117 | 117.967 | 2423 |
| OH | 4.395 | 101.33 | 2821 |
| ClO | 2.73 | 63 | ~4540 |
| HCl | 4.44 | 102 | ~2800 |
| HF | 5.86 | 135 | ~2115 |
| $H_2O \rightarrow H + OH$ | 5.117 | 117.98 | 2423 |
| $HO_2 \rightarrow O + OH$ | 2.73 | 63 | 4540 |
| $CO_2 \rightarrow CO + O$ | 5.455 | 125.750 | 2273 |
| $NO_2 \rightarrow NO + O$ | 3.117 | 71.86 | 3977 |
| $N_2O \rightarrow N_2 + O$ | 1.677 | 38.66 | 7393 |
| $O_3 \rightarrow O_2 + O$ | 1.052 | 24.25 | 11,785 |
| $CH_4 \rightarrow CH_3 + H$ | 4.45 | 103 | ~2785 |
| $H_2O_2 \rightarrow H_2O + O$ | 1.43 | 33 | ~8700 |
| $NO_3 \rightarrow NO_2 + O$ | 2.17 | 50 | ~5700 |
| $HNO_3 \rightarrow OH + NO_2$ | 2.30 | 53 | ~5400 |
| $ClNO_3 \rightarrow ClO + NO_2$ | 1.13 | 26.1 | ~10,970 |

[a] Wavelength equivalents of the dissociation energy are not necessarily the threshold wavelengths at which photodissociation actually occurs, because the cross section for photodissociation may be negligibly small at the wavelength corresponding to the dissociation energy.

TABLE 3.2 *Daily Mean Photolysis Rates*[a] ($sec^{-1}$) *(Averaged Day-Night Rates at* 30°N *Latitude, Solar Declination* +12°)

| Height $z$ (km) | $J(O_2)$ (1.5.1) | $J(O_3)$ (1.5.4) | $J(NO_2)$ (3.2.29) | $J(N_2O)$ (3.2.49) | $J(HNO_3)$ (3.2.32) |
|---|---|---|---|---|---|
| 60 | 5.7(−10) | 4.0(−3) | 4.42(−3) | 2.9(−7) | 5.3(−5) |
| 50 | 3.7(−10) | 2.9(−3) | 4.41(−3) | 2.4(−7) | 4.4(−5) |
| 40 | 1.5(−10) | 5.9(−4) | 4.35(−3) | 1.4(−7) | 2.4(−5) |
| 30 | 1.1(−11) | 8.6(−5) | 4.23(−3) | 1.6(−8) | 3.2(−6) |
| 20 | 4.7(−14) | 3.2(−5) | 4.12(−3) | 7.7(−11) | 2.2(−7) |

[a] Numbers in parentheses are powers of 10. Calculations from a 1973 program by McElroy et al. (1974). Revised cross sections (Johnston and Selwyn, 1975) reduce $J(N_2O)$ at 20 km and below. Scattering of sunlight will increase the $J$'s in some cases.

that bear on the chemistry discussed in this chapter. Appendix IV presents a model atmosphere for average conditions at 45° latitude.

## 3.2 Catalytic Destruction of Ozone

### 3.2.1 Pure Oxygen Chemistry

To recapitulate from Section 1.5.1 the four Chapman reactions involving only the allotropes of oxygen are (1.5.1–1.5.4) and have dissociation rates or rate coefficients of $J_2$, $k_{12}$, $k_{13}$, and $J_3$, respectively. The rates of change of $[O]$ and $[O_3]$ are given by (1.5.8–1.5.9) and the change in odd oxygen is

$$\frac{d([O]+[O_3])}{dt} = 2J_2[O_2] - 2k_{13}[O][O_3] \tag{3.2.1}$$

Throughout the stratosphere, O is in equilibrium with $O_3$ and $O_2$ because of the relatively small amounts of O that are present. However, $O_3$ is not necessarily in photochemical equilibrium in the lower stratosphere.

Setting $d[O]/dt = 0$ in (3.2.1) and (1.5.8), we can write from (3.2.1)

$$\frac{d[O_3]}{dt} = -\frac{2k_{13}J_3[O_3]^2}{k_{12}[O_2][M]} + 2J_2[O_2] \tag{3.2.2}$$

where we have used the empirical fact that $k_{13}[O_3] \ll k_{12}[O_2][M]$. An integration for $[O_3]$ increasing with $t$ yields

$$t - t_0 = \left(\frac{k_{12}[M]}{16k_{13}J_2J_3}\right)^{1/2} \ln\left\{\left(\frac{[O_3]_{eq}+[O_3]}{[O_3]_{eq}-[O_3]}\right)\left(\frac{[O_3]_{eq}-[O_3]_0}{[O_3]_{eq}+[O_3]_0}\right)\right\} \tag{3.2.3}$$

where $[O_3]_0$ is the abundance at $t_0$ and $[O_3]_{eq}$ is the equilibrium abundance of Eq. (1.5.12). Let $\tau_{eq}$ be the time for $O_3$ to increase from zero to $0.5[O_3]_{eq}$ (or, what is equivalent, the time to increase from 50 percent to 80 percent of $[O_3]_{eq}$—in either case the logarithm factor is $\ln 3 = 1.10$). Then

$$\tau_{eq}(O_3) = \frac{0.275[O_3]_{eq}}{[O_2]J_2} \tag{3.2.4}$$

Figure 3.1 shows this $\tau_{eq}$ as a function of height computed with $J_2$'s for different zenith angles. The figure has two important messages:

(1) Below 30 km the lifetime $\tau_{eq}$ is measured in weeks; below 25 km, in years. With such long photochemical lifetimes we shall find that dynamics (horizontal and vertical mixing) controls the distribution of ozone. A simple, local photochemical equilibrium—indeed, even a steady state—does not exist for $O_3$ in the lower stratosphere. We will return to this matter in Section 3.3.2.

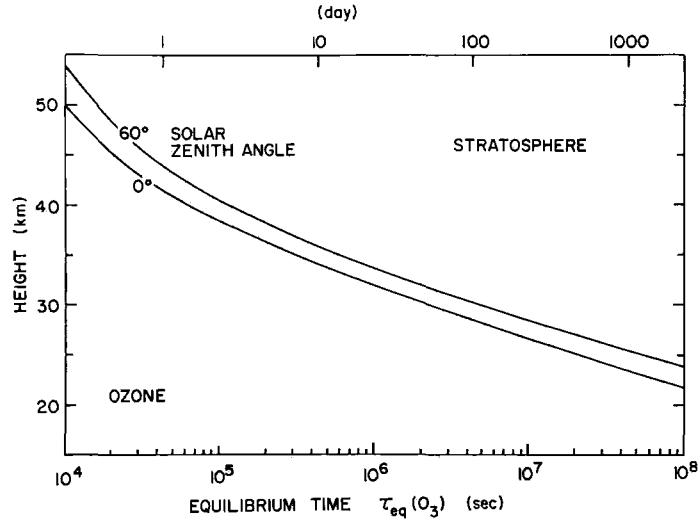



**Fig. 3.1** Equilibrium time $\tau_{eq}(O_3)$ for ozone in a pure oxygen atmosphere, as calculated from (3.2.4), where $J_2$ is a function of zenith angle and height. [Adapted from NICOLET (1975).]

(2) Above 40 km a chemical equilibrium should exist, because the lifetimes for adjustment become less than a day. Nevertheless, the observed values of $[O_3]$ are less than the values calculated from (1.5.12) and there seems to be no possibility that $k_{12}$, $k_{13}$, $J_2$, or $J_3$ can be in error enough to make up the discrepancy. Thus pure oxygen chemistry presents an internal inconsistency. We will find the resolution of this dilemma in catalytic chemistry with minor constituents $HO_x$, $NO_x$, and $ClO_x$. In this notation, for $x$ read 0, 1, or 2; just as O and $O_3$ constitute "odd oxygen" molecules, $HO_x$, $NO_x$, and $ClO_x$ represent odd hydrogen, odd nitrogen, and odd chlorine molecules.

### 3.2.2 Photochemistry of Hydrogenous Radicals

The free radicals $HO_x$ are H, OH (hydroxyl), and $HO_2$ (hydroperoxyl). The single H atom makes these radicals highly reactive with odd oxygen O and $O_3$. The production of $HO_x$ originates with the dissociation of $H_2O$ and $CH_4$ (methane).

The straightforward photodissociation of water into H + OH is shielded in the stratosphere by the Herzberg dissociation continuum of $O_2$ and is less important than the following indirect process: Metastable $O(^1D)$ is produced in the photodissociation of ozone,

$$O_3 + h\nu(\lambda < 3100\,\text{Å}) \rightarrow O_2 + O(^1D) \tag{3.2.5}$$

with rate $J_3$. The $O(^1D)$ atom has a radiative lifetime of nearly two minutes, but it is deactivated in much less than a second by collisions with $O_2$ and $N_2$ by

$$O(^1D) + M \rightarrow O(^3P) + M \qquad (3.2.6)$$
$$\alpha_6 = 4.5 \times 10^{-11} \text{ cm}^3/\text{sec}$$

where $\alpha$ represents the rate coefficient. Nevertheless, there is always a daytime trace abundance of

$$[O(^1D)] = \frac{J_3[O_3]}{\alpha_6[M]} \qquad (3.2.7)$$

which is about 10 $\text{cm}^{-3}$ at 30 km (cf. Fig. 3.3).

The $O(^1D)$ atoms, carrying 1.96 eV of excitation energy, then form OH by

$$O(^1D) + H_2O \rightarrow 2OH + 1.24 \text{ eV} \qquad (3.2.8)$$
$$\alpha_8 = 3 \times 10^{-10} \text{ cm}^3/\text{sec}$$

and

$$O(^1D) + CH_4 \rightarrow OH + CH_3 + 1.91 \text{ eV} \qquad (3.2.9)$$
$$\alpha_9 = 3 \times 10^{-10} \text{ cm}^3/\text{sec}$$

The latter reaction is the dominant destruction mechanism of $CH_4$ in the stratosphere; it becomes photodissociated only in the mesosphere, although at lower altitudes OH also destroys $CH_4$ [cf. reaction (3.3.2)]. Additional $HO_x$ is formed by the further oxidation of $CH_3$ (methyl radical).

Once formed, $HO_x$ destroys odd oxygen with no destruction of $HO_x$ itself. At the stratopause level ($\sim 50$ km), where $HO_x$ is most important, the reactions on O and $O_3$ are

$$H + O_3 \rightarrow OH(v \leq 9) + O_2 + 3.34 \text{ eV} \qquad (3.2.10)$$
$$\alpha_{10} = 1.2 \times 10^{-10} e^{-560/T} \text{ cm}^3/\text{sec}$$

$$OH + O_3 \rightarrow HO_2 + O_2 + 1.68 \text{ eV} \qquad (3.2.11)$$
$$\alpha_{11} = 1.6 \times 10^{-12} e^{-1000/T} \text{ cm}^3/\text{sec}$$

$$HO_2 + O_3 \rightarrow OH + 2O_2 + 1.34 \text{ eV} \qquad (3.2.12)$$
$$\alpha_{12} \approx 1 \times 10^{-13} e^{-1250/T} \text{ cm}^3/\text{sec}$$

$$OH + O \rightarrow H + O_2 + 0.72 \text{ eV} \qquad (3.2.13)$$
$$\alpha_{13} = (5 \pm 2) \times 10^{-11} \text{ cm}^3/\text{sec}$$

and

$$HO_2 + O \rightarrow OH(v \leq 6) + O_2 + 2.39 \text{ eV} \qquad (3.2.14)$$
$$\alpha_{14} = (4 \pm 3) \times 10^{-11} \text{ cm}^3/\text{sec}$$

In addition, $HO_x$ is internally redistributed without destroying odd oxygen by

$$H + O_2 + M \rightarrow HO_2 + M + 2.01 \text{ eV} \qquad (3.2.15)$$
$$\alpha_{15} = 2.1 \times 10^{-32} e^{290/T} \text{ cm}^3/\text{sec}$$

To complete the $HO_x$ cycle in the stratosphere, the family is destroyed by

$$OH + OH \rightarrow H_2O + O + 0.72 \text{ eV} \qquad (3.2.16)$$
$$\alpha_{16} \approx 2 \times 10^{-12} \text{ cm}^3/\text{sec}$$

$$OH + HO_2 \rightarrow H_2O + O_2 + 3.11 \text{ eV} \qquad (3.2.17)$$
$$\alpha_{17} = 2 \times 10^{-11} \text{ cm}^3/\text{sec}$$

and

$$HO_2 + HO_2 \rightarrow H_2O_2 + O_2 + 1.8 \text{ eV} \qquad (3.2.18)$$
$$\alpha_{18} \sim 5 \times 10^{-12} \text{ cm}^3/\text{sec}$$

The $H_2O_2$ (hydrogen peroxide) formed in the last reaction may be photodissociated into 2OH or it may be attacked by OH,

$$OH + H_2O_2 \rightarrow HO_2 + H_2O + 1.30 \text{ eV} \qquad (3.2.19)$$
$$\alpha_{19} = 8 \times 10^{-12} e^{-600/T} \text{ cm}^3/\text{sec}$$

Destruction of $HO_x$ may also occur through $CH_4$; see (3.3.2) below.

Analysis of these reactions is best handled in two parts: first, the catalytic reactions on O and $O_3$, and second, the production and loss of $HO_x$. The catalytic reactions are shown schematically in Fig. 3.2. Where photochemical equilibrium prevails, we can readily write down the equilibrium abundance

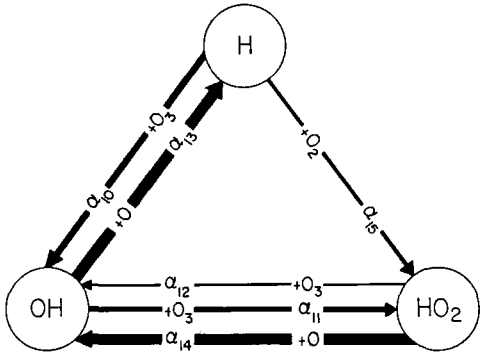



**Fig. 3.2** Internal reactions of the $HO_x$ family. The widths of the arrows are rough indicators of the relative importance of the reactions near the stratopause ($z \sim 50$ km), where the influence of $HO_x$ on odd oxygen is most important.

ratios. The [H] equilibrium gives

$$\frac{[OH]}{[H]} = \frac{\alpha_{10}[O_3] + \alpha_{15}[O_2][M]}{\alpha_{13}[O]} \tag{3.2.20}$$

With the low rate coefficients $\alpha_{11}$ and $\alpha_{12}$, we have

$$\frac{[HO_2]}{[H]} = \frac{\alpha_{15}[O_2][M]}{\alpha_{14}[O]} \tag{3.2.21}$$

from the $[HO_2]$ equilibrium.

Adding the hydrogenous reactions to the Chapman reactions, we find that $[O_3]$ equilibrium yields

$$[O_3] = \frac{k_{12}[O][O_2][M]}{J_3 + k_{13}[O] + \alpha_{12}[HO_2] + \alpha_{11}[OH] + \alpha_{10}[H]} \tag{3.2.22}$$

where $k_{13}[O] \ll J_3$ below 60 km, and $[O + O_3]$ equilibrium gives

$$[O] = \frac{2J_2[O_2] - \alpha_{10}[H][O_3] - \alpha_{11}[OH][O_3] - \alpha_{12}[HO_2][O_3]}{2k_{13}[O_3] + \alpha_{13}[OH] + \alpha_{14}[HO_2]} \tag{3.2.23}$$

Anticipating our results somewhat, we can see that [H] will be very small because of the small [O] in (3.2.20–3.2.21). Also, we will be expecting other $HO_x$ abundances of the order of $10^7$ cm$^{-3}$ or less. These values allow us to discard the terms in $\alpha_{10}$, $\alpha_{11}$, and $\alpha_{12}$ obtaining

$$\frac{[O_3]^2}{[O_2]^2} = \frac{k_{12}J_2[M]}{k_{13}J_3(1 + A)} \tag{3.2.24}$$

which is identical to the Chapman solution (1.5.12) except for the correction term,

$$A = \frac{\alpha_{13}[OH] + \alpha_{14}[HO_2]}{2k_{13}[O_3]} \tag{3.2.25}$$

Typical values at the 50 km stratopause are $k_{13} \sim 5 \times 10^{-15}$ cm$^3$/sec, $[O_3] \sim 5 \times 10^{10}$ cm$^{-3}$, and $\alpha_{13} \sim \alpha_{14} \sim 4 \times 10^{-11}$ cm$^3$/sec, although there is considerable uncertainty in $\alpha_{14}$. To have $A \gtrsim 1$ requires $[HO_x] \gtrsim 10^7$ cm$^{-3}$.

Returning now to the balance of production and loss of $HO_x$, we will omit the $CH_4$ chemistry for simplicity. Then production is primarily by (3.2.8) and loss occurs mainly through (3.2.17), or

$$[O(^1D)][H_2O]\alpha_8 = [OH][HO_2]\alpha_{17} \tag{3.2.26}$$

With $[O(^1D)]$ given in terms of the $[O_3]$ mixing ratio by (3.2.7) and with [OH] and $[HO_2]$ related by (3.2.20 and 3.2.21), we may solve for $[HO_x]$.

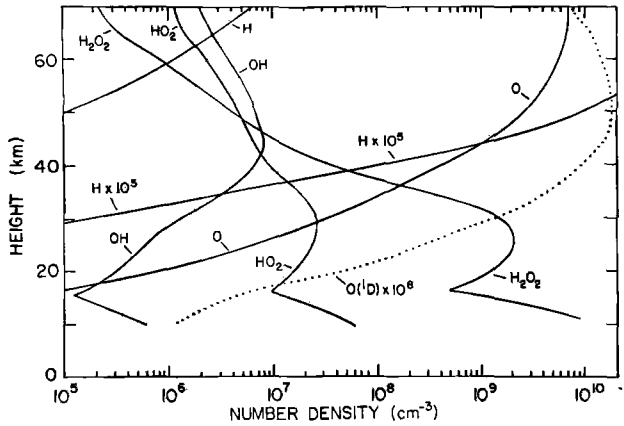



**Fig. 3.3** Model calculations of the distribution of hydrogenous radicals ($HO_x$), $H_2O_2$, $O(^3P)$, and $O(^1D)$ for the atmosphere at 30° latitude (24-hr averages). [After McElroy *et al.* 1974).]

Detailed calculations (including vertical mixing; see Section 3.3.1) as shown in Fig. 3.3 indicate that in the stratopause region ($z \sim 50$ km) the OH and $HO_2$ densities are of the order of $10^7$ cm$^{-3}$ and that $HO_x$ is indeed an important sink for O and $O_3$ in this region.

Referring again to Fig. 3.2 we see that the main reactions in the catalytic $HO_x$ cycle proceed clockwise around the triangle. Both $\alpha_{14}$ and $\alpha_{13}$ destroy O and then the $HO_x$ is rejuvenated by $\alpha_{15}$, which does not itself destroy odd oxygen. This three-way mechanism is in contrast to the $NO_x$ and $ClO_x$ cycles (to be discussed presently), which operate by two-way cycles. Thus $NO_x$ operates on the analogs to $\alpha_{11}$ and $\alpha_{14}$; $ClO_x$ operates on mechanisms like $\alpha_{10}$ and $\alpha_{13}$. Although the main $HO_x$ reactions in the stratosphere destroy O, these reactions must be regarded as destructive of $O_3$ since O is a potential $O_3$ molecule.

### 3.2.3 Oxides of Nitrogen

Odd oxygen is destroyed in a catalytic cycle by NO (nitric oxide) and $NO_2$ (nitrogen dioxide):

$$\begin{gathered} NO + O_3 \rightarrow NO_2 + O_2 + 2.06 \text{ eV} \\ \beta_{27} = 1.2 \times 10^{-12} e^{-1250/T} \text{ cm}^3/\text{sec} \end{gathered} \tag{3.2.27}$$

and

$$\begin{gathered} NO_2 + O \rightarrow NO + O_2 + 2.00 \text{ eV} \\ \beta_{28} = 9.3 \times 10^{-12} \text{ cm}^3/\text{sec} \end{gathered} \tag{3.2.28}$$

where $\beta$'s represent the rate coefficients. The first reaction does not always result in a net loss of ozone, because most often $NO_2$ is photodissociated,

$$NO_2 + h\nu(\lambda < 3977\ \text{Å}) \rightarrow NO + O \tag{3.2.29}$$

which is followed by

$$O + O_2 + M \rightarrow O_3 + M \tag{3.2.30}$$

The sequence of reactions (3.2.27, 3.2.29, and 3.2.30) produces no net change in the products (NO and $O_3$). Thus the process that limits the rate of $O_3$ destruction is (3.2.28). Whenever it occurs not only is an O atom (which represents a potential $O_3$ molecule) destroyed but a second $O_3$ is destroyed by (3.2.27).

The interaction of minor, catalytic substances can be important if the rate coefficients are high or if there is no other route to reach a specific end result. Thus $HO_x$ and $NO_x$ are rearranged by

$$HO_2 + NO \rightarrow OH + NO_2 \tag{3.2.31a}$$
$$\beta_{31a} \approx 8 \times 10^{-12}\ \text{cm}^3/\text{sec}$$

But the reaction is important mainly because (especially at night) a major fraction of $NO_2$ in the lower stratosphere is bound as $HNO_3$ (nitric acid) by

$$OH + NO_2 + M \rightarrow HNO_3 + M + 2.30\ \text{eV} \tag{3.2.31b}$$
$$\beta_{31b} = 1.1 \times 10^{-31} e^{900/T}\ \text{cm}^6/\text{sec}$$

This $NO_x$ is thereby removed temporarily from the catalytic cycle. In the daytime, photolysis of $HNO_3$ restores $NO_2$ by

$$HNO_3 + h\nu(\lambda < 3200\ \text{Å}) \rightarrow OH + NO_2 \tag{3.2.32}$$

Also, $NO_x$ in the form of $NO_2$ can be photodissociated, with the O atom producing an $O_3$. Thus the sequence (3.2.11, 3.2.31a, 3.2.29, and 3.2.30) is a *do-nothing cycle*, with all the original constituents recovered. Another reservoir for $NO_2$, especially at night, is $N_2O_5$ (nitrogen pentoxide) by

$$NO_2 + O_3 \rightarrow NO_3 + O_2 + 1.13\ \text{eV} \tag{3.2.33}$$
$$\beta_{33} = 1.3 \times 10^{-13} e^{-2450/T}\ \text{cm}^3/\text{sec}$$

followed by

$$NO_3 + NO_2 + \dot{M} \rightarrow N_2O_5 + M + 0.95\ \text{eV} \tag{3.2.34}$$
$$\beta_{34} = 2.8 \times 10^{-30}\ \text{cm}^6/\text{sec}$$

In the day $NO_2$ is recovered by

$$N_2O_5 + h\nu(\lambda < 3800\ \text{Å}) \rightarrow NO_2 + NO_3 \tag{3.2.35}$$

and

$$NO_3 + h\nu \rightarrow NO + O_2 \tag{3.2.36}$$

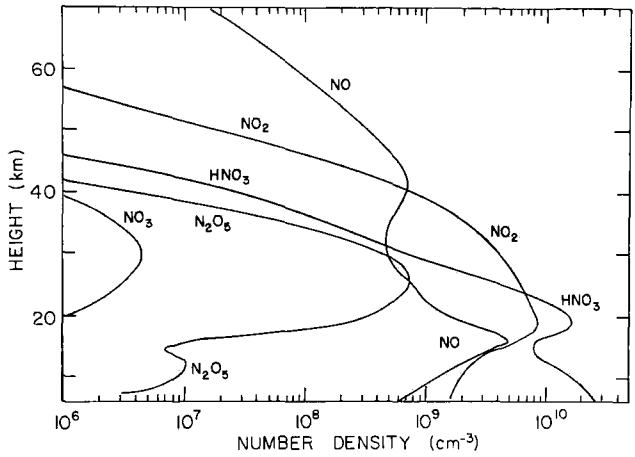



**Fig. 3.4** Model calculations of the distribution of nitrogen oxides ($NO_x$), $HNO_3$, and $N_2O_5$ at 30° latitude (24-hr averages). [After McElroy *et al.* (1974).]

Because most of the radiation photolyzing $NO_2$ by (3.2.29) is unattenuated by $O_2$ or $O_3$, the daytime dissociation rate is high, $J_{29} \approx 4 \times 10^{-3}$ sec$^{-1}$ or a lifetime of $\tau(NO_2) = 1/J_{29} \sim 2 \times 10^2$ sec. Hence $NO_2$ is in a daytime photoequilibrium of

$$\frac{[NO_2]}{[NO]} = \frac{\beta_{27}[O_3]}{\beta_{28}[O] + J_{29}} \approx \frac{\beta_{27}[O_3]}{J_{29}} \tag{3.2.37}$$

Figure 3.4 shows equilibrium abundances calculated with 24-hour averages of the dissociation rates $J$ and with vertical eddy diffusion.

The net rates of production of O and $O_3$, including the Chapman reactions and the hydrogenous and nitrogenous chemistry, are

$$\frac{d[O]}{dt} = 2J_2[O_2] + J_3[O_3] - k_{13}[O][O_3] - k_{12}[O][O_2][M] - \alpha_{14}[HO_2][O] - \alpha_{13}[OH][O] + J_{29}[NO_2] - \beta_{28}[NO_2][O] \tag{3.2.38}$$

and

$$\frac{d[O_3]}{dt} = k_{12}[O][O_2][M] - k_{13}[O][O_3] - J_3[O_3] - \alpha_{12}[HO_2][O_3] - \alpha_{11}[OH][O_3] - \alpha_{10}[H][O_3] - \beta_{27}[NO][O_3] \tag{3.2.39}$$

which yield

$$\frac{d[O + O_3]}{dt} = 2J_2[O_2] - [O_3](\alpha_{12}[HO_2] + \alpha_{11}[OH] + \alpha_{10}[H]) - [O](2k_{13}[O_3] + \alpha_{14}[HO_2] + \alpha_{13}[OH] + 2\beta_{28}[NO_2]) \tag{3.2.40}$$

Discarding terms that will be numerically small (see Fig. 3.5), we have in equilibrium from (3.2.39)

$$[O_3] \approx \frac{k_{12}[O_2][M][O]}{J_3} \tag{3.2.41}$$

where, from (3.2.40),

$$[O] \approx \frac{J_2[O_2]}{[O_3]k_{13} + \beta_{28}[NO_2] + \frac{1}{2}(\alpha_{14}[HO_2] + \alpha_{13}[OH])} \tag{3.2.42}$$

Combining the last two expressions in the form of (3.2.24) gives the correction term to the Chapman solution as

$$A = \frac{\alpha_{13}[OH] + \alpha_{14}[HO_2] + 2\beta_{28}[NO_2]}{2k_{13}[O_3]} \tag{3.2.43}$$

which indicates the relative importance of $O-O_3$ self-destruction and catalytic loss through $HO_x$ and $NO_x$.

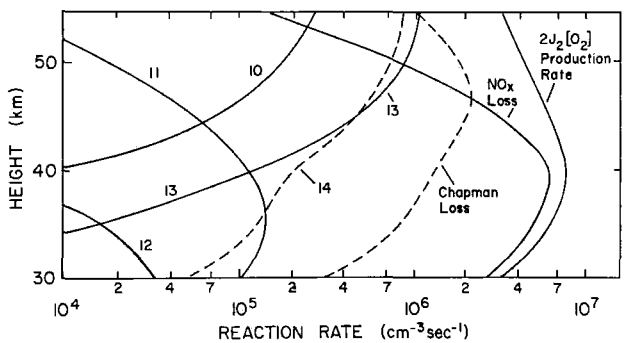



**Fig. 3.5** Rates of production and loss of odd oxygen. The (24-hr average) "production rate" by photodissociation is $2J_2[O_2]$ and the "Chapman loss" is $2k_{13}[O][O_3]$; see (3.2.1). The curve "$NO_x$ loss" gives $2\beta_{28}[NO_2][O]$; see the discussion following (3.2.30). The numbered curves give the contributions to the loss from $HO_x$ reactions (3.2.10–3.2.14). For example, curve "10" is $\alpha_{10}[H][O_3]$. [Adapted from McElroy *et al.* (1974).]

### 3.2.4 Natural and Artificial Sources of Nitrogen Oxides

**Biological Production of $N_2O$** The production and loss of the total $NO_x$ in the stratosphere depends critically on vertical transport. The principal production mechanism is from $N_2O$ (nitrous oxide), which is formed from biological activity on the surface, reacting with metastable oxygen:

$$O(^1D) + N_2O \rightarrow 2NO + 3.54\ \text{eV} \qquad (3.2.44)$$
$$\beta_{44} = 1.1 \times 10^{-10}\ \text{cm}^3/\text{sec}$$

The reaction

$$O(^3P) + N_2O \rightarrow 2NO \qquad (3.2.45)$$

is exothermic but has a rate coefficient $\beta_{45} \sim 10^{-31}$ cm$^3$/sec, which corresponds to an activation energy of about 1.1 eV. The $O(^1D)$ produced by ozone photolysis (3.2.5) evidently supplies this activation energy and the rate $\beta_{44}$ is very rapid.

In addition, $N_2O$ is destroyed in the stratosphere by

$$O(^1D) + N_2O \rightarrow N_2 + O_2 + 5.4\ \text{eV} \qquad (3.2.46)$$
$$\beta_{46} = 0.7 \times 10^{-10}\ \text{cm}^3/\text{sec}$$

for which the same remarks apply. However, the principal destruction mechanism is photodissociation,

$$N_2O + h\nu(\lambda < 2300\ \text{Å}) \rightarrow N_2 + O(^1D) \qquad (3.2.47)$$

Thus the rate of production of $NO_x$ depends on the rate of production of $N_2O$ at the surface and its subsequent transport into the stratosphere.

The principal loss mechanism for $NO_x$ from the stratosphere is downward transport into the troposphere, where $HNO_3$, being soluble in water, will rain out. Thus continuity (2.1.1) requires that

$$\frac{\partial[NO_x]}{\partial t} = -\frac{\partial}{\partial z}([NO_x]w) + 2\beta_{44}[N_2O][O(^1D)] \qquad (3.2.48)$$

Continuity equations involving chemical sources and sinks as well as mass flow are discussed further in Section 3.3.

**Aircraft Exhaust** The importance of $NO_x$ to the ozone balance has created concern over the role of high-altitude aircraft, and especially the prospective effects of large fleets of supersonic aircraft, because $NO_x$ is a principal constituent of engine exhausts. Deleterious effects might occur within a few decades of the onset of heavy contamination directly within the

stratosphere. The residence time depends on the height of injection. For $NO_x$ injected into the stratosphere, it is several years, and for injection above 20 km it is tens of years, with $NO_x$ increasing over long periods. Engine exhausts in the troposphere are not particularly harmful to stratospheric ozone because most of the $NO_x$ rains out eventually and is not readily transported into the stratosphere. Indeed, engine exhausts in the troposphere lead to the production of ozone (cf. Section 7.4.2).

**Industrial Fertilizers** The major source of $NO_x$ being plant activity in the manufacture of $N_2O$ has raised the unsettling question of whether artificial fertilizers, being applied to soils in enormous quantities the world over, will eventually contribute to deterioration of ozone. The first process of converting nitrogen into a form suitable for plant use is *fixation*, basically the conversion of even nitrogen $N_2$ into odd nitrogen $NH_4$ (ammonium). Fixation occurs naturally through certain bacteria operating on the roots of legumes and through combustion. Fixed nitrogen is used by plants to synthesize proteins and nucleic acids. Micro-organisms may then convert the organic $NH_4$ into inorganic $NO_3^-$ (nitrate), a process called *nitrification*. The final step in this biologic cycle is *denitrification*, the reduction, again with microbial assistance, of $NO_3^-$ to $N_2$ and $N_2O$. The net result is that atmospheric $N_2$ going through the cycle does not all return as $N_2$, but some of it becomes $N_2O$.

It appears that industrial fertilizers now account for an amount of fixed nitrogen fed to the soil that is comparable to the natural production. However, there are major uncertainties in attempting to extrapolate this result into an eventual increase in $N_2O$ flowing into the stratosphere. The role of fixed nitrogen in the sea is poorly understood, even as to whether it represents a source or sink of atmospheric $N_2O$. The branching ratio in denitrification to yield $N_2$ and $N_2O$ is clearly a critical number, but it must depend on the chemical processes involved, which are poorly understood. The subject of the biogeochemistry of $N_2O$ is likely to attract much attention in the future.

**Nuclear Explosions** A possible artificial source of $NO_x$ in the stratosphere is the explosion of nuclear weapons in or below the stratosphere. Defensive engagements with nuclear devices launched to intercept offensive missiles in the high atmosphere form commonplace scenarios for military planning by nuclear equipped nations. The explosive power of the interceptor must be in the megaton (MT) range to assure an adequate production of radiation to "kill" the intruder warhead. It seems clear that any such massive release of energy in the stratosphere will seriously wound the ozone layer and expose the entire world to intense solar ultraviolet radiation. (One MT of TNT explosive is equivalent to $4.2 \times 10^{22}$ erg.)

In 1962 the United States conducted megaton atmospheric tests in the tropics and in 1961–62 the Soviet Union exploded several large bombs in the Arctic, including one thought to exceed 50 MT of yield. In calculating what effects would be expected, there are two critical questions: How much $NO_x$ is produced? and How high would the explosion debris rise?

The first question seems to be well answered. At high temperatures within the fireball, the $NO_x$ abundance can be rather accurately estimated by thermodynamic equilibrium distributions of molecular energies. For $T \lesssim 2000°K$, the reactons that destroy NO become very slow, and the [NO] stabilizes at about 1 percent.

The second question is more ambiguous. Most of the United States' experience with large tests has been in the tropics, where the tropopause is higher and colder than in the fall in the Arctic (where the largest Soviet tests occurred). Possibly the fireball did not rise as high as scaling from low-altitude experience would suggest. If the injected $NO_x$ did not reach photochemical altitudes, the catalytic cycle (3.2.27–3.2.28) could not operate. At lower Arctic altitudes the [O] is already depleted and $[O_3]$ is essentially inert. Hence the destructive catalytic effects would await the transport of the weapon-produced $NO_x$ to lower latitudes and higher altitudes.

Analysis of ozone observations around the period of the tests by several groups has led to inconclusive results of whether the bombs' effects were evident or not. Possibly as much as 5 percent effect could have been expected. The data, always noisy and hard to average satisfactorily, may have been confused also by massive eruptions of the Mt. Agung volcano in early 1963 and an apparent breakdown of the so-called *quasi-biennial oscillation* of the atmosphere and the associated fluctuation of total ozone.

**Galactic Cosmic Rays** Still another type of variation is that produced by variable cosmic-ray ionization. It is potentially important in that (a) it may reveal observationally how the stratospheric ozone responds to a known change in $NO_x$ production, (b) it may be the physical explanation for the long suspected 11-year variation in total ozone, and (c) it has led to some speculative mechanisms for long-term changes in terrestrial climate controlled by the stratosphere.

The basic source of variations in the low-energy component of Galactic cosmic rays (GCR) striking the Earth is the heliomagnetic field imbedded in the solar wind. During times of high solar activity, the solar wind and its magnetic field are strong. Then low-energy GCR's are shielded from the Earth, especially over the polar cap, where the terrestrial magnetic field is most receptive to cosmic rays. The small amount of ionization produced in the stratosphere is thus considerably less than at sunspot minimum (see Fig. 3.6).

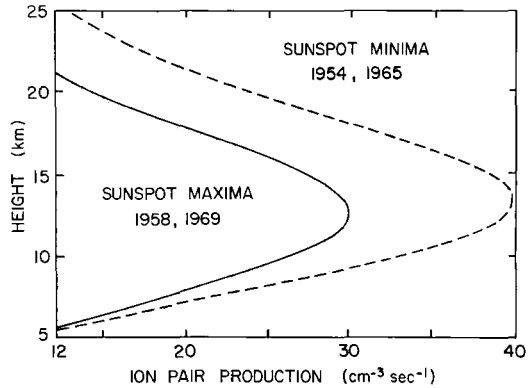



**Fig. 3.6** Height distribution of the ionization produced in the atmosphere by cosmic rays over Thule, Greenland (near the north geomagnetic pole) during times of high and low solar activity. The reduction in Galactic cosmic rays during high solar activity is attributed to screening of the Earth by magnetic fields associated with the solar wind. [Adapted from G. BRASSEUR and M. NICOLET (1973), *Planet Space Sci.* 21, 939; based on data from H. V. NEHER (1971), *J. Geophys. Res.* 76, 1637.]

Cosmic rays produce one ion-pair for each 35 eV lost in transit through the atmosphere. By various collision processes, this ionization results in the production of free N atoms, about 1.4 atom/ion-pair. These atoms may either destroy $NO_x$ by

$$N + NO \rightarrow N_2 + O + 3.25\ \text{eV} \qquad (3.2.49)$$
$$\beta_{49} = (8.2 \pm 1.4) \times 10^{-11} e^{-(460 \pm 60)/T}\ \text{cm}^3/\text{sec}$$

or create it by

$$N + O_2 \rightarrow NO + O + 1.39\ \text{eV} \qquad (3.2.50)$$
$$\beta_{50} = 5.5 \times 10^{-12} e^{-3200/T}\ \text{cm}^3/\text{sec}$$

The same reaction with excited N atoms,

$$N(^2D) + O_2 \rightarrow NO + O + 3.76\ \text{eV} \qquad (3.2.51)$$
$$\beta_{51} = 7 \times 10^{-12}\ \text{cm}^3/\text{sec}$$

goes much faster, with the 2.37 eV of the $^2D$ term countering the high activation energy. The matter is of some importance, because perhaps half the N atoms produced will be in a metastable state and will not be readily deactivated before a chemical encounter.

The $NO_x$ is thus modulated with an 11-year period, mainly over the polar caps. With this fluctuating component being transported out of the polar cap by horizontal motions on a time scale $\tau$ of a few months or a year, the

$[NO_x]$ varies as

$$\frac{\partial[NO_x]}{\partial t} = \mathscr{J}_0 + \mathscr{J}_1 e^{i\omega t} - \frac{[NO_x]}{\tau} \tag{3.2.52}$$

The $\mathscr{J}$'s represent production rates by cosmic rays and the frequency is $\omega = 2\pi/11\ \text{yr}^{-1}$. We assume a solution varying as $e^{i(\omega t - \delta)}$, where $\delta$ is the lag of maximum $NO_x$ concentration behind maximum production (or sunspot minimum). The solution for the oscillating component is

$$[NO_x] = \frac{\mathscr{J}_1 \tau}{(1 + \omega^2\tau^2)^{1/2}} \cos(\omega\tau - \delta) \tag{3.2.53}$$

where

$$\delta = \tan^{-1} \omega\tau \tag{3.2.54}$$

At lower latitudes, outside the polar cap, there is an additional phase lag due to meridional transport and eventual loss to the troposphere. Finally, there is a lag between sunspot maximum and solar wind maximum. Hence the $[NO_x]$ minimum (i.e., $[O_3]$ maximum) occurs at a time following sunspot maximum given by

$$\Delta t = \delta/\omega + \delta_\theta/\omega + (\Delta t)_{sw} \tag{3.2.55}$$

where $\delta/\omega \sim 1$ yr is the lag between production and maximum NO concentration, $\delta_\theta/\omega$ is the additional latitude-dependent lag for meridional transport, and $(\Delta t)_{sw} \sim 1$ yr is the solar-wind lag.

Observations of total ozone variations indicate that the phase shifts are about as predicted and that the effect is stronger at higher latitudes. So many factors evidently enter into ozone variability that it is difficult to extract convincing periodicities from examining even smoothed data, and it is sometimes helpful to seek periodicities and phase effects suggested by physical theory.

**Solar Cosmic Rays** A different kind of cosmic-ray effect likely arises from bursts of particles from solar flares. These solar cosmic rays occur mainly at peaks of solar activity and therefore will be out of phase with the maximum GCR's, which peak during sunspot minimum. Further, they have lower energies than even low-energy GCR's and are deposited higher in the stratosphere. In the mesosphere they are associated with polar-cap radio absorption (PCA events). The ionization occurs in rather short bursts of a few hours, but solar cosmic rays in that short period produce ionization equivalent to the excess ionization from several months of GCR modulation. Thus there are two distinct cosmic-ray effects, with quite different characteristics, that may produce natural and observable variations in ozone.

**Impact of $NO_x$ on Faunal Extinction or Climate?** The awareness that particle bombardment of the stratosphere could reduce its ozone abundance has led to speculative proposals for causes of rather sudden (geologically speaking) faunal extinctions or climate modification. One line of thinking sees the atmosphere as a target of radiation from a nearby supernova, with catastrophic reductions in ozone, thereby exposing the surface to intense solar ultraviolet. Another approach places the primary mechanism for faunal extinction in reversals of the geomagnetic field, which would allow an increase in the bombardment of the stratosphere by solar cosmic rays.

Still another proposal concentrates on the climatic effects that would accompany changes in the solar and terrestrial magnetic fields. For example, drastic reductions in the geomagnetic field strength could admit substantial increases in the GCR radiation and reduce the protective ozone cover. The diminished ozone would lead to a cooler stratosphere, which would normally be accompanied by cooler surface temperatures because of reduced greenhouse heating in the 9.6 $\mu$m band of $O_3$. The faunal extinctions that are correlated with geomagnetic field reversals may hence be due to the climate changes, rather than the direct action of ultraviolet light. A further speculative suggestion is that a stratosphere cooler than the present one might lead to greater condensation of ice haze, which would increase the Earth's albedo, or reduced stratospheric $H_2O$ vapor, which would permit more tropospheric cooling. Such climate mechanisms always involve complex feedbacks, both positive and negative, a fact that makes the subject exceedingly difficult to handle quantitatively.

### 3.2.5 Chlorine and the Halomethanes

Once released in the stratosphere a free chlorine atom Cl reacts immediately with ozone, starting the $ClO_x$ cycle, by

$$\begin{aligned} &Cl + O_3 \rightarrow ClO + O_2 + 1.68\ \text{eV} \\ &\gamma_{56} = (3.6 \pm 0.4) \times 10^{-11} e^{-(318 \pm 50)/T}\ \text{cm}^3/\text{sec} \end{aligned} \tag{3.2.56}$$

where $\gamma$ is the rate coefficient. Then ClO (chlorine monoxide) may either complete the catalytic cycle directly,

$$\begin{aligned} &ClO + O \rightarrow Cl + O_2 + 2.39\ \text{eV} \\ &\gamma_{57} = 5.3 \times 10^{-11}\ \text{cm}^3/\text{sec} \end{aligned} \tag{3.2.57}$$

or it may branch to

$$\begin{aligned} &ClO + NO \rightarrow Cl + NO_2 + 0.39\ \text{eV} \\ &\gamma_{58} = 1.7 \times 10^{-11}\ \text{cm}^3/\text{sec} \end{aligned} \tag{3.2.58}$$

The latter branch may complete the cycle either by (3.2.28), with the net result of destroying two odd oxygens, or by (3.2.29)–(3.2.30), which restores the lost $O_3$.

Chlorine is lost from the stratosphere mainly in the form of inactive HCl (hydrogen chloride), which is transported downward into the troposphere and is removed by rainout. The main reactions transforming Cl into HCl are

$$\mathrm{Cl} + \mathrm{CH_4} \rightarrow \mathrm{HCl} + \mathrm{CH_3} + 0.1\ \mathrm{eV} \qquad (3.2.59)$$
$$\gamma_{59} \approx 5 \times 10^{-12} e^{-(1114 \pm 38)/T}\ \mathrm{cm^3/sec}$$

$$\mathrm{Cl} + \mathrm{H_2} \rightarrow \mathrm{HCl} + \mathrm{H} - 0.04\ \mathrm{eV} \qquad (3.2.60)$$
$$\gamma_{60} = 5.6 \times 10^{-11} e^{-2250/T}\ \mathrm{cm^3/sec}$$

$$\mathrm{Cl} + \mathrm{HO_2} \rightarrow \mathrm{HCl} + \mathrm{O_2} + 2.43\ \mathrm{eV} \qquad (3.2.61)$$
$$\gamma_{61} \sim 10^{-11} \quad \text{to} \quad 10^{-10}\ \mathrm{cm^3/sec}$$

$$\mathrm{Cl} + \mathrm{H_2O_2} \rightarrow \mathrm{HCl} + \mathrm{HO_2} + 0.69\ \mathrm{eV} \qquad (3.2.62)$$
$$\gamma_{62} \sim 10\alpha_{19}$$

(i.e., 10 times the rate of $\mathrm{OH} + \mathrm{H_2O_2}$).

Active Cl is regenerated from HCl by

$$\mathrm{HCl} + \mathrm{OH} \rightarrow \mathrm{Cl} + \mathrm{H_2O} + 0.68\ \mathrm{eV} \qquad (3.2.63)$$
$$\gamma_{63} = 3 \times 10^{-12} e^{-400/T}\ \mathrm{cm^3/sec}$$

and

$$\mathrm{HCl} + h\nu(\lambda < 2200\ \text{Å}) \rightarrow \mathrm{H} + \mathrm{Cl} \qquad (3.2.64)$$

Incidentally, each of the catalytic cycles depends crucially on [OH]. Thus OH plays a direct role in the $HO_x$ cycle through reactions $\alpha_{11}$ and, more importantly, $\alpha_{13}$; it is the principal sink for $HO_x$ through reaction $\alpha_{17}$; it converts active $NO_2$ into inactive $HNO_3$ by reaction $\beta_{31b}$; and it releases Cl from HCl by reaction $\gamma_{63}$.

Purely chemical equilibrium of HCl does not exist in the stratosphere because the chemical sink terms are so small that the dynamical flow has to be included everywhere. The equation of continuity (2.1.1), with the chemical sources and sinks added, is

$$\frac{\partial[\mathrm{HCl}]}{\partial t} = -\nabla([\mathrm{HCl}]w_{\mathrm{HCl}}) - [\mathrm{HCl}](J_{64} + \gamma_{63}[\mathrm{OH}]) + [\mathrm{Cl}](\gamma_{59}[\mathrm{CH_4}] + \gamma_{60}[\mathrm{H_2}] + \gamma_{61}[\mathrm{HO_2}] + \gamma_{62}[\mathrm{H_2O_2}]) \qquad (3.2.65)$$

Because $[O_3] \gg [O]$ and the $Cl + O_3$ reaction is fast, chlorine is about evenly divided between the forms ClO and HCl; [Cl] is very small. At

35 km the lifetime of Cl is $\tau(\mathrm{Cl}) = 1/\gamma_{56}[\mathrm{O_3}] \sim 1/20$ sec, while $\tau(\mathrm{ClO}) = (\gamma_{57}[\mathrm{O}] + \gamma_{58}[\mathrm{NO}])^{-1} \sim 40$ sec and $\tau(\mathrm{HCl}) = 1/\gamma_{64}[\mathrm{OH}] \sim 3 \times 10^5$ sec.

Another reservoir for Cl is $\mathrm{ClNO_3}$ (chlorine nitrate). It is formed and removed by

$$\mathrm{ClO} + \mathrm{NO_2} + \mathrm{M} \rightarrow \mathrm{ClNO_3} + \mathrm{M} + 1.1 \text{ eV} \tag{3.2.66}$$

$$\gamma_{66} = 3 \times 10^{-31} \text{ cm}^6/\text{sec}$$

$$\mathrm{ClNO_3} + h\nu \rightarrow \begin{cases} \mathrm{ClO} + \mathrm{NO_2} \\ \mathrm{ClNO_2} + \mathrm{O} \end{cases} \tag{3.2.67}$$

$$\mathrm{ClNO_3} + \mathrm{O} \rightarrow \begin{cases} \mathrm{ClO} + \mathrm{NO_3} \\ \mathrm{ClNO_2} + \mathrm{O_2} \end{cases} \tag{3.2.68}$$

$$\gamma_{68} = 2.1 \times 10^{-13} \text{ cm}^3/\text{sec}$$

The importance of these reactions is not only that they reduce slightly the amount of active $\mathrm{ClO}_x$, but that they also reduce $\mathrm{NO}_x$, especially in the lower stratosphere. Hence we have the interesting phenomenon of two ozone-destroying catalysts tending to cancel one another out. Other examples of *nonlinear* or *coupled catalytic chemistry* may involve such substances as HOCl and $\mathrm{HO_2NO_2}$, as well as $\mathrm{HNO_3}$ (see 3.2.31a,b).

At the present time it is not clear whether Cl is a major sink for $\mathrm{O_3}$, compared with $\mathrm{NO}_x$. The natural abundance of atmospheric Cl (mainly in the forms HCl and $\mathrm{CH_3Cl}$), arising principally from volcanic emission, is probably small and the artificial (anthropogenic) sources have not yet become serious. With an estimated abundance at 35 km of $[\mathrm{ClO}] \approx 2 \times 10^7 \text{ cm}^{-3}$, the loss rate of $\mathrm{O_3}$ from the $\mathrm{ClO}_x$ cycle is only

$$d[\mathrm{O_3}]/dt = -2\gamma_{57}[\mathrm{ClO}][\mathrm{O}] \approx 4 \times 10^5 \text{ sec}^{-1} \tag{3.2.69}$$

which is smaller by a factor 2 than the Chapman loss rate, $2k_{13}[\mathrm{O_3}][\mathrm{O}]$, and an order of magnitude smaller than the $\mathrm{NO}_x$ loss rate, $2\beta_{28}[\mathrm{NO_2}][\mathrm{O}]$ (see Fig. 3.5). There are, however, major uncertainties in these estimates, and the destruction of $\mathrm{O_3}$ due to natural $\mathrm{NO}_x$ now seems less important, according to detailed model calculations, than it did a few years earlier, partly because of the nonlinear catalyst interactions.

The principal sources of Cl in the stratosphere are the *halomethanes* (or *halogenated methanes*), molecules in which halogens take the place of one or more hydrogen in $\mathrm{CH_4}$ (methane). The most important of these molecules in the troposphere are $\mathrm{CFCl_3}$ (trichlorofluoromethane, often called "*F*-11"), $\mathrm{CF_2Cl_2}$ (dichlorofluoromethane—"*F*-12"), $\mathrm{CCl_4}$ (carbon tetrachloride), and $\mathrm{CH_3Cl}$ (methyl chloride). [In this shorthand identifying system, *F* apparently originated from E. I. duPont de Nemours and Company's trademark name *Freon*, which is in common use, although over 20 companies the world over manufacture these substances; the first digit is the number of H atoms plus

one; the second is the number of fluorine atoms. For more complex fluorocarbons a "hundreds" digit is prefixed to specify the number of carbon atoms minus one.]

In recent years *F*-11 and *F*-12 have attracted wide popular and scientific attention as being potentially harmful to stratospheric ozone if their manufacture and release into the atmosphere continues at the rates existing in the mid-1970s. The molecule $CF_2Cl_2$ is the most widely used halomethane, both as a refrigerant and as a propellant in aerosol sprays, whereas $CFCl_3$ is used primarily in sprays. These *chlorofluoromethanes* or *fluorocarbons*, as they are often called, were developed in 1930 in a chemical search for a refrigerant that was both nontoxic and nonflammable, and being chemically inert is precisely the reason they cause a stratospheric problem. The molecules are practically insoluble so that the oceans are not an effective sink. Free molecules of these synthetic gases cannot be photodissociated in the troposphere, but upon diffusing into the stratosphere, they are photolyzed by

$$CFCl_3 + h\nu(\lambda < 2260\ \text{Å}) \rightarrow CFCl_2 + Cl \tag{3.2.70}$$

and

$$CF_2Cl_2 + h\nu(\lambda < 2150\ \text{Å}) \rightarrow CF_2Cl + Cl \tag{3.2.71}$$

(These wavelengths are the observed ones; the bond dissociation energies are considerably smaller.) Also, $CCl_4$, being fully halogenated, is highly inert in the troposphere and in the stratosphere undergoes

$$CCl_4 + h\nu \rightarrow CCl_2 + Cl_2 \tag{3.2.72}$$

The radicals thus formed immediately react with $O_2$ to release another Cl atom. While $CCl_4$ may arise in part naturally, its main origin seems to be its industrial manufacture, which has a very long history. There have been some laboratory experiments that indicate visual sunlight can dissociate the halomethanes if the molecules become first adsorbed onto sand or quartz. This area of surface photochemistry is likely to become an important one.

The partially halogenated $CH_3Cl$ may be produced naturally in the troposphere and it may be a minor source of Cl for the stratosphere. However, the molecule can be attacked by OH in the troposphere and probably has a lifetime against chemical destruction there of under a year. The halomethanes are part of a broader class of substance widely used industrially, the *halocarbons* (*halogenated hydrocarbons*, e.g., $CH_2ClCH_2Cl$). However, only the fully halogenated methanes, $CFCl_3$, $CF_2Cl_2$, and $CCl_4$, are now thought to pose a problem because of their tropospheric indestructibility. The partially halogenated methanes $CHFCl_2$,(*F*-21) and $CHF_2Cl$ (*F*-22) are attacked by OH in the troposphere and may serve as substitutes to *F*-11 and *F*-12 in some uses. On the other hand, OH is showing signs of becoming

overworked as a tropospheric cleansing agent, largely for the demands placed on it by $CH_4$ and CO. A shortage of OH would lead to a new set of problems; see Section 7.4.2.

The fluorine atoms in the halomethanes are in principle just as destructive of $O_3$ as is Cl. Thus, as in reactions $\gamma_{56}$ and $\gamma_{57}$ we have

$$\begin{aligned} F + O_3 &\rightarrow FO + O_2 \\ FO + O &\rightarrow F + O_2 \end{aligned} \tag{3.2.73}$$

In practice this $FO_x$ cycle is unimportant because the analog to $\gamma_{59}$,

$$\begin{aligned} F + CH_4 &\rightarrow HF + CH_3 + 1.41\ \text{eV} \\ \gamma_{74} &= 6 \times 10^{-11}\ \text{cm}^3/\text{sec} \end{aligned} \tag{3.2.74}$$

proceeds very rapidly. And once formed, HF (hydrogen fluoride) is not attacked by the analog to $\gamma_{63}$ to regenerate F because

$$HF + OH \rightarrow H_2O + F - 0.74\ \text{eV} \tag{3.2.75}$$

is comfortably endothermic.

On the other hand, bromine atoms released into the atmosphere could be an efficient catalytic destroyer of ozone. The sources of bromine in the air are marine aerosols, combustion of leaded gasoline, and the use of the halomethane $CH_3Br$ (methyl bromide) as an agricultural fumigant. In the stratosphere bromine chemistry is similar to chlorine chemistry.

## 3.3 Stratospheric Motions

### 3.3.1 Vertical Mixing

With the concept of eddy diffusion (see Section 2.3.2) the degree of mixing can be ascertained from the measured vertical distribution of some minor substance whose chemistry is understood. Then the derived coefficient for eddy diffusion, $K(z)$, may be used to model the distributions of other minor constituents. This approach is based on the supposition that vertical and horizontal transport are independent, which would require that the atmosphere be uniform horizontally (i.e., zonally and meridionally). This is not generally true, of course, but the *one-dimensional dynamic model* can be very useful if its limitations are appreciated. Thus if $K(z)$ is derived from a substance that has a latitudinal or day-night variation, it may not be appropriate for another substance whose horizontal variations would provide different dynamical sources or sinks (which are absorbed in the coefficient $K$).

In the troposphere convective mixing is very rapid and $K \sim 3 \times 10^5\ \text{cm}^2/\text{sec}$. In the stratosphere the stable temperature gradient inhibits mixing, which

then occurs only because of planetary-scale pressure waves. Just above the tropopause $K \sim 2 \times 10^3$ cm$^2$/sec. However, at 40 km these waves have associated vertical wavelengths of about 5 km; if we take the mixing length $l$ (Section 2.2.1) to be half a wavelength and vertical motions of $w \sim 0.5$ cm/sec, we have $K \sim wl \sim 10^5$ cm$^2$/sec. As waves are propagated upward their energy varies as the square of their amplitude, or as $\rho l^2$, and hence for energy to be approximately conserved, we expect $K$ to vary roughly as $\rho^{-1/2}$. (In practice the wave energy is not strictly conserved if some of it is dissipated into eddy motions.)

For a chemical constituent that is not conserved, the equation of continuity (2.1.1) must include chemical production ($P$) and loss ($L$) terms. For a minor constituent of density $N_1$,

$$\frac{\partial N_1}{\partial t} = -\frac{\partial}{\partial z}(N_1 w_1) + P(N_1) - L(N_1)$$

$$= \frac{\partial}{\partial z}\left[K(z)N(z)\frac{\partial}{\partial z}\left(\frac{N_1}{N}\right)\right] + \left(\frac{\partial N_1}{\partial t}\right)_{\text{chem}} \tag{3.3.1}$$

Putting the height dependence of the chemistry into the net chemical term $(\partial N_1/\partial t)_{\text{chem}}$ and averaging over a day or a year, we can integrate this equation for a steady state and solve for $N_1(z)$.

For example, $CH_4$ (methane) is produced at the surface, diffuses upward, and eventually penetrates through the tropopause (see Section 3.3.2). The principal loss in the stratosphere occurs through oxidation by (3.2.9) forming OH, through destruction of OH by

$$\text{OH} + \text{CH}_4 \rightarrow \text{CH}_3 + \text{H}_2\text{O} + 0.67 \text{ eV} \tag{3.3.2}$$

$$\alpha_{77} = 3.5 \times 10^{-12} e^{-1800/T} \text{ cm}^3/\text{sec}$$

and, in the mesosphere, by photodissociation. Thus the chemical term is

$$\left(\frac{\partial[\text{CH}_4]}{\partial t}\right)_{\text{chem}} = -[\text{CH}_4]\{\alpha_9[\text{O}(^1\text{D})] + \alpha_{77}[\text{OH}] + J(\text{CH}_4)\} \tag{3.3.3}$$

where $J(CH_4)$ is the diurnally averaged rate of photolysis. Measurements of $[CH_4]$ versus height then yield values of $K(z)$.

In Eq. (3.3.3) the term in curly brackets is the reciprocal chemical lifetime $\tau_{\text{chem}}^{-1}$. The dynamical time associated with mixing (see Section 2.3.1) is $\tau_{\text{mix}} \sim H^2/K$. Figure 3.7 compares these lifetimes for an eddy mixing coefficient derived from $CH_4$ measurements. It appears that $K$ is nearly constant ($2 \times 10^3$ cm$^2$/sec) just above the tropopause; it then starts to increase upward. There may be a real change in dynamical regime around 20 km because the lower stratosphere is driven by tropospheric dynamics, but the middle stratosphere is driven by its own internal heat sources. In any case

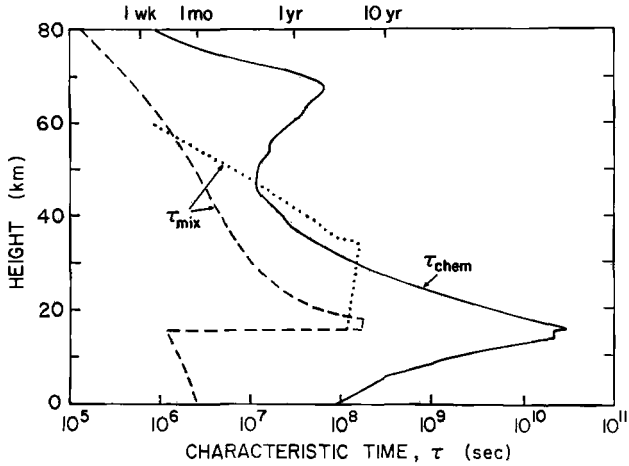



**Fig. 3.7** Comparison of the characteristic time for vertical mixing, $\tau_{mix} = H^2/K$, with the chemical time for destruction of $CH_4$ given by (3.3.3). The two curves for mixing are for two assumed height dependences of $K$. [Adapted from WOFSY and MCELROY (1973).]

$CH_4$ is mixed throughout the stratosphere in times that are short compared with the chemical lifetime and it thus serves (as does $N_2O$) as an effective tracer, being essentially inert between 10 and 30 km.

The one-dimensional ("1-D") model can be extended to take account of diurnal changes in dissociating radiation. For substances that change rapidly at sunrise and sunset, this kind of approach may be necessary, as daily averages could give misleading results.

### 3.3.2 Meridional Circulation and Stratospheric Exchange with the Troposphere

That meridional circulation in the stratosphere is extremely important to the distribution of ozone can be readily appreciated by examination of Fig. 3.8. The region of production of ozone by photolysis of $O_2$ [reaction (1.5.1) followed by (1.5.2)] lies above 20 km at the equator and above 25 km over the polar caps. The ozone concentrations are a maximum in the spring polar regions, which lie well outside the region of production. Large concentrations also occur in the temperate zones, but well below the photochemically active altitudes. During the southern hemisphere spring, Fig. 3.8 is more or less reversed, so there are clearly major meridional shifts of stratospheric air mass over periods of a few months.

Figure 3.1 shows the photochemical time scales of Eq. (3.2.4), which gives the characteristic time to approach photochemical equilibrium. Figure 3.9

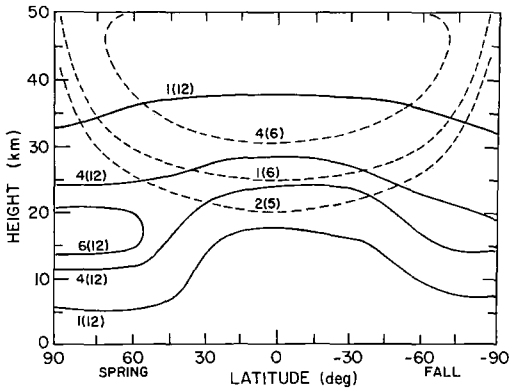



**Fig. 3.8** Regions of ozone production and residence. Zonal average ozone concentrations (solid lines) are shown in molecules/cm$^3$. Zonal average production rates (dashed lines) are given in molecules/cm$^3$ sec. An average production rate of $3 \times 10^4$ molecules/cm$^3$ sec is equivalent to $1 \times 10^{12}$ molecule/cm$^3$ yr. A number such as $4 \times 10^{12}$ is written 4(12). [Adapted from JOHNSTON (1975).]

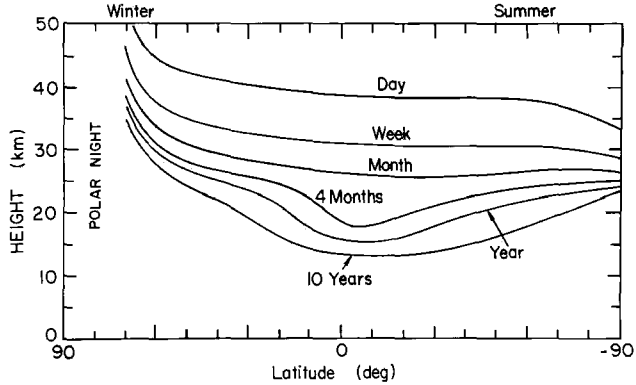



**Fig. 3.9** Photochemical times $\tau_{phot}(O_3)$ for local replacement of ozone, zonally averaged for January 15. [After JOHNSTON (1975).]

gives a slightly different calculation, the instantaneous photochemical lifetime, defined by

$$\tau_{phot}^{-1}(O_3) = \frac{1}{[O_3]} \frac{d[O_3]}{dt} = \frac{2J_2[O_2]}{[O_3]} \tag{3.3.4}$$

where the concentrations and rates are daily averages for a particular day. Where the photochemical lifetime exceeds a month (below about 30 km at mid-latitudes), the ozone is under dynamical control. At higher altitudes than 30 km (except in the polar night), ozone is primarily under photochemical control. We could obtain average vertical and meridional ("2-D") flow from observing how ozone or some other substance flows through the stratosphere, making the photochemical adjustments as in 1-D analyses.

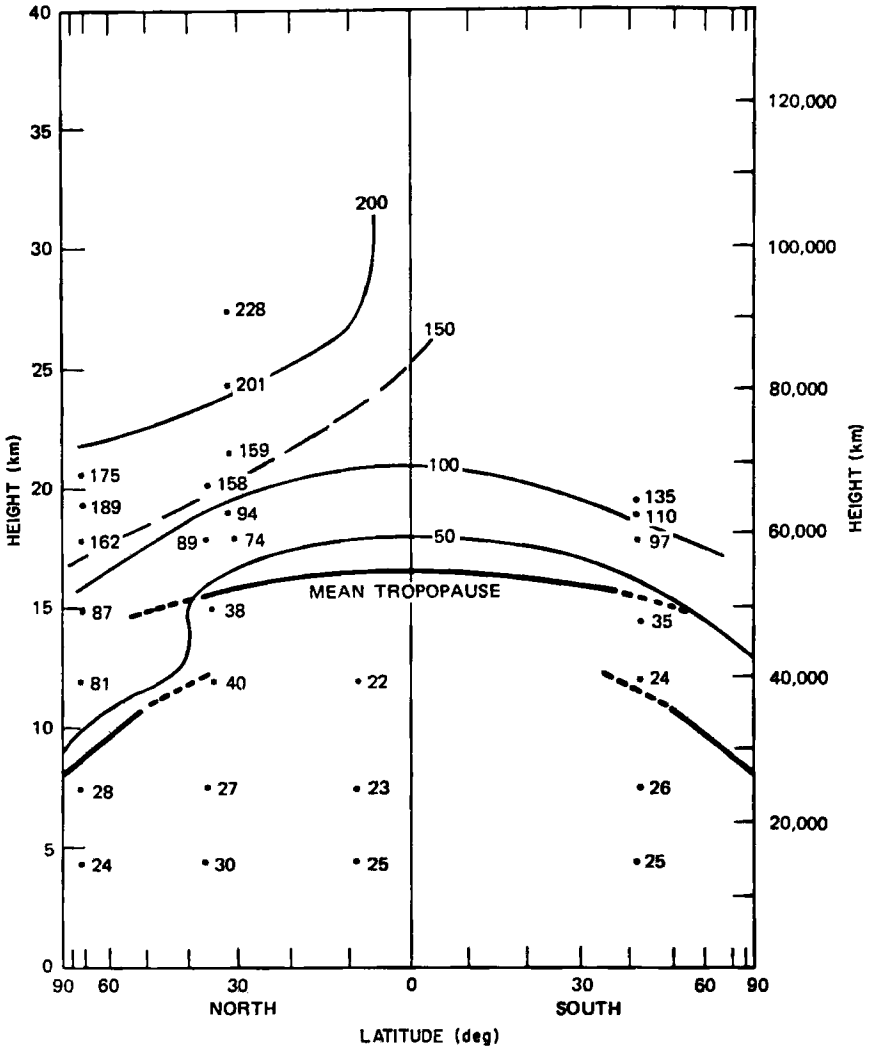



**Fig. 3.10** Atmospheric $^{14}C$ sampling and extrapolation of measurements for March–May, 1961. [After K. TELEGADAS (1973), U.S.A.W.C., Health and Safety Laboratory Report No. 243.]

Most of the direct quantitative evidence for stratospheric motions comes from monitoring radioactive debris injected into the stratosphere from nuclear explosions. This motion is a combination of a mean circulation and eddy diffusion and of settling of particulate matter. Meridional winds are highly variable but the mean circulation is very small, and a large number of accurate measurements of tracers would be necessary to reconstruct it. Adequate measurements are not available for most of the stratosphere (see Figs. 3.10–3.13) but temperature measurements are (e.g., Fig. 2.5). Hence

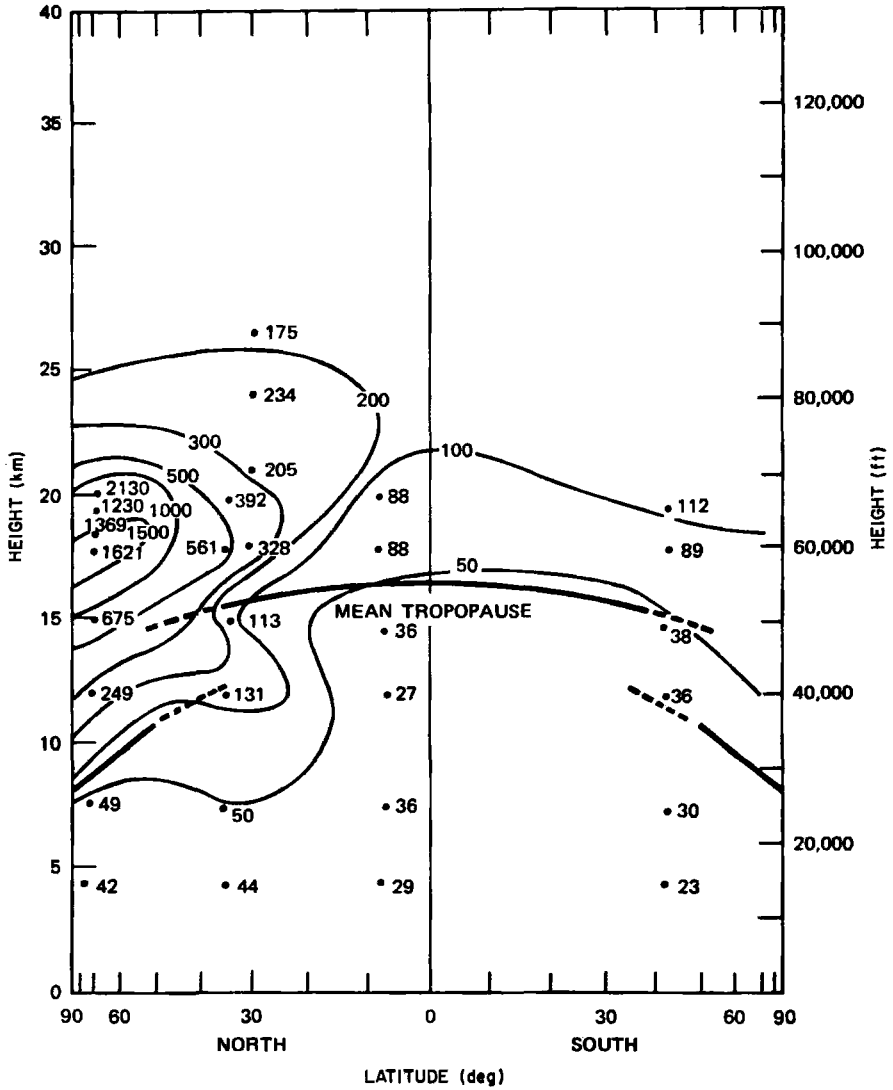



**Fig. 3.11** Atmospheric $^{14}C$ sampling and extrapolation of measurements for March–May, 1962. [After K. TELEGADAS (1973), U.S.A.W.C., Health and Safety Laboratory Report No. 243.]

one can utilize the thermodynamic equation (2.1.11) with radiative heating and cooling and eddy heat transport on the right. When averaged zonally and over one season, say, this equation is

$$\frac{\partial T}{\partial t} + v\frac{\partial T}{r\cos\phi\, d\phi} + w\left(\frac{\partial T}{\partial z} + \frac{g}{C_p}\right) = \frac{Q}{\rho C_p} \tag{3.3.5}$$

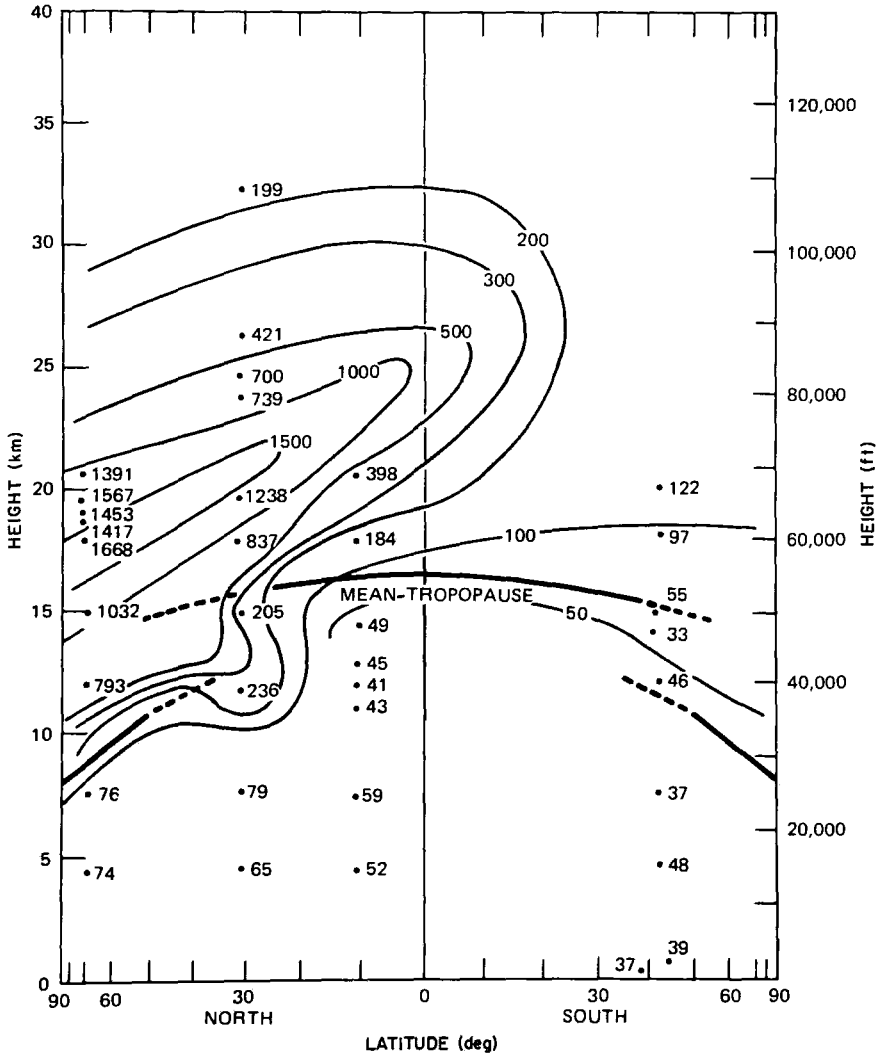



Fig. 3.12 Atmospheric $^{14}C$ sampling and extrapolation of measurements for March–May, 1963. [After K. TELEGADAS (1973), U.S.A.W.C., Health and Safety Laboratory Report No. 243.]

where $\phi$ is latitude and $Q$ is the net rate of heat supply. Combining this equation with continuity (2.1.1),

$$\frac{\partial(\rho v \cos\phi)}{r \cos\phi \, \partial\phi} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{3.3.6}$$

allows one to solve for the velocities of heat advection and so construct the stream lines of mass flow.

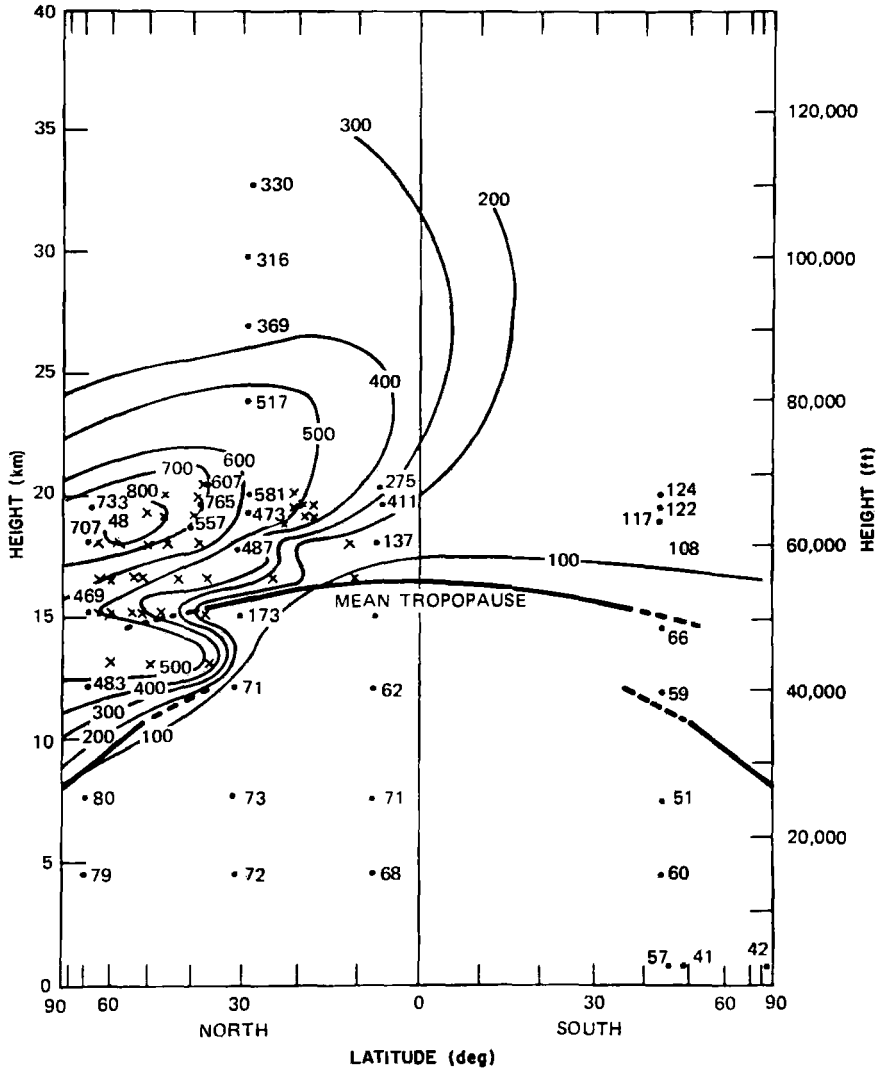



**Fig. 3.13** Atmospheric $^{14}C$ sampling and extrapolation of measurements for March–May, 1964. [After K. TELEGADAS (1973), U.S.A.W.C., Health and Safety Laboratory Report No. 243.]

Results of such a calculation are shown in Figs. 3.14–3.17 for the four seasons. It seems clear from these figures that the meridional circulation in the lower stratosphere, to above the 25 km level, is governed by tropospheric circulation. The spring and fall maps clearly show the Hadley, Ferrel, and polar convection cells, with the first two (in each hemisphere) protruding well

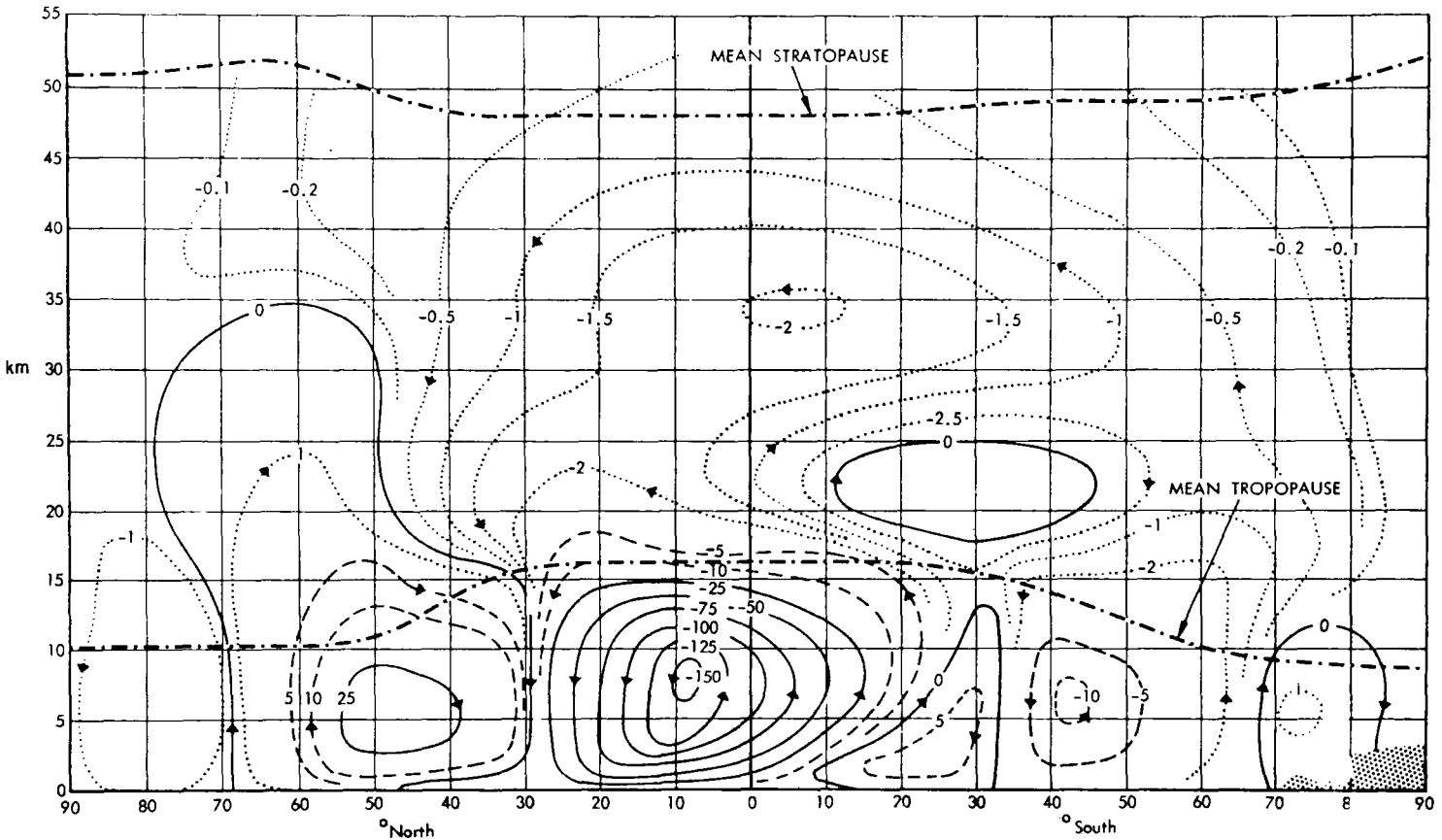



**Fig. 3.14** Mean meridional circulation for December–February. Mass flow is given in units of $10^{12}$ gm/sec. [After J. F. LOUIS *in* E. R. REITER *et al.* (1975), Chap. 6, CIAP Monograph No. 1, U.S. Dept. Transport., Washington, D.C.]

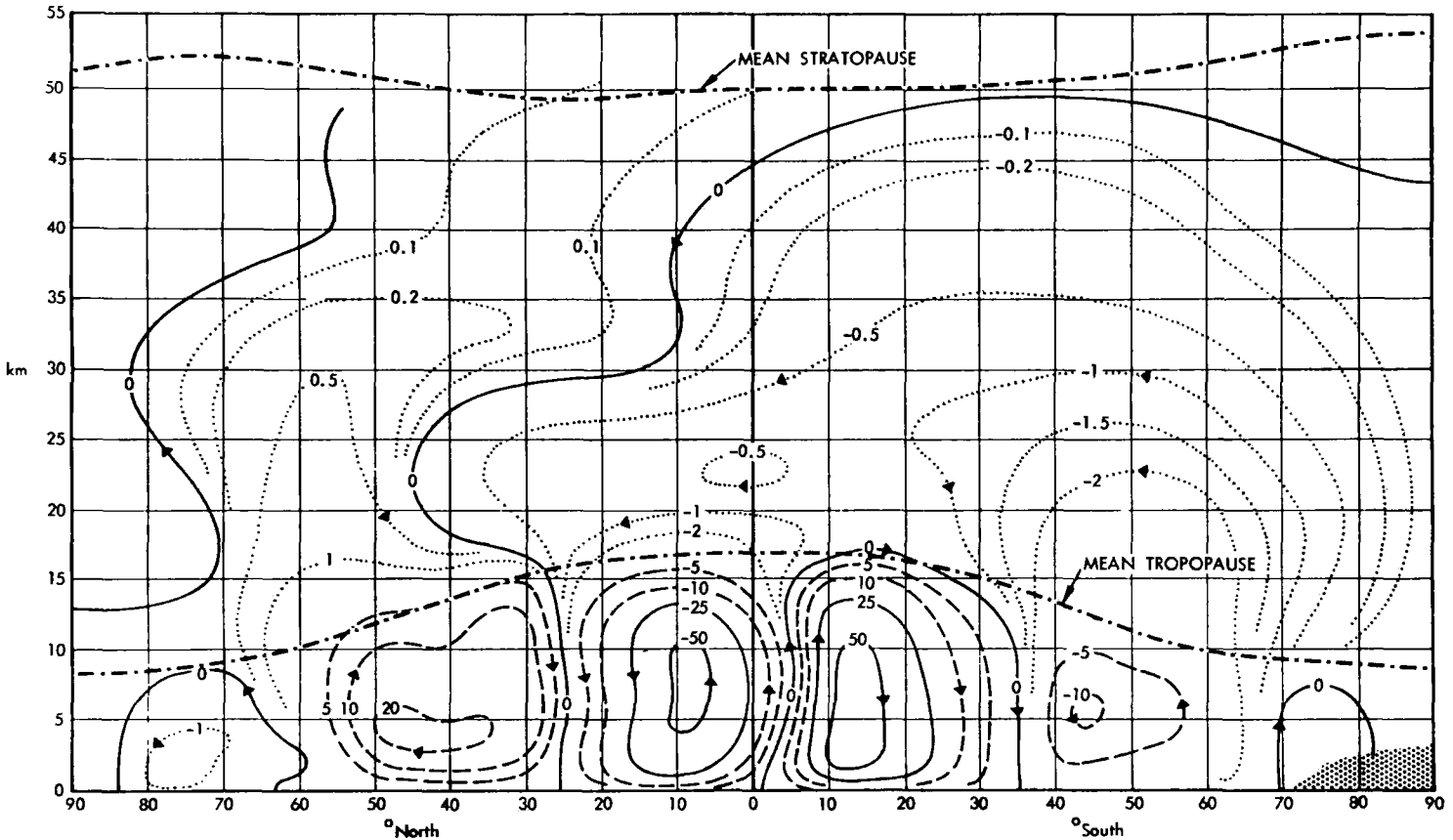



**Fig. 3.15** Mean meridional circulation for March–May. Mass flow is given in units of $10^{12}$ gm/sec. [After J. F. Louis *in* E. R. Reiter *et al.* (1975), Chap. 6, CIAP Monograph No. 1, U.S. Dept. Transport., Washington, D.C.]

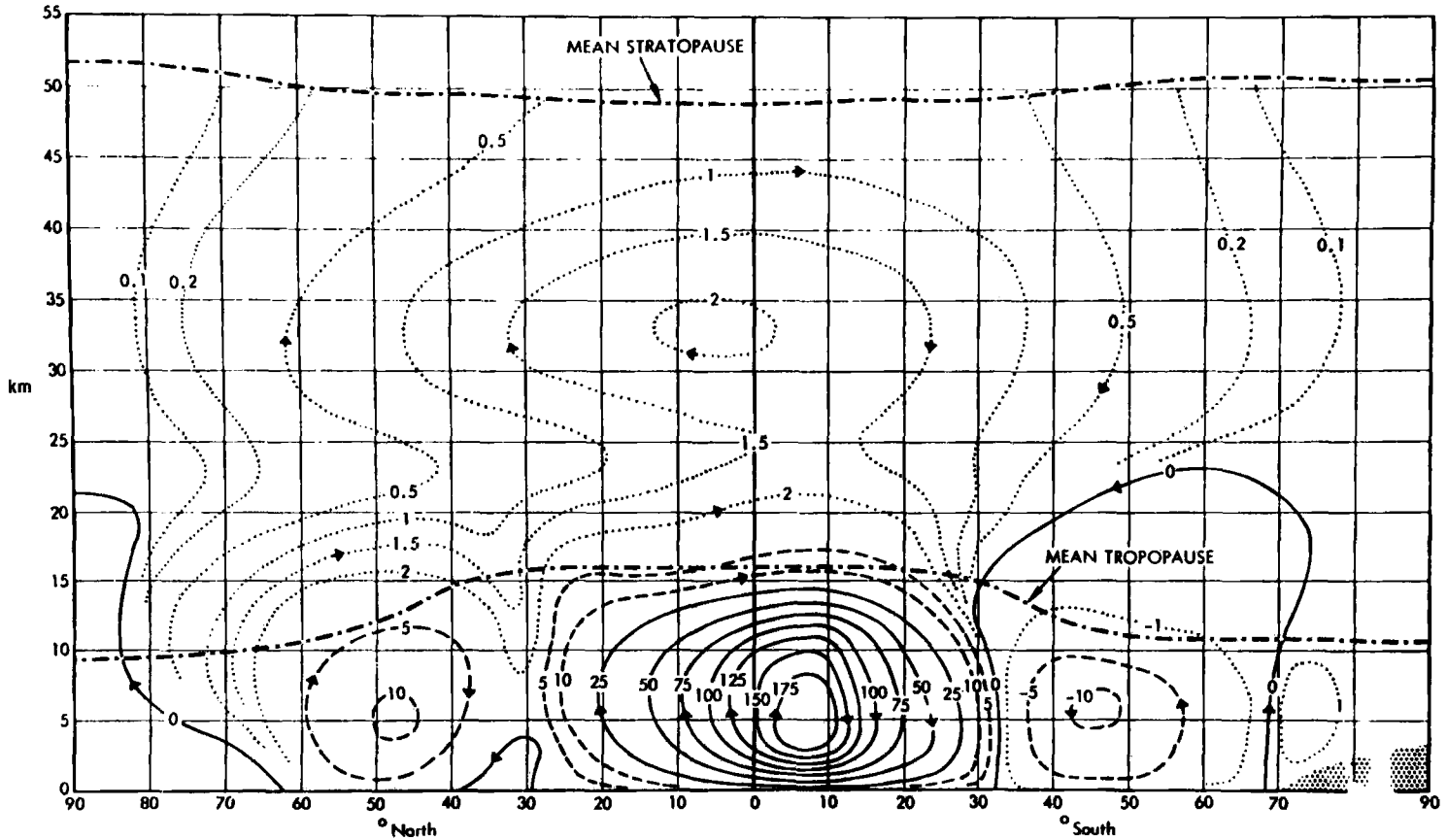



Fig. 3.16 Mean meridional circulation for June–August. Mass flow is given in units of $10^{12}$ gm/sec. [After J. F. Louis *in* E. R. Reiter *et al.* (1975), Chap. 6, CIAP Monograph No. 1, U.S. Dept. Transport., Washington, D.C.]

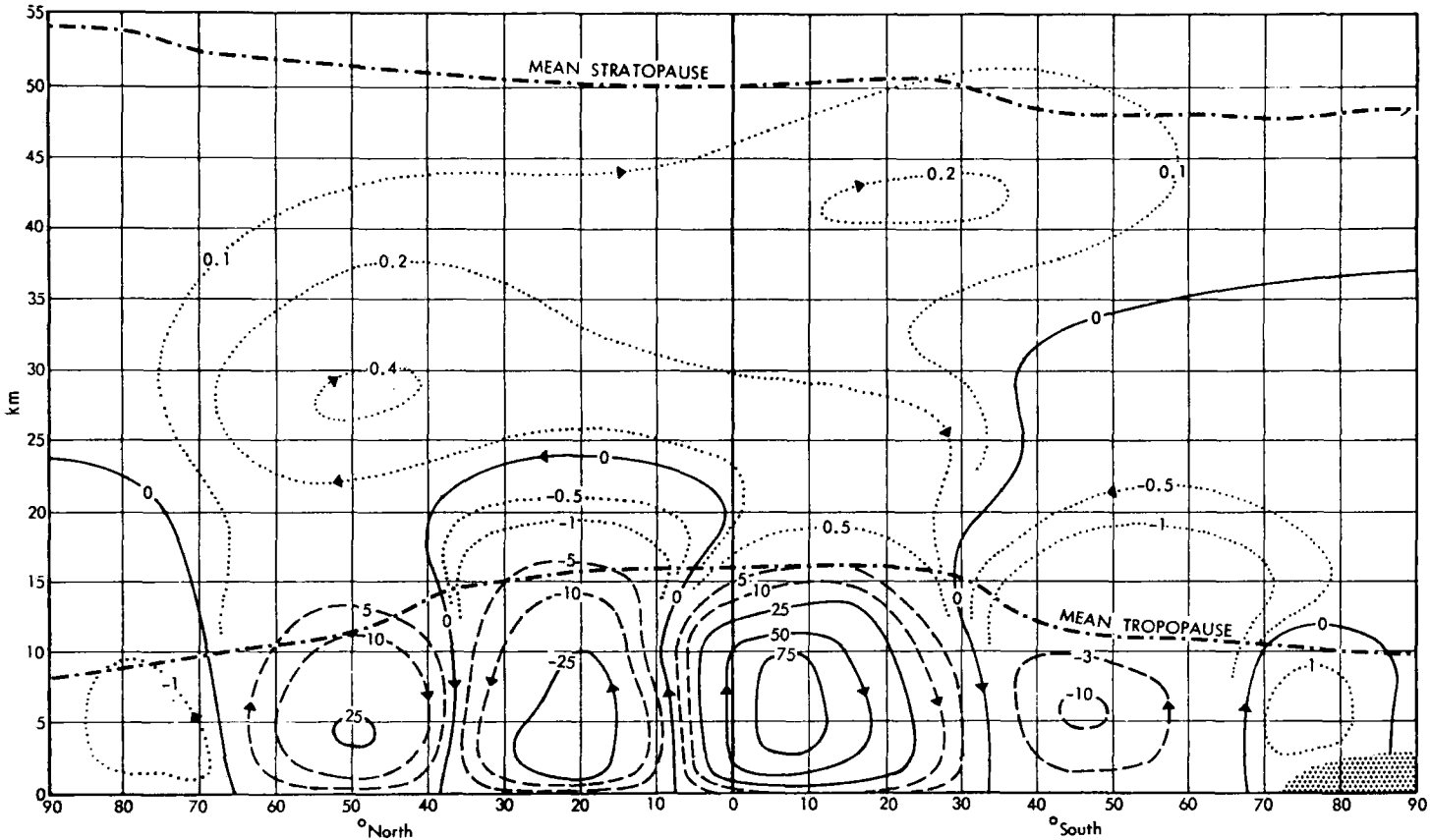



**Fig. 3.17** Mean meridional circulation for September–November. Mass flow is given in units of $10^{12}$ gm/sec. [After J. F. Louis *in* E. R. Reiter *et al.* (1975), Chap. 6, CIAP Monograph No. 1, U.S. Dept. Transport., Washington, D.C.]

into the stratosphere. The summer Hadley cell almost disappears and its mass flux is negligible, but the winter Hadley cell transfers tropospheric air upward and poleward, at the rate of $8 \times 10^{12}$ gm/sec between the tropopause and the 25 km level. This three month season thus interchanges $6 \times 10^{19}$ gm of air between troposphere and stratosphere, which is 15% of the stratospheric air mass in one hemisphere. Altogether the Hadley cells transfer 35 to 40 percent of the mass equivalent of the stratosphere through the tropopause per year.

The upward flow from the tropical branch of the Hadley cell is not a slow continuous flow, like the mean flow of Figs. 3.14–3.17. The tropical updrafts occur mainly as cumulus towers that penetrate the tropopause. The downward branch in the subtropics is more uniform, but there is also some downward flow around the tropical cumulus towers.

Were the mean circulation the only exchange mechanism between tropospheric and stratospheric air masses, the residence time, defined by

$$\tau_{\text{exch}}^{-1} = -\frac{1}{\mathcal{M}} \frac{d\mathcal{M}}{dt} \tag{3.3.7}$$

where $\mathcal{M}$ is the total stratospheric air mass ($10^{21}$ gm) would be three years. The actual mean residence time of pollutants between the tropopause and 25 km is 1 to 2 years, the difference being due largely to eddy diffusion on the size scale of cyclones and anticyclones. (Small-scale eddy diffusion at the tropopause is unimportant for purposes of mass transfer.)

Another mechanism for transporting mass across the tropopause is the seasonal shifting of the tropopause height (Figs. 3.14–3.17). Between 30° and 55° latitude the winter stratosphere gains mass and the summer stratosphere loses it.

The upper stratosphere is driven by radiative heating and cooling. The warm summer stratosphere rises and moves over the equator, descending into the winter hemisphere in a single-cell pattern.

The mass transferred from the summer to winter stratospheric hemisphere is $6 \times 10^{12}$ gm/sec, with one third this amount crossing the equator above 35 km. Thus the total mass transferred in six months is about $5–8 \times 10^{19}$ gm, or 10 to 15 percent of the hemispheric mass.

Water vapor distribution in the stratosphere poses special problems. The volume mixing ratios (i.e., the number density of $H_2O$ molecules to air molecules) is generally in the range $1–8 \times 10^{-6}$ up to about 40 km and it possibly increases gradually from the tropopause to the 40 km level. Beyond that, measurements by different workers are often seriously divergent from one another and consequently the questions of whether real, major fluctuations in humidity occur and whether latitudinal or secular trends exist are largely unresolved.

## BIBLIOGRAPHICAL NOTES

**Section 3.1 Principles of Photochemistry**

This subject is treated in depth in

CALVERT, J. G. and PITTS, J. N., JR. (1966), "Photochemistry," Wiley, New York.

The subject is developed with emphasis on planetary chemistry in

MCEWAN, M. J. and PHILLIPS, L. F. (1975), "Chemistry of the Atmosphere," Wiley, New York.

**Section 3.2 Catalytic Destruction of Ozone**

The definitive review of this topic is

NICOLET, M. (1975), Stratospheric ozone: An introduction to its study, *Rev. Geophys. Space Phys.* **13**, 593–636.

This paper provides references to over 350 research papers, including the sources of nearly all the reaction rates quoted here. Other compilations of chemical rate coefficients will be found in

HAMPSON, R. F. and GARVIN, D. (1975) "Chemical Kinetic and Photochemical Data for Modelling Atmospheric Chemistry," NBS Techn. Note 866, Nat'l. Techn. Infor. Serv., Springfield, Virginia 22151.

Panel on Stratospheric Chemistry (Committee on Impacts of Stratospheric Change) (1976) "Halocarbons: Effects on Stratospheric Ozone," Nat. Research Council, Nat. Acad. Sci., Washington, D.C.

ANDERSON, L. G. (1976), Atmospheric chemical kinetics data survey, *Rev. Geophys. Space Phys.* **14**, 151–171.

MCEWAN, M. J. and PHILLIPS, L. F. (1975), op. cit.

A "workshop" organized by NASA has collected together the best laboratory values (as of June 1977) of stratospheric rate constants, absorption cross sections and quantum yields, tropospheric abundances of constituents that can cause stratospheric contamination, and a few photodissociation rates. The latter include the effects of sunlight scattering by the atmosphere or clouds (which increases the effective $J$-values). The committee's report is available (for $9) as

HUDSON, R. D., ed. (1977), "Chlorofluoromethanes and the Stratosphere," NASA Reference Publ. 1010, 266 pp., Nat. Tech. Inform. Service, Springfield, Virginia 22151.

**Section 3.2.1 Pure Oxygen Chemistry**

The early (1930s) work on ozone photochemistry by S. CHAPMAN and by O. R. WULF and L. S. DEMING is documented in Chapter 1.

**Section 3.2.2 Hydrogenous Radicals**

The relationship between hydroxyl chemistry and ozone in the upper stratosphere and mesosphere was first developed by

BATES, D. R. and NICOLET, M. (1950), The photochemistry of water vapor, *J. Geophys. Res.* **55**, 301–327.

**Section 3.2.3 Oxides of Nitrogen**

The enormous impact of the nitrogen oxides on stratospheric ozone was first recognized by

CRUTZEN, P. J. (1970), The influence of nitrogen oxides on the atmospheric ozone content, *Quart. J. Roy. Meteorol. Soc.* **96**, 320–325.

Shortly afterwards, the potential for contamination of the stratosphere by $NO_x$ emission from aircraft was signaled by

JOHNSTON, H. S. (1971), Reduction of stratospheric ozone by nitrogen oxide catalysts from supersonic transport exhausts, *Science* **173**, 517–522.

The latter paper was largely responsible for setting in motion a chain of environmental studies by the U.S. Department of Transportation, the National Academy of Sciences, and other bodies

that has in turn greatly accelerated activity in stratospheric research. The role of $HO_x$ and $NO_x$ chemistry in ozone depletion has been developed most thoroughly for a given season and latitude in

McELROY, M. B.; WOFSY, S. C.; PENNER, J. E.; and McCONNELL, J. C. (1974), Atmospheric ozone: Possible impact of stratospheric aviation, *J. Atmos. Sci.* **31**, 287–303,

and for the global average in

JOHNSTON, H. S. (1975), Global ozone balance in the natural stratosphere, *Rev, Geophys. Space Phys.* **13**, 637–649.

The importance of the reaction of the minor constituents $HO_2$ and NO was established by the measurement of the large rate coefficient $\beta_{31a}$ by

HOWARD, C. J. and EVENSON, K. M. (1977), Kinetics of the reaction of $HO_2$ with NO, *Geophys. Res. Lettrs.* **4**, 437–440.

**Section 3.2.4 Natural and Artificial Sources of Nitrogen Oxides**

The role of aircraft exhausts in supplying the stratosphere directly with $NO_x$ has been widely studied since the original paper by

JOHNSTON, H. S. (1971), op. cit.

Particular reference should be made to the model calculations of

McELROY, M. B., *et al.* (1974), op. cit.

More recent two-dimensional (vertical and meridional) computer modeling with updated rate coefficients and chemical interactions between $NO_x$ and other minor constituents indicates much less $O_3$ reduction than did earlier estimates for $NO_x$ injection below 20 km according to

HIDALGO, H. and CRUTZEN, P. J. (1977), The tropospheric and stratospheric composition perturbed by $NO_x$ emissions of high-altitude aircraft, *J. Geophys. Res.* **82**, 5833–5866.

This revised chemistry would also reduce the ozone destruction in the lower stratosphere following a tropospheric explosion of a nuclear weapon. After the Johnston (1971) paper it was immediately noted in three independent papers that NO is formed chemically in the stratosphere from inert $N_2O$ produced in the soil:

CRUTZEN, P. J. (1971), Ozone production rates in an oxygen, hydrogen, nitrogen oxide atmosphere, *J. Geophy. Res.* **76**, 7311–7327,

McELROY, M. B. and McCONNELL, J. C. (1971), Nitrous oxide: A natural source of stratospheric NO, *J. Atmos. Sci.* **28**, 1095–1098,

NICOLET, M. and VERGISON, E. (1971), L'oxyde azoteux dans la stratosphere, *Aeronom. Acta* **90**, 1–16.

That the agricultural source of $N_2O$ might be stimulated by the wide use of chemical fertilizers was first noted by

CRUTZEN, P. J. (1974), Estimates of possible variations in total ozone due to natural causes and human activities, *Ambio* **3**, 201–210,

and has been investigated in detail by

McELROY, M. B.; ELKINS, J. W.; WOFSY, S. C.; and YUNG, Y. L. (1976), Sources and sinks for atmospheric $N_2O$, *Rev. Geophys. Space Phys.* **14**, 143–150,

LIU, S. C.; CICERONE, R. J.; DONAHUE, T. M.; and CHAMEIDES, W. L. (1977), Sources and sinks of atmospheric $N_2O$ and the possible ozone reduction due to industrial fixed nitrogen fertilizers, *Tellus* **29**, 251–263,

CRUTZEN, P. J. and EHHALT, D. H. (1977), Effects of nitrogen fertilizers and combustion on the stratospheric ozone layer, *Ambio* **6**, 112–117,

McELROY, M. B.; WOFSY, S. C.; and YUNG, Y. L. (1977), The nitrogen cycle: Perturbations due to man and their impact on atmospheric $N_2O$ and $O_3$, *Phil. Trans. Roy. Soc.* **B277**, 159–181.

The difficulties in assessing the effects of fertilizer on stratospheric ozone have been constructively addressed by

JOHNSTON, H. S. (1977), Analysis of the independent variables in the perturbation of stratospheric ozone by nitrogen fertilizers, *J. Geophys. Res.* **82**, 1767–1772.

Photodissociation of $N_2O$ was formerly thought to occur shortward of 3400 Å and therefore be important in the troposphere, but the wavelength limit of (3.2.47) was established by

JOHNSTON, H. S. and SELWYN, G. S. (1975), New cross sections for the absorption of near ultraviolet radiation by nitrous oxide ($N_2O$), *Geophys. Res. Lettrs.* **2**, 549–551.

The role of nuclear explosions was first noted by

FOLEY, H. M. and RUDERMAN, M. A. (1973), Stratospheric NO production from past nuclear explosions, *J. Geophys. Res.* **78**, 4441–4450,

and subsequent work has been evaluated by

BAUER, E. and GILMORE, F. R. (1975), Effect of atmospheric nuclear explosions on total ozone, *Rev. Geophys. Space Phys.* **13**, 451–458.

A sobering analysis of the devastating effects on the world-wide stratosphere that would be caused by a major nuclear exchange between two countries is given in

NATIONAL ACADEMY OF SCIENCES (1975), "Long-term worldwide effects of multiple nuclear-weapons detonations," N.A.S. Washington, D.C. 20418.

Galactic cosmic rays as a source of $NO_x$ was first noted by

WARNECK, P. (1972), Cosmic radiation as a source of odd nitrogen in the stratosphere, *J. Geophys. Res.* **77**, 6589–6591.

Extensive calculations of the distribution of this source in the atmosphere have been made by

NICOLET, M. (1975), On the production of nitric oxide by cosmic rays in the mesosphere and stratosphere, *Planet. Space Sci.* **23**, 637–649.

The established 11-year variation of cosmic rays was considered as a mechanism for causing the cyclic fluctuations in total ozone by

RUDERMAN, M. A. and CHAMBERLAIN, J. W. (1975), Origin of the sunspot modulation of ozone: Its implications for stratospheric NO injection, *Planet. Space Sci.* **23**, 247–268.

Solar cosmic rays have been proposed as an observable source of ozone reduction by

CRUTZEN, P. J.; ISAKSEN, I.S.A.; and REID, G. C. (1975), Solar proton events: Stratospheric sources of nitric oxide, *Science* **189**, 457–459.

and observations from Nimbus 4 have shown an ozone reduction associated with a major solar proton event according to

HEATH, D. F.; KRUEGER, A. J.; and CRUTZEN, P. J. (1977), Solar proton event: Influence on stratospheric ozone, *Science* **197**, 886–889.

Speculative proposals on how cosmic rays may have precipitated faunal extinctions through massive reductions in ozone are due to

RUDERMAN, M. A. (1974), Possible consequences of nearby supernova explosions for atmospheric ozone and terrestrial life, *Science* **184**, 1079–1081,

REID, G. C.; ISAKSEN, I.S.A.; HOLZER, T. E.; and CRUTZEN, P. J. (1976), Influence of ancient solar-proton events on the evolution of life, *Nature* **259**, 177–179.

A possible role of Galactic and solar cosmic rays in inducing climatic variations (and hence faunal extinctions) has been suggested by

CHAMBERLAIN, J. W. (1977), A mechanism for inducing climatic variations through the stratosphere: Screening of cosmic rays by solar and terrestrial magnetic fields, *J. Atmos. Sci.* **34**, 737–743.

**Section 3.2.5 Chlorine and the Halomethanes**

The potential importance of chlorine chemistry to stratospheric ozone was first declared publicly at an international symposium in Kyoto, September, 1973, by R. S. STOLARSKI and R. J. CICERONE. However, the chemistry in the published version of that paper is considerably revised from that in the original oral presentation. Another paper in the symposium proceedings discusses $ClO_x$ chemistry and was submitted for publication a few weeks after the Kyoto conference. Apparently the two groups were working independently on the problem and must share credit for recognition of the important reactions. The published papers are

WOFSY, S. C. and MCELROY, M. B. (1974), $HO_x$, $NO_x$, and $ClO_x$: Their role in atmospheric photochemistry, *Canad. J. Chem.* **52**, 1582–1591,
STOLARSKI, R. S. and CICERONE, R. J. (1974), Stratospheric chlorine: A possible sink for ozone, *Canad. J. Chem.* **52**, 1610–1615.

A third paper in the same proceedings discusses $ClO_x$ chemistry briefly as a result of the Stolarski and Cicerone presentation.

That the synthetic halomethanes evidently have no natural sink other than stratospheric photolysis and that they therefore pose a serious threat to the environment was first announced in

MOLINA, M. J. and ROWLAND, F. S. (1974), Stratospheric sink for chlorofluoromethanes: Chlorine atom catalyzed destruction of ozone, *Nature* **249**, 810–812.

An exceptionally thorough analysis of the Cl chemistry and projection of the environmental hazard is given in

ROWLAND, F. S. and MOLINA, M. J. (1975), Chlorofluoromethanes in the environment, *Rev. Geophys. Space Phys.* **13**, 1–35.

The formation of $ClNO_3$ (chlorine nitrate) was the first recognized example of self-cancellation of two ozone-destroying catalysts ($ClO_x$ and $NO_x$); see

ROWLAND, F. S.; SPENCER, J. E.; and MOLINA, M. J. (1976), Stratospheric formation and photolysis of chlorine nitrate, *J. Phys. Chem.* **80**, 2711–2713.

Volcanic gases as a possible source of stratospheric chlorine have been considered in

STOLARSKI, R. S. and CICERONE, R. J. (1974), op. cit.,
RYAN, J. A. and MUKHERJEE, N. R. (1975), Sources of stratospheric gaseous chlorine, *Rev. Geophys. Space Phys.* **13**, 650–658.

Measurements of photoabsorption cross sections for several halomethanes are reported in

ROWLAND, F. S. and MOLINA, M. J. (1975), op. cit.,
ROBBINS, D. E. and STOLARSKI, R. S. (1976), Comparison of stratospheric ozone distributions by fluorocarbons 11, 12, 21, and 22, *Geophys. Res. Lettrs.* **3**, 603–606.

Bromine in the atmosphere and its possible influence on the stratosphere is treated in

WOFSY, S. C.; MCELROY, M. B.; and YUNG, Y. L. (1975), The chemistry of atmospheric bromine, *Geophys. Res. Lettrs.* **2**, 215–218.

**Section 3.3 Stratospheric Motions**

The use of vertical profiles of trace constituents for the stratosphere to derive vertical mixing rates is developed in

WOFSY, S. C. and MCELROY, M. B. (1973), On vertical mixing in the upper stratosphere and lower mesosphere, *J. Geophys. Res.* **78**, 2619–2624.

Stratospheric circulation and exchange with the troposphere is reviewed in

REITER, E. R. (1975), Stratospheric–tropospheric exchange processes, *Rev. Geophys. Space Phys.* **13**, 459–474.

A theoretical discussion of stratospheric meteorology and general circulation is given in

HOLTON, J. R. (1975). "The Dynamic Meteorology of the Stratosphere and Mesosphere," American Meteorological Society, Boston.

The propagation of gravity waves in the stratosphere is reviewed by

WALLACE, J. M. (1973), General circulation of the tropical lower stratosphere, *Rev. Geophys. Space Phys.* **11**, 191–222.

The status of knowledge on the distribution of water vapor is summarized in

HARRIES, J. E. (1976), The distribution of water vapor in the stratosphere, *Rev. Geophys. Space Phys.* **14**, 565–575.

## PROBLEMS

**3.1** *Threshold kinetic energy.* If two molecules of mass $M_A$ and $M_B$ have relative velocities $g$ along a line connecting their centers, show that the threshold velocity for a chemical reaction is given by (3.1.6).

**3.2** *Reaction energies.* (a) Compare the heat energy released in the conversion of $O_3 + O$ into $2O_2$ by direct annihilation (1.5.3) with the catalytic processes using hydrogenous radicals (3.2.11 and 3.2.14) and the nitrogen oxides (3.2.27 and 3.2.28). (b) Which of the following reactions is exothermic? By how much?

(1) $NO_3 + NO_3 \rightarrow 2NO_2 + O_2$
(2) $N_2O_5 + O \rightarrow 2NO_2 + O_2$

**3.3** *Equilibrium ozone with $ClO_x$* [after NICOLET (1975)]. Write the equilibrium $O_3$ abundance in the form (3.2.24) and find the correction term for a linear combination of Chapman, $HO_x$, $NO_x$, and $ClO_x$ chemistry, considering only the principal reactions.

**3.4** *Eddy diffusion with a source function: (I) Analytic.* Solve Eq. (3.2.48) for $[NO_x] \equiv N_1(z)$ in a steady state where $\mathcal{J} \equiv [N_2O][O(^1D)]\beta_{44}$ is a constant value in the range $z_0 < z < z_t$ and zero outside. Take $H(z)$ and $K(z) = \text{const}$ in this range. At high altitudes $N_1(z)/N(z) =$ a constant mixing ratio (i.e., we neglect photodissociation) and the $z_0$ plane is a perfect sink, with $N_1(z \leq z_0) = 0$.

**3.5** *Eddy diffusion: (II) Numerical.* Approximate values for the lower stratosphere are $\mathcal{J} = 33\ \text{cm}^{-3}\ \text{sec}^{-1}$, $H = 6.7$ km, $K = 3 \times 10^3\ \text{cm}^2/\text{sec}$, and $z_0 = 16$ km. Also take $z_t = 40$ km.

(a) What is the altitude of the maximum $[NO_x]$?
(b) Sketch $[NO_x]$ between 16 and 40 km and compare with values estimated from Fig. 3.4.
(c) Explain qualitatively how $[O(^1D)]$ and $[N_2O]$ must vary with altitude and why, therefore, $\mathcal{J} = \text{const}$ is a fair approximation in this region.

**3.6** *Nonlinear interaction of $NO_x$ and $ClO_x$.* (a) Consider $[NO_x] = [NO] + [NO_2] + [HNO_3]$ to be in photochemical equilibrium with the following reactions and rates or rate coefficients:

$$\begin{aligned} NO + O_3 &\rightarrow NO_2 + O_2, & \beta_1 \\ NO_2 + O &\rightarrow NO + O_2, & \beta_2 \\ NO_2 + h\nu &\rightarrow NO + O, & J_3 \\ OH + NO_2 + M &\rightarrow HNO_3 + M, & \beta_4 \\ HNO_3 + h\nu &\rightarrow OH + NO_2, & J_5 \end{aligned}$$

The catalytic loss rate of $O_3$ depends on $[NO_2]$. Show that the ratio of $NO_2$ to total odd nitrogen is

$$\frac{[NO_2]}{[NO_x]} \equiv R = \left[1 + \frac{[OH][M]\beta_4}{J_5} + \frac{[O]\beta_2 + J_3}{[O_3]\beta_1}\right]^{-1}$$

(b) If Cl is added to the system the total odd nitrogen is

$$[NO_x] = [NO] + [NO_2] + [HNO_3] + [ClNO_3]$$

and we now consider, in addition to the above reactions,

$$ClO + NO_2 + M \rightarrow ClNO_3 + M, \qquad \gamma_6$$
$$ClNO_3 + h\nu \rightarrow ClO + NO_2, \qquad J_7$$

Show that if the total odd nitrogen remains fixed when Cl is added, the $[NO_2]$ is reduced such that

$$Q \equiv \frac{[NO_2](\text{with Cl})}{[NO_2](\text{without Cl})} = (1 + R \cdot S)^{-1}$$

where $S$ is proportional to $[ClO]$.

**3.7** *Upward diffusion with photochemical loss.* Over a height interval $z_0$ to $z$ an atmosphere has a constant scale height $H$ and vertical eddy diffusion coefficient $K$. In this region a minor constituent with number density $N_1(z)$ is photodissociated with a rate $J$, also independent of $z$. (a) Show that in equilibrium the density distribution has the solution

$$N_1(z) = N_1(z_0)e^{-\alpha(z-z_0)/H}$$

where $\alpha$ is a constant and $\alpha \geq 1$. (b) Relate $\alpha$ to the characteristic times of mixing $\tau_{mix}$ and photodissociation $\tau_{dis}$, and find $\alpha$ when $\tau_{dis} = \tau_{mix}$.

**3.8** *Diurnal variation.* Consider the formation of $NO_2$ in the day by

$$HNO_3 + h\nu \rightarrow OH + NO_2 \tag{1}$$

and its removal, day and night, by

$$NO_2 + OH + M \rightarrow HNO_3 + M \tag{2}$$

The rate coefficient $\beta_2$ for reaction (2) is constant. The production rate $J_1(t)$ in the 24-hour day of the Arctic summer is approximated by a sinusoidal variation,

$$J(t) = \tfrac{1}{2}J_{max}(1 - \cos \omega t)$$

where $t$ is measured from midnight and $\omega = (2\pi/24)\ \text{hr}^{-1}$. Regard all substances except $NO_2$ as constant. (a) How long after noon does $[NO_2]$ reach a maximum? Take $\beta_2 = 5 \times 10^{-30}\ \text{cm}^6/\text{sec}$, $[M] = 10^{18}\ \text{cm}^{-3}$, $[OH] = 10^7\ \text{cm}^{-3}$, $[HNO_3] = 10^{10}\ \text{cm}^{-3}$, $J_{max} = 2 \times 10^{-5}\ \text{sec}^{-1}$. (b) What is the ratio of the maximum to minimum $[NO_2]$?

# *Chapter 4*
# PLANETARY ASTRONOMY

This chapter deals with the remote sensing of planetary atmospheres by radiation. This radiation frequently contains the signatures of atmospheric characteristics (e.g., composition, temperature, pressure) integrated over a range of depth; proper interpretation of the data may be a formidable problem.

## 4.1 Radiative Transfer in an Optically Thick Atmosphere

### 4.1.1 Equation of Transfer

In the formation of a spectral absorption line by planetary gases, there may be scattering centers in the atmosphere as well as absorbers. We will suppose that the scattering, whether it be by gas (*Rayleigh scattering*) or aerosols (*Mie scattering*), does not change the frequency of the photon. In this sense the scattering is called *coherent*.

With a scattering and absorbing atmosphere illuminated by the sun (cf. Fig. 4.1), the general equation of transfer (1.2.11),

$$\mu \frac{dI}{d\tau} = I - \mathcal{J} \tag{4.1.1}$$

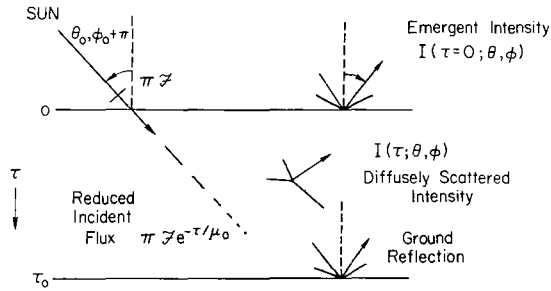



**Fig. 4.1** Scattering of sunlight in an optically thick atmosphere. The direction of propagation of diffuse intensity $I$ is toward $\theta$, $\phi$; however, the direction of the incident solar flux (measured across an area normal to the solar direction) is toward $\theta_0 + \pi$, $\phi_0$, since $\theta_0$ is conventionally taken as the solar zenith angle.

becomes (1.2.5) with a term added to the source function $\mathcal{J}$ to account for unscattered solar radiation:

$$\mu \frac{dI(\theta,\phi)}{d\tau} = I(\theta,\phi) - \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} I(\theta',\phi')p(\theta,\phi;\theta',\phi')\sin\theta'\,d\theta'\,d\phi'$$

$$-\frac{\pi\mathcal{F}}{4\pi} e^{-\tau/\mu_0} p(\theta,\phi;\theta_0+\pi,\phi_0) \tag{4.1.2}$$

Here $I$ (erg/cm$^2$ sec sr Hz) is the monochromatic specific intensity at optical depth $\tau$ (we will omit the subscript $\nu$ used in Section 1.2), the element of vertical optical thickness is $d\tau = -(\kappa + \sigma)\rho\,dz$, the direction cosines are $\mu = \cos\theta$, where $\theta$ is the zenith angle of the observer, $\mu_0 = \cos\theta_0$, where $\theta_0$ is the zenith angle of the sun, $p$ is the scattering phase function normalized by (1.2.4), $\sigma$ is the mass scattering coefficient (cm$^2$/gm) and $\kappa$ the absorption coefficient, and $\pi\mathcal{F}$ (erg/cm$^2$ sec Hz) is the solar flux crossing an area taken normal to the incident beam. The extra term in the solar flux enters as a matter of convenience. Where there is a parallel beam of incident radiation, its specific intensity is infinite. Hence we divide the radiation field into the unscattered part $\pi\mathcal{F}e^{-\tau/\mu_0}$ and the diffuse field $I$ that has experienced at least one scattering. That the flux term is equivalent to an additional intensity in the direction $\theta_0 + \pi$, $\phi_0$ may be verified by writing the solar intensity as $\pi\mathcal{F}\,\delta(\mu' - \mu_0)\delta(\phi' - \phi_0)$, where $\delta$ is the Dirac $\delta$-function.

With the single-scattering albedo defined as

$$\tilde{\omega} = \frac{\sigma}{\kappa + \sigma} \tag{4.1.3}$$

the equation for isotropic scattering ($p \equiv \text{const} = \tilde{\omega}$) becomes

$$\mu \frac{dI(\mu)}{d\tau} = I(\mu) - \frac{\tilde{\omega}}{2} \int_{-1}^{1} I(\mu)\, d\mu - \frac{\tilde{\omega}}{4} \mathscr{F} e^{-\tau/\mu_0} \tag{4.1.4}$$

We must bear in mind that $I$ is a function of depth $\tau$ as well as direction, even though the $\tau$ dependence is not explicitly stated.

### 4.1.2 Gaussian Quadrature Formula

Approximate solutions of the integro-differential equation (4.1.4) may be obtained by dividing the radiation field into $2n$ discrete streams and replacing the integral with a sum. Let $f(x)$ be a polynomial of degree $2m - 1$, completely specified by $2m$ numbers,

$$f(x) = c_0 + c_1 x + \cdots + c_{2m-1} x^{2m-1} \tag{4.1.5}$$

We wish to replace the integral with a sum such that

$$\int_A^B f(x)\, dx = \sum_{j=1}^{n} a_j f(x_j) \tag{4.1.6}$$

If we were to use Newton's method to evaluate the integral, we would have to specify the function at $n = 2$m evenly spaced points $x_j$, and the weighting factors in the summation would all be $a_j = (B - A)/2m$.

Gauss' method reduces the number of points $x_j$ required for an exact evaluation of (4.1.6) to only $n = m$ by selecting the optimum values of the *division points* for a $2m - 1$ degree polynomial. Thus the integral of an arbitrary function $f(x)$ specified at $n$ points, $x_j (j = 1 \cdots n)$, will be accurate for a polynomial of degree $2n - 1$ if the $x_j$ and $a_j$ are selected by Gauss' method.

To see how this method works, let us put (4.1.5) into (4.1.6) and evaluate the $l$th term. For the integral to be accurately represented by the summation, (4.1.6) must be exact for each term in the polynomial:

$$c_l \int_A^B x^l\, dx = c_l (a_1 x_1{}^l + a_2 x_2{}^l + \cdots + a_n x_n{}^l), \qquad l = 0, 1, \ldots, 2m - 1 \tag{4.1.7}$$

Thus we have $2m$ such equations ($l = 0$ to $2m - 1$) and we wish to solve them for $a_1 \cdots a_m$ and $x_1 \cdots x_m$. By scaling the function $f(x)$ and translating the origin we can always write the integral over a finite interval in terms of limits $A, B = -1, +1$. Writing

$$\alpha_l \equiv \int_{-1}^{1} x^l\, dx, \qquad l = 0, 1, \ldots, 2m - 1 \tag{4.1.8}$$

we define a set of $m$ constants $d_l$ by the $m$ equations

$$\alpha_{i+m} + \sum_{l=0}^{m-1} d_l \alpha_{i+l} = 0, \qquad i = 0, 1, \ldots, m-1 \tag{4.1.9}$$

From (4.1.7.) this equation becomes

$$\sum_{j=1}^{n} a_j x_j^{i+m} + \sum_{l=0}^{m-1} d_l \sum_{j=1}^{n} a_j x_j^{i+l} = \sum_{j=1}^{n} a_j x_j^{i} \left( x_j^{m} + \sum_{l=0}^{m-1} d_l x_j^{l} \right) = 0 \tag{4.1.10}$$

The terms in parentheses are of the form of a polynomial of degree $m$. Thus with the $m$ values of $d_l$ obtained from solving (4.1.9), we can find the $n\,(=m)$ *division points* $x_j$ from the $m$ solutions of the equation

$$x_j^{m} + \sum_{l=0}^{m-1} d_l x_j^{l} = 0 \tag{4.1.11}$$

Once the $d_l$'s and $x_j$'s are determined, we can find the $n$ Gaussian weighting factors from $n$ of the $2n$ equations (4.1.7),

$$\alpha_l = \sum_{j=1}^{n} a_j x_j^{l}, \qquad l = 0, 1, \ldots, 2n-1 \tag{4.1.12}$$

where $\alpha_l$ is found from (4.1.8). (Only $n$ of these equations are linearly independent.)

The division points and weighting factors are determined once and for all for a given degree $n$ and specified integration limits (conventionally $-1$ and $+1$) and are tabulated in various books on mathematical functions. An example of the calculation, use, and accuracy of Gaussian integration is given in Problem 4.1.

### 4.1.3 Solution for Isotropic Scattering in the First Approximation

As in our discussion of thermal radiation (see Section 1.2) we can gain an insight into radiative transfer problems with a very simple two-stream solution, even though quantitative accuracy may require more elaborate computations. The transfer equation for isotropic scattering, (4.1.4), with Gaussian summation is

$$\mu_i \frac{dI(\mu_i)}{d\tau} = I(\mu_i) - \frac{\tilde{\omega}}{2} \sum_j a_j I(\mu_j) - \frac{\tilde{\omega}}{4} \mathscr{F} e^{-\tau/\mu_0}, \quad i = \pm 1, \pm 2, \ldots, \pm n \tag{4.1.13}$$

For $n = 1$ the summation is a two-point Gaussian sum, with $a_{\pm 1} = 1$ and $\mu_{\pm 1} = \pm 3^{-1/2}$.

Our procedure is first to obtain a solution of the associated homogeneous equation and then add a particular solution for the full equation. The homogeneous equation

$$\mu_i \frac{dI_i}{d\tau} = I_i - \frac{\tilde{\omega}}{2}(I_1 + I_{-1}), \qquad i = \pm 1 \tag{4.1.14}$$

has a solution of the form $g_i e^{-k\tau}$, where

$$g_i = \frac{\text{const.}}{1 + \mu_i k} \tag{4.1.15}$$

However, this solution is compatible with (4.1.14) only for certain *characteristic values* of $k$. By putting the assumed solution, with $g_i$ as given above, back into (4.1.14), we obtain the *characteristic equation*

$$1 - {\mu_1}^2 k^2 = \tilde{\omega} \tag{4.1.16}$$

which has two solutions for $k$ that differ only in sign. Thus the homogeneous equation has the solution

$$I_i = \frac{\tilde{\omega}\mathscr{F}}{4}\left(\frac{Le^{-k\tau}}{1 + \mu_i k} + \frac{L'e^{k\tau}}{1 - \mu_i k}\right) \tag{4.1.17}$$

where $k = [3(1 - \tilde{\omega})]^{1/2}$ and where the constants $\tilde{\omega}\mathscr{F}/4$ have been extracted for later convenience.

A particular integral for the full equation (4.1.13) is

$$I_i = \frac{\tilde{\omega}\mathscr{F}}{4} h_i e^{-\tau/\mu_0} \tag{4.1.18}$$

Substitution gives

$$h_i = \frac{\gamma}{1 + \mu_i/\mu_0} \tag{4.1.19}$$

where $\gamma$ is a constant. However, this constant is not an arbitrary integration constant. The first approximation ($n = 1$) replaces an integro-differential equation with two coupled, first-order differential equations (4.1.14) and both integration constants have been obtained from the homogeneous solution. Putting (4.1.18) with $h_i$ given by (4.1.19) into (4.1.13) defines the constant,

$$\gamma = \left(1 - \frac{\tilde{\omega}}{1 - {\mu_1}^2/{\mu_0}^2}\right)^{-1} = \frac{1 - 3{\mu_0}^2}{1 - k^2{\mu_0}^2} \tag{4.1.20}$$

The particular solution is therefore

$$I_i = \frac{\tilde{\omega}\mathscr{F}}{4} \frac{(1 - \mu_1^2/\mu_0^2)e^{-\tau/\mu_0}}{(1 - \tilde{\omega} - \mu_1^2/\mu_0^2)(1 + \mu_i/\mu_0)} \tag{4.1.21}$$

The full solution consists of (4.1.17) and (4.1.21); these equations do not give the intensity $I(\mu)$ for arbitrary $\mu$ but only the two-stream intensities $I(\pm\mu_1)$. However, we can use these two-stream solutions to obtain the source function and then use the formal solution of the transfer equation (1.2.16) to obtain $I(\mu)$.

We first apply the boundary conditions to evaluate the integration constants $L$ and $L'$. Because there is no downward diffuse radiation at the top of the atmosphere, $I_1(\tau = 0; \mu_i = -\mu_1) = 0$, which gives

$$L + L' + (1 - \tilde{\omega})^{1/2}(L - L') = \frac{\gamma\tilde{\omega}u_0/\mu_1}{1 - \mu_0/\mu_1} = \frac{\tilde{\omega}\sqrt{3}\mu_0(1 + \sqrt{3}\mu_0)}{1 - k^2\mu_0^2} \tag{4.1.22}$$

The second boundary condition is that the ground albedo is some value $\Lambda$. If the incident solar flux reaching the ground (where $\tau = \tau_0$) is reflected by *Lambert's reflection law* (i.e., isotropically), then the albedo condition is

$$I_1(\tau = \tau_0; \mu_1 = +\mu_1) = \Lambda[I_1(\tau_0; -\mu_1) + \mu_0\mathscr{F}e^{-\tau_0/\mu_0}] \tag{4.1.23}$$

We are now ready to define the optically thick atmosphere as one in which $\exp(-\tau_0) \ll \exp(-k\tau_0)$, which in turn requires a high scattering albedo, or $1 - \tilde{\omega} \ll 1$. In this case, condition (4.1.23) yields

$$L' = Le^{-2k\tau_0}\zeta(\Lambda, \tilde{\omega}) \tag{4.1.24}$$

where

$$\zeta(\Lambda, \tilde{\omega}) = \frac{(1 - \tilde{\omega})^{1/2}(1 + \Lambda) - (1 - \Lambda)}{(1 - \tilde{\omega})^{1/2}(1 + \Lambda) + (1 - \Lambda)} \tag{4.1.25}$$

The source function is

$$\begin{aligned} \mathscr{J} &= \frac{\tilde{\omega}}{2}(I_{+1} + I_{-1}) + \frac{\tilde{\omega}}{4}\mathscr{F}e^{-\tau/\mu_0} \\ &= \frac{\tilde{\omega}^2\mathscr{F}}{4}\left[\frac{Le^{-k\tau} + L'e^{k\tau}}{1 - \mu_1^2k^2} + \frac{e^{-\tau/\mu_0}}{1 - \tilde{\omega} - \mu_1^2/\mu_0^2}\right] + \frac{\tilde{\omega}\mathscr{F}}{4}e^{-\tau/\mu_0} \end{aligned} \tag{4.1.26}$$

We can eliminate $\mu_1$ with (4.1.16) and (4.1.20). The formal solution (1.2.16) then yields the upward intensity

$$\begin{aligned} I(\tau, +\mu) &= \int_\tau^{\tau_0} e^{-(\tau' - \tau)/\mu}\mathscr{J}(\tau')\,d\tau'/\mu \\ &= \frac{\tilde{\omega}\mathscr{F}}{4}\left[\frac{Le^{-k\tau}}{1 + k\mu} + \frac{L'e^{k\tau}}{1 - k\mu} + \frac{\gamma e^{-\tau/\mu_0}}{1 + \mu/\mu_0}\right] \qquad (0 < \mu \leq 1) \end{aligned} \tag{4.1.27}$$

and the downward intensity

$$I(\tau, -\mu) = \int_0^\tau e^{-(\tau-\tau')/\mu} \mathscr{J}(\tau')\, d\tau'/\mu$$
$$= \frac{\tilde{\omega}\mathscr{F}}{4}\left[\frac{L}{1-k\mu}(e^{-k\tau} - e^{-\tau/\mu}) + \frac{L}{1+k\mu}(e^{k\tau} - e^{-\tau/\mu})\right.$$
$$\left. + \frac{\gamma}{1-\mu/\mu_0}(e^{-\tau/\mu_0} - e^{-\tau/\mu})\right] \qquad (0 < \mu \leq 1) \qquad (4.1.28)$$

The emergent intensity at the top of the atmosphere can be written in the convenient form

$$I(0, +\mu) = \frac{\tilde{\omega}\mathscr{F}}{4}\frac{\mu_0}{\mu+\mu_0}\frac{(1+3^{1/2}\mu)(1+3^{1/2}\mu_0)}{(1+k\mu)(1+k\mu_0)}[1 + P(\tilde{\omega}, \mu, \mu_0)] \qquad (4.1.29)$$

where

$$P = \frac{2k(\mu+\mu_0)[1-(1-\tilde{\omega})^{1/2}]\zeta(\Lambda, \tilde{\omega})e^{-2k\tau_0}}{(1-k\mu)(1-k\mu_0)[1+(1-\tilde{\omega})^{1/2}]} \qquad (4.1.30)$$

Except for the flux projection factor $\mu_0$ in the numerator of (4.1.29), the emergent intensity is completely symmetrical in $\mu$ and $\mu_0$. This feature is common to radiative transfer problems and is known as the *principle of reciprocity*. Defining

$$H_1(\tilde{\omega}, \mu) \equiv \frac{1+\sqrt{3}\mu}{1+k\mu} \qquad (4.1.31)$$

we may write the emergent intensity as

$$I(0, +\mu) = \frac{\tilde{\omega}\mathscr{F}}{4}\frac{\mu_0}{\mu+\mu_0}H_1(\tilde{\omega}, \mu)H_1(\tilde{\omega}, \mu_0)(1 + P) \qquad (4.1.32)$$

In the limit $\tau_0 \to \infty$, then $P = 0$ and the $H_1$ functions give the solution for a semi-infinite atmosphere. With the Ambartzumian–Chandrasekhar *principles of invariance* the transfer equation (4.1.4) can also be formulated as an integral equation. Exact solutions of the integral equation for the emergent intensity in the case of semi-infinite atmospheres are given in terms of certain tabulated *Chandrasekhar H functions*, of which $H_1$ is the leading term. Indeed, an idea of the accuracy of the first approximation is afforded by the consideration that $H_1(\tilde{\omega}, \mu)$ is accurate to within about 10 percent. For example, $H(\tilde{\omega} = 1, \mu = 1) = 2.91$, which is 7 percent higher than (4.1.31). Somewhat better accuracy can be achieved while still retaining the analytic simplicity of the first approximation by replacing (4.1.31) with

$$H_1(\tilde{\omega}, \mu) = \frac{H(1, \mu)}{1+k\mu} \qquad (4.1.33)$$

where $H(1, \mu)$ is the tabulated exact $H$ function for conservative scattering.

With $H_1$ so defined, Eq. (4.1.32) is accurate to a few percent for $\tau_0 \gtrsim 3$ and $\tilde{\omega} \gtrsim 0.9$. Note that the additional term $P$ for the finiteness of the atmosphere is by (4.1.30) proportional to $\zeta$ of (4.1.25) and can be either positive or negative. We obtain $P = \zeta = 0$ when the ground albedo is

$$\Lambda = \frac{1 - (1 - \tilde{\omega})^{1/2}}{1 + (1 - \tilde{\omega})^{1/2}} \tag{4.1.34}$$

Therefore a finite but thick atmosphere with this ground albedo will reflect radiation in the same manner as a semi-infinite one. It is apparent physically why condition (4.1.34) is independent of $\tau_0$: It simulates the albedo for a semi-infinite atmosphere that is diffusely illuminated from above (Problem 4.4).

### 4.1.4 Anisotropic Scattering

In general we can express any phase function $p(\cos\Theta)$ as a series of Legendre polynomials with arguments

$$\cos\Theta = \mu\mu' + (1 - \mu^2)^{1/2}(1 - \mu'^2)^{1/2}\cos(\phi - \phi') \tag{4.1.35}$$

In the simplest example of $p = \tilde{\omega}(1 + a\cos\Theta)$, for constants $\tilde{\omega}$ and $a$ $(-1 \le a \le 1)$, we may represent the scattered intensity as the sum of two components,

$$I(\tau, \mu, \phi) = I^{(0)}(\tau, \mu) + I^{(1)}(\tau, \mu)\cos(\phi - \phi_0) \tag{4.1.36}$$

The equation of transfer (4.1.2) now becomes two independent equations (Problem 4.6). The azimuth-independent intensity is given by

$$\mu \frac{dI^{(0)}(\mu)}{d\tau} = I^{(0)}(\mu) - \frac{\tilde{\omega}}{2}\int_{-1}^{+1} I^{(0)}(\mu')\,d\mu' - \frac{\tilde{\omega}a\mu}{2}\int_{-1}^{+1} I^{(0)}(\mu')\mu'\,d\mu' - \frac{\tilde{\omega}(1 - a\mu\mu_0)\mathscr{F}e^{-\tau/\mu_0}}{4} \tag{4.1.37}$$

and the azimuth-dependent term has a coefficient satisfying

$$\mu \frac{dI^{(1)}(\mu)}{d\tau} = I^{(1)}(\mu) - \frac{\tilde{\omega}a(1 - \mu^2)^{1/2}}{4}\int_{-1}^{+1} I^{(1)}(\mu')(1 - \mu'^2)^{1/2}\,d\mu' - \frac{\tilde{\omega}a\mathscr{F}}{4}(1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}e^{-\tau/\mu_0} \tag{4.1.38}$$

Both $I^{(0)}$ and $I^{(1)}$ may be solved separately just as for the isotropic case above. For a semi-infinite atmosphere the emergent intensity $I^{(0)}$ is

$$I^{(0)}(0, \mu) = \frac{\tilde{\omega}\mathscr{F}}{4}\frac{\mu_0}{\mu + \mu_0}H^{(0)}(\mu)H^{(0)}(\mu_0)[1 - c_0(\mu + \mu_0) - a(1 - \tilde{\omega})\mu\mu_0] \tag{4.1.39}$$

where

$$c_0 = \frac{\tilde{\omega}\alpha_1^{(0)} a(1-\tilde{\omega})}{2-\tilde{\omega}\alpha_0^{(0)}} \tag{4.1.40}$$

and where $\alpha_n^{(0)}$ is the $n$th moment of $H^{(0)}$ (see expression in Problem 4.5).

The solution for $I^{(1)}$ is

$$I^{(1)}(0,\mu) = \frac{\tilde{\omega} a \mathscr{F}}{4} \frac{\mu_0}{\mu+\mu_0} (1-\mu^2)^{1/2}(1-\mu_0{}^2)^{1/2} H^{(1)}(\mu) H^{(1)}(\mu_0) \tag{4.1.41}$$

In the first approximation these two sets of $H$ functions are given by equations like (4.1.31) but where the characteristic equations yield $k$'s given by

$$k^{(0)} = [(1-\tilde{\omega})(3-\tilde{\omega}a)]^{1/2}, \qquad k^{(1)} = (3-\tilde{\omega}a)^{1/2} \tag{4.1.42}$$

Solutions of these equations, even in the first approximation, can readily yield information on differences in scattering for forward ($a > 0$) and backward ($a < 0$) phase functions.

More realistic phase functions, such as those appropriate to *Mie scattering* by droplets (see Section 4.3.2), are much more difficult to handle and it is frequently necessary to resort to complex computer programs. However, in many cases a satisfactory approximation is a solution expressed in terms of the *asymmetry factor* $\langle \cos\Theta \rangle$ of the scattering phase function. Thus the degree of anisotropy in $p(\cos\Theta)$ can be crudely described by the mean,

$$g \equiv \langle \cos\Theta \rangle = \frac{1}{2\tilde{\omega}} \int_{-1}^{1} p(\cos\Theta) \cos\Theta \, d(\cos\Theta) \tag{4.1.43}$$

where $p(\cos\Theta)$ is normalized to $\tilde{\omega}$ by (1.2.4). Thus $g = 0$ corresponds to isotropic scattering, $g \to 1$ for strongly forward-elongated phase functions that are typical of small drops, and $g \to -1$ for strongly backward scattering.

For the function $p(\cos\Theta) = \tilde{\omega}(1 + a\cos\Theta)$, we find $g = a/3$ is restricted to the range $\pm\frac{1}{3}$. A fairly simple analytic function that gives the full range of possible $g$'s is the *Henyey–Greenstein phase function*

$$p(\cos\Theta) = \frac{\tilde{\omega}(1-g^2)}{(1+g^2-2g\cos\Theta)^{3/2}} \tag{4.1.44}$$

The reflected radiation from an optically thick atmosphere is not strongly dependent on the precise phase function but does vary with $g$. Hence two different phase functions will produce similar radiation fields if their $g$'s (and the parameters $\tilde{\omega}$ and $\tau_0$; see below) are the same. Indeed, the similarity of two atmospheres may be carried further. If the $g$ is changed, it is possible to adjust the albedo $\tilde{\omega}$ to compensate for that change. In this way problems

may frequently be reduced to an *equivalent problem* for isotropic scattering. The similarity relations are

$$\tilde{\omega}' = \frac{(1-g)\tilde{\omega}}{1-g\tilde{\omega}} \tag{4.1.45}$$

$$\tau_0' = (1-g\tilde{\omega})\tau_0 \tag{4.1.46}$$

where primes refer to the isotropically scattering case and $\tau_0$ is the total optical thickness of the atmosphere.

### 4.1.5 Numerical Methods in Radiative Transfer

There are many approaches possible for solving the transfer equation (4.1.2). One procedure in common use is to solve the equation rigorously for a very thin layer, then apply a mathematical algorithm to extend the solution for the thin layer to layers of any desired optical thickness. If we choose the optical thickness of an atmospheric layer to be so small ($\tau \sim 10^{-5}$) that only single scattering occurs, the integral in (4.1.2) is zero, and the transfer equation has the form

$$dy/dx + f(x)y = Q(x) \tag{4.1.47}$$

By multiplying through by the integrating factor $\exp[\int f(x)\,dx]$, we can put (4.1.47) in the form of an exact differential,

$$d(ye^{\int f(x)\,dx}) = e^{\int f(x)\,dx}Q(x)\,dx \tag{4.1.48}$$

Introducing the physical variables in (4.1.48), we find

$$d(Ie^{-\tau/\mu}) = -e^{-\tau/\mu}\frac{\pi\mathscr{F}}{4\pi\mu}p(\cos\Theta)e^{-\tau/\mu_0}\,d\tau \tag{4.1.49}$$

Then, applying the boundary conditions of

$$\begin{aligned} I(\tau = 0) &= 0, \qquad \mu < 0 \\ I(\tau) &= 0, \qquad \mu > 0 \end{aligned} \tag{4.1.50}$$

we obtain

$$I = \frac{\mu_0\mathscr{F}}{4(\mu+\mu_0)}\{1 - e^{-\tau/\mu - \tau/\mu_0}\}p(\cos\Theta) \tag{4.1.51}$$

for the reflected intensity, and

$$I = \frac{\mu_0\mathscr{F}}{4(\mu-\mu_0)}\{e^{-\tau/\mu} - e^{-\tau/\mu_0}\}p(\cos\Theta) \tag{4.1.52}$$

for the transmitted intensity.

The expressions (4.1.51) and (4.1.52) for single scattering by a very thin layer are exact in the sense that no approximation regarding the phase function $p$ has been introduced. The next step in the procedure calls for evaluating the single-scattering expressions at specified Gaussian quadrature points and inserting these values in a *doubling equation algorithm.* The doubling procedure accounts for multiple scattering between the two thin layers that are being added, and returns values for the reflected and transmitted intensities valid for an optical thickness of $2\tau$. By successive doublings of the initial layer, the scattering properties for layers of large optical thickness can be rapidly obtained. Similarly, a vertically inhomogeneous atmosphere can be treated by combining layers of different properties.

## 4.2 Spectroscopy

### 4.2.1 Optically Thin Continuum

Consider the radiation entering a thin atmosphere defined by $\tau_c \ll 1$ at continuum frequencies $\nu$—well removed from spectral absorption lines. The light is diffusely reflected from the surface, or possibly from cloud tops, and observed at some angle $\theta$ from the local zenith (see Fig. 4.2). The *absorption profile* in a spectral line relative to the neighboring continuum is

$$s_\nu \equiv \frac{I_c - I_\nu}{I_c} = 1 - \exp\left[-\tau_\nu\left(\frac{1}{\mu_0} + \frac{1}{\mu}\right)\right] \tag{4.2.1}$$

The factor $(\mu_0^{-1} + \mu^{-1})$ for observation at a particular point on the disk is the air-mass factor $\eta$. For observations that cover the entire disk at full phase the mean air-mass factor is $\langle\eta\rangle = 4$ in the optically thin limit (see below).

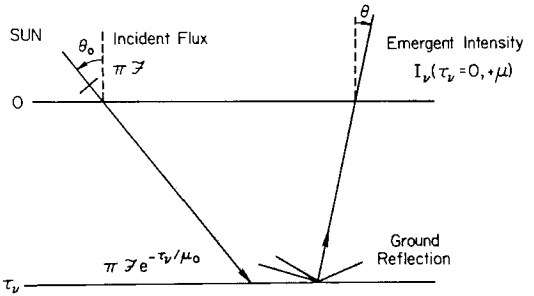



**Fig. 4.2** Transmission of sunlight through an atmosphere that has negligible scattering ($\tau_c \ll 1$) but absorption in the lines (where the thickness is $\tau_\nu$).

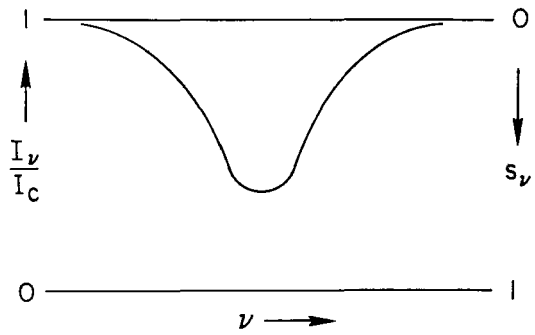



Fig. 4.3 Schematic profile of a spectral absorption line.

The entire area absorbed in one spectral line in Fig. 4.3 is the *equivalent width* $W$, so called because it is the width (in $\nu$ or $\lambda$ units) of a totally black line ($I_\nu = 0$) having the same area as the real profile. Without high spectral resolution it is not possible to measure profiles directly, but the equivalent width represents total energy loss and is independent of the spectral resolution.

In the limit of very weak lines such that the optical thickness in the line center $\tau_0 \ll 1$, we have

$$W \equiv \int_0^\infty s_\nu \, d\nu \approx \eta \int_0^\infty \tau_\nu \, d\nu = \eta \mathcal{N} S \tag{4.2.2}$$

where $\mathcal{N}$ is the *integrated overhead density* (1.1.6) and $S$ (cm$^2$/sec) is the line *strength* or integrated cross section (1.6.5).

For stronger absorption, as the center of the line profile approaches blackness or *saturation*, $W$ can clearly not continue to increase linearly with $\eta \mathcal{N} S$. The plot of $W$ vs. $\eta \mathcal{N} S$ is called the *curve of growth*. Its form depends in turn on the shape of the line absorption coefficient. For collisional broadening, usually the most important type of line broadening in planetary absorption spectra, the absorption cross section follows the Lorentz profile,

$$\alpha_\nu = \frac{S\Gamma/4\pi^2}{(\nu - \nu_0)^2 + (\Gamma/4\pi)^2} \tag{4.2.3}$$

where $\Gamma$ is the *collisional damping constant*. As $\eta \mathcal{N} S$ is increased (say, for different lines in a spectrum), the curve of growth begins to level out as the line core becomes saturated. However, the broad wings can still absorb additional radiation. Somewhat arbitrarily let us define the absorbing wings as beginning at the point $\nu\,(\equiv \nu_1)$, where the effective optical thickness drops to unity, $\eta\tau_\nu = \eta \mathcal{N} \alpha_\nu = 1$, or

$$\nu_1 - \nu_0 \approx \pm(\eta \mathcal{N} S \Gamma/4\pi^2)^{1/2} \tag{4.2.4}$$

Thus after the line's core becomes saturated the equivalent width is given by total absorption from the center $\nu_0$ to $\nu_1$ and linearly from $\nu_1$ onward:

$$W \approx 2\int_0^{|\nu_1-\nu_0|} d\nu + \frac{2\eta\mathcal{N}S\Gamma}{4\pi^2}\int_{|\nu-\nu_0|}^{\infty}\frac{d(\nu-\nu_0)}{(\nu-\nu_0)^2}$$

$$= \frac{2}{\pi}(\eta\mathcal{N}S\Gamma)^{1/2} \tag{4.2.5}$$

and strong, pressure-broadened lines are said to be on the square-root portion of the curve of growth.

A more accurate derivation of (4.2.5) replaces $2/\pi$ with $1/\sqrt{\pi}$, a correction of 13 percent. Writing $\mathcal{N}\alpha_\nu$ from (4.2.3) for $\tau_\nu$ in (4.2.1) yields the *Ladenberg–Reiche formula*,

$$W = \eta\mathcal{N}Se^{-2\eta\mathcal{N}S/\Gamma}[J_0(i2\eta\mathcal{N}S/\Gamma) - iJ_1(i2\eta\mathcal{N}S/\Gamma)] \tag{4.2.6}$$

where $J_0$ and $J_1$ are Bessel functions of the first kind. Limiting cases follow from asymptotic expressions for $J_0$ and $J_1$. For $\eta\mathcal{N}S/\Gamma \ll 1$, we recover (4.2.2); for $\eta\mathcal{N}S/\Gamma \gg 1$ we obtain

$$W \approx \left(\frac{\eta\mathcal{N}S\Gamma}{\pi}\right)^{1/2} \tag{4.2.7}$$

in place of (4.2.5).

The air mass factor $\eta$ is still $\mu_0^{-1} + \mu^{-1}$ for observations of a single region on the disk, and for observations of the entire disk at full phase, $\langle\eta^{1/2}\rangle$ is proportional to $\langle\mu^{-1/2}\rangle$. That is, $\eta^{1/2}$ in (4.2.7) must be replaced with

$$\langle\eta^{1/2}\rangle = 2\int_0^1 \eta^{1/2}\mu\, d\mu$$

$$= 2\int_0^1 \left(\frac{2}{\mu}\right)^{1/2}\mu\, d\mu = \frac{4\sqrt{2}}{3} \tag{4.2.8}$$

So far, we have ignored the fact that the optical path length traverses a range of pressure, but $\Gamma$ varies as the pressure and an effective pressure for the entire atmosphere is needed. As noted earlier, the problem does not arise if there is no saturation. If the lines are strong, the *Strong–Plass formula* gives the absorption $W$ in terms of the damping constant $\Gamma_0$ at the ground (or other reflecting surface):

$$W = \frac{\Gamma_0}{2\sqrt{\pi}}\frac{\gamma(2\eta\mathcal{N}S/\Gamma_0 + \frac{1}{2})}{\gamma(2\eta\mathcal{N}S/\Gamma_0)} \tag{4.2.9}$$

where $\gamma$ is the gamma function. In the limit of $\eta\mathcal{N}S/\Gamma \ll 1$, we have $\gamma(\frac{1}{2}) =$

$\sqrt{\pi}$ and $\gamma(2\eta\mathcal{N}S/\Gamma_0) = (2\eta\mathcal{N}S/\Gamma_0)^{-1}$, which yields $W = \eta\mathcal{N}S$, a result in agreement with (4.2.2) and independent of the pressure. For strong lines, $\eta\mathcal{N}S/\Gamma_0 \gg 1$, the asymptotic dependence of the gamma function gives

$$W \approx \left(\frac{\eta\mathcal{N}S\Gamma_0}{2\pi}\right)^{1/2} \tag{4.2.10}$$

Equation (4.2.10) is sometimes called the *Curtis–Godson approximation.* Comparing this formula with (4.2.7) indicates that $\Gamma = \Gamma_0/2$. The effective pressure for pressure-broadened lines in a barometric atmosphere is one-half the ground pressure (see Problem 4.7).

Absorption in a strong band does not distinguish between effects of the abundance $\mathcal{N}$ and the pressure. Weak bands, on the other hand, are sensitive only to $\mathcal{N}$ but are simply harder to observe. Thus telescopic observations of the strong 1.6 $\mu$m bands on Mars indicated a number of years ago that

$$f(CO_2)p_0^2 = 90\,(\text{mb})^2 \tag{4.2.11}$$

if the pressure broadening were mainly due to $N_2$. (Here $f$ is the $CO_2$ mixing ratio.) However, self-broadening by $CO_2$ is about 2.2 times as effective as broadening by $N_2$. With this correction, (4.2.11) becomes

$$f(CO_2)[1 + 1.2f(CO_2)]p_0^2 = 90\,(\text{mb})^2 \tag{4.2.12}$$

which gives $p_0 \gtrsim 6.4$ mb as the lower pressure limit for a pure $CO_2$ atmosphere. On the other hand, with the $CO_2$ abundance estimated telescopically from the weak 8689 Å band, it appeared that $p_0$ was in the range 10–20 mb. More refined spectroscopic analyses and direct spacecraft measurements have now confirmed the lower pressure limit and the preponderance of $CO_2$ in the Martian atmosphere.

Relative absorption of the different lines in a band may also be used to derive *rotational temperatures* for an atmosphere at the level of band absorption. The relative absorption of the $J$th rotational level in the ground state follows the Boltzmann distribution,

$$N(J) = \text{const.}(2J + 1)\exp[-BJ(J + 1)hc/kT] \tag{4.2.13}$$

where $B$ is the rotational constant and $2J + 1$ is the statistical weight of the level. The relative line strengths for a particular branch of the band may be written $S(JJ')/(2J + 1)$. Hence we have, for an optically thin atmosphere,

$$W(JJ') = \text{const.}\,S(JJ')\exp[-BJ(J + 1)hc/kT] \tag{4.2.14}$$

Thus a plot of $\log[W(JJ')/S(JJ')]$ against $J(J + 1)$ gives a straight line whose slope is inversely proportional to the temperature. In the square-root limit

of (4.2.7) we have $W(JJ') = \text{const.}(\mathcal{N}S)^{1/2}$ and the temperature is found from plotting $\log[W/S^{1/2}]$ against $\frac{1}{2}J(J+1)$.

### 4.2.2 Optically Thick Continuum

For an optically thick (semi-infinite) atmosphere the absorption profile of a spectral line, Eq. (4.2.1), has the form

$$\begin{aligned} s_\nu \equiv \frac{I_c - I_\nu}{I_c} &= 1 - \frac{\tilde{\omega}_\nu H(\tilde{\omega}_\nu, \mu) H(\tilde{\omega}_\nu, \mu_0)}{\tilde{\omega}_c H(\tilde{\omega}_c, \mu) H(\tilde{\omega}_c, \mu_0)} \\ &\approx \sqrt{3}(\mu + \mu_0)[(1 - \tilde{\omega}_\nu)^{1/2} - (1 - \tilde{\omega}_c)^{1/2}] \\ &\quad + (1 + 3\mu\mu_0)[(1 - \tilde{\omega}_\nu) - (1 - \tilde{\omega}_c)] \\ &\quad - 3(\mu + \mu_0)^2[(1 - \tilde{\omega}_\nu) - (1 - \tilde{\omega}_\nu)^{1/2}(1 - \tilde{\omega}_c)^{1/2}] \end{aligned} \tag{4.2.15}$$

where we have used (4.1.32) with $P = 0$ and expanded (4.1.33) for the $H$ functions for the case $1 - \tilde{\omega} \ll 1$ (i.e., for weak lines and a weakly absorbing continuum). This equation is more complicated than that for a simple reflecting surface with a thin atmosphere. Note that the angular dependence is in the opposite sense, with the strength of absorption increasing with increasing $\mu$ and $\mu_0$ (instead of increasing with increasing $1/\mu$ and $1/\mu_0$). The reason for this behavior is due to the fact that larger values of $\mu$ and $\mu_0$ are associated with radiation deeper in the atmosphere since the first (or last) scattering occurs at about $\tau/\mu_0$ (or $\tau/\mu$) = 1. Radiation that penetrates deeper will have more chances for scatterings, therefore a higher probability of becoming absorbed.

A second difference is that the degree of absorption no longer depends only on the gaseous absorption coefficient $\alpha_\nu$, or even the total albedo in the line $\tilde{\omega}_\nu$, but it depends as well on the absorption in the continuum $\tilde{\omega}_c$.

Let us now write the albedo in terms of absorption coefficients. Since we have two albedos in the problem, we need to divide the absorption coefficient $\kappa$ into two parts, so that $\kappa$ now becomes $\kappa + a_\nu$. If $\alpha_\nu$ is the molecular absorption cross section of (4.2.3), then $a_\nu = N\alpha_\nu/\rho$ is the gaseous mass absorption coefficient ($\text{cm}^2/\text{gm}$), where $N$ is the number density of absorbing molecules and $\rho$ is the mass density (including particles and gas). Similarly let $\sigma$ and $\kappa$ be the mass scattering and absorption coefficients of aerosols or other scattering particles in the atmosphere. Then the albedo within the absorption line is

$$\tilde{\omega}_\nu = \frac{\sigma}{a_\nu + \kappa + \sigma} \tag{4.2.16}$$

and in the continuum,

$$\tilde{\omega}_c = \frac{\sigma}{\kappa + \sigma} \tag{4.2.17}$$

It is convenient to refer the line albedo to the value at the line center,

$$\tilde{\omega}_0 = \frac{\sigma}{a_0 + \kappa + \sigma} \tag{4.2.18}$$

and write $\alpha_\nu = \alpha_0/x^2$, where

$$x^{-2} = \frac{(\Gamma/4\pi)^2}{(\nu - \nu_0)^2 + (\Gamma/4\pi)^2} \tag{4.2.19}$$

The radiative transfer theory is developed, of course, for the parameters $\tilde{\omega}_0$ and $\tilde{\omega}_c$. Our concern with the gaseous absorption in the atmosphere makes it desirable to work in parameters proportional to $a_0$. Thus line absorption can be expressed in terms of a dimensionless *line absorption parameter* [see (4.2.3)]:

$$u \equiv \frac{a_0}{\kappa + \sigma} = \frac{\tilde{\omega}_c - \tilde{\omega}_0}{\tilde{\omega}_0} = \frac{4S}{\Gamma}\frac{N}{\rho(\kappa + \sigma)} \tag{4.2.20}$$

The ratio

$$\mathcal{N}_s \equiv N/\rho(\kappa + \sigma) \text{ (molecules/cm}^2\text{)} \tag{4.2.21}$$

is the *specific abundance*, the amount of absorbing gas in a column of unit cross section and with a length of unit optical thickness in the continuum $[= 1/\rho(\kappa + \sigma)]$. This quantity is the scattering atmosphere's analog to gaseous abundance above a reflecting surface in a transparent atmosphere.

A second auxiliary parameter to accompany $u$ (in place of using $\tilde{\omega}_c$ or $\tilde{\omega}_0$) is the *absorption-ratio parameter*,

$$q \equiv \frac{a_0}{\kappa} = \frac{u}{1 - \tilde{\omega}_c} \tag{4.2.22}$$

which measures the relative importance of line and continuum absorption. Any two of the four parameters $\tilde{\omega}_0, \tilde{\omega}_c, u$, and $q$ define the profile. In the case of weak lines wherein $u \ll 1$, we may replace $(1 - \tilde{\omega}_\nu)$ with

$$(1 - \tilde{\omega}_\nu) \approx (1 - \tilde{\omega}_c) + u/x^2 \tag{4.2.23}$$

Then with (4.2.22), the equivalent width obtained from (4.2.15) may be

written in terms of the two parameters $\tilde{\omega}_c$ and $q$:

$$W = \int_0^\infty s_\nu \, d\nu$$
$$= \frac{\Gamma}{2\pi} \int_1^\infty \frac{dx\, x}{(x^2-1)^{1/2}} \{\sqrt{3}(\mu+\mu_0)(1-\tilde{\omega}_c)^{1/2}[(1+q/x^2)^{1/2}-1]$$
$$+ (1+3\mu\mu_0)(1-\tilde{\omega}_c)q/x^2 - 3(\mu+\mu_0)^2(1-\tilde{\omega}_c)[(1+q/x^2)$$
$$- (1+q/x^2)^{1/2}]\} \qquad (4.2.24)$$

We now examine two limiting sub-cases of the weak-line limit; we will suppose that either the continuum absorption is much weaker or much stronger than the line absorption. If the continuum absorption is the stronger, then $q \ll 1$; the profile is Lorentzian,

$$s_\nu = \{\tfrac{1}{2}\sqrt{3}(\mu+\mu_0)(1-\tilde{\omega}_c)^{1/2}q + (1-\tfrac{3}{2}(\mu^2+\mu_0{}^2)](1-\tilde{\omega}_c)q\}/x^2 \qquad (4.2.25)$$

and the equivalent width is linear in $\mathcal{N}_s$ and independent of $\Gamma$:

$$W = [(1-\tilde{\omega}_c)^{-1/2}\tfrac{1}{2}\sqrt{3}(\mu+\mu_0) + 1 - \tfrac{3}{2}(\mu^2+\mu_0{}^2)]\mathcal{N}_s S \qquad (4.2.26)$$

The dominant term is clearly the first one, except at $\mu \approx 0 \approx \mu_0$. For a given position ($\mu$ and $\mu_0$) on the planet this linear curve of growth is indistinguishable from the weak-line case in the reflecting-surface model, although the variation with the geometry is different. The cause of the linearity of (4.2.26) is a bit more subtle than the linearity in optically thin atmospheres. But before we discuss the physical mechanics of line formation (see Section 4.2.3), we shall examine the other weak-line case.

If the continuum absorption is very weak compared with that at the line's center $q \gg 1$, then, except in the far wings (i.e., except where $x \gtrsim \sqrt{q}$), the profile is

$$s_\nu \approx \sqrt{3}(\mu+\mu_0)(1-\tilde{\omega}_c)^{1/2}[1+\sqrt{3}(\mu+\mu_0)(1-\tilde{\omega}_c)^{1/2}]\left(1+\frac{x^2}{2q}-\frac{x}{\sqrt{q}}\right)\frac{\sqrt{q}}{x}$$
$$+ [1-3(\mu^2+\mu\mu_0+\mu_0{}^2)](1-\tilde{\omega}_c)\frac{q}{x^2} \qquad (4.2.27)$$

The dominant term varies as $|\nu-\nu_0|^{-1}$ over most of the profile (Fig. 4.4). This situation is unique to the scattering, optically thick atmosphere. It is most clearly seen in the limit of $\tilde{\omega}_c = 1$, where (except for $\mu \approx 0 \approx \mu_0$)

$$s_\nu \to \sqrt{3}(\mu+\mu_0)(u^{1/2}/x) \qquad (4.2.28)$$

The latter profile gives an equivalent width that diverges logarithmically. With a finite continuum absorption, however, the profile at large $x$ varies as $1/x^2$ and the equivalent width is finite. Nevertheless, the broad, slowly diminishing profile can cause measurements of the equivalent width to be uncertain because of uncertainty in where the line merges into the continuum.

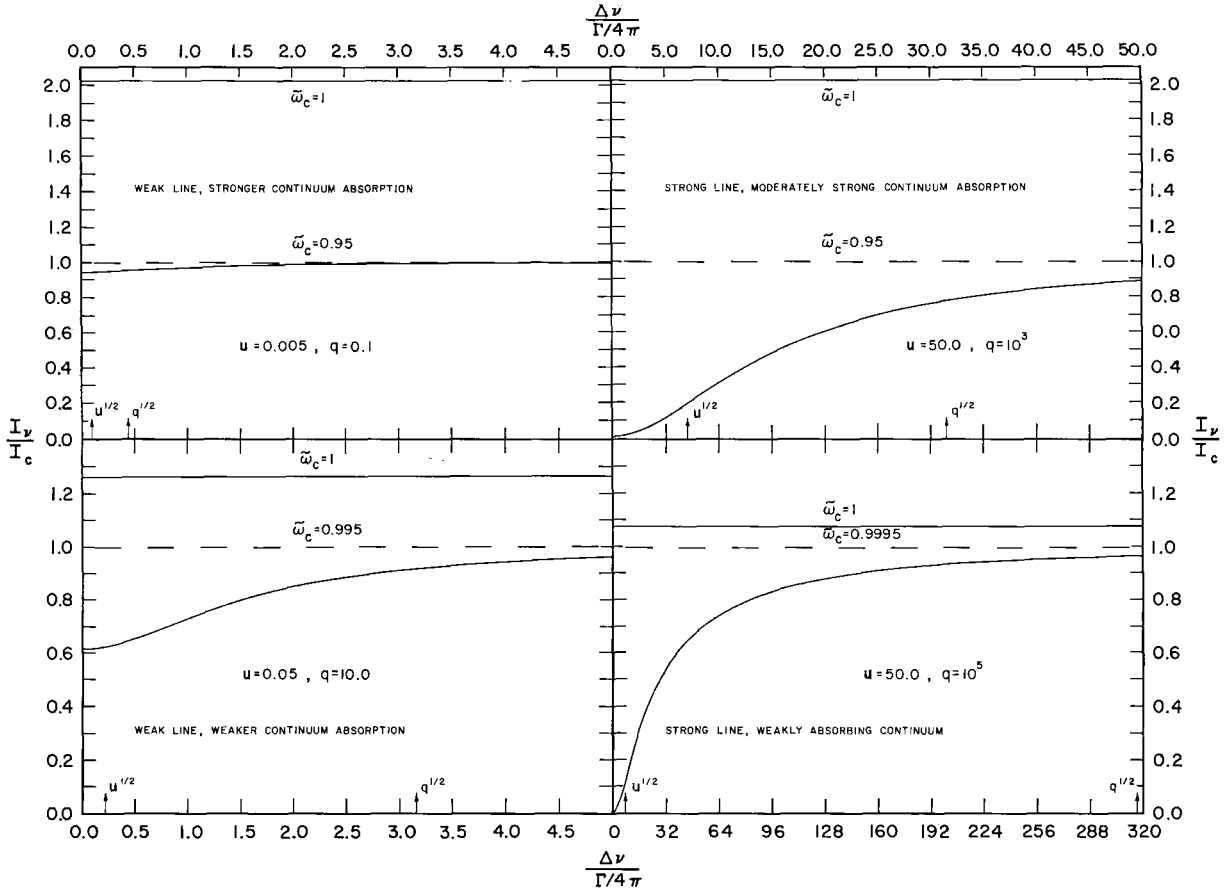



**Fig. 4.4** Line profiles for different combinations of the continuum albedo $\tilde{\omega}_c$ and line absorption parameter $u$. Note the distance from the line center $\Delta\nu$ is given in dimensionless units. The profiles are for semi-infinite atmospheres with $\mu = \mu_0 = 1$. In each case the intensity that would be observed for a continuum albedo of unity is shown relative to the actual continuum for albedo $\tilde{\omega}_c$. Values of $u^{1/2}$ and $q^{1/2}$ are indicated by arrows on the abscissa scale. [After CHAMBERLAIN (1970).]

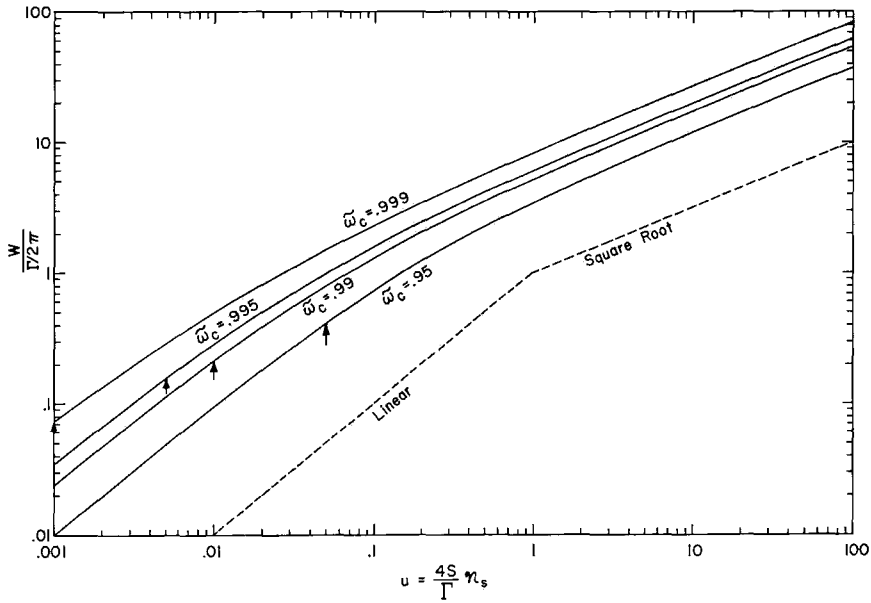



**Fig. 4.5** Curves of growth for semi-infinite atmospheres with $\mu = \mu_0 = 1$. Dashed lines are parallel to the asymptotic dependences. Arrows denote the points where $q = 1$, which is roughly the point where the changeover from square-root to linear dependence occurs. [After CHAMBERLAIN (1970).]

Integrating (4.2.27) yields

$$W \approx (\mu + \mu_0)\{1 + [3(1 - \tilde{\omega}_c)]^{1/2}(\mu + \mu_0)\} \ln\left(\frac{4S\mathcal{N}_s}{\Gamma(1 - \tilde{\omega}_c)}\right)\left(\frac{3}{4\pi^2}\mathcal{N}_s S\Gamma\right)^{1/2}$$
$$+ \mathcal{N}_s S[1 - 3(\mu^2 + \mu\mu_0 + \mu_0{}^2)] \tag{4.2.29}$$

Thus the curve of growth (Fig. 4.5) is in the square-root regime (except for $\mu \approx 0 \approx \mu_0$), but for quite a different reason than we found for (4.2.7). There the line saturation in the core was responsible for the diminished rate of increase in the curve of growth. Here the lines are assumed to be unsaturated and the square-root law is due to the $|\nu - \nu_0|^{-1}$ shape of the profile, which is due in turn to the diffuse scattering within the atmosphere (see discussion in the following section).

In (4.2.26) and (4.2.29) the air-mass factor is

$$\eta = \mu + \mu_0 \tag{4.2.30}$$

which gives a distinctive variation of the spectrum across the disk and with phase. For integrated observations of the whole disk at full phase, $\langle\eta\rangle = \frac{4}{3}$ if the continuum reflection is uniform over the disk, and $\langle\eta\rangle = \frac{3}{2}$ if the continuum is weighted by $\mu_0$ (which would give the continuum limb dark-

ening of a Lambertian surface) with the center of the image showing the strongest absorption. We can get a rough idea of the phase variation from its trend near the center of the planetary image, viz., at $\mu = \mu_0 = \cos(\Phi/2)$, where $\Phi$ is the *phase angle* (the angle between the sun and observer as seen from the planet). Thus with $\langle \eta \rangle \approx \text{const} \cos(\Phi/2)$, the absorbing spectrum is strongest at full phase ($\Phi = 0$) and weakest at the crescent ($\Phi \sim \pi$) (cf. Fig. 4.6).

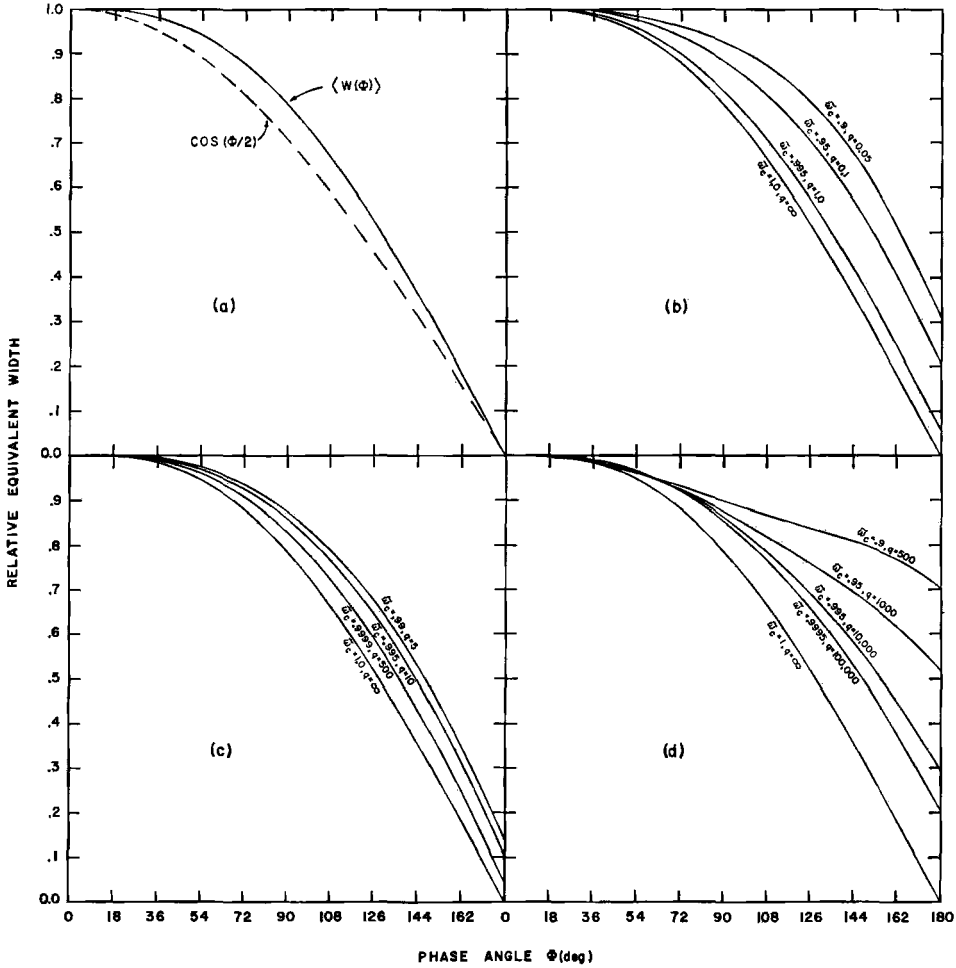



**Fig. 4.6** Phase variations of equivalent width for absorption lines in the integrated light of a planet. Angle 0° corresponds to full phase; 180° is new phase. Curves in (a) are for $W$ varying as $\mu + \mu_0$ and the continuum reflected according to Lambert's law. Dashed line is the rough approximation of $\mu = \mu_0 = \cos(\Phi/2)$; solid line is computed with (4.2.31). Curves in (b), (c), and (d) are for isotropic scattering and apply, respectively, to very weak lines ($u = 0.005$), moderately weak lines ($u = 0.05$), and very strong lines ($u = 50$). [After CHAMBERLAIN (1970).]

A more accurate integration, obtained by weighting the continuum with a limb-darkening factor $\mu_0$, is

$$\langle \eta \rangle = \frac{\iint (\mu + \mu_0)\mu\mu_0 \sin\theta \, d\theta \, d\phi}{\iint \mu\mu_0 \sin\theta \, d\theta \, d\phi}$$
$$= \frac{3\pi}{8} \frac{(1 + \cos\Phi)^2}{(\pi - \Phi)\cos\Phi + \sin\Phi} \tag{4.2.31}$$

Here $\theta$ is co-latitude and $\phi$ is longitude. Then $\mu = \sin\theta\cos\phi$ and $\mu_0 = \sin\theta\cos(\Phi - \phi)$; $\phi$ is integrated between $\Phi - \frac{1}{2}\pi$ and $\frac{1}{2}\pi$ and $\theta$ from 0 to $\pi$. It is possible to calibrate one spectrum relative to another taken at a different phase by the superimposed solar (Fraunhofer) lines. A long series of such spectra first disclosed that Venus' clouds do not reflect as a sharp deck of cumulus, but act instead like a thin haze in which unit optical thickness extends over a linear thickness the order of a scale height or more. Were the clouds more compact and dense, sunlight would traverse more of Venus' air above the cloud tops than below and the analog to $\langle \eta \rangle \approx \text{const}\cos(\Phi/2)$ would be $\langle \eta \rangle \approx \text{const}\sec(\Phi/2)$, which is opposite to the observed trend.

An approximate analytic theory has been developed for strong, saturated lines, as well, but curves of growth are best calculated numerically. Once the curve of growth reaches the square-root regime, strong lines continue to vary as $(\mathcal{N}S\Gamma)^{1/2}$. The physical cause for this variation is, for the saturated core, the same as for the optically thin case; that is, the transition frequency $\nu$ between core and wing varies as $(\mathcal{N}S\Gamma)^{1/2}$. In addition, the wide, intermediate region between saturated core and linear wing follows the square-root absorption because of multiple scattering. For strong lines

$$W \approx \tfrac{1}{2}(\mathcal{N}_s S\Gamma)^{1/2} f(\tilde{\omega}_c, \mu, \mu_0) \tag{4.2.32}$$

where $f$ is a monotonically increasing function of $\mu + \mu_0$ and is unity at $\mu = 0 = \mu_0$.

### 4.2.3 Radiative Transfer as a Random Walk of Photons

The square-root dependence of the absorption on the amount of absorber, (4.2.29), even for weak, unsaturated lines, is a pervasive feature of radiative transfer problems. To see physically how this dependence arises, and why a linear absorption law emerges for the case where the continuum absorption dominates that in the line, (4.2.26), it is useful to look at the simple one-dimensional flight of a photon initially deposited at point $s = 0$ in a medium wherein it can be scattered a distance of unit length either upward or downward every time it meets a scattering particle. By a succession of $n$ such scatterings it eventually arrives at point $s$ (see Fig. 4.7).

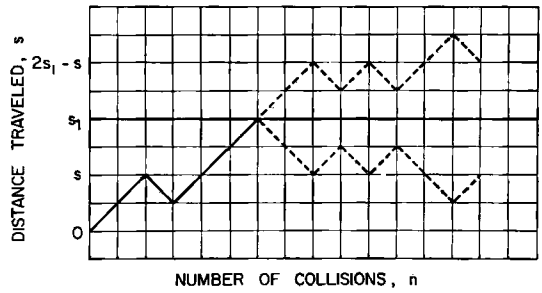



Fig. 4.7 A particle starts a random walk at $s = 0$. If $s_1$ is an absorbing screen, the particles that would otherwise reach $s$ after $s_1$ (by the lower dashed path) must be subtracted from $p(s, n)$ to give the effect of the screen. Trajectories above and below $s_1$ are symmetric.

Stochastic theory gives the probability for this photon arriving at point $s$ after $n$ scatterings as

$$P(s,n) = \frac{n!}{[\frac{1}{2}(n+s)]![\frac{1}{2}(n-s)]!2^n} \approx \left(\frac{2}{\pi n}\right)^{1/2} e^{-s^2/2n} \tag{4.2.33}$$

where the approximation is the asymptotic value for large $n$ obtained by Stirling's formula.

Let us now introduce an absorbing screen at position $s_1$. We must now subtract all those trajectories that would otherwise intersect $s_1$ before reaching $s$ on the $n$th collision. But the unbounded scattering medium is symmetric in the trajectories on either side of $s_1$. Therefore, the number to be subtracted from $P(s,n)$ is just the number that would have reached the point at $s_1 + (s_1 - s)$ after $n$ scatterings, since all such particles have necessarily crossed $s_1$ on the way. Thus we have the probability for the photon at $s$ after $n$ scatterings with a sink at $s_1$ to be

$$P(s,n;s_1) = P(s,n) - P(2s_1 - s, n) = \left(\frac{2}{\pi n}\right)^{1/2} [e^{-s^2/2n} - e^{-(2s_1 - s)^2/2n}] \tag{4.2.34}$$

In the atmosphere a distance $s$ measured from the source corresponds to $s_1 - \tau/\mu_1$, since the mean scattering distance in the vertical direction (that is, the unit of length) is $\mu_1 = 1/\sqrt{3}$, and since the top of the atmosphere is at $s_1$ from the source. The incident photons are deposited at $\tau/\mu_0 = 1$ or $\tau = \mu_0$; this depth corresponds to $s = 0$, so that $s_1$ becomes $\mu_0/\mu_1$. Similarly, photons leaving the atmosphere have their last scattering at $\tau = \mu$ or $s = (\mu_0 - \mu)/\mu_1$. The probability that a photon will arrive at the top of the atmosphere after

$n$ scatterings must be multiplied by the probability that it was not destroyed in one of those scatterings. Hence the probability of a photon emerging from the atmosphere after $n$ scatterings is

$$P[\sqrt{3}(\mu-\mu_0), n; \sqrt{3}\mu_0] = \left(\frac{2}{\pi n}\right)^{1/2} \tilde{\omega}^n \left[\exp\left(-\frac{3(\mu_0-\mu)^2}{2n}\right) - \exp\left(-\frac{3(\mu_0+\mu)^2}{2n}\right)\right] \tag{4.2.35}$$

To find the number of scatterings per incident photon we compare (4.2.35) with the value for $\tilde{\omega} = 1$, since in that limit all incident photons eventually re-emerge. Thus the mean number of scatterings is, for large $n$ and $\tilde{\omega} \sim 1$,

$$\begin{aligned}\langle n\rangle &\approx \frac{\int_1^\infty dn\, n^{-1/2} e^{n \ln \tilde{\omega}}[1 - 3(\mu_0-\mu)^2/2n]}{\int_1^\infty dn\, n^{-3/2}[1 - 3(\mu_0-\mu)^2]} \\ &\approx \frac{\sqrt{\pi}[1 + 3(\mu_0-\mu)^2(1-\tilde{\omega})]}{(1-\tilde{\omega})^{1/2}[2 - (\mu_0-\mu)^2]}\end{aligned} \tag{4.2.36}$$

When the absorption is weak, the total absorption is the mean number of scatterings times the fractional energy lost per scattering, or

$$\langle n\rangle(1-\tilde{\omega}) = \frac{[\pi(1-\tilde{\omega})]^{1/2}}{2 - (\mu_0-\mu)^2} \tag{4.2.37}$$

This is an approximate mean-free-path theory for the $\langle n\rangle$ of photons that can escape, but it does not give the correct angular distribution; it demonstrates, however, the origin of the square-root absorption law as being due to the decreasing number of scatterings with increasing $(1-\tilde{\omega})$. In the case of negligible continuum absorption ($\tilde{\omega}_c = 1$), the leading term of (4.2.15) varies as $(1-\tilde{\omega}_\nu)^{1/2}$. We can define a *sink function* analogously to the source function (4.1.26) but with $(1-\tilde{\omega})$ replacing $\tilde{\omega}$. This sink function is simply the amount of absorption occurring at each depth. Figure 4.8 shows two sink functions that are nearly proportional to $1-\tilde{\omega}_\nu$ high in the atmosphere, but at great depths the amount of absorption is diminished as the absorption coefficient is increased, causing the total absorption to vary as $(1-\tilde{\omega})^{1/2}$.

We can now understand why the line absorption varies as $\mathcal{N}_s S/(1-\tilde{\omega}_c)^{1/2}$ instead of as $(\mathcal{N}_s S)^{1/2}$ for the case where the line absorption is much weaker than the continuum absorption, (4.2.25–4.2.26). In that case the number of scatterings as given by (4.2.36) is governed by $(1-\tilde{\omega}_c)^{-1/2}$, but the line absorption is now only $\langle n\rangle u$, where $u$ is given by (4.2.20). Thus the statistical wanderings of photons through the atmosphere are affected very little by the line absorption. The dominant process of radiation transfer is now the scattering as modified by the photon sink due to *continuum* absorption. The line absorption makes only a small perturbation and hence has a linear

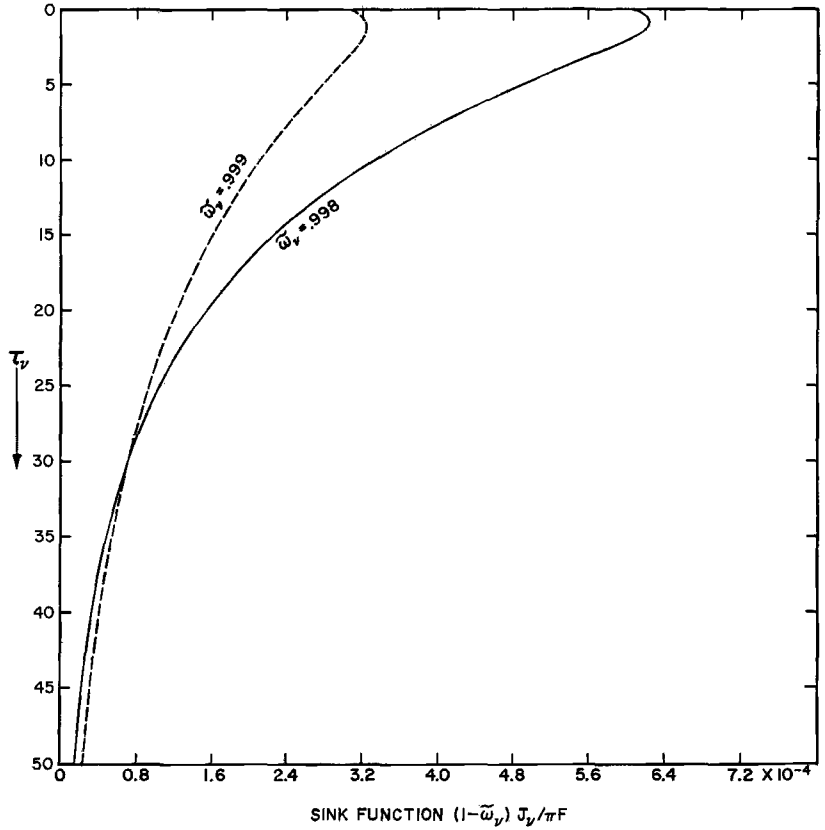



**Fig. 4.8** Comparison of two sink functions (the rate of absorption per unit optical thickness). [After CHAMBERLAIN (1970).]

effect on the number of photons escaping. This linearity is thus physically distinct from that produced by a thin atmosphere, a conclusion also emphasized by the qualitatively different air-mass dependence $\eta$.

### 4.2.4 Atmospheric Spectra and Abundances of the Elements

The absorption spectra of Mars and Venus are dominated by $CO_2$ bands. Strong bands in the far infrared totally dominate Venus' spectrum, and even the weak bands in the near infrared (see Fig. 4.9) appear clearly. The absorption occurs in a scattering atmosphere and the penetration of sunlight into the atmosphere depends on the albedo $\tilde{\omega}_\nu$. In the visual continuum $\tilde{\omega}_c \approx 1$ and appreciable light (the order of a few percent of the solar flux above the atmosphere) reaches the surface of Venus.

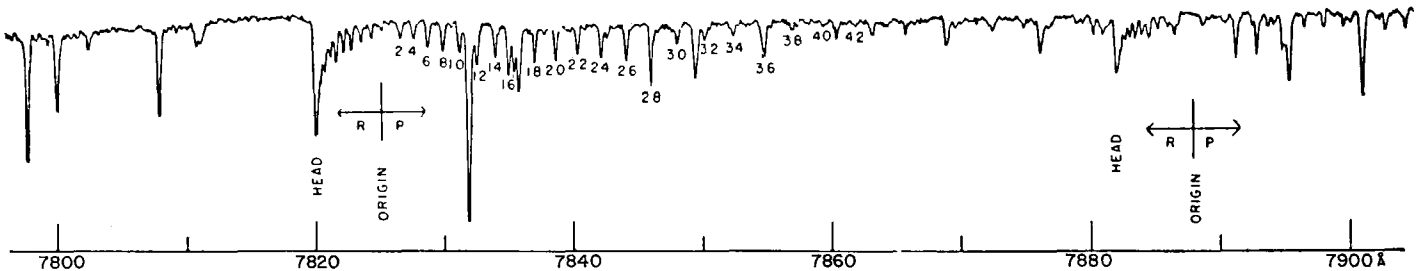



**Fig. 4.9** Vibration-rotation bands of $CO_2$ in spectrum of Venus. The two bands have heads at 7882.8 and 7820.0 Å arising from transitions from the ground vibrational level to $[5\nu_3 + (\nu_1, 2\nu_2)]$, where the $\nu_1$ and $2\nu_2$ levels are in Fermi resonance (see Fig. 1.12). Only the even numbered rotational lines are present. Plate obtained by T. Dunham, Jr. with the 2.5-m (100-in.) telescope at Mt. Wilson in 1938; plate dispersion, 4.4 Å/mm.

In the centers of the absorption lines, the penetration is only to a scale height or two below the apparent cloud tops. The $CO_2$ bands appear very roughly as though they were absorbed by 10–20 km of $CO_2$ at about 0.1 atmospheric pressure. Hence it is common to see quoted in the older literature $CO_2$ abundances "above the cloud tops" of a few atm-km. In a purely scattering atmosphere such a concept is meaningless.

In recent years high-resolution *Fourier spectroscopy* with a Michelson interferometer (see Fig. 4.10) has disclosed a number of minor constituents, and entry probes from Veneras 4–8 have produced in situ measurements of abundances (see Table 4.1).

The spectrum of Mars is more straightforward to interpret, but, having a total abundance equivalent to less than $10^{-2}$ atmospheres, its absorptions are very weak. For example, from the Earth $H_2O$ appears as weak satellite lines that are Doppler shifted from the telluric absorptions (Fig. 4.11). The Vikings 1 and 2 carried mass spectrometers to the surface, providing additional abundance ratios (Table 4.1.).

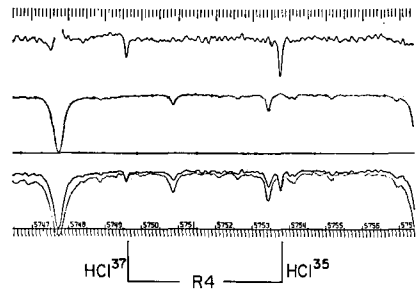



**Fig. 4.10** The $J = 4$ line in the $R$ branch of the 1.7 $\mu$m band of HCl on Venus, showing the isotope splitting, obtained with a Michelson interferometer. The frequency scale is $cm^{-1}$. From the bottom upward, the spectra are of Venus (two tracings), the sun, and the ratio Venus/sun. [After CONNES *et al.* (1969).]

**TABLE 4.1** *Fractional Composition of Planetary Atmospheres*[a]

| Species | Molecular Wt. | Earth | Venus | Mars | Jupiter |
|---|---|---|---|---|---|
| $H_2$ | 2 | 5(−7) | — | 5(−5) | 0.8–0.95(?) |
| He | 4 | 5(−6) | — | — | 0.05–0.2(?) |
| $CH_4$ | 16 | 2(−6) | — | <1(−3) | ~1(−3) |
| $NH_3$ | 17 | 4(−9) | — | <2(−4) | ~2(−4) |
| $H_2O$ | 18 | 1(−3)–1(−2) | ~1(−3) | ~4(−4) | 1(−6) |
| HF | 20 | — | 5(−9) | — | — |
| CO | 28 | 2(−7) | 5(−5) | 8(−4) | 2(−9) |
| $N_2$ | 28 | 0.78 | — | 2.4(−2) | — |
| NO | 30 | 5(−10) | — | 1(−9)–1(−8) | — |
| $O_2$ | 32 | 0.21 | — | 1.6(−3) | — |
| HCl | 36; 38 | — | 6(−7) | — | — |
| A | 40 | 9(−3) | — | 1.5(−2) | — |
| $CO_2$ | 44 | 3(−4) | 0.97 | 0.95–0.99(?) | — |
| $O_3$ | 48 | 4(−7) | - - | ~1(−8) | — |

[a] A number $a \times 10^b$ is written $a(b)$. See Appendix III for more detail on Earth's atmosphere.

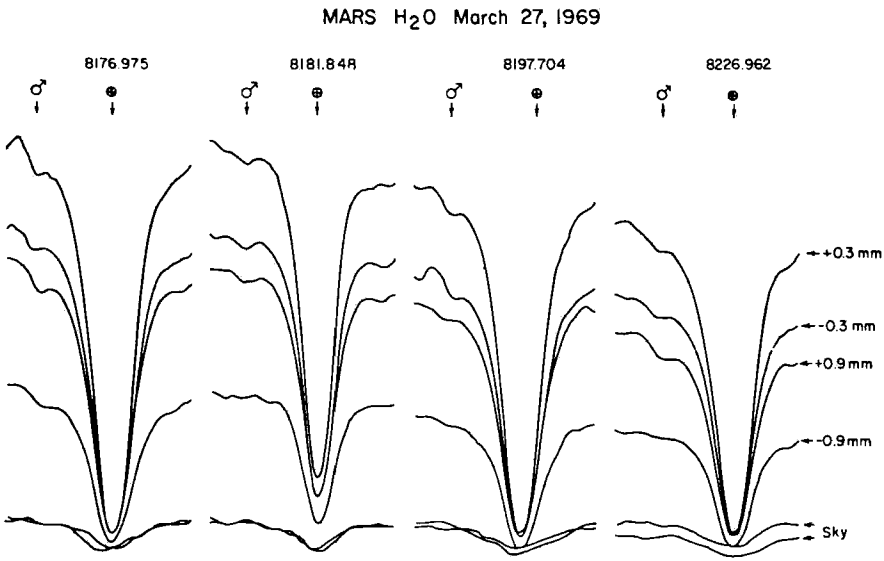



**Fig. 4.11** Water vapor in the spectrum of Mars. These tracings of four of the strongest lines in the 8200 Å band were made at different positions on the photographic plate, corresponding to different latitudes on Mars. The Martian absorption shows as a weak, Doppler-shifted indentation on the shoulder of the strong telluric absorption. Spectra obtained with the 2.7-m telescope at McDonald Observatory; dispersion, 1.9 Å/mm. [After R. G. TULL (1970) *Icarus*, **13**, 43.]

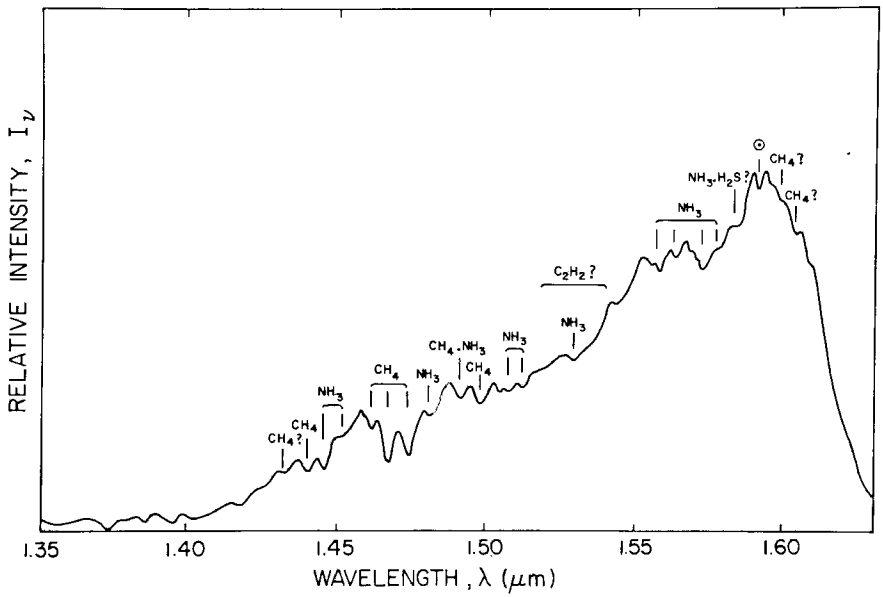



**Fig. 4.12** Infrared spectrum of Jupiter, 1.35–1.63 $\mu$m. [Adapted from D. P. CRUIKSHANK and A. B. BINDER (1969), Comm. Lunar Planetary Lab. No. 103, Univ. Arizona, Tucson.]

The red and infrared spectrum of Jupiter (see Fig. 4.12) is dominated by $CH_4$ (methane) and $NH_3$ (ammonia). The bands are complex and often difficult to interpret because the molecular structure involved in the transitions is not completely understood. However, it was long apparent, from thermodynamic considerations of the relative amounts of different molecules that would exist in equilibrium at Jovian temperatures, that the main constituent must be $H_2$.

Being a *homonuclear molecule*, $H_2$ (like $N_2$ or $O_2$) has no permitted vibration-rotation spectrum, since the two nuclei have a zero dipole moment about the center of mass. Such a rotating, vibrating molecule emits no classical (or quantum-mechanical) dipole radiations. The electric-quadrupole term in the field does not vanish, however, and the molecule can make transitions with the absorption of *forbidden bands*, which are weak but present and confirm the dominance of $H_2$. In addition, at high pressure $H_2$ may absorb by *enforced dipole transitions* (*pressure-induced absorption*). These vibration-rotation lines are very broad, being produced by molecular collisions, and are difficult to observe.

With the spectrographic slit focused along Jupiter's equator, it is easy to distinguish Jovian absorptions from Fraunhofer (solar) lines. Because of its large radius and 10-hr rotation period, the Doppler shift between the east and west limbs gives a definite slant to the lines. The solar radiation, being reflected from a moving object, has twice the Doppler shift that is characteristic of the rotational speed. Hence the Fraunhofer lines have twice the tilt of the Jovian absorptions.

Interpretation of the Jovian spectrum is confounded by several factors other than the complexity of the main absorptions. The principal high clouds are formed by $NH_3$ and the atmosphere is a scattering one. The banded structure of these clouds makes the depth of penetration of sunlight highly variable from place to place. Also different wavelengths penetrate to different depths giving large differences in determinations of the $[NH_3]/[CH_4]$ ratio.

The $H_2$ quadrupole lines are complicated by the phenomenon of *collisional narrowing*. Ordinary *pressure broadening* occurs (classically) because the phase of a passing wave train is suddenly altered; a Fourier analysis transforms the altered wave trains into a broadened frequency distribution. However, in certain cases this effect occurs only infrequently. In addition for molecules that are not isolated, the ability to absorb at a certain Doppler displacement depends not only on the molecule having the proper translational velocity, but on its capability to change its translational energy so as to conserve energy and momentum with the photon. For a molecule with a mean free path the order of the wavelength or smaller, this factor (it turns out, from a quantum treatment) increases the likelihood of absorbing a

photon near the line center. A scattering atmosphere with this narrowing has a complicated curve of growth, which hampers interpretation.

The presence of He has also long been inferred, but its direct observation was made possible only by the observation from Pioneer 10 of the airglow scattering in the resonance transition at 584 Å (see discussion in Section 6.3.1).

## 4.3 Photometry and Polarimetry

### 4.3.1 Determination of the Scattering Function from Spectral Variations

The variation of a planetary spectrum with the planet's phase (see Fig. 4.6) will depend (for an optically thick atmosphere) on the nature of the scattering phase function $p(\cos\Theta)$. One could in principle solve the transfer equation for various functions $p(\cos\Theta)$, fit the results with observations across the disk or at different phases, and so deduce the correct $p$. In practice this has not been a fruitful endeavor. Among the various parameters entering the spectroscopic theory, there are compensating trade-offs so that isotropic scattering can be forced to fit most of a given set of observations by adjusting unknowns. The *similarity relations* (Section 4.1.4) give, for example, the line albedo $\tilde{\omega}'$ and total thickness $\tau_0'$ of an isotropically scattering atmosphere that will mimic the reflected intensity from an atmosphere scattering anisotropically and with a different albedo $\tilde{\omega}_\nu$ and thickness $\tau_0$. There is also the possibility that the vertical mixing of aerosols and gas is inhomogeneous, which complicates interpretation.

In any case there are more sensitive methods for deriving the scattering phase function. Of some use is the photometry of the planet at different scattering angles and in different colors. More powerful is the method of polarimetry.

Once having a scattering function the number of adjustable parameters in the theory is helpfully narrowed but still not necessarily uniquely fixed because observations inherently are not as homogeneous as one could wish. The theoretical spectra may be obtained through large computing programs with accurate Mie-scattering phase functions, but for most purposes it is sufficient to use simplified functions $p(\cos\Theta)$ that can be solved more readily but which contain the essential features of the Mie scattering.

### 4.3.2 Scattering by Particles (Mie Scattering)

In discussions of scattering, either of two angles may be used. Phase functions are usually expressed in terms of a phase angle $\Theta$, which is the

angle at the scattering center between the directions of incidence and emergence, whereas a scattering angle $\alpha$ is the deviation of a scattered ray relative to the direction of an unscattered ray. Thus $\alpha = \pi - \Theta$.

The scattering diagram for a particular particle depends on its size, shape, and orientation, the wavelength of light, and on the real and imaginary parts of the index of refraction. However, for particles that are large compared with the wavelength, we can make a simplification by calculating the scattering diagram with ray optics. This is not to imply that the answer will be entirely correct, since interference between various emerging rays still must be considered.

Figure 4.13 shows the different components of light scattered by a sphere. The $l = 0$ ray is diffraction, which is concentrated in the forward direction and is unpolarized (for natural incident light). The intensity and polarization of light reflected and refracted may be computed separately for components of the intensity perpendicular and parallel to the plane of scattering from the well-known *Fresnel reflection coefficients*. The external reflection ($l = 1$) amounts to only a few percent for an index of refraction $n \lesssim 1.5$, but it is highly polarized. The most important internal rays are those refracted twice with no internal reflections ($l = 2$). These rays account for the strong forward scattering that is prevalent in *Mie scattering* (scattering by spherical drops) when the size parameter,

$$x \equiv 2\pi r/\lambda \tag{4.3.1}$$

is large. The light that is internally reflected once ($l = 3$) or twice ($l = 4$) accounts for only a small fraction of the scattered intensity but for spheres the resulting concentration of the light into a small angle gives rise to the primary and secondary *rainbows*.

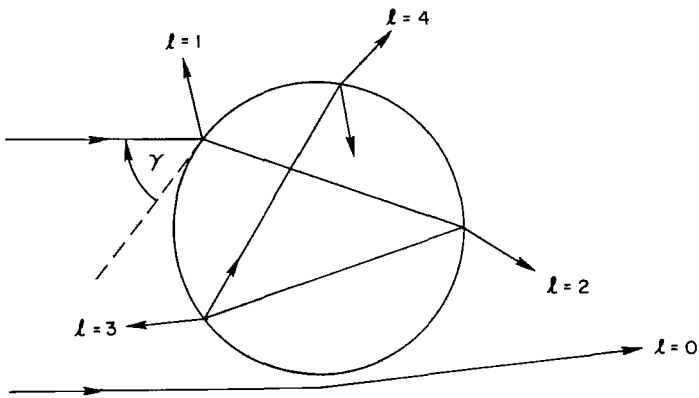



**Fig. 4.13** Paths of light rays scattered by a sphere according to geometrical optics. [Adapted from HANSEN and TRAVIS (1974).]

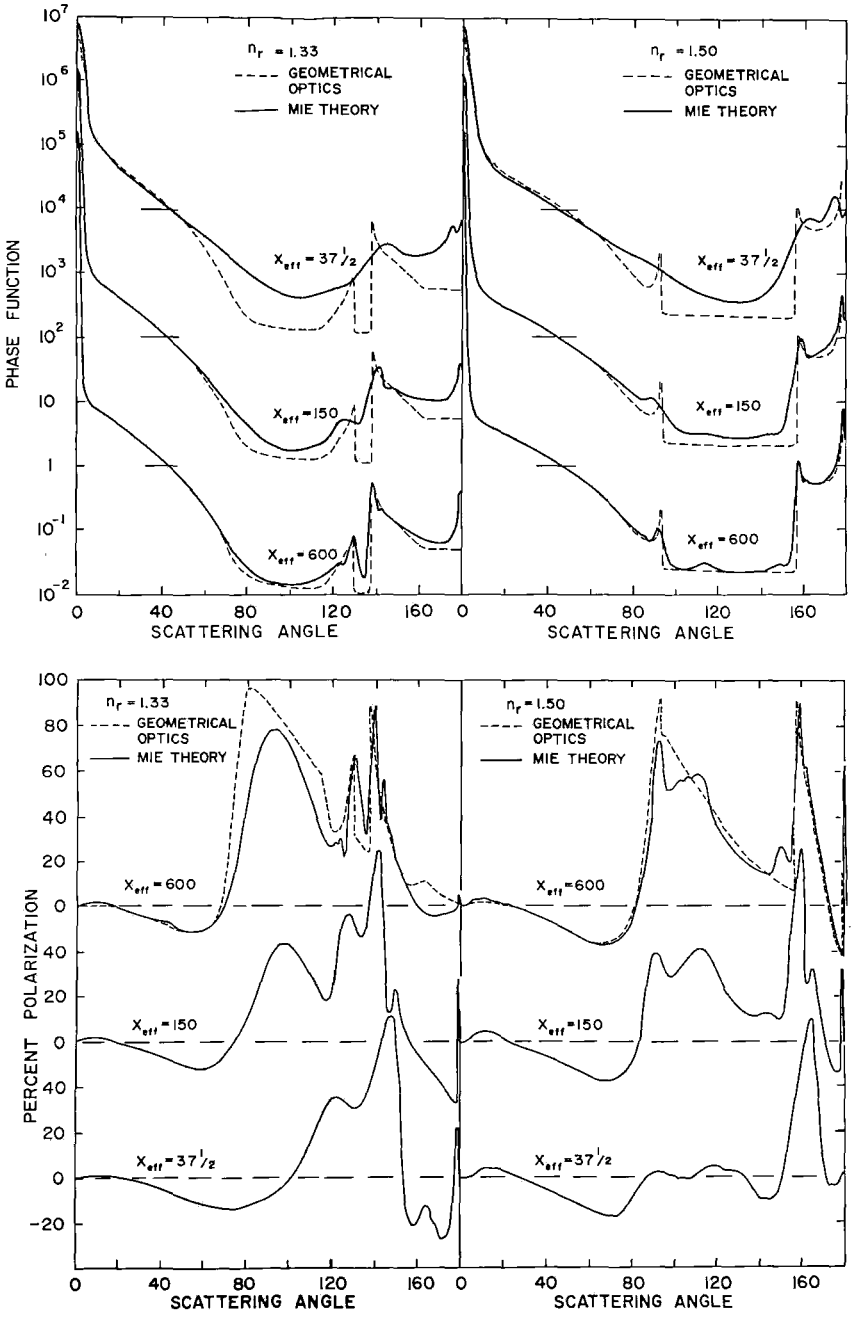



**Fig. 4.14** Comparison of geometrical optics with Mie theory. The phase functions (upper figure) and polarization (lower) are computed for single scattering by spheres with a distribution of sizes given by $N(x) = x^6 \exp(-9x/x_{\text{eff}})$, and real indices of refraction $n_r$. For the phase function the scale applies to the curves with $x_{\text{eff}} = 600$; the other curves are successively displaced upward by factors of $10^2$. [After HANSEN and TRAVIS (1974).]

Rainbows in general arise from a concentration of internally reflected rays at a maximum or minimum angle of deviation. Thus for $n = \frac{4}{3}$ (appropriate for water) as the angle $\gamma$ in Fig. 4.13 varies from normal ($\gamma = 90°$) to grazing incidence, the $l = 3$ ray decreases from a scattering angle $\alpha = 180°$ (back-scattering) to 137°, the angle of minimum deviation, and then increases. The concentration of scattering around $\alpha = 137°$ is the primary rainbow. Spherical scatterers tend to focus reflected rays at $\alpha \sim 180°$. This phenomenon is the *glory*, which is often seen from aircraft flying over sunlit clouds and appears as a brightening of the clouds near the edge of the aircraft shadow. Occasionally colored rings are visible.

Precise values for the phase function and polarization generally require the use of Mie computations. However, a number of features, such as the rainbow, that do not involve interference phenomena are adequately represented by geometrical optics. This is especially true when the particles are substantially larger than the wavelength of light and also when the dispersion in particle sizes is sufficiently large to wash out the interference effects. Figure 4.14 shows an example of calculations with Mie theory compared with geometrical optics.

### 4.3.3 Photometry of Planets

The scattering phase functions in common use because of their elementary nature are (a) isotropic, $p(\cos\Theta) = \tilde{\omega}$; (b) the Rayleigh phase function,

$$p(\cos\Theta) = \tfrac{3}{4}(1 + \cos^2\Theta) \tag{4.3.2}$$

which takes no account of the polarization produced by the scattering; (c) Rayleigh scattering,

$$p_l(\cos\Theta) = \tfrac{3}{4}\cos^2\Theta, \qquad p_r(\cos\Theta) = \tfrac{3}{4} \tag{4.3.3}$$

which treats the electric vectors in the plane of scattering ($E_l$) and perpendicular to it ($E_r$) separately, since the scattered light is polarized; and (d) the first-order anisotropic phase function,

$$p(\cos\Theta) = \tilde{\omega}(1 + a\cos\Theta), \qquad -1 < a < 1 \tag{4.3.4}$$

From solutions of the transfer equation, one can compute for any of these functions the total flux reflected by the planet toward the Earth as a function of planetary phase $\Phi$. Integrating the reflected intensity over co-latitude $\theta$ and longitude $\phi$ (both co-ordinates referred to the Earth-planet-sun plane rather than to a rotational axis), we have the flux referred to unit planetary radius,

$$j(\Phi) = \int_{\Phi-\pi/2}^{\pi/2} d\phi \int_0^{\pi} d\theta \sin\theta\mu I(\mu, \psi; \mu_0) \tag{4.3.5}$$

where $\mu = \sin\theta\cos\phi$ is the factor projecting the spherical area onto the disk and $\psi$ is the local azimuthal angle that enters for anisotropic scattering.

Knowing $j(\Phi)$ we can integrate it over a sphere to find the total radiant flux to all space. The ratio of the total reflected flux in all directions to the incident solar flux received by a sphere of unit radius is the *Bond albedo*:

$$\begin{aligned}\Lambda_B &\equiv \frac{2\pi \int_0^{\pi} j(\Phi)\sin\Phi\, d\Phi}{\pi\mathscr{F}\int_0^{2\pi}\int_0^{1}\mu_0\, d\mu_0\, d\psi_0} \\ &= \left[\frac{j(0)}{\pi\mathscr{F}}\right]\left[2\int_0^{\pi}\frac{j(\Phi)}{j(0)}\sin\Phi\, d\Phi\right] \\ &= \lambda(0)q \qquad (4.3.6)\end{aligned}$$

where $\lambda(0)$ and $q$ represent the two expressions in square brackets, respectively.

The factor $\lambda(0)$ is the *geometric albedo*; for a highly reflecting atmosphere $\lambda(0)$ is in the range 0.65–0.80, depending on the phase function $p(\cos\Theta)$ (see Problem 4.10).

The *phase integral* $q$ can be obtained theoretically from (4.3.5). However, ground-based measurement of the phase variation $j(\Phi)$ is difficult, and for the outer planets is limited to rather small angles. Thus spacecraft measurements are necessary to obtain reliable values of $j(\Phi)$, and, in turn, the Bond albedo, which is necessary for estimating the thermal balance of a planet (see Section 1.2).

Center-to-limb photometry is also capable in principle of yielding the scattering phase function, which is of value as an indicator of the size of scattering particles as well as necessary to derive $\Lambda_B$. The difficulty in practice is that the information from near the limb of the planet is hard to obtain with precision, and the data do not therefore give enough information to discriminate among a variety of possible forms of $p(\cos\Theta)$.

The thermal infrared emission of Jupiter has been investigated from the ground (mainly in the 2.8–13 $\mu$m wavelength region), from high-altitude aircraft (20–500 $\mu$m), and from Pioneers 10 and 11 (20 and 45 $\mu$m). The effective radiating temperature of the planet is $T_e = 134°\text{K}$ according to Earth-based data and 125°K according to the Pioneer measurements. Both values are significantly higher than can be accounted for by absorption and reradiation of sunlight. The planet emits about twice the energy it absorbs and the conclusion must be that Jupiter generates this energy from its interior by gravitational contraction, much as does a newly forming star.

### 4.3.4 Polarimetry of Venus

Because the electric (and magnetic) vibrations of light can have preferred directions, radiation toward a given direction is not a simple scalar with

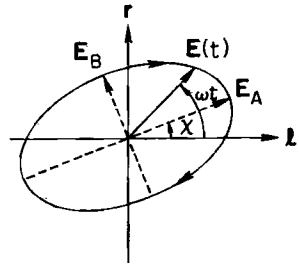



**Fig. 4.15** The polarization ellipse in the plane perpendicular to wave propagation. The axes $l$ and $r$ are fixed. The angle $\chi$ gives the direction of maximum linear polarization. The vector $\mathbf{E}(t)$ is the resultant of the electric oscillations along the major and minor axes. The arrows on the ellipse give the convention for right-hand polarization when the wave propagation is into the paper.

magnitude $I$. In general, a radiation field is the sum of an unpolarized and an elliptically polarized component. Elliptical polarization has a phase relationship between the electrical oscillations along the major and minor axes of the ellipse. Four quantities are required to define the field. They are the intensities $(I_l, I_r)$ polarized in specified directions (cf. Fig. 4.15), the position $\chi$ of the plane of polarization (the major axis of the ellipse), and the phase lag $\delta$ between the electrical oscillations $\mathbf{E}_l(t)$ and $\mathbf{E}_r(t)$. (When the major and minor axes of the polarization ellipse are not along the reference axes $l$ and $r$, this phase difference will not generally be $\pi/2$.)

In place of two intensities $(I_l, I_r)$ we could use total intensity $(I = I_l + I_r)$ and the difference $Q \equiv I_l - I_r$ or the polarization $P = Q/I$. There are major conveniences in handling multiple scattering of polarized light if the angles $(\chi, \delta)$ are also replaced by quantities having intensity units. Thus we shall develop the four *Stokes parameters*:

$$\begin{aligned} I &= I_l + I_r \\ Q &= I_l - I_r \\ U &= Q \tan 2\chi \\ V &= E_l^{(0)} E_r^{(0)} \sin \delta \end{aligned} \tag{4.3.7}$$

for amplitudes $E^{(0)}$. Defining ellipticity as

$$\tan \beta = E_B^{(0)} / E_A^{(0)} \tag{4.3.8}$$

we shall find that the plane of polarization $\chi$, phase lag $\delta$, and ellipticity angle $\beta$ are all related. This interrelationship makes it possible to work with $\beta$ in (4.3.7) in lieu of $\delta$. If we express the plane of polarization by

$$\begin{aligned} E_A(t) &= E^{(0)} \cos \beta \sin \omega t \\ E_B(t) &= E^{(0)} \sin \beta \cos \omega t \end{aligned} \tag{4.3.9}$$

we can obtain amplitudes in the longitudinal and perpendicular coordinate system in the form

$$\begin{aligned} E_l &= E_A \cos \chi + E_B \sin \chi \\ E_r &= E_A \sin \chi - E_B \cos \chi \end{aligned} \tag{4.3.10}$$

Substituting (4.3.9) and averaging over an oscillation yields

$$\begin{aligned} I_l &\equiv \langle E_l{}^2 \rangle = \tfrac{1}{2}E^{(0)2}(\cos^2\beta\cos^2\chi + \sin^2\beta\sin^2\chi) \\ I_r &\equiv \langle E_r{}^2 \rangle = \tfrac{1}{2}E^{(0)2}(\cos^2\beta\sin^2\chi + \sin^2\beta\cos^2\chi) \end{aligned} \tag{4.3.11}$$

With $I = \frac{1}{2}E^{(0)2}$, we readily verify the first Stokes parameter (4.3.7). The second takes the form

$$Q \equiv I_l - I_r = I\cos 2\chi\cos 2\beta \tag{4.3.12}$$

If on the $l$ and $r$ axes the phases of oscillation are correlated, we may write those oscillations as

$$\begin{aligned} E_l(t) &= E_l^{(0)}\sin(\omega t - \varepsilon_l) \\ E_r(t) &= E_r^{(0)}\sin(\omega t - \varepsilon_r) \end{aligned} \tag{4.3.13}$$

where $\delta = \varepsilon_r - \varepsilon_l$. One may show (Problem 4.11) that this representation is consistent with (4.3.10) if

$$\begin{aligned} \tan\varepsilon_l &= -\tan\beta\tan\chi \\ \tan\varepsilon_r &= \tan\beta\,\mathrm{ctn}\,\chi \end{aligned} \tag{4.3.14}$$

and

$$E_l^{(0)}E_r^{(0)}\cos\delta = \tfrac{1}{2}\mathrm{E}^{(0)2}\sin 2\chi\cos 2\beta \tag{4.3.15}$$

Thus from (4.3.12) the third Stokes parameter is

$$\begin{aligned} U &\equiv Q\tan 2\chi = I\sin 2\chi\cos 2\beta \\ &= E_l^{(0)}E_r^{(0)}\cos\delta \end{aligned} \tag{4.3.16}$$

and the fourth is

$$\begin{aligned} V &\equiv E_l^{(0)}E_r^{(0)}\sin\delta = I\sin 2\beta \\ &= Q\tan 2\beta\sec 2\chi \end{aligned} \tag{4.3.17}$$

Conversely, in terms of the four parameters $I$, $Q$, $U$, and $V$, we may obtain the plane of polarization from

$$\tan 2\chi = U/Q \tag{4.3.18}$$

the polarization from

$$P = Q/I \tag{4.3.19}$$

the ellipticity from

$$\sin 2\beta = \frac{V}{(Q^2 + U^2 + V^2)^{1/2}} \tag{4.3.20}$$

and the phase lag from

$$\tan\delta = \tan 2\beta\sec 2\chi \tag{4.3.21}$$

The polarization of a planet for scattered sunlight can thus be treated as we would treat scattering of the intensity without polarization, except that all four intensity components of the "vector intensity,"

$$\begin{aligned}\mathbf{I} &= (I, Q, U, V) \\ &= (I_l, I_r, U, V)\end{aligned} \tag{4.3.22}$$

must be treated separately. For example, the scattering through angle $\Theta$ for Rayleigh scattering is given by

$$\left(\frac{d\omega'}{4\pi}\,\sigma\right)\mathbf{R}\mathbf{I}\,d\omega \tag{4.3.23}$$

where the scattering matrix is

$$\mathbf{R} = \tfrac{3}{2}\begin{pmatrix} \cos^2\Theta & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\Theta & 0 \\ 0 & 0 & 0 & \cos\Theta \end{pmatrix} \tag{4.3.24}$$

This matrix replaces the partial phase functions of (4.3.3). For incident natural light, $I_{\|} = I_{\perp} = \frac{1}{2}I$ and $U = V = 0$, and the matrix reduces to the simpler case. But for multiple scattering the complete matrix is necessary to follow the polarization vector.

Venus provides a particularly illuminating example of the information contained in polarized light and the numerical difficulty in retrieving this information. Until the 1920s it was debatable whether sunlight reflected by Venus was in fact polarized. However, when B. Lyot developed a precision polarimeter capable of accuracies of about 0.1 percent, it was revealed that Venus has well-defined polarization features as a function of phase angle. Moreover, the fortuitous location of Venus as an interior planet provides a nearly complete phase angle coverage as viewed from Earth. But it was not until recently that sufficiently large computers became available to perform the Mie-scattering computations to adequately model the polarization by the Venus cloud particles. The results (see Fig. 4.16) showed conclusively that: (1) the cloud particles in the visible cloud-top were spherical; (2) the effective radius of the particles was $\sim 1\ \mu$m; (3) the particle size distribution was very narrow with an effective variance of only $\sim 0.07$; (4) the refractive index decreased from 1.46 at $\lambda = 0.365\ \mu$m to 1.43 at $\lambda = 0.99\ \mu$m; and (5) that the cloud optical thickness of unity occurs at a pressure of $\sim 50$ mb. The precise limits on the refractive index had the effect of eliminating all previously suggested cloud particle compositions except for concentrated sulfuric acid, $H_2SO_4 \cdot (H_2O)$, which provides a good fit to the observed polarization.

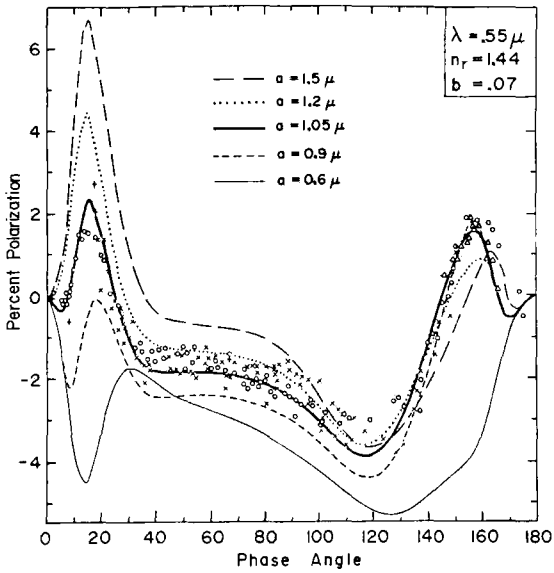



**Fig. 4.16** Polarization of Venus. Observational points and calculations centered at $\lambda = 0.55$ $\mu$m. All curves for spheres with refractive index of 1.44. The different curves show effect of radius of sphere. Distribution of sizes is $n(r) = \text{const}\ r^{11.3}e^{-r/0.07a}$. Rayleigh fractional contribution to scattering is about 0.009. [After HANSEN and HOVENIER (1974).]

The successful analysis of Venus polarization, however, is not likely to be repeated for other planets in the solar system. The precision and uniqueness of the results obtained for Venus depends largely on the fact that the cloud particles proved to be spherical—as indicated by the characteristic features such as the rainbow and the anomalous diffraction and their wavelength dependence. On planets such as Mars and Jupiter we would expect to find irregular dust and ice particles rather than spherical (liquid) particles. This irregularity would have the effect of introducing additional parameters with compensating trade-offs, making the analysis much more complicated and less unique.

## BIBLIOGRAPHICAL NOTES

**Section 4.1 Radiative Transfer in an Optically Thick Atmosphere**

The basic text is

CHANDRASEKHAR, S. (1950), "Radiative Transfer," Clarendon Press, Oxford, (reprinted by Dover, New York, 1960).

This classic treatise develops solutions with exact $H$-functions for simple phase functions, as well as solutions in the $n$th approximation. Finite atmospheres are also treated there with the exact formalism and solved in terms of so-called $X$- and $Y$-functions.

A review and extensive bibliography of theoretical calculations for scattering with a variety of phase functions is given in

IRVINE, W. M. (1975), Multiple scattering in planetary atmospheres, *Icarus* **25**, 175–204.

The similarity relations for scattering by atmospheres with different phase functions, albedos, and thicknesses are discussed in

VAN DE HULST, H. C. and GROSSMAN, K. (1968), Multiple light scattering in planetary atmospheres, *in* "The Atmospheres of Venus and Mars," (J. C. Brandt and M. B. McElroy, eds.), pp. 35–55, Gordon and Breach, New York.

For the accurate solution of realistic problems of planetary atmospheres, with scattering phase functions such as (4.1.44), the function introduced by L. G. HENYEY and J. L. GREENSTEIN [(1941), *Astrophys. J.* **93**, 70] it is necessary to employ numerical techniques. Computational methods in wide use are the *doubling method* and the related *adding method* in which the single-scattering properties of an optically thin layer are used to derive the cumulative scattering of a large number of layers (see Section 4.1.5, due to A. A. LACIS). The concept originated with

STOKES, G. G. (1862), On the intensity of the light reflected from or transmitted through a pile of plates, *Proc. Roy. Soc.* (*London*) **11**, 545–556,

and was introduced to atmospheric physics by

BELLMAN, R. E.; KALABA, R.; and UENO, S. (1963), Invariant imbedding and diffuse reflection from a two-dimensional flat layer, *Icarus* **1**, 297–303;
TWOMEY, S.; JACOBOWITZ, H.; and HOWELL, H. B. (1966), Matrix methods for multiple-scattering problems, *J. Atmos. Sci.* **23**, 289–296;
VAN DE HULST, H. C. and GROSSMAN, K. (1968), Multiple light scattering in planetary atmospheres, *in* "The Atmospheres of Venus and Mars," (J. C. Brandt and M. B. McElroy, eds.), pp. 35–55, Gordon and Breach, New York.

The numerical technique has been developed for planetary atmospheres by

HANSEN, J. E. (1969), Radiative transfer by doubling very thin layers, *Astrophys. J.* **155**, 565–573;
HANSEN, J. E. (1971), Multiple scattering of polarized light in planetary atmospheres, Part I. The doubling method, *J. Atmos. Sci.* **28**, 120–125,
HANSEN, J. E. and TRAVIS, L. D. (1974), Light scattering in planetary atmospheres, *Space Sci. Rev.* **16**, 527–610.

Because clouds may be optically thin over distances of a scale height or because two or more distinct cloud levels may be present, some numerical models have been developed for vertically inhomogeneous atmospheres—meaning that the relative absorption and scattering varies with depth. Among these models are

DANIELSON, R. E. and TOMASKO, M. G. (1969), A two-layer model of the Jovian clouds, *J. Atmos. Sci.* **26**, 889–897;
COCHRAN, W. D. (1977), Jupiter: An inhomogeneous atmospheric model analysis of spatial variations of the $H_2$(4-O)S(1) line, *Icarus* **31**, 325–347;
DANIELSON, R. E.; COCHRAN, W. D.; WANNIER, P. G.; and LIGHT, E. S. (1977), A saturation model of the atmosphere of Uranus, *Icarus* **31**, 97–109.

**Section 4.2 Spectroscopy**

The solution (4.58) for the equivalent width of an absorption line with Lorentz broadening in a nonscattering atmosphere was first obtained by

LADENBERG, R. and REICHE, F. (1913), Über Selektive Absorption, *Ann. Physik* **42**, 181–209.

For a barometric atmosphere the solution (4.2.9) was obtained by

STRONG, J. and PLASS, G. N. (1950), The effect of pressure broadening of spectral lines on atmospheric temperature, *Astrophys. J.* **112**, 365–379.
[The strong-line limit (4.2.10) was later given by CURTIS, A. R. (1952), *Q. J. Roy. Met. Soc.* **78**, and by GODSON, W. L. (1953), *ibid.* **79**, 367.]

The complications inherent in both the spectroscopic and the photometric techniques used to estimate the total pressure of the Martian atmosphere from ground-based observations are reviewed in

CHAMBERLAIN, J. W. and HUNTEN, D. M. (1965), Pressure and $CO_2$ content of the Martian atmosphere: A critical discussion, *Rev. Geophys.* **3**, 299–317.

With refined observations of $CO_2$, relative altitudes of surface features on Mars were first obtained by

BELTON, M. J. S. and HUNTEN, D. M. (1971), The distribution of $CO_2$ on Mars: Spectroscopic determination of surface topography, *Icarus* 15, 204–232.

It was first noted that a weak line formed in an atmosphere scattering with $\tilde{\omega}_\epsilon \approx 1$ should vary as the square root of the absorption coefficient (4.2.28) in

VAN DE HULST, H. C. (1949), Scattering in the atmospheres of the Earth and the planets, *in* "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), pp. 49–111, Univ. Chicago Press, Chicago [2nd ed., 1952].

The phase variation and spectroscopic temperatures of a long series of near-infrared $CO_2$ spectra of Venus (obtained by Kuiper) were interpreted in terms of multiple scattering in a hazy atmosphere by

CHAMBERLAIN, J. W. and KUIPER, G. P. (1956), Rotational temperature and phase variation of the carbon dioxide bands of Venus, *Astrophys. J.* **124**, 399–405.

The curve-of-growth theory presented here follows

CHAMBERLAIN, J. W. (1970), Behavior of absorption lines in a hazy planetary atmosphere, *Astrophys. J.* **159**, 137–158.

Carbon dioxide was found in the spectrum of Venus with infrared-sensitive photographic plates that recorded the same bands shown in Fig. 4.8 by

ADAMS, W. S. and DUNHAM, T., JR. (1932), Absorption bands in the infra-red spectrum of Venus, *Publ. Astr. Soc. Pacific* **44**, 243–245.

A number of planetary spectra are reproduced and their interpretation discussed by

HUNTEN, D. M. (1971), Composition and structure of planetary atmospheres, *Space Sci. Rev.* **12**, 539–599.

The highest resolution planetary spectra are those obtained with a Michelson interferometer by the technique commonly called *Fourier spectroscopy*. In particular, the spectra of the brighter planets are recorded in

CONNES, P.; CONNES, J.; and MAILLARD, J. P. (1969), "Atlas des spectres infrarouges de Vénus, Mars, Jupiter et Saturn," Centre National dela Recherche Scientifique, Paris.

The atmospheric abundances in Table 4.1 are based on tabulations and summaries in

ALLEN, C. W. (1973), "Astrophysical Quantities," 3d. ed., Athlone Press, London, p. 119;
HUNTEN, D. M. (1971), op. cit.;
MCELROY, M. B.; KONG, T. Y.; and YUNG, Y. L. (1976), Composition and structure of the Martian upper atmosphere: Analysis of results from Viking, *Science* **194**, 1295–1298;

RIDGEWAY, S. T.; LARSON, H. P.; and FINK, U. (1976), The infrared spectrum of Jupiter, *in* "Jupiter," (T. Gehrels, ed.), pp. 384–417, Univ. Arizona Press, Tucson.

The problems of interpreting Jupiter's spectrum are treated by

TEIFEL, V. G. (1976), Morphology of molecular absorption on the disk of Jupiter, *in* "Jupiter," (T. Gehrels, ed.), pp. 441–485, Univ. Arizona Press, Tucson,

WALLACE, L. (1976), The thermal structure of Jupiter in the stratosphere and upper troposphere, *in* "Jupiter," (T. Gehrels, ed.), pp. 284–303, Univ. Arizona Press, Tucson.

WALLACE, L. and HUNTEN, D. M. (1978), The Jovian spectrum in the region 0.4–1.1 $\mu$: The C/H ratio, *Rev. Geophys. Space Phys.* **16** (No. 3, Aug.), *in press.*

The phenomenon of collisional narrowing of the Doppler profile was developed by

DICKE, R. H. (1953), The effects of collisions on the Doppler width of spectral lines, *Phys. Rev.* **89**, 472–473

and applied to a curve of growth for $H_2$ in a scattering model of Jupiter's atmosphere by

FINK, U. and BELTON, M. J. S. (1969), Collision-narrowed curves of growth for $H_2$ applied to new photoelectric observations of Jupiter, *J. Atmos. Sci.* **26**, 952–962.

The quadrupole and pressure-induced dipole spectra of $H_2$ are discussed in

HERZBERG, G. (1952), Laboratory absorption spectra obtained with long paths, *in* "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), pp. 406–416, Univ. Chicago Press, Chicago, 2nd ed.,

BIRNBAUM, A. and POLL, J. D. (1969), Quadrupole transitions in the $H_2$, HD, and $D_2$ molecules, *J. Atmos. Sci.* **26**, 943–945.

**Section 4.3 Photometry and Polarimetry**

Scattering by particles is most thoroughly treated in VAN DE HULST, H. C. (1957), "Light Scattering by Small Particles," Wiley, New York.

On a more elementary level is

MCCARTNEY, E. J. (1976), "Optics of the Atmosphere: Scattering by Molecules and Particles," Wiley, New York.

Mie scattering by an atmosphere in which the particles are not of uniform size is treated in

DEIRMENDJIAN, D. (1969), "Electromagnetic Scattering on Spherical Polydispersions," American Elsevier, New York,

WICKRAMASINGHE, N. C. (1973), "Light Scattering Functions with Applications in Astronomy," Wiley, New York.

The fundamentals of planetary photometry are thoroughly reviewed in

HARRIS, D. L. (1961), Photometry and colorimetry of planets and satellites, *in* "Planets and Satellites," (G. P. Kuiper and B. M. Middlehurst, eds.), pp. 272–342, Univ. Chicago Press, Chicago.

That the "surface" temperatures of Jupiter and Saturn were extremely low—comparable to the values that would be maintained by solar radiation—was first shown (from a reduction of Lowell Observatory radiometric measurements) by

MENZEL, D. H. (1923), Water-cell transmissions and planetary temperatures, *Astrophys. J.* **58**, 65–74.

On the basis of these results Menzel later gave the first convincing arguments that the major planets are composed largely of hydrogen (rather than having terrestrial compositions with hot, gaseous interiors to account for their low densities) in

MENZEL, D. H. (1930), Hydrogen abundance and the constitution of the giant planets, *Publ. Astr. Soc. Pacific* **42**, 228–232.

Thermal infrared observations from Pioneer, their interpretation in terms of an effective temperature, and references to earlier Earth-based measurements are given in

INGERSOLL, A. P.; MÜNCH, G.; NEUGEBAUER, G.; and ORTON, G. S. (1976), Results of the infrared radiometer experiment on Pioneers 10 and 11, *in* "Jupiter," (T. Gehrels, ed.), pp. 197–205, Univ. Arizona Press, Tucson.

That Jupiter is heated from below was inferred from the cloud activity and potential energy considerations by

KUIPER, G. P. (1952), Planetary atmospheres and their origin, *in* "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), pp. 306–405, Univ. Chicago Press, Chicago, 2nd ed. (cf. pp. 326, 376).

Kuiper was reluctant to attach too much importance to radiometric temperatures measured in the 8–12 $\mu$m window. That a genuine discrepancy with the solar-equilibrium temperature existed and must indicate internal heating was first emphasized by

ÖPIK, E. J. (1962), Jupiter: Chemical composition, structure, and origin of a giant planet, *Icarus* **1**, 200–257.

High-resolution photometry of Venus from Mariner 10, with emphasis on the mysterious dark markings that appear in the ultraviolet wavelengths, is discussed by

HAPKE, B. (1976), Photometry of Venus from Mariner 10, *J. Atmos. Sci.* **33**, 1803–1815.

The formulation of radiative transfer in terms of the parameters originally devised in 1852 by G. G. STOKES is due to

CHANDRASEKHAR (1950), op. cit. pp. 24–35.

Scattering by planetary atmospheres, with emphasis on the polarization produced by Mie scattering, is reviewed in

HANSEN, J. E. and TRAVIS, L. D. (1974), op. cit.

Multiple-scattering calculations, compared with observations, established that the index of refraction for the Venus aerosols is $1.46 \pm 0.015$ in the blue and $1.43 \pm 0.015$ in the near infrared and that the sizes of the droplets are concentrated around 1.05 $\mu$m radius. This polarization analysis is due to

HANSEN, J. E. and HOVENIER, J. W. (1974), Interpretation of the polarization of Venus, *J. Atmosph. Sci.* **31**, 1137–1160.

On the basis of this analysis, sulfuric acid, $H_2SO_4$, was proposed as the source of Venus' visual clouds independently by

SILL, G. T. (1972), Sulfuric acid in the Venus clouds, *Com. Lunar Planet. Lab.* Univ. Arizona, Tucson, **9**, No. 171, 191–198,

YOUNG, A. T. (1973), Are the clouds of Venus sulfuric acid? *Icarus* **18**, 564–582.

The power of polarization studies of scattering atmospheres was first demonstrated by

LYOT, B. (1929), Recherches sur la polarisation de la lumière des planètes et de quelques substances terrestres. *Ann Observ. Paris (Meudon)* **8**, 161 pp. [English translation, NASA TT T-187 (1964)].

## PROBLEMS

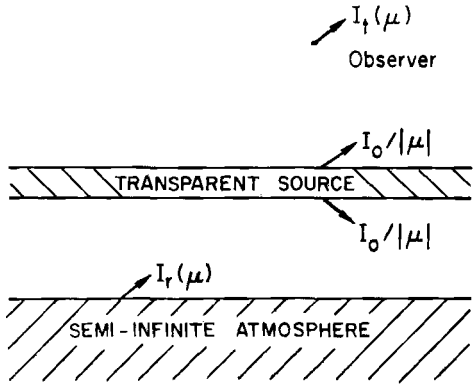
**4.1** *Gaussian quadrature.* (a) Find the division points and weighting factors for a three-point quadrature. Show that the three unused equations (4.1.12) satisfy the $a_j$'s found from the other three. (b) Use the Gaussian quadrature to evaluate $\int_{-1}^{1} e^x\,dx$ and compare the result with the exact value and with the approximate value obtained from a series expansion of $e^x$ to terms in $x^5$.

**4.2** *Diffuse reflection.* A semi-infinite atmosphere is illuminated by a parallel beam of light with incident flux $\pi\mathscr{F}$ (measured across an area normal to the beam). The beam is incident from a zenith angle $\cos^{-1}\mu_0$. (a) What is the albedo of the atmosphere for this incident radiation in the first approximation when the scattering albedo is $\tilde{\omega} = 1$? (b) Evaluate this albedo for $\mu_0 = 1, 1/\sqrt{3}, \frac{1}{2}$, and 0.1 and comment on what these results connote about the accuracy of the first approximation.

**4.3** *Physical meaning of the H-functions* [after CHANDRASEKHAR (1950)]. A layer that emits airglow in all directions, with an angular distribution of intensity $I_s(\mu) = I_0/|\mu|$, lies above the semi-infinite, isotropically scattering atmosphere. An observer in space sees a total intensity $I_t(0, +\mu) = I_s(\mu) + I_r(0, +\mu)$, where $I_r$ is the intensity diffusely reflected by the atmosphere. Solve the transfer equation in the first approximation for $I_r$ and show that the relative enhancement to the observed intensity provided by the atmosphere is

$$I_t(+\mu)/I_s(\mu) = H_1(\mu)$$

where $H_1$ is given by (4.1.31). (Note: This result actually holds to all orders of approximation. Hence the reflectivity of an atmosphere, for the source distribution assumed here, provides a physical meaning for the Chandrasekhar $H$-functions.)




**4.4** *Atmospheric albedo (I).* A semi-infinite atmosphere is illuminated from above by isotropic radiation of total flux $\pi\mathscr{F}$ (measured across a surface parallel to the top of the atmosphere). (a) Solve the transfer equation in the first approximation and apply the boundary conditions to fix the two integration constants. (b) Find the source function and the upward radiation field $I(\tau, +\mu)$. (c) What is the albedo $\Lambda$ of the atmosphere for this incident radiation field? Evaluate $\Lambda$ to order $(1 - \tilde{\omega})$ for the case when $(1 - \tilde{\omega}) \ll 1$, and show that $\Lambda$ is almost, but not quite, the same as (4.1.34).

**4.5** *Atmospheric albedo (II).* Show that the albedo $\Lambda$ in Problem 4.4 can be written as

$$\Lambda = 1 - 2\alpha_1(1 - \tilde{\omega})^{1/2}$$

where $\alpha_n$ is the $n$th moment of the $H$-functions. In the first approximation

$$\alpha_n = \int_0^1 \mu^n H_1(\mu)\, d\mu$$

where $H_1$ is given by (4.1.31).

**4.6** *Anisotropic scattering.* (a) Show that the transfer equation with $I(\tau, \mu)$ divided into azimuth-dependent and -independent parts, (4.1.36), may be written as the two equations (4.1.37) and (4.1.38). (b) Solve the transfer equation (4.1.2) in the optically thin limit (where only single scattering need be considered) and obtain expressions (4.1.51) and (4.1.52) for the reflected and transmitted intensities.

**4.7** *Effective damping constant.* Consider a barometric (constant temperature) atmosphere in which the absorption cross section of a well-mixed absorber everywhere has a Lorentz profile (4.2.3) with $\Gamma$ proportional to the local pressure. The continuum is optically thin. (a) Show that the effective optical thickness as seen from outside the atmosphere is

$$\eta\tau_\nu = \frac{2\eta S}{\Gamma_0} \ln\left[1 + \left(\frac{\Gamma_0/4\pi}{\nu - \nu_0}\right)^2\right]$$

where $\Gamma_0$ is the value at the ground, and compare with $\eta\tau_\nu$ for an atmosphere of constant density. (b) In the manner of Eq. (4.2.4) find the frequency displacement where the optical thickness is unity and show that for strongly absorbing lines the equivalent width is the same as for a constant-density atmosphere, except that $\Gamma$ is replaced with $\Gamma_0/2$.

**4.8** *Air-mass factor.* (a) Show that $\langle\eta\rangle$, as given by the first equality of (4.2.31), has the same phase variation, relative to the value at full phase, for any zonal strip (i.e., a band with boundaries of constant latitude). (b) Show that within any zonal strip $\eta$ is a maximum where $\mu = \mu_0$.

**4.9** *Spectra with anisotropic scattering.* Use the first approximation to compare the variation of a weak line from center to limb for a planet at full phase $\Phi = 180°$ when $\tilde{\omega}_e = 1.00$ for forward scattering ($a = 1$) and backward scattering ($a = -1$). Explain why the line absorption is stronger for $a = +1$ in the center of the image.

**4.10** *Bond albedo.* In the case of a planet reflecting as a Lambertian surface (i.e., isotropically) with a ground albedo $\Lambda$, show that the geometric albedo is $\lambda(0) = \frac{2}{3}\Lambda$, the phase integral is $q = \frac{3}{2}$, and hence the Bond albedo is $\Lambda_B = \Lambda$.

**4.11** *Stokes parameters.* (a) Show that phase lag $\delta$, polarization angle $\chi$, and ellipticity angle $\beta$ are related by (4.3.15) and that $U$ is given by (4.3.16). (b) Derive (4.3.17).

# *Chapter 5*
# IONOSPHERES

## 5.1 Formation of Ionospheric Regions

Discovery of the ionosphere has an interesting history that is summarized on pp. 203–4. The Earth's ionosphere is divided into several regions designated by the letters D, E, and F, the latter being subdivided into F1 and F2. Historically, the division arose from the successive plateaus of electron density $N_e$ observed on records of the time delay (i.e., *virtual height*) of radio reflections as the transmitted signal was swept through frequency; see Fig. 5.1. (The *critical frequency* at which reflection occurs varies as $N_e^{1/2}$. Thus higher frequencies penetrate farther into the ionosphere and are reflected by higher $N_e$; cf. Section 5.2.) The E "layer" was the first to be detected and was so labeled as being the atmospheric layer reflecting the E vector of the radio signal. Later the lower D and higher F layers were discovered.

Distinct ionospheric regions develop because (a) the solar spectrum deposits its energy at various heights depending on the absorption characteristics of the atmosphere, (b) the physics of recombination depends on the density, and (c) the composition of the atmosphere changes with height. Thus the four main ionospheric regions can be associated with different governing physical processes, and this physics (rather than simple height differentiation) is the basis for labeling an ionospheric region on another planet as a D, E, F1, or F2 region.

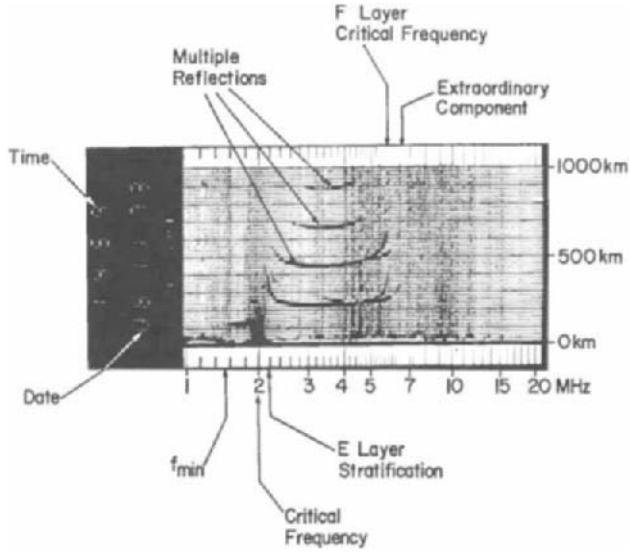



**Fig. 5.1** Ionogram, giving the virtual height of reflection *versus* the transmitted frequency ($f$). As increasing frequency approaches the critical frequency for F-region reflection, the signal rises steeply on the record and then disappears as reflection becomes impossible. This record was obtained at Slough in England U.T. 1459 hours, Nov. 6, 1973, and thus indicates daytime conditions. Below about 1.6 MHz (in the AM radio band) D-region absorption prevents the signal from being reflected. The "ordinary" and "extraordinary" polarization components transmit differently because of differing refractive indices. The echo reflections at about two, three, and four times the height of the main reflections arise from multiple reflections between ionosphere and ground. (Courtesy, H. RISHBETH and Science Research Council, Appleton Laboratory, Slough, United Kingdom.)

Even though ionospheric recombination does not necessarily vary as the square of the electron density $N_e$, that dependence is common under uncomplicated circumstances. Hence it is phenomenologically useful to consider recombination as occurring at a rate described by an *effective recombination coefficient* $\alpha_{\text{eff}}$ so that $N_e$ can be described by

$$\frac{dN_e}{dt} = q - \alpha_{\text{eff}} N_e^2 \tag{5.1.1}$$

where $q$ is the production rate of free electrons. Table 5.1 gives a thumbnail sketch of the Earth's ionospheric regions with characteristic values, determined observationally, of $\alpha_{\text{eff}}$. Tables 5.2 and 5.3 show the frequency ranges used for different communication purposes, which are largely dictated by the properties of the ionosphere. In the following sections we will develop the basic physical processes that give the unique characteristics to each ionospheric region.

TABLE 5.1 *The Ionosphere*

| Region | Nominal height of layer peak (km) | $N_e^{(max)}$ ($cm^{-3}$) | $\alpha_{eff}$ ($cm^3$/sec) | Ion production | Recombination |
|---|---|---|---|---|---|
| D | 90<br>Lower following solar flare | $1.5 \times 10^4$ (noon);<br>absent at night | $3 \times 10^{-8}$ | Ionization by solar X-rays, or Ly $\alpha$ ionization of NO. Enhanced ionization following solar flares due to X-ray ionization of all species. Electron attachment to O and $O_2$ forms negative ions; ratio of negative ions to electrons increases with depth and at night. | Electrons form negative ions, which are destroyed by photodetachment (daytime only), associative detachment ($O + O^- \rightarrow O_2 + e$), and mutual neutralization ($O^- + X^+ \rightarrow O + X$). |
| E | 110 | $1.5 \times 10^5$ (noon);<br>$< 1 \times 10^4$ (night) | $10^{-8}$ | Ionization of $O_2$ may occur directly by absorption in the first ionization continuum ($h\nu > 12.0$ eV). Coronal X-rays also contribute, ionizing O, $O_2$, and $N_2$. Nighttime E and sporadic E (thin patches of extra ionization) are due to electron and meteor bombardment. Some $E_s$ radio reflections may be due to turbulence in normal E layer. | Dissociative recombination $O_2^+ + e \rightarrow O + O$ and $NO^+ + e \rightarrow N + O$. |
| F1 | 200 | $2.5 \times 10^5$ (noon);<br>absent at night | $7 \times 10^{-9}$ | Ionization of O by Lyman "continuum" or by emission lines of He. This ionization probably accompanied by $N_2$ ionization, which disappears rapidly after sunset. | $O^+$ ions readily transfer charge to NO and perhaps to $O_2^+$. Most of the ionization is thus in molecular form and disappears by dissociative recombination. |
| F2 | 300<br>Height and electron density highly variable. Large daily, seasonal, and sunspot-cycle variations are combined with general erratic behavior. | $10^6$ (noon)<br>$10^5$ (midnight) | $10^{-10}$–$10^{-9}$<br>Variable; probably decreases with increasing height. | Ionization of O by same process producing F1; F2 formed because $\alpha_{eff}$ decreases with increasing height; F2 region produces little attenuation of radiation. Additional ionization processes may contribute in F2 that are attenuated in F1. | Recombination of molecular ions as in F1; but limiting process is here charge transfer, giving an attachment-like recombination law. |

TABLE 5.2 *Designations of Radio Bands*

| Frequency designation | Frequency range | Wavelength designation | Common uses in communications |
|---|---|---|---|
| Very low (VLF) | 3–30 kHz | ten-kilometric | Maritime mobile; radio navigation |
| Low (LF) | 30–300 kHz | kilometric | Maritime mobile; radio navigation |
| Medium (MF) | 300–3000 kHz | hectometric | Commercial AM broadcast |
| High (HF) | 3–30 MHz | decametric | Short-wave broadcast |
| Very high (VHF) | 30–300 MHz | metric | TV; FM radio; space to Earth |
| Ultra high (UHF) | 300–3000 MHz | decimetric | TV channels 14–69; navigation; space to Earth; radar |
| Super high (SHF) | 3–30 GHz | centimetric | Spacecraft communication; radar |
| Extremely high (EHF) | 30–300 GHz | millimetric | Spacecraft communication; radar |

TABLE 5.3 *Microwave Bands*[a]

| Designation | Frequency range (GHz) |
|---|---|
| P | 0.225– 0.390 |
| L | 0.390– 1.550 |
| S | 1.55 – 5.20 |
| X | 5.20 – 10.90 |
| K | 10.90 – 36.00 |
| Q | 36.0 – 46.0 |
| V | 46.0 – 56.0 |
| W | 56.0 – 100.0 |

[a] See Westman, H. P. (ed.) (1970), "Reference Data for Radio Engineers", 5th ed.

### 5.1.1 Chapman Layer: The E and F1 Regions

Let us write the cross section for photoionization of a particular atom or molecule with number density $N(z)$ at a given frequency as $\sigma_\nu$ (cm$^2$) and the total absorption coefficient as $\kappa_\nu$ (cm$^{-1}$). If the ionizable gas is the only constituent absorbing, then $\kappa_\nu \equiv N(z)\sigma_\nu$. However, in the more general

case the rate of photoionization for incident monochromatic radiation is

$$q(z, \mu_0) = N(z)\sigma_\nu \pi \mathscr{F}_\nu \exp\left[ -\int_z^\infty \kappa_\nu(z')\, dz'/\mu_0 \right] (\mathrm{cm}^{-3}\ \mathrm{sec}^{-1}) \quad (5.1.2)$$

where $\pi\mathscr{F}_\nu$ is the solar photon flux (photon/cm$^2$ sec) outside the atmosphere, and $\mu_0 = \cos\chi_0$ for solar zenith angle $\chi_0$. If in addition we take the atmosphere to be isothermal and consider only a single, ionizable constituent, we have

$$q(z, \mu_0) = \pi\mathscr{F}_\nu \sigma_\nu N(z_0) \exp\left[ -\frac{z - z_0}{H} - \frac{\sigma_\nu N(z_0) H}{\mu_0}\, e^{-(z-z_0)/H} \right] \quad (5.1.3)$$

where $H$ is the scale height of the absorber and $z_0$ is an arbitrary reference height. The first term in brackets is the decrease of the ionizable constituent with height and the second term represents the increase of unattenuated solar flux with height. Differentiating places the maximum production at $z_{\max}$ given by

$$e^{(z_{\max} - z_0)/H} = \frac{N(z_0) H \sigma_\nu}{\mu_0} \quad (5.1.4)$$

Since the vertical optical thickness is

$$\tau_\nu(z) = N(z_0) H \sigma_\nu e^{-(z-z_0)/H} \quad (5.1.5)$$

the maximum ionization occurs where the slant optical thickness is $\tau_\nu/\mu_0 = 1$. Hence from (5.1.3) we can write the maximum production as

$$\begin{aligned} q(z_{\max}, \mu_0) &= \pi\mathscr{F}_\nu \sigma_\nu N(z_0) \exp\left[ -\frac{z_{\max} - z_0}{H} - 1 \right] \\ &= \frac{\pi\mathscr{F}_\nu \mu_0}{H}\, e^{-1} \end{aligned} \quad (5.1.6)$$

where we make use of (5.1.4).

We now write the maximum production for an overhead sun as

$$q_M \equiv q(z_{\max}, \mu_0 = 1) = \frac{\pi\mathscr{F}_\nu}{He} \quad (5.1.7)$$

and let $z_M$ be the height $z_{\max}$ for $\mu_0 = 1$. Then (5.1.3) takes the convenient form

$$q(z, \mu_0) = q_M \exp\left[ 1 - \frac{z - z_M}{H} - \frac{1}{\mu_0}\, e^{-(z-z_M)/H} \right] \quad (5.1.8)$$

If the recombination varies as $N_e^2$ with a coefficient $\alpha$ as in (5.1.1), the electron density at any height in equilibrium is

$$N_e(z, \mu_0) = \left(\frac{q_M}{\alpha}\right)^{1/2} \exp\left\{\frac{1}{2}\left[1 - \frac{z - z_M}{H} - \frac{1}{\mu_0} e^{-(z-z_M)/H}\right]\right\} \quad (5.1.9)$$

We can now note three important characteristics for this idealized *Chapman layer*. Expanding the exponential about $z_M$ for small distances compared with $H$ gives for an overhead sun ($\mu_0 = 1$)

$$\begin{aligned} N_e(z, 1) &\approx \left(\frac{q_M}{\alpha}\right)^{1/2} \exp\left[-\frac{1}{4}\left(\frac{z - z_0}{H}\right)^2\right] \\ &\approx \left(\frac{q_M}{\alpha}\right)^{1/2} \left[1 - \left(\frac{z - z_M}{2H}\right)^2\right] \end{aligned} \quad (5.1.10)$$

Thus the electron distribution is parabolic around the peak of the layer.

Secondly, from (5.1.6) the production varies as $\cos \chi_0$; hence the daily, seasonal, and latitudinal variation of $N_e$ will vary as $(\cos \chi_0)^{1/2}$ for a layer in quasi-equilibrium with the ionizing flux.

Finally, by setting the arbitrary reference height $z_0 = z_{\max}$ in (5.1.4) we find

$$N(z_{\max}) = \mu_0 / H \sigma_v \quad (5.1.11)$$

Thus in principle the measurement of $z_{\max}$ through the day could, if the theory were rigorously correct, lead to a determination of $N$ (density of the ionizable constituent) versus $z$ in the neighborhood of the layer peak. In practice, such an analysis becomes complicated by vertical oscillations due to atmospheric tides.

The elementary theory given above requires many refinements to be useful:

(1) The plane-parallel geometry in which the absorbing path varies as sec $\chi_0$ is clearly invalid near sunrise and sunset.

(2) The actual atmosphere has a time lag in responding to the solar zenith angle and is not in quasi-equilibrium.

(3) The thermosphere is not isothermal, so that $H$ varies with $z$.

(4) The ionizing radiation usually covers a wide range of frequency and is not monochromatic as assumed.

(5) The ionization does not remain at the altitude where it is produced, but moves up or down by ambipolar diffusion (with positive ions and electrons traveling together as a single gas).

(6) Recombination is not strictly proportional to $N_e^2$; this point will be examined further in the next section, for in some situations it is a very poor approximation.

The E and F1 regions are basically Chapman layers. The E region (90–140 km) is physically the simplest. Ions are produced mainly from $O_2$ molecules by *photoionization*,

$$O_2 + h\nu \rightarrow O_2^+ + e, \qquad \lambda < 1027\ \text{Å} \tag{5.1.12}$$

Additional ionization by coronal $X$-rays leads to $O_2^+$ and $NO^+$ ions by rapid *charge-exchange*, such as

$$N_2^+ + O_2 \rightarrow N_2 + O_2^+, \qquad k_{13} \approx 5 \times 10^{-10}\ \text{cm}^2/\text{sec} \tag{5.1.13}$$

or *atom-ion interchange*,

$$N_2^+ + O \rightarrow NO^+ + N, \qquad k_{14} = 1 \times 10^{-10}\ \text{cm}^2/\text{sec} \tag{5.1.14}$$

The molecular ions $O_2^+$ and $NO^+$ are of comparable density in the daytime E region (Fig. 5.2). Recombination occurs through *dissociative recombination*,

$$\left.\begin{aligned} O_2^+ + e &\rightarrow O + O \\ NO^+ + e &\rightarrow N + O \end{aligned}\right\} k_{15} \approx 3 \times 10^{-7}\ \text{cm}^3/\text{sec} \tag{5.1.15}$$

Thus $k_{15}$ essentially accounts for $\alpha_{\text{eff}} \sim 10^{-8}$ cm$^3$/sec in the E region (Table 5.1). The E region also has a minor, long-lived component of metallic ions,

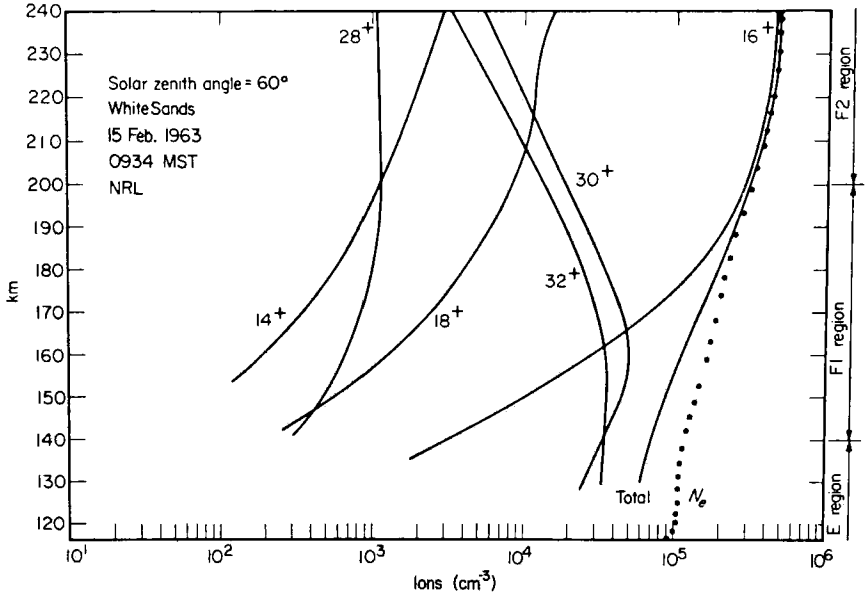



**Fig. 5.2** Daytime positive ion concentrations obtained by a rocket-borne mass spectrometer. Probable identifications of the ionic mass numbers are $N^+(=14^+)$, $O^+(=16^+)$, $H_2O^+(=18^+$—probably carried by the rocket), $N_2^+(=28^+)$, $NO^+(=30^+)$, $O_2^+(=32^+)$. The $N_e$ curve was obtained from an ionogram. [After HOLMES, J. C., JOHNSON, C. Y.; and YOUNG, J. M. (1965), *in* "Space Research," (D. G. King-Hele *et al.*, eds.), North Holland, Amsterdam, 5, 756.]

responsible for most mid-latitude sporadic E—thin patches of ionization superimposed on the regular E layer (cf. Section 5.3.2).

In the F1 region (140–200 km) the principal ion formed is O+ by

$$O + h\nu \rightarrow O^+ + e, \qquad \lambda < 911 \text{ Å} \tag{5.1.16}$$

with some contribution from

$$N_2 + h\nu \rightarrow N_2^+ + e, \qquad \lambda < 796 \text{ Å} \tag{5.1.17}$$

The sun is a fairly strong emitter in the Lyman continuum, $\lambda < 911$ Å, but emission lines of HeI (584 Å) and HeII (304 Å) also contribute. The distinguishing feature between E and F1 is that the atomic ions constituting F1 must transfer their charge to molecules prior to recombination, since *radiative recombination*,

$$O^+ + e \rightarrow O + h\nu, \qquad k_{18} = 3 \times 10^{-12} \text{ cm}^3/\text{sec} \tag{5.1.18}$$

is extremely slow. Thus F1 recombines by a two-step mechanism: *atom-ion interchange*,

$$\left.\begin{aligned} O^+ + O_2 &\rightarrow O_2^+ + O \\ O^+ + N_2 &\rightarrow NO^+ + N \end{aligned}\right\} k_{19} \approx \begin{cases} 2 \times 10^{-11} \text{ cm}^3/\text{sec} \\ 1 \times 10^{-12} \text{ cm}^3/\text{sec} \end{cases} \tag{5.1.19}$$

followed by dissociative recombination, (5.1.15).

The lifetime for an $O^+$ ion against destruction by (5.1.19) at 180 km is

$$\tau(O^+) = \frac{1}{k_{19}[N_2 + O_2]} \approx \frac{1}{3 \times 10^{-12} \times 5 \times 10^9}$$

$$\sim 60 \text{ sec} \tag{5.1.20a}$$

The lifetime in the day of the molecular ion is

$$\tau(NO^+) = \frac{1}{k_{15} N_e}$$

$$\approx \frac{1}{3 \times 10^{-7} \times 2.5 \times 10^5} \sim 10 \text{ sec} \tag{5.1.20b}$$

Thus the ions are locked up in the atomic form, with (5.1.19) being the rate-limiting process in the upper F1 region. Consequently, $\alpha_{\text{eff}} < k_{15}$ in F1 (Table 5.1).

### 5.1.2 Low Attenuation Layer and Bradbury Recombination: The F2 Region

In the F2 region three things occur that cause the ionization profile to differ from that in F1. First, the atmosphere becomes optically thin to most

ionizing radiations. Thus the second term in brackets in (5.1.3)—the second exponential—vanishes.

Second, of the two processes atom-ion interchange (5.1.19) and dissociative recombination (5.1.15), the former becomes totally dominant at high altitudes in limiting the rate of recombination. Consequently, the electron density changes at the rate

$$\frac{dN_e}{dt} = \pi \mathscr{F}_\nu \sigma_\nu [\mathrm{O}]_0 e^{-(z-z_0)/H(\mathrm{O})} - N_e k_{19} [\mathrm{X}]_0 e^{-(z-z_0)/H(\mathrm{X})} \quad (5.1.21)$$

where X represents $N_2$ and $O_2$ and where $[O]_0$ and $[X]_0$ are densities at $z_0$.

The recombination term thus varies as the first power of $N_e$. The height variation of $N_e$ for a neutral atmosphere in diffusive equilibrium and for ions in local chemical equilibrium is then

$$\begin{aligned} N_e(z) &= \text{const.} \exp\left\{(z - z_0)\left[\frac{1}{H(\mathrm{X})} - \frac{1}{H(\mathrm{O})}\right]\right\} \\ &= \text{const.} \exp\left\{(z - z_0)\frac{g}{kT}[M(\mathrm{X}) - M(\mathrm{O})]\right\} \end{aligned} \quad (5.1.22)$$

Hence the electron density in chemical equilibrium would increase indefinitely with height.

This awkward state of affairs brings up the third point about F2. Clearly $N_e$ cannot increase upward indefinitely and the mechanism that stops the increase is ambipolar diffusion. At some point the dynamical time for diffusive flow (Section 2.3.1) will become shorter than the chemical time, (5.1.20). Then the upward trend of $N_e(z)$ will approach a diffusive equilibrium with a scale height appropriate to the ion-electron gas. Therefore, F2 cannot be treated with chemical equilibrium only. Because the recombination rate varies as $\beta N_e$, where $\beta$ is a decreasing function of height, dynamics is an essential feature of an F2 region.

In general, the linear recombination law (which also applies for different reasons in the D region; see Section 5.1.4) is known as *Bradbury* or *$\beta$ recombination*, to distinguish it from the conventional $\alpha$ recombination of Eq. (5.1.1).

### 5.1.3 Ambipolar Diffusion: Nighttime Recombination at High Altitudes

We can most simply examine the role of dynamics in structuring a high-altitude F2 layer by removing the electron production term of (5.1.21) that is, by considering recombination at night. The diffusion equation (see

Section 2.3.2) will involve the scale height for the ion-electron gas in hydrostatic equilibrium. If we suppose that the neutral parent molecules, their positive ions, and the electrons all have the same temperature (as defined, say, by their mean kinetic energies), then the isothermal scale height of the ion-electron plasma ($H^*$) will be twice that of the parent neutral gas ($H_1$).

This result comes about from the fact that the electrons, with mass $m \ll M$ (where $M$ is the ion and neutral mass), tend to develop a very large scale height ($kT/mg$) but are restrained from doing so by the electrostatic attraction of the ions. Thus for the ions hydrostatic equilibrium (1.1.1) becomes, for equal electron and ion densities,

$$dp_i/dz = -gN_eM + N_eeE \tag{5.1.23}$$

and, for the electrons,

$$dp_e/dz = -gN_em - N_eeE \approx -N_eeE \tag{5.1.24}$$

where $E$ is the electrostatic field generated by the gravitational tendency for separation. Eliminating $E$ and using the perfect gas law (1.1.2), we obtain an isothermal hydrostatic equilibrium for the plasma of

$$\frac{1}{N_e}\frac{dN_e}{dz} = -\frac{gM}{2kT} \tag{5.1.25}$$

Comparison with (1.1.5) shows that $H^* = 2H_1$. (Also, see Problem 5.1.)

The equation of continuity (2.1.1) with a loss term due to recombination and one-dimensional diffusion thus gives

$$\frac{\partial N_e}{\partial t} = -\beta(z)N_e + \frac{\partial}{\partial z}\left[D(z)\left(\frac{\partial N_e}{\partial z} + \frac{N_e}{2H_1}\right)\right] \tag{5.1.26}$$

where the diffusion velocity is substituted from (2.3.1). The height dependence of the recombination coefficient is given by (5.1.21), which we write as

$$\beta(z) = k_{19}[\mathrm{X}]_0 e^{-(z-z_0)/H(\mathrm{X})} \equiv \beta_0 e^{-(z-z_0)/H(\mathrm{X})} \tag{5.1.27}$$

The diffusion coefficient varies inversely as the total density. We will assume that the major constituent is the ionizable one—in the case of Earth's F2 it is O with scale height $H_1$. A magnetic field will inhibit diffusion of ions across the field lines. For $O^+$ diffusing through O, the coefficient is

$$\begin{aligned} D(z) &= \frac{2.3 \times 10^{18} \sin^2 i}{[\mathrm{O}(z)]} \\ &= D_0 e^{(z-z_0)/H_1} \end{aligned} \tag{5.1.28}$$

where $i$ is the *inclination* (to the horizontal) of the field lines.

Thus we are concerned with two scale heights: $H_1$ for the ionizable neutral gas, which is also the major constituent, and $H(\mathrm{X})$ for the minor molecular constituent involved in the first step of recombination, atom-ion interchange

(or charge exchange) (5.1.19). Continuity (5.1.26) thus gives

$$\frac{\partial N_e}{\partial t} = D_0 e^{(z-z_0)/H_1}\left(\frac{\partial^2 N_e}{\partial z^2} + \frac{3}{2H_1}\frac{\partial N_e}{\partial z}\right) + N_e\left(\frac{D_0 e^{(z-z_0)/H_1}}{2{H_1}^2} - \beta_0 e^{-(z-z_0)/H(X)}\right) \tag{5.1.29}$$

This equation is linear in $N_e$, making it possible to extract the time dependence, leaving a characteristic-value equation for $N_e(z)$.

Assuming the solution

$$N_e(z,t) = \sum_{j=0}^{\infty} N_j(z) e^{-\lambda_j t} \tag{5.1.30}$$

for constants $\lambda_j$, we have the equation for $N_j(z)$,

$${H_1}^2 e^{(z-z_0)/H_1}\left(\frac{\partial^2 N_j}{\partial z^2} + \frac{3}{2H_1}\frac{\partial N_j}{\partial z}\right)$$

$$-\left(\frac{{H_1}^2\beta_0}{D_0} e^{-(z-z_0)/H(X)} - \tfrac{1}{2} e^{(z-z_0)/H_1} - \frac{{H_1}^2}{D_0}\lambda_j\right) N_j = 0, \qquad j = 0, 1, \ldots, \infty \tag{5.1.31}$$

A change of variables removes the first derivative term:

$$y_j = N_j e^{(z-z_0)/H_1}, \qquad x = e^{-(z-z_0)/H_1} \tag{5.1.32}$$

Also if we consider the special case of $H_1 = H(X)$ (which implies mixing and is not realistic for Earth's F2 region but nevertheless shows the kind of solution that emerges), we have

$$\frac{\partial^2 y_j}{\partial x^2} + \left(\frac{3}{16x^2} + \frac{{H_1}^2\lambda_j}{D_0 x} - \frac{{H_1}^2\beta_0}{D_0}\right) y_j = 0 \tag{5.1.33}$$

which is readily demonstrated by writing $N_j = Y_j x^{1/4}$ (Problem 5.2).

This equation has the same form as the familiar *radial wave equation* for the hydrogen atom. We seek a solution subject to two boundary conditions. First, as $z \to \infty$ the diffusion velocity must not be infinite. Hence the term in brackets in (5.1.26) must vanish, which in turn requires that all components of $N_e$ approach diffusive equilibrium or

$$N_j \to \text{const. } e^{-(z-z_0)/2H_1} = \text{const. } x^{1/2} \tag{5.1.34}$$

The second boundary condition requires that $N_e$ remain bounded at low altitudes ($z \to -\infty$). In terms of the $y_j(x)$ variable of (5.1.32), these conditions are

$$\begin{aligned} y_j &\to \text{const. } x^{1/4}, & x &\to 0 \\ y_j &\to 0, & x &\to \infty \end{aligned} \tag{5.1.35}$$

The eigenfunctions of (5.1.33) are given in terms of generalized Laguerre polynomials, $L_j^{-1/2}(v)$:

$$N_j = A_j v^{1/2} e^{-v/2} L_j^{-1/2}(v) \tag{5.1.36}$$

Here $A_j$ is a constant to be fixed by the initial electron distribution and the independent variable is

$$v = 2H_1 \left(\frac{\beta_0}{D_0}\right)^{1/2} x = 2H_1 \left(\frac{\beta_0}{D_0}\right)^{1/2} e^{-(z-z_0)/H_1} \tag{5.1.37}$$

The polynomial may be written in a power series,

$$L_j^{-1/2}(v) = 1 + a_1 v + a_2 v^2 + \cdots + a_j v^j \tag{5.1.38}$$

where the coefficients are related by the recursion relation,

$$a_{m+1} = \frac{m-j}{(m+1)(m+\frac{1}{2})} a_m, \qquad m = 0, 1, \ldots; \quad a_0 = 1 \tag{5.1.39}$$

The eigenvalues are

$$\lambda_j = \frac{2(\beta_0 D_0)^{1/2}}{H_1}(j + \tfrac{1}{4}), \qquad j = 0, 1, \ldots \tag{5.1.40}$$

From (5.1.30) and (5.1.40) it is apparent that solutions for large $j$ represent short-lived fluctuations. A short time after sunset the higher-order solutions will vanish and $N_e$ can be expressed as a linear combination of the first two or three eigenfunctions. Figure 5.3 shows $N_0$, $N_1$, and $N_2$ on an arbitrary

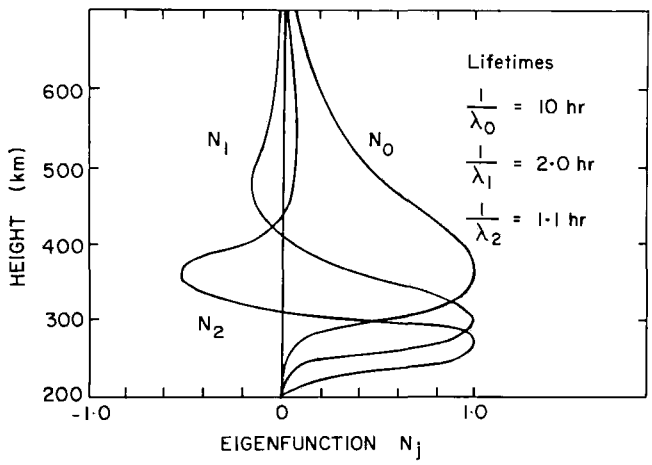



**Fig. 5.3** The first three eigenfunctions for the decay of an idealized nighttime F region. [Adapted from CHAMBERLAIN (1961).]

scale of unity for the maximum of each curve. We assume for illustration that $[O(z_0 = 300\text{ km})] = 3 \times 10^9\text{ cm}^{-3}$, $H_1 = 50$ km, $D_0 = 7 \times 10^8\text{ cm}^2/\text{sec}$, and $\beta_0 = 10^{-4}\text{ sec}^{-1}$.

As $N_2$ and $N_1$ decay with lifetimes, respectively, of $1/\lambda_2 = 1.1$ hr and $1/\lambda_1 = 2.0$ hr, the peak altitude rises provided $A_2$ and $A_1$ are positive. The relative importance of these "overtone" solutions will be governed by the initial distribution of $N_e$ and the $A_j$'s can be negative provided $N_e$ is everywhere and at all times positive.

After several hours

$$N_e \to N_0 e^{-\lambda_0 t} \tag{5.1.41}$$

The maximum height and shape of F2 then no longer changes with time. The peak has risen to 365 km and $N_e$ decays with a lifetime of $\lambda_0^{-1} = 10$ hr. The distribution with height from (5.1.36) is

$$N_0(z) = \text{const}\, e^{-(z-z_0)/2H_1} \exp[-H_1(\beta_0/D_0)^{1/2} e^{-(z-z_0)/H_1}] \tag{5.1.42}$$

which is in the form of a Chapman layer (5.1.9). Around the peak the electron distribution is parabolic and at high alltitudes it is in diffusive equilibrium.

Mathematically the eigenvalue nature of the solution is a result of the linearity of (5.1.26). Physically this linearity means that for each eigenfunction the relative decay rate, $N_j^{-1}\,\partial N_j/\partial t = -\lambda_j$, is a constant with time and height. Thus when only the zeroth-order solution remains, the flux convergence of diffusive flow is exactly compensated, at every height, by recombination loss—except for a constant diffusive loss rate $\lambda_0$ arising from a constant downward flow velocity (see Problem 5.2). Consequently, the layer retains a constant shape.

This simplified treatment ignores the important effects of horizontal winds which, by moving the ions across magnetic-field lines, produce a vertical component of the Lorentz force, $e\mathbf{v} \times \mathbf{B}/c$.

### 5.1.4 Ion Chemistry in the D Region

The distinguishing feature of the D region (75–90 km) is the predominance of negative ions. The $N_2$ molecule does not form a stable negative ion but $O_2$ does. These ions are formed by *three-body attachment*,

$$O_2 + e + O_2 \to O_2^- + O_2, \qquad k_{43} \approx 5 \times 10^{-31}\text{ cm}^6/\text{sec} \tag{5.1.43}$$

(The third body may also be $N_2$ but the rate coefficient is then substantially lower.) The three-body process is more important at D-region densities ($[O_2] \approx 10^{13}$–$10^{14}\text{ cm}^{-3}$) than *radiative attachment*,

$$O_2 + e \to O_2^- + h\nu, \qquad k_{44} \sim 10^{-19}\text{ cm}^3/\text{sec} \tag{5.1.44}$$

The electrons may be removed from $O_2^-$ in the daytime by *photodetachment* (the inverse of the above process),

$$O_2^- + h\nu \rightarrow O_2 + e, \qquad J_{45}(z = \infty) \sim 0.3 \text{ sec}^{-1} \tag{5.1.45}$$

and in day and night by *collisional detachment* [the inverse of (5.1.43)],

$$O_2^- + O_2 \rightarrow O_2 + e + O_2, \qquad k_{46} \sim 10^{-22} \text{ cm}^3/\text{sec} \tag{5.1.46}$$

The electron affinity (i.e., ionization potential) of $O_2^-$ is 0.43 eV. (The rate of $k_{46}$ is extremely sensitive to temperature and the quoted value is an estimate for 200°K obtained from considerations of detailed balancing; see Problem 5.3.) Because this reaction is so slow, similar reactions with minor constituents (or molecules in metastable states) may be more significant detachment processes at night.

Positive ions and electrons are created through photoionization of $N_2$ and $O_2$ by X-rays and by Lyman alpha ionization of the minor constituent NO:

$$X + h\nu \rightarrow X^+ + e, \qquad J_{47} \sim 10^{-7} \text{ sec}^{-1} \tag{5.1.47}$$

The ultimate loss of ionization occurs when the positive ions are neutralized, not simply when the free electrons disappear by attachment. The main recombination process when electrons are attached to molecules forming negative ions is *mutual neutralization*,

$$X^+ + Y^- \rightarrow X + Y, \qquad k_{48} \sim 2 \times 10^{-7} \text{ cm}^3/\text{sec} \tag{5.1.48}$$

In addition to recombining by attachment (5.1.43) followed by mutual neutralization (5.1.48), free electrons can disappear directly through dissociative recombination (5.1.15), and this reaction must be considered for the daytime equilibrium.

On the basis of this limited chemistry we can set up equations for the time variation of the density of free electrons $N_e$, positive ions $N^+$, and negative ions $N^-$. We write the ratio

$$\lambda \equiv N^-/N_e \tag{5.1.49}$$

If the free-electron population is produced by (5.1.45), (5.1.46), and (5.1.47) and destroyed mainly by (5.1.43), we have

$$\frac{dN_e}{dt} = J_{45}\lambda N_e + k_{46}\lambda[O_2]N_e + J_{47}[X] - k_{43}[O_2]^2 N_e \tag{5.1.50}$$

The loss process for free electrons (5.1.43) is linear with $N_e$ and thus follows the $\beta$- or Bradbury-recombination law discussed in Section 5.1.2. Positive

ions are destroyed only by (5.1.48) so that

$$\frac{dN^+}{dt} \equiv (1+\lambda)\frac{dN_e}{dt} + N_e\frac{d\lambda}{dt}$$

$$= J_{47}[\mathrm{X}] - k_{48}\lambda(1+\lambda)N_e^2 \tag{5.1.51}$$

or

$$\frac{dN_e}{dt} = \frac{J_{47}[\mathrm{X}]}{1+\lambda} - k_{48}\lambda N_e^2 - \frac{N_e}{1+\lambda}\frac{d\lambda}{dt} \tag{5.1.52}$$

Written in this form the *effective production rate* of electrons, for constant $\lambda$, is [see (5.1.1)]

$$q_{\mathrm{eff}} = \frac{J_{47}[\mathrm{X}]}{1+\lambda} \tag{5.1.53}$$

and the *effective recombination rate* is

$$\alpha_{\mathrm{eff}} = k_{48}\lambda \tag{5.1.54}$$

At night the $J$'s vanish and Eq. (5.1.50) indicates that free electrons rapidly disappear after sunset. One can solve (5.1.50) and (5.1.52) simultaneously for quasi-equilibrium to find daytime values of $N_e$ and $\lambda$ as functions of height and solar zenith angle. For many years it seemed that the $O_2^-$ chemistry summarized above accounted for the D-region ionization in a satisfactory way. However, reaction (5.1.45) detaches the electron with visual light and detailed observations of two types failed to conform to expectations. First, twilight observations of the D region following ionization surges caused by particle bombardment from solar flares (*polar-cap absorption* or *PCA events*) showed the free electrons disappearing earlier than expected. Second, high-altitude nighttime tests of nuclear weapons in 1962 did not produce the extensive radio blackouts expected on the following mornings when sunlight could photodetach the high levels of ionization produced by gamma rays from the weapon. (Radio propagation effects are discussed in Section 5.2.) Both effects indicated that the electrons were bound to molecules that required ultraviolet rather than visual light to photodetach the electrons. The negative ions must therefore have electron affinities of several eV.

Laboratory work has greatly clarified the problem but has also shown it has surprising complexity. Figure 5.4 shows the chain of reactions that leads to the formation of $NO_2^-$ and $NO_3^-$ ions, which have electron affinities of 2.5 and 3.6 eV, respectively.

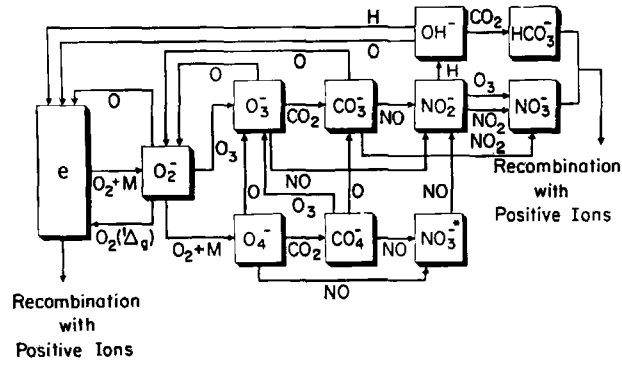



**Fig. 5.4** Schematic diagram of negative-ion reactions in the D region with hydration neglected. [After FERGUSON (1974).]

The $O_2^-$ ion when formed may react with O to release the electron,

$$O_2^- + O \rightarrow O_3 + e \tag{5.1.55}$$

which merely starts the process over again by (5.1.43). But $O_2^-$ may also react with $O_2$ or $CO_2$ or undergo charge transfer with $O_3$:

$$O_2^- + O_2 + M \rightarrow O_4^- + M + 0.54\text{ eV}, \qquad k_{56} \approx 5 \times 10^{-31}\text{ cm}^6/\text{sec} \tag{5.1.56}$$

$$O_2^- + CO_2 + M \rightarrow CO_4^- + M + 0.8\text{ eV}, \qquad k_{57} \approx 3 \times 10^{-29}\text{ cm}^6/\text{sec} \tag{5.1.57}$$

$$O_2^- + O_3 \rightarrow O_2 + O_3^- + 1.5\text{ eV}, \qquad k_{58} \approx 4 \times 10^{-10}\text{ cm}^3/\text{sec} \tag{5.1.58}$$

The $O_4^-$ of (5.1.56) is connected to $CO_4^-$ by

$$O_4^- + CO_2 \rightarrow CO_4^- + O_2 + 0.36\text{ eV}, \qquad k_{59} \approx 4.3 \times 10^{-10}\text{ cm}^3/\text{sec} \tag{5.1.59}$$

and $CO_4^-$ reacts with ozone by

$$CO_4^- + O_3 \rightarrow O_2 + CO_2 + O_3^- + 0.6\text{ eV}, \qquad k_{60} = 1.3 \times 10^{-10}\text{ cm}^3/\text{sec} \tag{5.1.60}$$

The terminal ion is probably formed by

$$O_3^- + NO_2 \rightarrow NO_3^- + O_2, \qquad k_{61} = 2 \times 10^{-11}\text{ cm}^3/\text{sec} \tag{5.1.61}$$

and

$$O_3^- + CO_2 \rightarrow CO_3^- + O_2 \tag{5.1.62}$$

followed by

$$CO_3^- + NO_2 \rightarrow NO_3^- + CO_2 \tag{5.1.63}$$

The ion chemistry is further complicated by hydration. The large dipole moment of $H_2O$ allows it to become attached readily to positive and negative ions alike, and clusters of even three or more water molecules can develop. The presence of the hydration affects the electron affinity and reaction rates of the molecule.

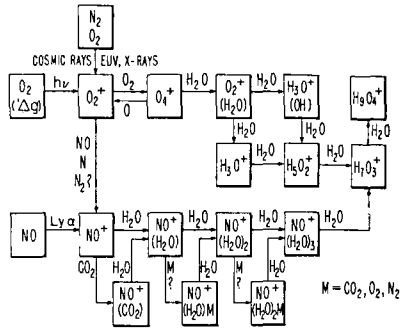



Fig. 5.5 Schematic diagram of positive-ion reactions in the D region with hydration neglected. [After FERGUSON (1971).]

Positive-ion chemistry is also complicated. The $N_2^+$ and $O_2^+$ will readily transfer positive charge to NO, which has a lower ionization potential. Mass spectrometers flown on rockets have shown, however, that water-derived ions $H_3O^+$, $H_5O_2^+$, and $H_7O_3^+$ are dominant. The principal reactions appear to be ion-molecule association,

$$O_2^+ + O_2 + M \rightarrow O_4^+ + M \tag{5.1.64}$$

followed by hydration,

$$O_4^+ + H_2O \rightarrow O_2^+ \cdot (H_2O) + O_2 \tag{5.1.65}$$

and either

$$O_2^+ \cdot (H_2O) + H_2O \rightarrow H_3O^+ + OH + O_2 \tag{5.1.66}$$

or

$$O_2^+ \cdot (H_2O) + H_2O \rightarrow H_3O^+ \cdot (OH) + O_2 \tag{5.1.67}$$

followed by

$$H_3O^+ \cdot (OH) + H_2O \rightarrow H_3O^+ \cdot (H_2O) + OH \tag{5.1.68}$$

Continued interaction with $H_2O$ builds up the heavier ions. Figure 5.5 shows alternate sequences of positive-ion reactions.

## 5.2 Radio Waves in an Ionized Atmosphere

In an ionized medium, the real index of refraction is less than unity and can even drop to zero. This characteristic, which occurs in the medium- and high-frequency (MF and HF) ranges of the spectrum (0.3–30 MHz) in the Earth's ionosphere, makes radio sounding of the ionosphere the powerful technique it is. The index of refraction is also affected by a magnetic field, which greatly complicates the theory.

For later reference we write Maxwell's equations in Gaussian units (where current density **J** is given in electromagnetic units and conductivity $\sigma$ is in

electrostatic units). The electric field is **E**; magnetic induction is **B**; permeability, $\mu$; dielectric constant or permittivity, $\kappa$; and $e$ is the positive elementary charge. Ampere's law is

$$\nabla \times \left(\frac{\mathbf{B}}{\mu}\right) = 4\pi\mathbf{J} + \frac{1}{c}\frac{\partial(\kappa\mathbf{E})}{\partial t} \tag{5.2.1}$$

Faraday's law is

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{5.2.2}$$

the magnetic lines of force are continuous, or

$$\nabla \cdot \mathbf{B} = 0 \tag{5.2.3}$$

and electrostatic lines of force start and stop on charges, or

$$\nabla \cdot (\kappa\mathbf{E}) = e4\pi(N_i - N_e) \tag{5.2.4}$$

where $N_i$ and $N_e$ are singly-charged ion and electron number densities. In a stationary atmosphere with no magnetic field, Ohm's law is

$$\mathbf{J} = \sigma\mathbf{E}/c \tag{5.2.5}$$

When there is no net charge ($N_i = N_e$) and when $\mu$ and $\kappa$ are independent of time and position, we obtain the common wave equation by taking the curl of (5.2.2). Then eliminating the resulting term in $\nabla \cdot \mathbf{E}$ with (5.2.4), $\nabla \times \mathbf{B}$ with (5.2.1), and **J** with (5.2.5) yields

$$\nabla^2\mathbf{E} - \frac{\mu\kappa}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} - \frac{4\pi\sigma\mu}{c^2}\frac{\partial \mathbf{E}}{\partial t} = 0 \tag{5.2.6}$$

When $\sigma$ is sufficiently large, the second term is negligible and E is governed by a heat-conduction or diffusion equation. When $\sigma = 0$ and $\mu\kappa$ is real, (5.2.6) is the ordinary wave equation, and the phase velocity is $c/n$, where $n = (\kappa\mu)^{1/2}$ is the ordinary index of refraction of the medium.

### 5.2.1 An Ionosphere with No Magnetic Field: Critical Reflection

Let us examine propagation of a radio wave in a uniform, ionized medium (where uniformity means that any departures from homogeneity are on a scale large compared with the wavelength). With $\mu = 1$ the wave equation is

$$\nabla^2\mathbf{E} = \frac{\kappa}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} \tag{5.2.7}$$

The oscillating **E** field sets electrons in motion. When these electrons collide with the gas molecules with a collision frequency $\nu_c$, two related things occur. The amount of energy re-emitted as electromagnetic radiation is diminished and the electron's motion lags behind the phase of the incident wave. Thus using complex exponentials to treat the phase will lead to a real (negative) exponential that we can identify with the absorption coefficient of the atmosphere.

Writing $\mathbf{E}(t) = \mathbf{E}_0 e^{-i\omega t}$, where $\omega = 2\pi\nu$ and $\nu$ is the radiation frequency (cycle/sec $\equiv$ hertz), the equation of motion of an electron is

$$m\frac{d^2\mathbf{x}}{dt^2} = -e\mathbf{E}_0 e^{-i\omega t} - m\nu_c\frac{d\mathbf{x}}{dt} \tag{5.2.8}$$

where **x** is the electron displacement from its neutral position and $m$ is the electron mass. We are supposing that in a collision the electron loses all its momentum. The steady-state solution must clearly vary as $e^{-i\omega t}$ and is

$$\mathbf{x} = \frac{-i}{\nu_c - i\omega}\frac{e}{\omega m}\mathbf{E}_0 e^{-i\omega t} \tag{5.2.9}$$

To see how the **E** field itself varies, we relate the dielectric constant to the polarization **P**, here caused by the displacement of electrons from their electrically neutral position. Thus we have

$$\kappa \equiv 1 + 4\pi\frac{\mathbf{P}}{\mathbf{E}} = 1 - \frac{4\pi N_e e\mathbf{x}}{\mathbf{E}} = 1 + \frac{i4\pi\sigma_P}{\omega} \tag{5.2.10}$$

where we define

$$\sigma_P = \frac{N_e e^2(\nu_c + i\omega)}{m(\omega^2 + \nu_c^2)} \tag{5.2.11}$$

That the quantity $\sigma_P$ is a complex conductivity can be seen from Ohm's law, (5.2.5), where the polarization current, $(1/c)\,d\mathbf{P}/dt = -(N_e e/c)\,d\mathbf{x}/dt$, replaces the conduction current **J** (also see Problem 5.5).

The wave equation (5.2.7) has a plane-wave solution

$$E(z,t) = E_0 \exp[-i\omega(t - \kappa^{1/2}z/c)] \tag{5.2.12}$$

where $z$ is the distance along the path of propagation. From (5.2.10) and (5.2.11), the real part of $\kappa$ is

$$\kappa_0 = 1 - \frac{4\pi N_e e^2}{m(\nu_c^2 + \omega^2)} = 1 - \frac{\omega_0^2}{\omega^2 + \nu_c^2} \tag{5.2.13}$$

where the *plasma frequency* is

$$\omega_0 = (4\pi N_e e^2/m)^{1/2} \tag{5.2.14}$$

To obtain the real index of refraction and the absorption coefficient, we write $\kappa^{1/2} = n + ig$ or $\kappa = n^2 - g^2 + i2ng$. Then we have from (5.2.10).

$$\kappa = \kappa_0 + i4\pi\sigma_0/\omega \tag{5.2.15}$$

where the real part of $\sigma_P$ is

$$\sigma_0 = \frac{N_e e^2 \nu_c}{m(\omega^2 + \nu_c^2)} \tag{5.2.16}$$

Solving simultaneously for $n^2$ and $g$ in terms of $\kappa_0$ and $\sigma_0$ we obtain the real index of refraction $n$ from

$$n^2 = \frac{\kappa_0}{2} + \left(\frac{\kappa_0^2}{4} + \frac{4\pi^2\sigma_0^2}{\omega^2}\right)^{1/2} \tag{5.2.17}$$

The absorption coefficient is obtained by considering the attenuation with distance of the mean intensity (averaged over an oscillation). Thus from (5.2.12) $I = \frac{1}{2}\mathbf{E} \cdot \mathbf{E}^* = I_0 \exp(-2\omega g z/c)$ and the volume absorption coefficient is

$$k = \frac{2\omega g}{c} = \frac{4\pi\sigma_0}{cn} \text{ (cm}^{-1}\text{)} \tag{5.2.18}$$

At high-collision frequencies $\nu_c \gg \omega$, (5.2.16) gives the kinetic theory value of $\sigma_0$ proportional to $1/\nu_c$. When the collision frequency is low ($\nu_c \ll \omega$), the conductivity $\sigma_0$ varies directly as $\nu_c$; that is, collisions are required for the electrons to carry any net conduction current. But since $\sigma_0$ is small,

$$n \equiv \kappa_0^{1/2} = (1 - \omega_0^2/\omega^2)^{1/2} \tag{5.2.19}$$

and waves with a circular frequency $\omega < \omega_0$ cannot propagate in the medium. In the ionosphere waves will be prohibited from entering the region where $\omega_0 \geq \omega$ by refraction or reflection. If the incident angle, measured from the normal, is $\theta_0$, the wave will, according to Snell's law, be turned back at an electron density corresponding to $n = \sin\theta_0$. For the particular case of vertical incidence ($\theta_0 = 0$), *critical reflection* occurs when $\omega_0 = \omega$ and $n = 0$. The electron density required to reflect a wave with frequency $\nu = \omega/2\pi$ is, from (5.2.14),

$$N_e = \frac{\pi m \nu^2}{e^2} = 1.24 \times 10^4 \nu^2 \tag{5.2.20}$$

where $\nu$ is measured in megahertz and $N_e$ in electron/cm$^3$. The absorption coefficient for $\nu_c \ll \omega$ is

$$k = \frac{4\pi N_e e^2 \nu_c}{mc\omega(\omega^2 - \omega_0^2)^{1/2}} \tag{5.2.21}$$

At $\omega \approx \omega_0$ the coefficient is large, high attenuation being associated with a low $n$. At frequencies well above the local plasma frequency ($\omega \gg \omega_0$), the absorption varies as $\omega^{-2}$. Thus high-frequency (short-wave) radio is less susceptible to daytime D-region absorption than is the medium-frequency commercial broadcast band.

The cyclic plasma frequency $\omega_0/2\pi$ is the natural (plane-wave) frequency for coherent oscillation of electrons in the plasma. The *critical frequency* $v_0$ of an ionized layer is the maximum frequency ($\omega/2\pi$) reflected and hence the maximum value of $\omega_0/2\pi$ in the layer.

The theory for the dispersion relation (5.2.17) in the presence of an external magnetic field will be examined in Section 5.2.5.

### 5.2.2 Refraction for Grazing Incidence: Spacecraft Occultations

One of the most powerful tools for sensing atmospheres of other planets with spacecraft has been the occultation of SHF and UHF radio waves. The technique is especially valuable when it has been possible to use two frequencies, the higher one (the X band at 8400 MHz) being nearly unaffected by the ionosphere and the lower UHF frequency (the S band at 2300 MHz) being somewhat altered. These frequencies are coherent at a ratio of 3/11.

In the following we will consider the radio signals to be transmitted from Earth to the spacecraft (the uplink mode) with the information received by the spacecraft being read out at a later time. This was in fact the procedure used on Mariner 5 (Venus). Later occultation experiments were designed for downlink operation, or sometimes an uplink-downlink round trip, with a transponder on the spacecraft returning a signal to Earth. The basic theory is, of course, independent of the sense of transmission.

Suppose the signal transmitted from Earth is received and recorded by the spacecraft. As the spacecraft passes behind the planet, (Fig. 5.6) the

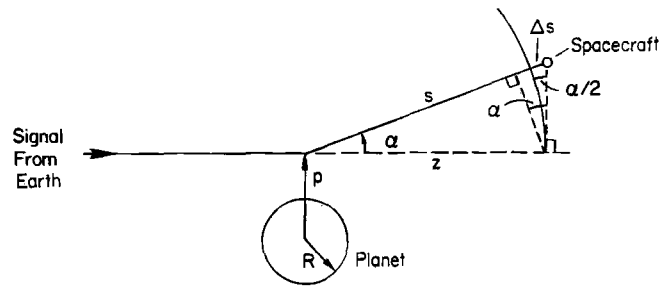



**Fig. 5.6** Geometry of refraction of a radio wave passing through an ionosphere to a spacecraft receiver. The angle of refraction is $\alpha$ for a ray that has an impact parameter $p$.

atmosphere affects the signal in two related ways: the changing index of refraction causes the signal to be deflected and its phase to be shifted. The phase shift is the quantity measured.

Since the phase velocity is $c/n$, an unmodified wave will arrive at the spacecraft at time $t$ and distance $s$ from the transmitter with phase $\phi_0 = -(\omega t - \omega s/c)$, whereas the wave passing through an atmosphere with index $n$ will have an altered phase of $\phi_1 = -(\omega t - n\omega s/c)$. Thus the phase shift is

$$\phi \equiv \phi_1 - \phi_0 = \frac{\omega}{c} \int_{-\infty}^{\infty} (n - 1)\, ds \tag{5.2.22}$$

For $n < 1$ the phase velocity of the wave is advanced and therefore the phase shift is negative (a given phase arrives earlier than it would for an unmodified wave).

The phase shift and angle of refraction are related through the index of refraction, but the refraction angle $\alpha$ also affects the total path length and thereby produces an additional phase shift. Therefore, we must first express $\alpha$ in terms of the phase shift (5.2.22) caused by the changed phase velocity. Then we can find the additional effect due to the path length. The angle of refraction is obtained from Snell's law, but because $n$ steadily changes through the atmosphere, the total deviation is a bit complicated. Consider the refraction occurring as a ray enters a thin layer between $r$ and $r'$ (Fig. 5.7). From the triangle OAB we have

$$\frac{r}{\sin\psi} = \frac{r'}{\sin\theta} \tag{5.2.23}$$

and Snell's law gives

$$n' \sin\theta' = n \sin\psi \tag{5.2.24}$$

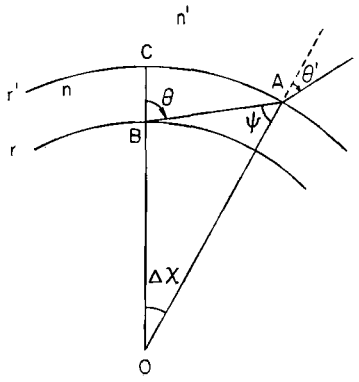



**Fig. 5.7** Refraction of a ray as it passes from a region of refractive index $n'$ into a thin spherical shell with index $n$.

Eliminating $\psi$ gives

$$nr \sin\theta = n'r' \sin\theta' \tag{5.2.25}$$

With $n' = n + \Delta n$, and so forth, the differential form of (5.2.25) becomes

$$\frac{dr}{r} + \frac{dn}{n} + \frac{d\theta}{\tan\theta} = 0 \tag{5.2.26}$$

The differential refraction is

$$\Delta\alpha \equiv \theta' - \psi \equiv \theta' - (\theta - \Delta\chi) \tag{5.2.27}$$

and from the small triangle ABC in Fig. 5.7 we have

$$r' \Delta\chi \approx \Delta r \tan\theta \tag{5.2.28}$$

Removing $dr/r$ with (5.2.28) and writing $d\theta + d\chi$ in terms of $d\alpha$ by (5.2.27), we obtain from (5.2.26)

$$d\alpha = -\tan\theta \, dn/n \tag{5.2.29}$$

Now suppose that in traversing a shell of thickness $dr$ the ray covers a total distance $ds$ and experiences a change in index of $dn$. We wish to find the change in index of refraction over a given ray path $ds$ when the *impact parameter* $p$ is changed by amount $dp$, since $p$ is the coordinate that observationally is known. From Fig. 5.8 the segment of ray path $ds$ through the shell

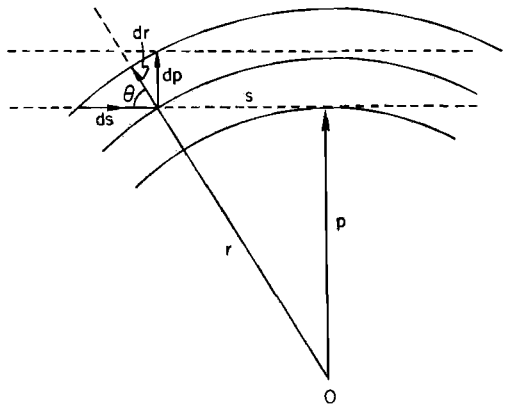



**Fig. 5.8** Geometry associated with the change in the refractive index accompanying a change $dp$ in the impact parameter as the spacecraft moves behind (or out from) the planet. (Note that $ds$ is always positive in the direction of propagation; for instance, we have $0 \leq \theta \leq \pi/2$, but for emergence the angle as shown must be regarded as $\theta + \pi$.)

$dr$ is displaced out of the shell if the impact parameter is increased by $dp$, or

$$\frac{dn}{ds} = -\frac{dn}{dr}\cos\theta = -\left(\frac{dn}{dp}\frac{1}{\sin\theta}\right)\cos\theta \tag{5.2.30}$$

Consequently, the total angle of refraction when $n \approx 1$ (i.e., for $\omega$ far from the critical frequency) is given by

$$\begin{aligned}\alpha &= -\int \frac{dn}{n}\tan\theta = -\int_{-\infty}^{\infty} \frac{\tan\theta}{n}\frac{dn}{ds}ds \\ &= \int_{-\infty}^{\infty} \frac{1}{n}\frac{dn}{dp}ds = \frac{d}{dp}\int_{-\infty}^{\infty}(n-1)\,ds \\ &= \frac{c}{\omega}\frac{d\phi}{dp}\end{aligned} \tag{5.2.31}$$

This expression relates the refraction angle to the phase advancement due to the change in phase velocity caused by the lower index $n$. As noted above, offsetting this effect is a phase retardation due to the bending of the ray, which increases the distance traversed compared with the straight-line distance. In Fig. 5.6 the increased distance of transmission due to the deflection $\alpha$ is

$$\Delta s = \alpha z(\alpha/2) = \tfrac{1}{2}z\alpha^2 \tag{5.2.32}$$

The total or integrated phase lag due to the velocity change and the path length change is, from (5.2.31) and (5.2.32), respectively,

$$\phi = \frac{\omega}{c}\int (n-1)\,ds + \frac{\omega}{c}\frac{z\alpha^2}{2} \tag{5.2.33}$$

With $\alpha$ related to $n$ by (5.2.31) and $n$ given by (5.2.19) and (5.2.14), we obtain for $\omega \gg \omega_0$

$$\phi = -\frac{\omega}{c}I + \frac{z\omega}{2c}\left(\frac{dI}{dp}\right)^2 \tag{5.2.34}$$

where

$$I \equiv -\int (n-1)\,ds = \frac{e^2}{2\pi m v^2}\int_{-\infty}^{\infty} N_e(s)\,ds \tag{5.2.35}$$

The first term in (5.2.34) is normally the dominant one, and hence the phase shift is negative. The use of two frequencies simplifies the determination of the phase shift. If only one frequency is used the spacecraft's position must be known with high accuracy. Thus the simultaneous measurement of $\phi(t)$ with position $p(t)$ allows one to infer the tangentially integrated electron

density, or $I(p)$, for varying impact parameters $p$, and hence reconstruct $N_e$ *versus* height $r$ above the center of the planet. A simple example is given in Problem 5.7.

In addition to the phase shift and deflection, differential refraction causes focusing and defocusing as the ray passes through the ionosphere and defocusing in the lower atmosphere (where $n > 1$). Thus the signal strength, as well as the phase, changes during occultation entry and exit.

The above discussion in terms of phase emphasizes the physical effects at work. For actually reducing observations a much more practical approach is to measure the Doppler shift in frequency at the spacecraft. The relative motion of Earth and spacecraft is known from the spacecraft's ephemeris. Because of the bending of the ray through angle $\alpha$ the spacecraft will detect a slight Doppler shift in the direction perpendicular to the Earth-spacecraft direction. This Doppler shift gives the deflection angle $\alpha(p)$, which in turn is related to the index $n$ by (5.2.31), and it is not necessary to work with the two-term solution for $\phi$ (5.2.34).

### 5.2.3 Partial Reflection from Sharp and Diffuse Surfaces

Critical reflection, developed in Section 5.2.1, is a highly efficient mechanism. However, just as in ordinary optics, some reflections occur whenever a wave crosses a boundary between two indices of refraction. The effect is important in reflection by relatively thin ionospheric layers, such as one type of sporadic $E$. At one time it aroused interest in accounting for radar reflections from so-called *radio-aurorae*: areas of ionization aligned along the magnetic field and created by the same particle bombardment that creates visual aurorae. Radio-aurorae are observed at quite high frequencies. If a 400 MHz signal were due to critical reflection, it would require $N_e \approx 2 \times 10^9$ cm$^{-3}$. On the other hand, much lower densities can, under the right circumstances, reflect such waves. The right circumstances are that the sheet of ionization have a rather distinct "surface" and that this surface be aligned perpendicular to the radar beam. Hence radio-aurorae are most commonly seen poleward of the radar. As we shall see, however, partial reflection from a diffuse surface tends to be very inefficient because of interference. (Radio-aurorae probably arise from inhomogeneous patches of ionization and not plane sheets; cf. Section 5.2.4.)

The Fresnel formulae for reflected and refracted amplitudes in the general case of arbitrary angles of incidence and polarization are developed in texts on electromagnetism. For our purposes a brief review for normal incidence will suffice. The formulae arise from the requirements that the tangential components of $\mathbf{E}$ and $\mathbf{H}\,(=\mathbf{B}/\mu)$ be continuous across a boundary. In general we can write $\mathbf{H} = n\hat{\mathbf{k}} \times \mathbf{E}$, where $\hat{\mathbf{k}}$ is a unit vector parallel to the direction of

propagation, since E and H both have plane-wave solutions and are related by Faraday's law (5.2.2).

The incident electric amplitude (from a region with index $n_1$) is $E_0$; the transmitted (to a region with $n_2$) is $E_0'$; the reflected, $E_0''$. Thus the two conditions require

$$\begin{aligned} E_0 + E_0'' &= E_0' \\ n_1 E_0 - n_1 E_0'' &= n_2 E_0' \end{aligned} \tag{5.2.36}$$

Eliminating $E_0'$ we have the reflection coefficient for the intensity,

$$R = \left(\frac{E_0''}{E_0}\right)^2 = \left(\frac{1 - n_2/n_1}{1 + n_2/n_1}\right)^2 \tag{5.2.37}$$

When $n_2 \to 0$ the boundary is totally (i.e., critically) reflecting.

Suppose now the wave enters from a region where $n_1 = 1$ into one where $n_2$ is given by (5.2.19). For the case of $n_2 \approx 1$ (i.e., $\omega$ far above the critical frequency $\omega_0$), (5.2.14) gives

$$1 - n_2 = \frac{N_e e^2}{2\pi m \nu^2} \tag{5.2.38}$$

and the reflection coefficient for a sharp surface is

$$R = \frac{N_e{}^2 e^4}{16\pi^2 m^2 \nu^4} = 4 \times 10^{-10} \frac{N_e{}^2}{\nu^4} \tag{5.2.39}$$

where the numerical value applies when $N_e$ is in electron/cm$^3$ and $\nu$ in MHz. Thus at 100 MHz and $N_e = 5 \times 10^6$ cm$^{-3}$, we have $R = 10^{-4}$, but $R$ decreases rapidly toward higher frequencies.

The difficulty with auroral reflections is that the boundaries between different indices of reflection are diffuse, not sharp. This complication does not generally enter the problem in optical reflection and we shall examine it in some detail. So long as there are no irregularities in electron density in a plane perpendicular to the direction of propagation, we may use the wave equation (5.2.7). Again, ignoring absorption and taking $\kappa = \kappa_0$ from (5.2.13), we have

$$\frac{\partial^2 \mathbf{E}(z,t)}{\partial z^2} = \frac{\omega^2 - \omega_0{}^2}{\omega^2 c^2} \frac{\partial^2 \mathbf{E}(z,t)}{\partial t^2} \tag{5.2.40}$$

Removing the time dependence by $E(z,t) = E_1(z)\exp(-i\omega t)$ gives the amplitude equation

$$\frac{\partial^2 E_1(z)}{\partial z^2} + \frac{\omega^2}{c^2} E_1(z) = \frac{\omega_0{}^2(z)}{c^2} E_1(z) \tag{5.2.41}$$

If $\omega_0^2(z)$ were a step function with value $\omega_0^2(z) = 0$ for $z < 0$ and $\omega_0^2 =$ const for $z > 0$, then (5.2.41) admits plane-wave solutions. Outside the scattering medium ($z < 0$) we have simply

$$E_1(z) = E_0 e^{i\omega z/c} + E_0'' e^{-i\omega z/c} \tag{5.2.42}$$

With $\omega_0^2(z)$ gradually changing with $z$, the perturbed wave is the summation of the perturbations produced by every individual element of volume. Thus we have

$$E_1(z) = E_0 e^{i\omega z/c} - \int_{-\infty}^{\infty} G(z|z_0) \frac{\omega_0^2(z)}{c^2} E_1(z_0)\, dz_0 \tag{5.2.43}$$

where $G(z|z_0)$ is the perturbation in the wave amplitude at $z$ produced by a single perturbation in the index of refraction at $z_0$. This *Green's function* is a solution of the equation

$$\frac{d^2G}{dz^2} + \frac{\omega^2}{c^2} G = -\delta(z - z_0) \tag{5.2.44}$$

which describes a simple oscillator with a $\delta$-function driving force. Its solution is

$$G(z|z_0) = -\frac{c}{2i\omega} e^{i\omega|z - z_0|/c} \tag{5.2.45}$$

From (5.2.43) the solution, with $\omega_0$ given in terms of $N_e$ by (5.2.14), is

$$E_1(z) = E_0 e^{i\omega z/c} + \frac{2\pi e^2}{i\omega mc} \int_{-\infty}^{z} e^{i\omega(z - z_0)/c} E_1(z_0) N_e(z_0)\, dz_0$$

$$+ \frac{2\pi e^2}{i\omega mc} \int_{z}^{\infty} e^{i\omega(z_0 - z)/c} E_1(z_0) N_e(z_0)\, dz_0 \tag{5.2.46}$$

The first term on the right side is the plane wave incident from $z = -\infty$. In the limit of $z \to -\infty$, (5.2.46) gives both the incident and reflected wave. Suppose the scattering region has only a minor perturbation on the incident wave (as when $\omega \gg \omega_0$), then we can adopt the *Born approximation* and write

$$E_1(z_0) \approx E_0 e^{i\omega z_0/c} \tag{5.2.48}$$

and the reflected wave amplitude at $z \to -\infty$ is

$$E_0'' = \frac{2\pi e^2 E_0}{i\omega mc} \int_{-\infty}^{\infty} e^{i2\omega z_0/c} N_e(z_0)\, dz_0 \tag{5.2.49}$$

The reflection coefficient becomes, with an integration by parts,

$$R = \left(\frac{E_0''}{E_0}\right)^2 = \frac{e^4}{16\pi^2 m^2 \nu^4}\left|\int_{-\infty}^{\infty} e^{i2\omega z_0/c}\,\frac{dN_e(z_0)}{dz_0}\,dz_0\right|^2 \tag{5.2.50}$$

Unless the boundary is much sharper than a wavelength, interference between waves reflected from different distances into the boundary drastically reduces the reflected amplitude. When $dN_e/dz_0 = N_e\delta(z_0)$—when the boundary is sharp—equation (5.2.50) reduces to the Fresnel formula (5.2.39).

If $\nu$ is near the critical frequency $\omega_0/2\pi$ for the maximum $N_e$ in the reflecting volume, as it is in observations of *sporadic E*, the Born approximation (5.2.48) is not valid. As in the analogous situation for scattering in quantum-mechanical problems, one may then resort to the so-called WKB approximation.

### 5.2.4 Incoherent Scattering and Small-Scale Irregularities

The ionosphere so far has been treated as a fluid medium—a *plasma*. Under certain circumstances the ionosphere ceases to behave as a unified plasma and electrons or ions respond individually to an electromagnetic impulse. The coherent response of a plasma to a wave arises because all electrons in a small region of dimension $a \ll \lambda$, where $\lambda$ is the wavelength, are subjected to the same external driving force. If, however, the wavelength is sufficiently small—say, smaller than interparticle spacings ($\lambda \lesssim N_e^{-1/3}$)—the concept of a fluid becomes meaningless and the wave interacts with individual electrons. In most situations, and certainly for the ionosphere, another limit, the Debye screening length $\lambda_D$, is a more severe restriction. That is, for the plasma to behave coherently the wave must have $\lambda \gtrsim \lambda_D$ and usually we find $\lambda_D > N_e^{-1/3}$.

With the wavelength related to frequency by $\lambda\omega/2\pi = c/n$ (the phase velocity), in a collisionless medium the dispersion relation (5.2.19) for the wave becomes

$$\frac{c}{\lambda} = \frac{(\omega^2 - \omega_0{}^2)^{1/2}}{2\pi} \tag{5.2.51}$$

Similarly, if we consider the mean motion of electrons in more detail, a pressure gradient term [see (2.1.5)] enters the derivation and competes with the electric force, leading to a dispersion relation for the electron oscillations:

$$\frac{\langle 3v^2\rangle_{\text{th}}^{1/2}}{\lambda_e} \approx \frac{(\omega^2 - \omega_0{}^2)^{1/2}}{2\pi} \tag{5.2.52}$$

where $\lambda_e$ is the wavelength of electron oscillations and $\langle v^2 \rangle_{\text{th}}^{1/2}$ is their rms thermal velocity. For large electron oscillations $\lambda_e$, the frequency is $\omega \approx \omega_0$. But at high frequencies, $\omega \gg \omega_0$, the electrons have much shorter oscillations. The Debye screening length is the order of $\langle v^2 \rangle_{\text{th}}^{1/2}/\omega_0$; specifically it is defined as

$$\lambda_D = \left(\frac{kT}{4\pi e^2 N_e}\right)^{1/2} = 6.90\left(\frac{T}{N_e}\right)^{1/2} \tag{5.2.53}$$

For $T = 10^{3\circ}$K and $N_e = 10^5$ cm$^{-3}$, we have $\lambda_D = 0.7$ cm, whereas $N_e^{-1/3} = 0.02$ cm. Within the distance $\lambda_D$ the Coulomb field of an individual charge is distinguishable and the charge responds to an electromagnetic wave as an individual. Outside this distance the Coulomb field of an electron is damped by the cloud of other charges surrounding it and it is merely part of this element of fluid, responding to a wave as a body.

Electrons scatter unpolarized incident light through an angle $\Theta$, according to the Thomson differential cross section (Problem 5.9),

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{mc^2}\right)^2 \tfrac{1}{2}(1 + \cos^2\Theta) \text{ (cm}^2\text{/sr)} \tag{5.2.54}$$

where the factor $\frac{1}{2}$ accounts for the equal division of incident flux into two perpendicular polarizations. If the total flux incident on a volume $V$ is $\pi F$ and if the electrons are far apart ($\lambda \ll N_e^{-1/3}$), the total scattered intensity is simply additive:

$$I_{\text{sca}}(\Theta) = \frac{N_e V(1 + \cos^2\Theta)}{2r^2}\left(\frac{e^2}{mc^2}\right)^2 (\pi F) \tag{5.2.55}$$

where $r$ is the distance from the electron. The length $e^2/mc^2 = 2.82 \times 10^{-13}$ cm is the classical radius of the electron.

Incoherent scatter from high-power radars, whose frequency penetrates the F-region peak, is a means of sounding the topside of the ionosphere from the ground. (Another means involves the use of satellites viewing downward.) The velocity dispersion causes a narrow frequency band to be Doppler broadened on its return, providing information on the electron temperature. If the probing wavelength is $\lambda \gtrsim \lambda_D$, the electrons will not incoherently scatter, as noted above, and the scattering occurs mainly from ion-acoustic and ion-cyclotron waves. In this fashion it is possible to investigate the positive-ion species as well.

The scattering by a small volume $V$ containing free electrons whose scattered waves interfere may be obtained by supposing that each electron Thomson-scatters independently and then adding **E** vectors, with regard

for phase interference, much as we did in (5.2.49) for partial reflection. This technique is the *Rayleigh–Gans approximation*, valid for small phase shifts. Thus the scattered intensity is

$$I(\Theta) = \frac{V^2(1 + \cos^2\Theta)}{2r^2} \frac{N_e{}^2 e^4}{m^2 c^4} (\pi F) |\mathscr{R}(\theta, \phi)|^2 \tag{5.2.56}$$

where the *interference factor* (which depends on the particular geometry) is

$$\mathscr{R}(\theta, \phi) = \frac{1}{V} \int e^{i\delta} \, dV \tag{5.2.57}$$

and $\delta$ is the phase lag for each point in the cylinder referred to some origin. A simple example is given as Problem 5.10.

As another example, for a cylinder of length $l$ and diameter $d$ scattering backward ($\Theta = \pi$), the interference factor is

$$\mathscr{R} = \frac{2}{x} J_1(x) \left(\frac{\pi}{2y}\right)^{1/2} J_{1/2}(y) \tag{5.2.58}$$

where $J_1$ and $J_{1/2}$ are Bessel functions, $x = (2\pi d/\lambda)\cos\psi$, $y = (2\pi l/\lambda)\sin\psi$, and $\pi/2 - \psi$ is the angle between the direction of incidence and the axis of the cylinder. To further simplify, suppose we observe at right angles to a cylinder. For example, a tube of ionization along a magnetic field line is probed with a radar beam approximately perpendicular to the field; then $\psi \sim 0$. For small dimensions ($x \ll 1$ and $y \ll 1$) and thin rods ($d \ll l$ or $x \ll y$), we have, by expanding the Bessel functions,

$$\mathscr{R} \approx \left(1 - \frac{x^2}{8}\right)\left(1 - \frac{y^2}{6}\right) \tag{5.2.59}$$

The intensity back-scattered by a single cylinder is then

$$I(\pi) = \frac{\pi^2 d^4 l^2 N_e{}^2 e^4}{16 r^2 m^2 c^4} \left(1 - \frac{\pi^2 d^2}{\lambda^2}\right)\left(1 - \frac{4}{3}\frac{\pi^2 l^2 \psi^2}{\lambda^2}\right)(\pi F) \tag{5.2.60}$$

If the cylinders of density $N_e + \Delta N_e$ are embedded in a uniform background $N_e$, we simply replace $N_e{}^2$ above with $(\Delta N_e)^2$. Irregularities in $N_e$ play a major role in ionospheric observations. Certain types of sporadic $E$, the spread $F$ phenomenon, and most radio-aurorae are almost certainly due to inhomogeneous patches of ionization, even though the origin of the irregularities may not be well understood.

Note that for the cylindrical geometry the dependence of the reflection on angle $\psi$ enters with the length $l$. In the case of auroral reflections, the $\psi$ dependence is called *aspect sensitivity*. The strongest reflections usually

occur when the beam is perpendicular to the field lines ($\psi = 0$) and the intensity falls off rapidly away from this geometry.

### 5.2.5 Magneto-Ionic Theory for a Collisionless Ionosphere

The complete *Appleton formula* describes the propagation of a wave in an arbitrary direction in an ionized medium with an external magnetic field, including the effects of collisions. We have previously (see Section 5.2.1) examined the role of collisions when there is no external field, but in the interest of simplicity we will here omit collisional damping. In addition we will omit the second-order effects of the magnetic field perpendicular to the direction of propagation. The result, (5.2.70), is a simplified version of the Appleton dispersion relation.

The wave is assumed to be propagating in the $z$ direction and the magnetic field $\mathbf{B}$ is in an arbitrary direction, with components $B_x$, $B_y$, $B_z$. We take the permeability as $\mu = 1$. The equations of motion for an electron accelerated by the electric vector of the wave $\mathbf{E}$ and subject to the constraint provided by the *Lorentz force* $e\mathbf{v} \times \mathbf{B}/c$, are

$$\frac{d^2x}{dt^2} = -\frac{e}{m} E_x(t) - \frac{e}{mc}(\mathbf{v} \times \mathbf{B})_x$$

$$= -\frac{e}{m} E_x(t) - \omega_z \frac{dy}{dt} \tag{5.2.61}$$

and

$$\frac{d^2y}{dt^2} = -\frac{e}{m} E_y(t) + \omega_z \frac{dx}{dt} \tag{5.2.62}$$

where

$$\omega_z = eB_z/mc \tag{5.2.63}$$

is the *gyrofrequency*. We have neglected here a second-order effect caused by the $B_x$ and $B_y$ components and the magnetic vector of the wave; they give rise to an oscillating $z$-component of electron motion. The important motion occurs in the $xy$ plane; hence we can simplify the treatment by using complex quantities:

$$u \equiv x + iy, \qquad E = E_x + iE_y, \qquad B = B_x + iB_y \tag{5.2.64}$$

The equations of motion become

$$\frac{\partial^2 u}{\partial t^2} \equiv \frac{\partial^2 x}{\partial t^2} + i\frac{\partial^2 y}{\partial t^2} = -\frac{e}{m} E(t) + i\omega_z \frac{\partial u}{\partial t} \tag{5.2.65}$$

Faraday's law (5.2.2) is

$$\frac{\partial E}{\partial z} = \frac{i}{c}\frac{\partial B}{\partial t} \tag{5.2.66}$$

and Ampère's law (5.2.1) is

$$\frac{\partial B_z}{\partial z} = i\frac{4\pi N_e e}{c}\dot{u} - \frac{i\kappa}{c}\frac{\partial E}{\partial t} \tag{5.2.67}$$

A solution to this set of three simultaneous equations is

$$E = E_0 e^{\pm i\omega(t-nz/c)}, \qquad B = B_0 e^{\pm i\omega(t-nz/c)}$$
$$u = u_0 e^{\pm i\omega(t-nz/c)} \tag{5.2.68}$$

These waves in $E$, $B$, and $u$ travel in the $+z$ direction. The ambiguity of sign in the exponent does not matter in the absence of a magnetic field, but the option gives rise to two solutions when $B \neq 0$. Thus substituting (5.2.68) into (5.2.65–5.2.67) yields

$$\begin{aligned} \frac{e}{m}E_0 - (\omega^2 \mp \omega\omega_z)u_0 &= 0 \\ nE_0 + iB_0 &= 0 \\ \kappa E_0 + inB_0 - 4\pi N_e e u_0 &= 0 \end{aligned} \tag{5.2.69}$$

For this set of equations to have a nontrivial solution, their determinant must vanish. Expansion of the determinant gives the two solutions for the index of refraction:

$$n^2 = \kappa - \frac{{\omega_0}^2}{\omega^2 \mp \omega\omega_z} \tag{5.2.70}$$

where the plasma frequency $\omega_0$ is given by (5.2.14) and where $\kappa = 1$ for the collisionless case. For a field of $B_z = 1$ G, $eB_z/mc = 17.6$ Mrad/sec. The Earth's field is 0.315 G at the magnetic equator and twice as strong at the poles. Thus the gyrofrequency $\omega_z$ is of the same order as the plasma frequency $\omega_0$ for the lower ionosphere and therefore of the same order as the frequencies $\omega$ used to probe the ionosphere. [For $B = 0.5$ G, $\omega_0 = \omega_z$ corresponds to $N_e = 2.43 \times 10^4\ cm^{-3}$; see [5.2.20).]

The radio wave probing an ionosphere with a magnetic field splits into two modes, as does an optical wave in an anisotropic double-refracting crystal. The ionosphere thus has *magnetic birefringence*. For $\kappa = 1$, the *ordinary wave* is defined by

$${n_+}^2 = 1 - \frac{{\omega_0}^2}{\omega(\omega + \omega_z)} \tag{5.2.71}$$

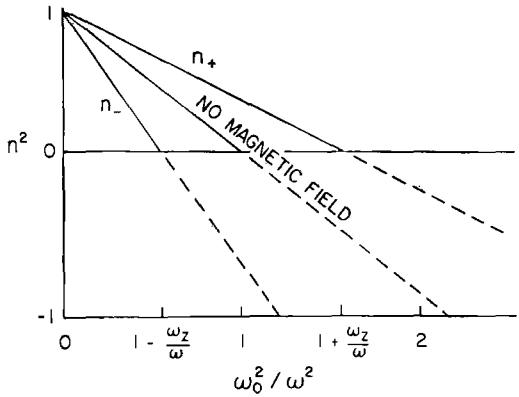



**Fig. 5.9** Variation of index of refraction with the plasma frequency $\omega_0$ for propagation along the magnetic field and when $\omega > \omega_z$ (wave frequency exceeds the gyrofrequency). In this situation the extraordinary wave would be critically reflected at a lower plasma frequency than the ordinary one. Compare Fig. 5.1.

which gives critical reflection ($n_+ = 0$) for

$$\omega = \left(\omega_0{}^2 + \frac{\omega_z{}^2}{4}\right)^{1/2} - \frac{\omega_z}{2} \tag{5.2.72}$$

The *extraordinary wave* has

$$n_-{}^2 = 1 - \frac{\omega_0{}^2}{\omega(\omega - \omega_z)} \tag{5.2.73}$$

and this index becomes infinite for $\omega = \omega_z$. A linearly polarized wave transmitted into the ionosphere becomes split into right- and left-handed circularly polarized components, propagating with the two different velocities. (Right-handed polarization is a clockwise rotation when the wave is viewed toward the direction of propagation.) The wave returning to Earth may be drastically changed from the transmitted wave, frequently containing the separate circularly polarized components (see Fig. 5.1). Figure 5.9 illustrates dispersion relations for the ionosphere when propagation is along the magnetic field, which gives the main, first-order splitting of modes.

## 5.3 Ionospheres of Venus, Mars, and Jupiter

### 5.3.1 Venus' and Mars' Chapman Layers

Both Venus' and Mars' atmospheres consist mainly of $CO_2$ and their dayside ionospheres are somewhat similar. The models shown in Fig. 5.10 and 5.11 are also each based on exospheric temperatures of about 350°K.

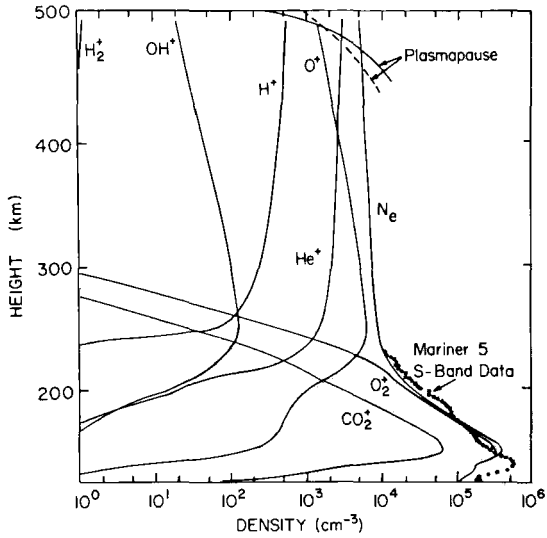



**Fig. 5.10** Model of the Venus ionosphere for an exosphere temperature of 350°K. [The Mariner 5 S-band data are from KLIORE, A. *et al.* (1967), *Science* **158**, 1683; the plasmapause was deduced with 49.8 MHz data from Mariner 5 by FJELDBO, G. and ESHLEMAN, V. R. (1969), *Radio. Sci.* 4, 879; model developed by KUMAR and HUNTEN (1974).]

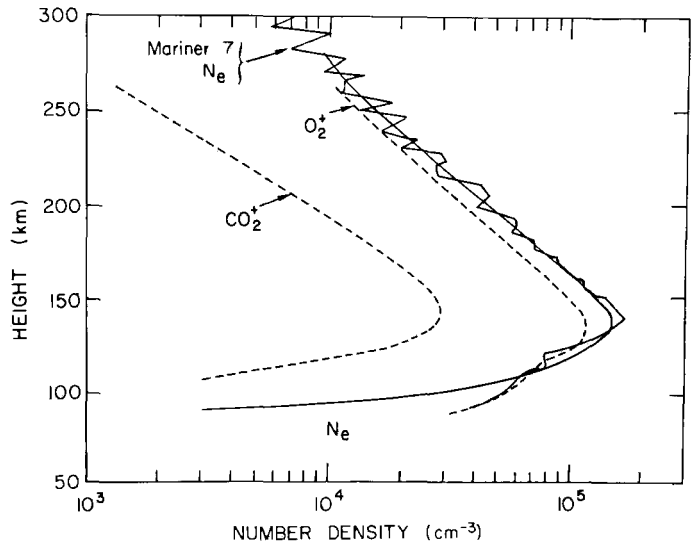



**Fig. 5.11** Model of the Mars ionosphere for an exosphere temperature of 364°K. [The Mariner 7 data are from RASOOL, S. I. and STEWART, R. W. (1971), *J. Atmos. Sci.* **28**, 869; adapted from MCCONNELL (1976).]

Both ionospheres are basically of the Chapman type—at least in the region of Mariner spacecraft observations. There has been some dispute as to whether they should be regarded as E or F1. Because $CO_2$ remains the dominant constituent far above the layers' peaks, it is the main ionizable source to high altitude. Solar X-rays and ultraviolet longward of 900 Å (being absorbed by minor constituents) will be deposited much lower, and one could argue that this deposition produces an ionization enhancement corresponding to an E region; hence the higher Chapman layer should be identified as F1. However, photoionization of $CO_2$ corresponds to the direct $O_2$ and $N_2$ ionization in the E region on Earth and, more importantly, an atomic ion does not assume the role that $O^+$ does in Earth's F1 by slowing the chain of recombination reactions. Thus in spite of the fact that the molecular densities at the ionospheres' peaks suggest F1, the physics of ion production and recombination indicates E-type regions. That the Martian ionosphere is indeed Chapman-like is substantiated by the daily variation of the height and electron density of the peak [see Eq. (5.1.6)].

Ion production occurs by

$$CO_2 + h\nu \rightarrow CO_2^+ + e, \qquad \lambda < 900\ \text{Å} \tag{5.3.1}$$

which has a rate outside the atmosphere at Venus of $J_1 = 7.5 \times 10^{-7}\ \text{sec}^{-1}$, and

$$O + h\nu \rightarrow O^+ + e, \qquad \lambda < 911\ \text{Å} \tag{5.3.2}$$

with $J_2 = 4 \times 10^{-7}\ \text{sec}^{-1}$.

The main ambient ion is $O_2^+$, which is quickly formed by atom-ion interchange,

$$O + CO_2^+ \rightarrow O_2^+ + CO, \qquad k_3 = 1.6 \times 10^{-10}\ \text{cm}^3/\text{sec} \tag{5.3.3}$$

or by charge transfer

$$O + CO_2^+ \rightarrow O^+ + CO_2, \qquad k_4 = 9.6 \times 10^{-11}\ \text{cm}^3/\text{sec} \tag{5.3.4}$$

quickly followed by

$$O^+ + CO_2 \rightarrow O_2^+ + CO, \qquad k_5 = 1.1 \times 10^{-9}\ \text{cm}^3/\text{sec} \tag{5.3.5}$$

Both $O_2^+$ and $CO_2^+$ disappear by dissociative recombination,

$$\left.\begin{aligned} CO_2^+ + e &\rightarrow CO + O \\ O_2^+ + e &\rightarrow O + O \end{aligned}\right\} \quad k_6 \approx 3 \times 10^{-7}\ \text{cm}^3/\text{sec} \tag{5.3.6}$$

The small degree of $CO_2$ dissociation in the ionospheres of both planets is surprising but well established from the intensities of the O($\lambda$1302) and CO (fourth positive and Cameron systems) airglow emissions (see Section

6.3). The O emission is excited primarily by fluorescence, and its brightness gives directly the O abundance and distribution. Thus Mariner observations indicate that on Mars at the electron peak the $[O]/[CO_2]$ and $[CO]/[CO_2]$ mixing ratios are only the order of 1 percent. On Venus the ratio $[O]/[CO_2]$ could be as high as 10–20 percent, which is still not enough to give a substantial $O^+$ abundance at the peak.

This comparative lack of dissociation at densities of $[CO_2] \sim 10^{10}$ cm$^{-3}$ is not well accounted for. Dissociation occurs by

$$CO_2 + h\nu \rightarrow CO + O, \qquad \lambda < 1690 \text{ Å} \tag{5.3.7}$$

and recombines, albeit very inefficiently, by

$$CO + O + M \rightarrow CO_2 + M, \qquad k_8 \approx 2 \times 10^{-37} \text{ cm}^6/\text{sec} \tag{5.3.8}$$

Without a more efficient way of restoring $CO_2$, the O produced by (5.3.7) would become abundant enough to self-associate by

$$O + O + M \rightarrow O_2 + M, \qquad k_9 \sim 3 \times 10^{-34} \text{ to } 3 \times 10^{-33} \text{ cm}^6/\text{sec} \tag{5.3.9}$$

which is then balanced by

$$O_2 + h\nu \rightarrow O + O, \qquad \lambda < 1750 \text{ Å} \tag{5.3.10}$$

Thus one would have expected to find copious amounts of O, CO, and $O_2$ in the upper atmospheres of Venus and Mars. In the absence of some overlooked and rapid in situ chemistry, it appears necessary to postulate rapid vertical transport (see Section 2.3). An eddy diffusion coefficient of $K \sim 10^8$ cm$^2$/sec is required for Mars (compared with $K \sim 10^6$ to $5 \times 10^6$ cm$^2$/sec for comparable density levels on Earth). Why that should be is not clear. One possibility is that turbulent heating from the dissipation of planetary (Rossby) waves is more important at high altitudes than on Earth, where dissipation occurs in the stratosphere and mesosphere. Once transported to low altitudes O and CO can associate catalytically with the aid of OH chemistry (Section 6.4).

From (5.1.9) we see that, well above the electron peak,

$$N_e(z) \rightarrow \text{const} \exp\left(-\frac{z - z_M}{2H}\right) \tag{5.3.11}$$

or in a Chapman layer the electron density (fortuitously) approaches the diffusive-equilibrium value (for equal electron, ion, and neutral temperatures) in which the ion scale height is twice that of the neutral parent. The $CO_2$-related airglow measurements (Section 6.3) and radio occultations confirm for Mars that this scale-height relationship is at least approximately satisfied with average values of $H(CO_2) \approx 18$ km and $H(e^-) \approx 38.5$ km.

The situation for Venus is less satisfactory for several reasons: (1) The exospheric temperature that thermospheric models must fit is uncertain because the data seem ambiguous (Section 7.3.3). (2) There is somewhat more $CO_2$ dissociation, but just how much is uncertain. Hence the mean molecular weight and the degree to which O becomes a major source for ionization by sunlight above 200 km are uncertain. (3) The dayside ionosphere has a rather sharp ionopause, but it was different as seen by Mariners 5 and 10. The ionopause seen by M5 was $N_e \sim 10^4$ cm$^{-3}$ at 500 km altitude; for M10, $N_e \sim 10^3$ cm$^{-3}$ at 350 km. These results suggest direct interaction of the solar wind with the upper atmosphere, possibly with an energy source feeding the thermosphere. It is not clear whether the plasmapause changes primarily because of a change in solar-wind momentum or because of a variability in thermospheric structure that alters the atmospheric densities at very high altitudes. For these reasons, and perhaps others, the models for Venus have not fit the available data above 200 km in a very satisfactory way.

### 5.3.2 Nightside Ionization on Venus

A major surprise from the Venus Mariners was the detection of a nightside ionosphere with a peak $N_e \sim 10^4$ cm$^{-3}$ at 140 km altitude. Venus' sidereal period of rotation is $P_{\text{sid}} \sim -244.3$ days (the minus sign meaning westward or retrograde); its year is $P_{\text{rev}} = 225$ days; hence its solar rotation period, given by

$$\frac{1}{P_{\text{sol}}} = \frac{1}{P_{\text{sid}}} - \frac{1}{P_{\text{rev}}} \tag{5.3.12}$$

is $P_{\text{sol}} = -117$ day $= -1.01 \times 10^7$ sec. The time required for rotation of a point on the terminator to the region of the Mariner 10 observations is about 10 percent of the rotation period, giving a co-rotation time scale of

$$\tau_{\text{CR}} \sim 10^6 \text{ sec} \tag{5.3.13}$$

Even a four-day co-rotation period, sometimes inferred for the middle atmosphere (see Section 2.4), gives a time scale of

$$\tau_{4\text{d}} \sim 3 \times 10^4 \text{ sec} \tag{5.3.14}$$

The dayside ionosphere contains predominantly the molecular positive ions $O_2^+$ and $CO_2^+$, which recombine with free electrons by dissociative recombination (5.1.45) with rate coefficients $\alpha_{\text{rec}} \sim 3 \times 10^{-7}$ cm$^3$/sec. Hence

an initial $N_e(0) \sim 5 \times 10^5$ cm$^{-3}$ at sunset will decay in four days to

$$N_e(\tau_{4d}) = \left(\frac{1}{N_e(0)} + \alpha_{\text{rec}}\tau_{4d}\right)^{-1} \approx (\alpha_{\text{rec}}\tau_{4d})^{-1}$$
$$= 10^2 \text{ cm}^{-3} \tag{5.3.15}$$

Thus to maintain the nightside ionization, it is necessary to transport the dayside ionization very rapidly, have a nonmolecular ion component that does not recombine so rapidly, or create ionization directly on the nightside. A variety of mechanisms has been examined in these categories, and the most likely seems to involve metallic ions ($Mg^+$, $Na^+$, $Si^+$, $Fe^+$).

Such ions as these are known to be formed in the Earth's lower ionosphere as a result on meteor ablation, and they probably form one type of sporadic $E$ seen most frequently in middle latitudes. Also, atomic ions recombine with free electrons by radiative recombination,

$$Mg^+ + e \rightarrow Mg + h\nu, \qquad \alpha_{16} \sim 3 \times 10^{-12} \text{ cm}^3/\text{sec} \tag{5.3.16}$$

If this mechanism were the sole means of recombination, it would increase the ion lifetime to $(N_e\alpha_{16})^{-1} \sim 10^6$ sec.

Two other considerations are important. Metallic ions have low ionization potentials and will not transfer their charge to other (molecular) constituents On the other hand $O^+$, for example, is not long lived because of ion-atom interchange,

$$O^+ + CO_2 \rightarrow O_2^+ + CO, \qquad k_{17} = 1.1 \times 10^{-9} \text{ cm}^3/\text{sec} \tag{5.3.17}$$

giving a lifetime for $O^+$ of less than 0.1 sec. Indeed, the metals pick up charge on the dayside by exothermic charge transfer,

$$Mg + O_2^+ \rightarrow Mg^+ + O_2, \qquad k_{18} \approx 2 \times 10^{-9} \text{ cm}^3/\text{sec} \tag{5.3.18}$$

The second consideration is that negative ions must not form rapidly. If they did, not only would free electrons (which are the charges observed by radio occultations) become depleted, but the positive ions would disappear rapidly by mutual neutralization,

$$Mg^+ + X^- \rightarrow Mg + X, \qquad k_{19} \sim 3 \times 10^{-7} \text{ cm}^3/\text{sec} \tag{5.3.19}$$

In the Earth's ionosphere the primary negative ion—the one most readily formed directly from electron attachment—is $O_2^-$. On Venus the $O_2$ abundance is very small so that $O^-$ becomes the primary ion, formed by radiative attachment,

$$O + e \rightarrow O^- + h\nu, \quad k_{20} = 1.3 \times 10^{-15} \text{ cm}^3/\text{sec} \tag{5.3.20}$$

But since O is also a minor constitutent, the rate of formation is slow and, once formed, $O^-$ is destroyed in seconds by

$$O^- + CO \rightarrow CO_2 + e, \qquad k_{21} = 6 \times 10^{-10}\ \text{cm}^3/\text{sec} \tag{5.3.21}$$

These two reactions alone would give $[O^-]/N_e \sim 2 \times 10^{-6}$. Alternatively, $O^-$ might transfer its electron to a constituent with a higher electron affinity, as $O_2^-$ does in the Earth's D region, so that the negative ions become relatively stable. The importance of mutual neutralization could then be increased. With charge transfer.

$$O^- + X \rightarrow O + X^-, \qquad k_{22} \sim 3 \times 10^{-10}\ \text{cm}^3/\text{sec} \tag{5.3.22}$$

we would have a density $[X^-] > [O^-]$ if the neutral parent had a mixing ratio $[X]/[CO_2] \gtrsim 10^{-4}$. In that case mutual neutralization would become more important than radiative recombination in removing metallic ions.

A number of factors are uncertain in this analysis, not the least of which is the concentration of minor constituents in the Venus atmosphere. The surprises encountered on the behavior of ionization in Earth's D region are sufficient to instill caution. Nevertheless, it seems reasonable to think that metal ionization produced by charge transfer on the dayside and transported by rotation to the nightside, plus ionization of metals produced on the nightside itself by meteor ablation, may account for the Venus nightside ionosphere.

Another suggestion is that the shock front in the solar wind, caused by its impact on the upper atmosphere, interacts viscously with the local ionosphere, and this mixed region of hot and cold plasma is forced into the nightside upper atmosphere. Only one or two percent of the solar-wind energy is necessary to account for the nighttime ionization, and since a mechanism for delivering energy into the umbra exits, a solar-wind source of the ionosphere can scarcely be dismissed. The theory of viscous transfer in the near-wake flow past the planet requires further development, however.

### 5.3.3 Jupiter's Giant Ionosphere

Pioneer 10 and 11 provided another example of a planetary ionosphere that did not fit theoretical expectations. Considering only the solar ultraviolet light that is absorbed in the thermosphere, we would estimate the temperature to rise only 15°K or so above the mesopause (see Section 1.8.3) to a value of about 165°K. In fact, the radio-occultation data indicate a rise to an exospheric temperature of about 750°K. Figure 5.12 shows the observed electron-density distribution with height.

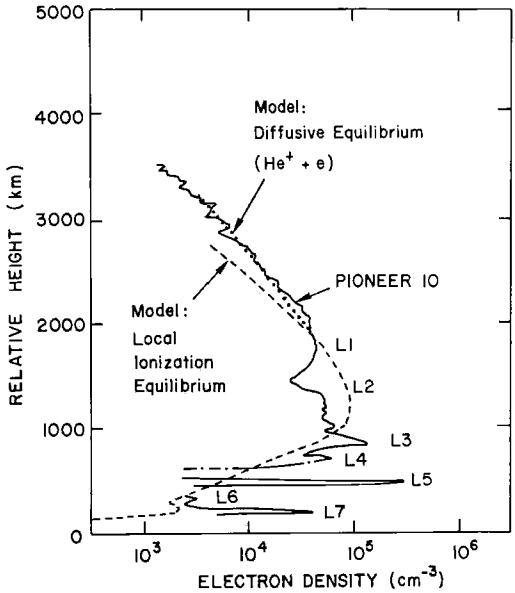



**Fig. 5.12** Model of the Jupiter ionosphere for an exosphere temperature of 1050°K. The model is calculated to fit observations of Pioneer 10 at latitude 26°N and a solar zenith angle of 81° (afternoon). The observed peaks (L1, L2, etc.) may or may not be real layers of ionization. [Adapted from ATREYA and DONAHUE (1976).]

The unexpected source of heating at the low densities of the thermosphere is not definitely identified. The most reasonable hypothesis is that the lower atmosphere generates gravity waves that propagate into the thermosphere, where they are dissipated into heat. Possibly the neutral temperature is lower than the mean plasma temperature, but a heat source is needed all the same.

The principal reactions producing ions in the Jovian atmosphere are photoionization,

$$H_2 + h\nu \rightarrow H_2^+ + e \tag{5.3.23}$$

$$CH_3 + h\nu \rightarrow CH_3^+ + e \tag{5.3.24}$$

$$He + h\nu \rightarrow He^+ + e \tag{5.3.25}$$

and *dissociative photoionization*,

$$H_2 + h\nu \rightarrow H^+ + H + e \tag{5.3.26}$$

The latter process is important as the main source of *primary* $H^+$ ions.

Of these primary ions, $H_2^+$ disappears by dissociative recombination,

$$H_2^+ + e \rightarrow H + H \tag{5.3.27}$$

and $CH_3^+$, which is produced at comparatively low altitudes by solar Lyman $\alpha$, reacts with methane by

$$CH_3^+ + CH_4 \rightarrow C_2H_5^+ + H_2 \tag{5.3.28}$$

which is followed by dissociative recombination.

Of more interest are the atomic ions $H^+$ and $He^+$, which may lend the lower Jupiter ionosphere an F1 character. The $He^+$ ion mainly reacts with $H_2$ to give a variety of products, the most important of which is

$$He^+ + H_2 \rightarrow H^+ + H + He \tag{5.3.29}$$

The protons react with $CH_4$ to form molecular ions at low altitude; somewhat higher, they form a molecular ion by three-body association,

$$H^+ + H_2 + H_2 \rightarrow H_3^+ + H_2 \tag{5.3.30}$$

Above the electron peak, the molecular densities are probably too low for chemical reactions to be important compared with the slow radiative recombination,

$$H^+ + e \rightarrow H + h\nu \tag{5.3.31}$$

which accounts for electron densities as high as $10^5$ cm$^{-3}$ on a planet so far (5 AU) from the sun.

The agreement of calculations with the Pioneer observations is only fair. The uncertainties include the eddy-diffusion coefficient $K$ (which is usually modified to obtain good fits to the data); the rate coefficients for many of the important reactions; the composition of the neutral atmosphere; and the temperature distribution with height (which can only be crudely estimated from the noisy radio-occultation records). The huge spikes in the empirical curve may represent sharp layers (they have been compared to sporadic $E$). They may also be artifacts in the data created by turbulent-like inhomogeneities or weak signal levels such that multiple paths of the signal are present without being so identified.

## BIBLIOGRAPHICAL NOTES

**Section 5.1 Formation of Ionospheric Regions**

The existence of a conducting region in the atmosphere was first inferred by

STEWART, B. (1882), Hypothetical views regarding the connection between the state of the Sun and terrestrial magnetism, *Encyclopedia Britannica* (9th ed.) **16**, 181–184. [The date

1878 appearing on the title page of the volume cannot apply to this article, which contains at least some material written later.]

Stewart's conclusion was based on daily changes in the geomagnetic field. Much earlier a mathematical method for separating the magnetic field into two parts—one internal and one external to the Earth's surface—had been developed by

GAUSS, C. F. (1839), Allgemeine Theorie des Erd-magnetismus, *in* "Resultate aus den Beobachtungen des magnetischen Vereins im Jahre 1838," publ. C. F. Gauss and W. Weber, Leipzig. (Reprinted in *Werke* **5**, 119–180, Göttingen, *Kgl. Ges. Wiss.* 1877; English translation by Mrs. Sabine, revised by Sir John Herschel, "Scientific Memoirs," (R. Taylor, ed.), **2**, 184–251, London, 1841.)

With adequate data at last accumulated, the method was further developed and applied by

SCHUSTER, A. (1889), The diurnal variation of terrestrial magnetism, *Phil Trans. Roy. Soc. (London)* **A180**, 467–512,

who demonstrated conclusively that electric currents must be flowing overhead.

This geomagnetic history was not widely known outside the field, so when Marconi transmitted radio waves across the Atlantic on Dec. 12, 1901, an overhead conducting layer was invoked anew independently by

KENNELLY, A. E. (1902), On the elevation of the electrically conducting strata of the Earth's atmosphere, *Elec. World Eng.* **39**, 473,

HEAVISIDE, O. (1902), Telegraphy I. Theory, *Encyclopedia Britannica* (10th ed.) **33**, 213–218.

Kennelly placed the layer at 80 km, much higher than had been thought likely by Stewart and Schuster.

Direct evidence for the existence of the Kennelly–Heaviside layer (called the E layer by Appleton) was obtained through comparing the fading of signals received on two types of directional antennas by

APPLETON, E. V. and BARNETT, M. A. F. (1925), On some direct evidence for downward atmospheric reflection of electric waves, *Proc. Roy. Soc. (London)* **A109**, 621–641,

SMITH-ROSE, R. L. and BARFIELD, R. H. (1927), Further measurements on wireless waves received from the upper atmosphere, *Proc. Roy. Soc. (London)* **A116**, 682–693.

The most direct means of sounding the ionosphere is the use of radio pulses of shorter duration than the transit time to the ionosphere and back. Measurement of the time delay then gives directly the apparent (*virtual*) height of reflection. The original experiments of this nature, which gave dramatic proof of the existence and location of ionized layers, were performed by

BREIT, G. and TUVE, M. A. (1926), A test of the existence of the conducting layer, *Phys. Rev.* **28**, 554–575 (also *Nature* **116**, 357 [1925]).

The theory of ionospheric layering was developed most thoroughly by

CHAPMAN, S. (1931), The absorption and dissociation or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, *Proc. Phys. Soc.* **43**, 26–45 and 483–501.

Linear or $\beta$ recombination was introduced in connection with D-region ionization and negative ion formation in

BRADBURY, N. E. (1938), Ionization, negative-ion formation, and recombination in the ionosphere, *Terr. Magn.* **43**, 55–66.

That $\beta$ recombination is also dominant in the high F2 region, where the positive ions are mainly atomic and where charge transfer is the rate-limiting process for ion destruction is a concept developed by

BATES, D. R. and MASSEY, H. S. W. (1946), The basic reactions in the upper atmosphere I, *Proc. Roy. Soc. (London)* **A187**, 261–296

BATES, D. R. and MASSEY, H. S. W. (1947), The basic reactions of the upper atmosphere II, The theory of recombination in the ionized layers, *Proc. Roy. Soc.* (*London*) **A192**, 1–16.

The importance of $N_2$ ionization and the rapid conversion of $N_2^+$ to $NO^+$ by atom-ion interchange (5.1.14) were first demonstrated from a model based on rocket observations of solar fluxes and neutral densities by

NORTON, R. B.; VANZANDT, T. E.; and DENISON, J. S. (1963), A model of the atmosphere and ionosphere in the E and F1 regions, *Proc. Intern. Conf. on the Ionosphere* pp. 26–34, *Inst. Phys.* and *Phys. Soc.* (*London*).

Nighttime recombination in F2 and its dependence on ion diffusion was first developed in a reasonably accurate way by

DUNGEY, J. W. (1956), The effect of ambipolar diffusion in the night-time F layer, *J. Atm. Terr. Phys* **9**, 90–102.

The nighttime F2 solution given here is not highly accurate but it is probably the most realistic one that can be treated completely analytically. This discussion is based on

CHAMBERLAIN, J. W. (1961), "Physics of the Aurora and Airglow," Section 13.2, Academic Press, New York.

Two types of observation in the early 1960s demonstrated that $O_2^-$ could not be the dominant negative ion in the D region, but that a terminal ion of greater electron affinity must be formed. First, early observations of twilight variations of polar-cap absorption (PCA) events were interpreted as requiring solar ultraviolet, rather than visual, photodetachment by

REID, G. C. (1961), A study of the enhanced ionization produced by solar protons during a polar cap absorption event, *J. Geophys. Res.* **66**, 4071–4085.

Later, analysis of enhanced D-region radio absorption, produced by gamma rays from atmospheric nuclear tests in the Pacific in 1962, led to $CO_3^-$, $NO_2^-$, and $NO_3^-$ being invoked as possible terminal ions by

LELEVIER, R. E. and BRANSCOMB, L. M. (1968), Ion chemistry governing mesospheric electron concentrations, *J. Geophys. Res.* **73**, 27–41.

The complexity of the positive-ion make-up of the D region was demonstrated from rocket measurements by

NARCISI, R. S. and BAILEY, A. D. (1965), Mass spectrometric measurements of positive ions at altitudes from 64 to 112 kilometers, *J. Geophys. Res.* **70**, 3687–3700.

Development of D-region chemistry has dependended crucially on the laboratory measurement of reaction rates. Reviews on the subject include

FERGUSON, E. E. (1971), D-region ion chemistry, *Rev. Geophys. Space Phys.* **9**, 997–1008,
FERGUSON, E. E. (1974), Laboratory measurements of ionospheric ion-molecure reaction rates, *Rev. Geophys. Space Phys.* **12**, 703–713,
REID, G. C. (1973), Ion chemistry of the D and E regions, *in* "Physics and Chemistry of Upper Atmospheres," (B. M. McCormac, ed.), pp. 99–109, D. Reidel Publ. Co., Dordrecht, The Netherlands.

The rate coefficients quoted here are taken from summary tables contained in

BORTNER, M. H. and BAURER, T. (eds.) (1972), "DNA Reaction Rate Handbook," Defense Nuclear Agency, Washington, D.C.

General and comprehensive surveys of ionospheric physics are provided by

BAUER, S. J. (1973), "Physics of Planetary Ionospheres," Springer-Verlag, Berlin,
RISHBETH, H. and GARRIOTT, O. K. (1969), "Introduction to Ionospheric Physics," Academic Press, New York.

**Section 5.2 Radio Waves in an Ionized Atmosphere**

The theory underlying wave propagation in the ionosphere, including the differences in propagation parallel and perpendicular to a magnetic field, is contained in the classical treatment of the electron by

LORENTZ, H. A. (1909), "The Theory of Electrons and its Applications to the Phenomena of Light and Radiant Heat," B. G. Teubner, Leipzig.

The basic theory for radio propagation by way of the Kennelly–Heaviside layer was developed by

ECCLES, W. H. (1912), On the diurnal variations of the electric waves occurring in nature and on the propagation of electric waves around the bend of the Earth, *Proc. Roy. Soc.* (*London*) **A87**, 79–99.

It was first shown that a plasma has a real index of refraction less than unity and consequently refracts a wave incident on the ionosphere away from the vertical, leading to reflection, in

LARMOR, J. (1924), Why wireless electric rays can bend around the Earth, *Phil. Mag.* **48**, 1025–1036.

This paper introduced the famous critical-reflection equation (5.2.19) as well as the role of electron collisions in attenuating the wave. The importance of the magnetic field on radio propagation was immediately realized by

APPLETON, E. V. (1925), Geophysical influences on the transmission of wireless waves, *Proc. Phys. Soc.* 37, 16D–22D.
NICHOLS, H. W. and SCHELLENG, J. C. (1925), Propagation of electric waves over the Earth, *Bell Syst. Tech. J.* 4, 215–234.

The latter paper introduced a tensor for the dielectric properties of the nonisotropic magnetoplasma. The complete magneto-ionic theory for propagation of radio waves through an ionized medium in the presence of an external magnetic field is due to

APPLETON, E. V. (1932), Wireless studies of the ionosphere, *J. Inst. Elec. Engrs.* **71**, 642–650.
[A summary of Appleton's derivation appears in
APPLETON, E. V. (1928), *U.R.S.I. Mém Sci.* **1**, 2–3 (Papers Presented to the 1927 October General Assembly of the Intern. Sci. Radio Union); reprinted in
BOOKER, H. G. (1974), *J. Atmos. Terr. Phys.* **36**, 2113]

Other papers of the period that considered the dispersion formula are

GOLDSTEIN, S. (1928), The influence of the Earth's magnetic field on electric transmission in the upper atmosphere, *Proc. Roy. Soc.* (*London*) **A121**, 260–285.
HARTREE, D. R. (1929), The propagation of electromagnetic waves in a stratified medium, *Proc. Phil. Soc.* **25**, 97–120,
HARTREE, D. R. (1931), The propagation of electro-magnetic waves in a refracting medium in a magnetic field, *Proc. Phil. Soc.* **27**, 143–162.

Some priority questions have been raised as to who first developed the complete dispersion formula [the extension of (5.2.70) to include the second-order effects]. An obscure paper by

LASSEN, H. (1927), Über den Einfluss des Erdmagnetfeldes auf die Fortpflanzung der elektrischen Wellen der drahtlosen Telegraphie in der Atmosphäre, *Elektr. Nachr. Technik.* **4**, 324–334,

contains a general dispersion relation, including some collisional and positive-ion effects, from which the Appleton formula can be deduced as a special case. The issue is clouded in that APPLETON (1932), following Hartree, incorrectly included the so-called "Lorentz polarization term," appropriate for a dielectric, while otherwise treating the medium as a conductor, although APPLETON's (1928) original note contains the correct expression. In addition, the birefringence

character of a magnetoplasma was contained in the NICHOLS–SCHELLENG (1925) tensor. For further details see

BOOKER, H. G. (1974), Fifty years of the ionosphere, *J. Atmos. Terr. Phys.* **36**, 2113–2136,
RAWER, K. and SUCHY, K. (1976), Remarks concerning the dispersion of electromagnetic waves in a magnetized cold plasma, *J. Atmos. Terr. Phys.* **38**, 395–398.

Most textbooks on electromagnetism do not treat the *Appleton* (or, as it is frequently but incorrectly called, the *Appleton–Hartree*) *formula*, but a complete account of propagation in ionized media appears in

RATCLIFFE, J. A. (1959), "The Magneto-Ionic Theory and its Application to the Ionosphere," Cambridge University Press, Cambridge,
MITRA, S. K. (1952), "The Upper Atmosphere," 2d ed., Asiatic Society, Calcutta.

The theory of radio occultation of spacecraft by planets was developed by

FJELDBO, G.; ESHLEMAN, V. R.; GARRIOTT, O. K.; and SMITH, F. L., III (1965), The two-frequency bistatic radar-occultation method for the study of planetary ionospheres, *J. Geophys. Res.* **70**, 3701–3710,
ESHLEMAN, V. R. (1973), The radio occulation method for the study of planetary atmospheres, *Planet. Space Sci.* **21**, 1521–1531.

The method has been highly successful for Venus', Mars', and Jupiter's ionospheres. There have been problems in deriving the pressure-temperature curves for Jupiter below the ionosphere (where $n > 1$) largely because of its oblateness. This difficulty was identified by

HUBBARD, W. B.; HUNTEN, D. M.; and KLIORE, A. (1975), Effect of the Jovian oblateness on Pioneer 10/11 radio occultations, *Geophys. Res. Lettrs.* 2, 265–268.

A discussion of the oblateness problem and other sources of error, including a simplified explanation of the error magnification, is given by

ESHLEMAN, V. R. (1975), Jupiter's atmosphere: Problems and potential of radio occultation, *Science* **189**, 876–878.

The drastic reduction of partial reflection of radio waves from a diffuse surface compared with Fresnel reflection from a sharp surface was noted by

HERLOFSON, N. (1947), Interpretation of radio echoes from polar auroras, *Nature* **160**, 867–868.

Incoherent scatter as a method for sounding the top-side ionosphere, making use of expected small-scale irregularities to return very high frequency signals, was proposed by

GORDON, W. E. (1958), Incoherent scattering by free electrons with applications to space exploration by radar, *Proc. Inst. Radio Eng.* **46**, 1824–1829.

The first successful experiments were conducted by

BOWLES, K. L. (1958), Observation of vertical incidence scatter from the ionosphere at 41 Mc/s, *Phys. Rev. Letters* **1**, 454–455.

The Rayleigh–Gans approximation used here to estimate backscattering from small irregularities is developed in depth by

VAN DE HULST, H. C. (1957), "Light Scattering by Small Particles," Chapter 7, Wiley, New York.

**Section 5.3 Ionospheres of Venus, Mars, and Jupiter**

A review of the ionospheres and upper atmospheres of Venus and Mars, with an extensive bibliography, is in

MCCONNELL, J. C. (1976), The Ionospheres of Mars and Venus, *Ann. Rev. Earth Planet. Sci.* **4**, 319–346.

Additional material on the Mars ionosphere is reviewed by

BARTH, C. A. (1974), The atmosphere of Mars, *Ann. Rev. Earth Planet. Sci.* **2**, 333–367.

The role of the solar wind in generating a flow of electric current in the ionospheres of Venus and Mars is examined in

CLOUTIER, P. A. and DANIELL, R. E., JR. (1973), Ionospheric currents induced by solar wind interaction with planetary atmospheres, *Planet. Space Sci.* **21**, 463–474.

The Venus model ionosphere discussed here is that of

KUMAR, S. and HUNTEN, D. M. (1974), Venus: An ionospheric model with an exospheric temperature of 350°K, *J. Geophys. Res.* **79**, 2529–2532.

The metallic-ion source of Venus' nighttime ionosphere and its low abundance of negative ions is treated in

BUTLER, D. M. and CHAMBERLAIN, J. W. (1976), Venus night-side ionosphere: Its origin and maintenance, *J. Geophys. Res.* **81**, 4757–4760.

The role of metallic ions in the E region of Earth's ionosphere has been reviewed by

REID, G. C. (1973), op. cit.,
FERGUSON, E. E. (1972), Atmospheric metal ion chemistry, *Radio Sci.* **7**, 397–401.

The possible role of the shocked solar wind in producing a nighttime ionosphere on Venus is treated in

PEREZ-DE-TEJADA, H.; DRYER, M.; and VAISBERG, O. L. (1977), Viscous flow in the near-Venusian plasma wake, *J. Geophys. Res.* **82**, 2837–2841.

The radio-occultation data for the Jovian ionosphere have been analyzed by

FJELDBO, G.; KLIORE, A.; SEIDEL, B.; SWEETMAN, D.; and WOICESHYN, P. (1976), The Pioneer 11 radio occultation measurements of the Jovian ionosphere, *in* "Jupiter," (T. Gehrels, ed.), pp. 238–246, Univ. Arizona Press, Tucson.

The Jovian ionosphere has been modeled from the occultation data by

ATREYA, S. K. and DONAHUE, T. M. (1976), Model ionospheres of Jupiter, *in* "Jupiter," (T. Gehrels, ed.), pp. 304–318, Univ. Arizona Press, Tucson.

The proposal that there is anomalous heating of the Jovian thermosphere due to the dissipation of gravity waves was made on the basis of a stellar occultation that preceded the Pioneers and anticipated the radio-occultation results. The observations and theory, respectively, are in

VEVERKA, J.; ELLIOTT, J.; WASSERMAN, L.; and SAGAN, C. (1974), The upper atmosphere of Jupiter, *Astron. J.* **179**, 73–84,
FRENCH, R. G. and GIERASCH, P. J. (1974), Waves in the Jovian upper atmosphere, *J. Atmos. Sci.* **31**, 1707–1712.

## PROBLEMS

**5.1** *Charged-particle scale height.* (a) What is the scale height for a plasma in a gravitational field when there are two ions with density, mass, and temperature of $N_1$, $M_1$, and $T_i$ and $N_2$, $M_2$, $T_i$? The electrons have density $N_1 + N_2$, $m_e \ll M_i$, and $T_e \neq T_i$. All ions have a single charge $+e$. (b) In the special case where $N_1 \gg N_2$ and $T_e = T_i$, show that the plasma has twice the scale height of the parent neutral of mass $M_1$. (c) Show that at high altitudes ($z \gg z_M$) the distribution of $N_e(z)$ given by (5.1.9) approaches diffusive equilibrium. Explain why (physically) this result comes about.

**5.2** *Decay of F2.* (a) Show that the change of variables (5.1.32) reduces (5.1.31) to the "radial wave equation" (5.1.33). (b) Show by substitution into (5.1.26) that the $j = 0$ solution (5.1.42) gives a diffusive flux divergence that arises from a constant downward velocity $w = -(\beta_0 D_0)^{1/2}$. Show further that this divergence is balanced everywhere by $\beta$-recombination, except for a constant relative loss rate $\lambda_0$.

**5.3** *Detailed balancing.* The Saha ionization equation for thermodynamic equilibrium is

$$\frac{N_{q+1}N_e}{N_q} = \frac{(2\pi mkT)^{3/2}}{h^3}\frac{2\tilde{\omega}_{q+1}}{\tilde{\omega}_q}\exp\left(-\frac{\chi_q}{kT}\right)$$

where $N_q$ is the density of a species in the $q$th stage of ionization, $N_{q+1}$ the density in the next higher stage, $\chi_q$ the ionization potential from the ground state of stage $q$, and the $\tilde{\omega}$'s are the statistical weights of the ground states. (Actually the $\tilde{\omega}$'s should be replaced with temperature-dependent partition functions summed over all the rotational levels in the molecules.) In thermodynamic equilibrium (5.1.43) would be balanced by the inverse process (5.1.46) involving $O_2(X\,^3\Sigma_g^-)$ and $O_2^-(X\,^2\Pi_g)$. With the rate coefficient $k_{43}$ adopted, compute the rate coefficient $k_{46}$ for $T = 200°K$ and $150°K$. [Approximate the partition functions with $\tilde{\omega}(O_2) = 1$ and $\tilde{\omega}(O_2^-) = 2$.]

**5.4** *D-region chemistry.* Equations (5.1.50) and (5.1.51) give the rate of change of electrons and positive ions. (a) Using the reactions (5.1.43) through (5.1.48) [omitting (5.1.44)], write the analogous equation for negative ions. (b) Express this rate as a rate of change $dN_e/dt$ in terms of $N_e$ and $\lambda$, in the manner of (5.1.52). (c) Show that this equation introduces no new information beyond what is already known.

**5.5** *Wave equation for a conducting fluid.* In the derivation of (5.2.11) for $\sigma_p$ we assumed that the atmosphere is a dielectric and that the conduction current $\mathbf{J}$ vanishes and $\sigma = 0$. Alternatively, consider the atmosphere to be a conductor with $\kappa = 1$ (for the neutral component of the atmosphere) and $\sigma \neq 0$ in (5.2.6), and regard the electron oscillations as contributing to the current $\mathbf{J}$. Show that then $\sigma$ is given by (5.2.11) and $\mathbf{E}$ is given by (5.2.12).

**5.6** *Critical reflection and absorption loss.* The electron density in a medium is given by $N_e(z) = az$, where $a$ is a constant gradient. A radio wave of circular frequency $\omega$ is transmitted from the point $z = 0$ along the $+z$ axis and is eventually reflected at distance $z = m\omega^2/4\pi e^2 a$. Assuming that the beam remains perfectly collimated, find the signal loss of returned/transmitted power due to absorption if $\nu_c \ll \omega$. (Regard $\nu_c$ as a constant.)

**5.7** *Phase lag for radio occultation.* Suppose the electron density in a planetary atmosphere follows an isothermal scale height with constant gravity ($H \ll r$). For a circular frequency $\omega \equiv 2\pi\nu$ [that everywhere meets the condition $\omega \gg \omega_0(r)$], show, using only the first term of (5.2.34), that the electron density is related approximately to the phase lag $\phi$ produced by grazing incidence, tangential at $r$, by

$$N_e(r) = -\frac{mvc\phi}{e^2(2\pi rH)^{1/2}}$$

**5.8** *Partial reflection.* The electron density across a diffuse "surface" has a gradient

$$\frac{dN_e}{dz} = \frac{N_0}{a\sqrt{\pi}}\exp\left(\frac{-z^2}{a^2}\right)$$

where $a$ is a constant and $N_0$ is $N_e(z)$ as $z \to \infty$; $N_0$ is well below the value required for critical reflection. Show that the reflected intensity is reduced from the value for a sharp surface by a factor $\exp(-8\pi^2 a^2/\lambda^2)$, where $\lambda$ is the wavelength in vacuum.

**5.9** *Thomson scattering.* (a) Given the total cross section for scattering by a single electron

$$\sigma = \frac{8\pi}{3}\left(\frac{e^2}{mc^2}\right)^2$$

show that the differential cross section is given by (5.2.54) when the incident radiation is unpolarized. (b) With this differential cross section, show that a beam incident along the $x$ axis and plane polarized with **E** parallel to the $z$ axis has the same total cross section as given above.

**5.10** *Rayleigh–Gans scattering.* Show that for backward scattering the interference factor (5.2.57) for a small sphere of radius $a \ll \lambda$ and index of refraction $n \approx 1$ is

$$R \approx 1 - \frac{8}{5}\frac{\pi^2 a^2}{\lambda^2}$$

(*Hint:* Divide the sphere into slices aligned perpendicular to the direction of incidence and emergence and first find the phase shift for such a thin disk.)

# *Chapter 6*
# AIRGLOWS AND AERONOMY

*Airglow* is the amorphous, optical radiation continuously emitted by a planetary atmosphere. It extends from the far ultraviolet into the near infrared, but the term excludes thermal emission in the long-wavelength infrared. We distinguish between airglow and aurora on the basis of the aurora's geographic confinement to (magnetic) polar and sub-polar regions and its sporadic occurrence.

The Earth's airglow arises mainly from discrete atomic and molecular transitions and thus has predominantly an emission-line and emission-band spectrum. An exception is a weak continuum in the green, probably due to *radiative association*, $NO + O \rightarrow NO_2 + h\nu$.

Historically airglow was discovered as being a localized component of the *light of the night sky*, which includes as well the diffuse galactic light due to the unresolved stars and the zodiacal light (sunlight scattered by interplanetary particles). Later, especially bright (by airglow standards) emissions were found in the twilight, arising from atmospheric regions illuminated directly by sunlight. In some cases airglow emissions can be observed from the ground in full daylight, but the instrumental techniques are difficult. In any case, to examine the airglows of the other planets, one observes from spacecraft flying by or orbiting the planet, and scattered light from the lower atmosphere need not pose a problem.

Airglow arises from a multitude of sources; each spectral emission has to be examined individually to ascertain its origin. From the point of view of atmospheric physics and chemistry, it is useful to classify emissions into three

categories: (a) those arising from direct scattering of sunlight, (b) emission associated with creation or destruction of the ionosphere, and (c) radiation resulting from the photochemistry of neutral constituents. Bear in mind that some radiations can be assigned to more than one category.

## 6.1 Airglow Photometry

### 6.1.1 The Rayleigh: A Specialized Photometric Unit

Suppose that radiation crossing a surface, as illustrated in Fig. 6.1, has a *specific intensity* depending on the direction of $I_\nu(\cos\theta, \phi)$ erg/cm$^2$ sec sr Hz. At a distance $r$ from the origin, the energy crossing a unit area aligned normal to the beam and also passing through a unit area normal to the beam at the emitting surface (shown in the figure) is $I_\nu\, d\omega$, where $d\omega = 1/r^2$ is the field of view of unit area at the receiver as seen from the origin. If the emitting surface is uniform in intensity, the observer at position **r** with a field or view $d\Omega$ will receive a total amount of energy across unit area of $I_\nu\, d\omega\, r^2\, d\Omega$, since $r^2\, d\Omega$ is the perceived area of the emitting surface projected onto a plane normal to the line of sight. Thus the *specific surface brightness* as seen at distance $r$ is the radiation crossing unit area per unit time from a cone of unit solid angle,

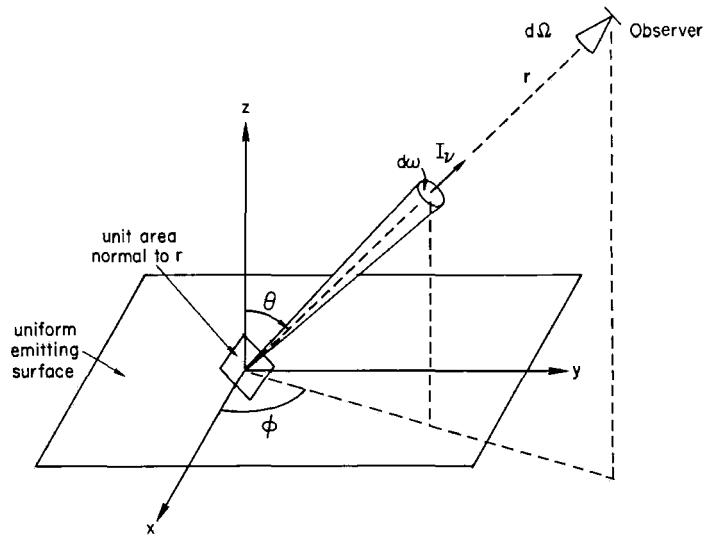



**Fig. 6.1** Specific surface brightness perceived by an observer at a distance $r$ is equivalent to the specific intensity $I_\nu$ and independent of $r$.

which is simply $I_\nu$. Thus the surface brightness for an extended source is independent of the distance of the observer and is identical to the intensity emitted by the surface.

Suppose now that a plane-parallel atmosphere everywhere emits radiation at a rate independent of horizontal position but dependent on height $z$. The directional rate of emission is $\varepsilon_\nu(z|\mu,\phi)$ erg/cm$^3$ sec sr Hz, where $\mu = \cos\theta$. Each unit volume contributes to the total specific intensity emitted by the atmosphere and, if there is no loss due to absorption or scattering, the observer at **r** will measure

$$I_\nu(\mu,\phi) = \int_0^\infty \varepsilon_\nu(z|\mu,\phi)\frac{dz}{\mu} \tag{6.1.1}$$

In airglow photometry one is usually concerned with the emission over a single spectral line or group of lines. Also, the energy emitted is usually of less intrinsic interest than the number of photons, which is a direct measure of the number of photochemical reactions occurring in the atmosphere. Thus the number of photons emitted in a given direction is

$$\mathscr{I}(\mu,\phi) = \int_0^\infty \mathscr{E}(z|\mu,\phi)\frac{dz}{\mu} \tag{6.1.2}$$

where

$$\mathscr{E}(z|\mu,\phi) = \int_{\text{line}} \frac{\varepsilon_\nu(z|\mu,\phi}{h\nu}d\nu \quad \text{photon/cm}^3\,\text{sec sr} \tag{6.1.3}$$

The total rate of photon emission in all directions is then

$$E(z) = \int_0^{2\pi}\int_{-1}^{1} \mathscr{E}(z|\mu',\phi')\,d\mu'\,d\phi' \quad \text{photon/cm}^3\,\text{sec} \tag{6.1.4}$$

For resonance scattering (see Section 6.2 and Problem 6.1), the emission rate $\mathscr{E}$ may indeed have a directional dependence. But for chemical excitation the emission from unit volume is isotropic and $E(z) = 4\pi\mathscr{E}(z|\mu,\phi)$. Then we have for plane-parallel geometry

$$4\pi\mathscr{I}(\mu) = \frac{1}{\mu}\int_0^\infty E(z)\,dz \quad \text{photon/cm}^2\,\text{sec} \tag{6.1.5}$$

The *rayleigh* unit (R) is a measure of the *omnidirectional emission rate* in a column of unit cross section along the line of sight, with 1 R $\equiv 10^6$ photon/cm$^2$ sec. Thus if the frequency-integrated intensity $\mathscr{I}(\mu)$ is expressed in units of megaphoton/cm$^2$ sec sr, $4\pi\mathscr{I}(\mu)$ is in rayleighs.

Observations of airglow from spacecraft generally are made along a tangent to a spherical shell with radius $r$ from the center of the planet. In that case the plane-parallel geometry is not valid and extracting the height distributions of emission is analogous to obtaining the electron densities $N_e(r)$ from radio occulations (see Problem 5.7).

Most of the Earth's airglow emissions are confined to layers that have thicknesses of the order of a scale height, and except in the F regions, this scale factor is $H \sim 10^6$ cm. Hence a total emission rate referred to the vertical of $\mu 4\pi \mathcal{I}(\mu) = 1$ R, say, corresponds to 1 photon/cm$^3$ sec emitted from the layer.

### 6.1.2 Brightness and Composition of Night Airglow

Figure 6.2 shows a tracing of the green-red spectrum of Earth's night airglow, which is dominated by the green and red forbidden lines of neutral atomic oxygen [O I] and the yellow resonance doublet of Na I. (In discussions of spectra, square brackets indicate transitions forbidden by electric-dipole selection rules, rather than concentrations, and Roman numerals indicate the state of ionization: O I ≡ O; O II ≡ O$^+$, etc.)

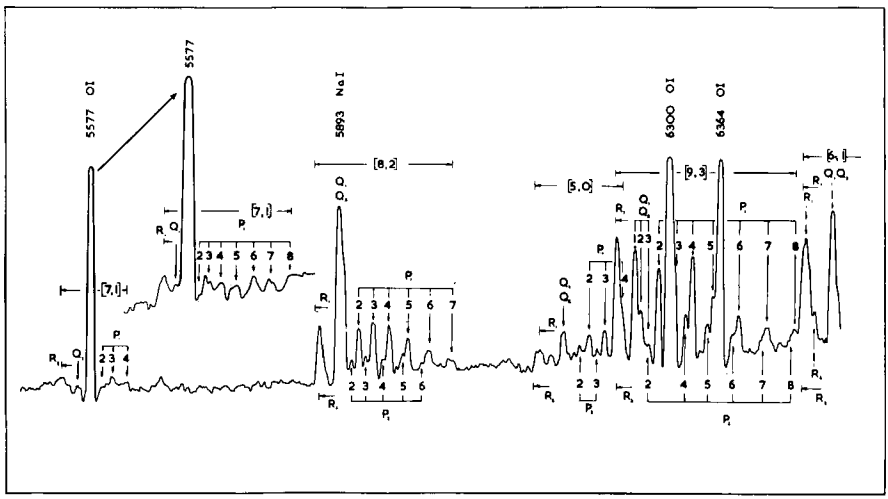



**Fig. 6.2** Microphotometer tracing of the visual airglow showing the strong forbidden lines of [O I] at 5577, 6300, and 6364 Å, the Na I lines, blended at this resolution at 5893 Å, and several of the weaker OH bands, with open $P_1$ and $P_2$ branches and blended $R$ and $Q$ branches. [After BLACKWELL, D. E.; INGHAM, M. F.; and RUNDLE, H. (1960), *Astrophys. J.* **131**, 15.]

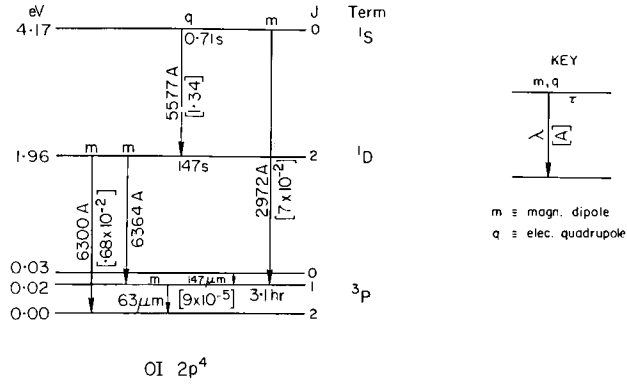



**Fig. 6.3** Ground configuration, $1s^2 2s^2 2p^4$, of O I, showing term and level splitting. The diagram gives transition probabilities $A$ ($sec^{-1}$) for the forbidden transitions, the dominant multipole term ($m$ or $q$) in the transitions, the wavelengths, and mean lifetime $\tau$ of the levels. The level splitting of the $^3P$ term is exaggerated. The transition probability for $^3P$–$^1D$ is the total: the ratio for $\lambda 6300/\lambda 6364$ is $\frac{3}{1}$. [Transition probabilities from R. H. GARSTANG: (1968), in "Planetary Nebulae," *I.A.U. Symp.* **34**, 143; (1969), *Liège Colloq.*, p. 35; (1951), *M.N.R.A.S.* **111**, 115.]

Figure 6.3 shows the energy-level diagram of the ground configuration of O I($2p^4$), along with lifetimes of the metastable levels and Einstein transition probabilities $A$. Much of the recombination and photochemical airglow arises from low-lying atomic and molecular levels because of the limited chemical energies available. These low levels are frequently metastable, with long lifetimes. Consequently, the altitude at which the emission occurs depends in part on whether collisional deactivation can occur within the radiative lifetime.

An exception to the above remarks is illustrated by the Na I resonance doublet, 5890–5896 Å, where the first excited electron level 3p($^2$P) lies 2.1 e$V$ above ground, 3s($^2$S). A similar situation exists for the other alkali metals, such as Li I at 6708 Å, K I at 7699 Å, and Ca II at 3933–3968 Å, which appear in twilight.

The atomic lines in the green and red are all of the same order of brightness, $10^2$ R, with Na having a strong seasonal variation and the red [O I] and Na lines having major twilight enhancements (for different reasons).

In twilight Na emits by resonance scattering and its abundance and height distribution can be ascertained. The maximum density of free Na is less than $10^4$ cm$^{-3}$ at 90 km—a mixing ratio of one part in $10^{10}$. Its prominence in the night airglow is evidently due to a high reactivity with oxygen (or possibly hydrogen) compounds. Whether or not those reactions account for

excitation of the D lines, the equilibrium abundance of Na is fixed mainly by

$$\begin{gathered}\mathrm{Na} + \mathrm{O_3} \rightarrow \mathrm{NaO} + \mathrm{O_2} \\ k_6 = 6.5 \times 10^{-12}\ \mathrm{cm^3/sec}\end{gathered} \tag{6.1.6}$$

followed by

$$\begin{gathered}\mathrm{NaO} + \mathrm{O} \rightarrow \mathrm{Na} + \mathrm{O_2} \\ k_7 = 4 \times 10^{-11}\ \mathrm{cm^3/sec}\end{gathered} \tag{6.1.7}$$

Consequently, the ratio of free sodium to sodium oxide is

$$\frac{[\mathrm{Na}]}{[\mathrm{NaO}]} = \frac{k_7[\mathrm{O}]}{k_6[\mathrm{O_3}]} \approx 6\,\frac{[\mathrm{O}]}{[\mathrm{O_3}]} \tag{6.1.8}$$

The green line of $[\mathrm{O\,I}]({}^1\mathrm{D}_2 - {}^1\mathrm{S}_0)$, $\lambda$5577, is produced mainly in the 100 km region as a by-product of oxygen association,

$$\begin{gathered}\mathrm{O} + \mathrm{O} + \mathrm{O} \rightarrow \mathrm{O_2} + \mathrm{O}^*({}^1\mathrm{S}_0) \\ k_9 \sim 6 \times 10^{-33}\ \mathrm{cm^6/sec}\end{gathered} \tag{6.1.9}$$

Additional but weaker emission arises as a result of ionospheric recombination, which also produces the major part of the nightglow [O I] red lines $({}^3\mathrm{P}_{2,1}{-}{}^1\mathrm{D}_2)$, $\lambda\lambda$6300–64, by

$$\mathrm{O_2}^+ + \mathrm{e} \rightarrow \mathrm{O}^*({}^1\mathrm{D}, {}^1\mathrm{S}) + \mathrm{O} \tag{6.1.10}$$

The product $^1$D has a yield of nearly unity and $^1$S, about 10 percent. The other main dissociative-recombination process in the F region produces $[\mathrm{N\,I}]({}^2\mathrm{D}^0_{5/2,3/2}{-}{}^4\mathrm{S}^0_{3/2})$, $\lambda$5200, by

$$\mathrm{NO}^+ + \mathrm{e} \rightarrow \mathrm{N}({}^2\mathrm{D}) + \mathrm{O} \tag{6.1.11}$$

This emission is barely detectable ($\sim$1 R) at night. The [O I] red lines have twilight enhancements due to the fast rate of F2 recombination after sunset (see Section 5.1.3), dissociation of $\mathrm{O_2}$ in the Schumann–Runge continuum,

$$\mathrm{O_2} + h\nu \rightarrow \mathrm{O}({}^1\mathrm{D}) + \mathrm{O}({}^3\mathrm{P}) \tag{6.1.12}$$

and photoelectron impact. The latter may well precede morning twilight itself because photoelectrons in the southern hemisphere, say, can travel along magnetic-field lines and impact O atoms in the northern hemisphere at a point farther west and in total darkness, since the geomagnetic pole is displaced from the rotational pole.

Figure 6.2 also shows several of the weaker Meinel bands of hydroxyl (OH). These bands arise from vibrational transitions within the ground electronic state, and consequently their dipole moment is small but not zero. A harmonic oscillator has allowed transitions only for $\Delta v = 1$ and the existence of sequences for $\Delta v > 1$ is due to the anharmonicity of the OH potential well. For OH the strongest bands ($\Delta v = 2$) appear in the $\lambda = 1.5$ $\mu$m region and have emission rates of the order of $10^5$ R, compared with $10^1$–$10^2$ R for the $\Delta v = 6$ sequence.

The ground electronic state of OH is $^2\Pi$; in addition to R and P branches, there is a Q branch in which the rotational lines are nearly superimposed and which appears as a strong spike at the band origin. In the open P branch, the spin doubling is shown for each rotational quantum number K as weak $P_2$ and strong $P_1$ lines, from the $^2\Pi_{1/2}$ and $^2\Pi_{3/2}$ levels, respectively.

The OH emission arises from the reaction

$$H + O_3 \rightarrow OH^*(v \leq 9) + O_2 \tag{6.1.13}$$

which is recycled by

$$OH + O \rightarrow H + O_2 \tag{6.1.14}$$

These reactions are part of the hydroxyl catalytic cycle for destruction of ozone (Section 3.2.2). The 3.34 eV available from (6.1.6) is adequate to excite $v = 9$ but no higher levels and this is precisely what is observed. Photochemical theory and rocket observations agree in placing the emission layer around 90 km, but spectroscopic temperatures, while highly variable, are usually scattered around 250°K, which is warm for the emitting layer. The reason for this discrepancy is not clearly understood. Possibly the rotational distribution in the higher vibrational levels is not completely in thermal equilibrium with the surrounding gas; perhaps deactivation by $O_2$ is important at lower altitudes for certain vibrational levels, which then emit more strongly from higher, warmer regions. In any case a long series of spectrophotometric recordings near Moscow shows much less discrepancy with the kinetic temperature of the emitting environment, giving rotational temperatures between 200 and 225°K for a variety of bands.

The low electronic states of $O_2$ are shown in Fig. 6.4. The C $^3\Sigma_u^-$ state dissociates into $O(^3P) + O(^1D)$ from absorption in the Schumann–Runge

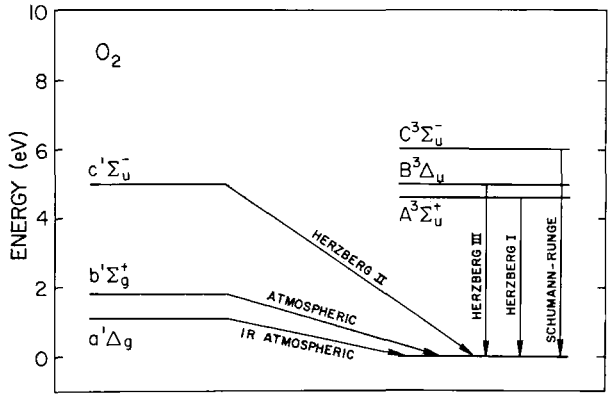



**Fig. 6.4** Energy-level diagram for $O_2$ showing transitions important in airglow spectra. Compare Fig. 1.9.

continuum. All the other states illustrated dissociate in the weak Herzberg continuum and give two $O(^3P)$ atoms. In the 100 km region where $O_2$ dissociation is important, the principal re-association mechanism is the three-body type,

$$O(^3P) + O(^3P) + X \rightarrow O_2^* + X \tag{6.1.15}$$

where $O_2^*$ indicates a molecule in an excited state. Thus the newly formed molecule may be formed in any of the states A, B, a, b, or c, as well as the ground $X\,^3\Sigma_g^-$ state.

The $O_2$ Atmospheric ($b\,^1\Sigma_g^+ \rightarrow X\,^3\Sigma_g^-$) and Infrared Atmospheric ($a\,^1\Delta_g \rightarrow X\,^3\Sigma_g^-$) systems appear at the ground in the $v' = 0 \rightarrow v'' = 1$ bands only. The 0–0 transitions are absorbed in the middle atmosphere. The total emission rate in the Atmospheric system is about 30 kR and in the IR Atmospheric, 75 kR. The relative intensities of the individual rotational lines are governed by the Boltzmann distribution of populations at the kinetic temperature of the emitting region.

The Herzberg I bands, $A\,^3\Sigma_u^+ \rightarrow X\,^3\Sigma_g^-$, violate the first-order dipole selection rule that $\Sigma^+$ does not combine with $\Sigma^-$. The bands appear weakly in the 3100–4000 Å region and are somewhat stronger shortward of the ozone cutoff. Their total emission rate is of the order of 1 kR. All three $O_2$ band systems observed in the terrestrial airglow give *rotational temperatures* of 200°K or under, consistent with their production in the 90–100 km region.

The continuum radiation in the yellow-green region is probably due to radiative association,

$$\begin{aligned} &NO + O \rightarrow NO_2 + h\nu \\ &\qquad k_{16} = 1.3 \times 10^{-21}\ \text{cm}^3/\text{sec} \end{aligned} \tag{6.1.16}$$

which is also commonly seen in air afterglows in the laboratory.

**Venus Nightglow.** The Venera 9 and 10 spacecraft made observations from planetary orbit of a night airglow on Venus in the 3000–8000 Å region. The emission bands have been identified as the $v' = 0$ progression of the Herzberg II ($c\,^1\Sigma_u^- – X\,^3\Sigma_g^-$) system of $O_2$. Laboratory afterglow studies demonstrate that this system is formed by reaction (6.1.15) where $CO_2$ is the third body, namely,

$$O(^3P) + O(^3P) + CO_2 \rightarrow O_2(^1\Sigma_u^-) + CO_2 \tag{6.1.17}$$

The $v' = 0$ level at 4.08 eV acquires the major part of the association energy (5.12 eV). Mapping this airglow on Venus can define the circulation in the upper atmosphere, since the local density of O atoms on the night side is controlled by the rates of association and transport.

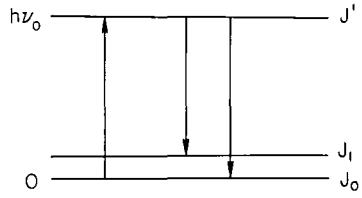



**Fig. 6.5** Simplified energy-level diagram with one level in the upper state and multiple levels in the ground state.

## 6.2 Resonant and Fluorescent Scattering of Sunlight

### 6.2.1 Photon Scattering Coefficient: The $g$-Value.

Consider a three-level atom or molecule, as illustrated in Fig. 6.5. The ground state has two levels $J_i$, either of which can radiatively connect with the upper level $J'$. (The levels $J_i$ may have further fine structure or hyperfine structure. Or they may, in the case of molecules, represent vibrational levels in the ground electronic states with each vibrational level being subdivided into rotational fine structure.)

The incident photon flux at energy $h\nu_0$ from the sun is $\pi\mathscr{F}_\nu(z) = \pi\mathscr{F}_\nu(\infty)\exp(-\tau_\nu/\mu_0)$, where $\pi\mathscr{F}_\nu(\infty)$ is the flux outside the atmosphere in units of photon/cm$^2$ sec Hz and the exponential $\tau_\nu(z)/\mu_0$ is the *slant optical thickness* due to extinction (absorption plus scattering) of the flux in traversing the overhead atmosphere. The photon intensity scattered by unit volume in a direction $(\mu, \phi)$ at an angle $\Theta$ to the solar direction $(\mu_0, 0)$ in transition $J' \to J_i$ is

$$\mathscr{E}(z|\mu,\phi) = \frac{E(z)p(\Theta)}{4\pi}$$

$$= N(z)\pi\mathscr{F}_\nu(z)\frac{\pi e^2}{mc} f(J_0J')\tilde{\omega}(J'J_i)\frac{p(\Theta)}{4\pi} \quad \text{photon/cm}^3\,\text{sec}\,\text{sr} \qquad (6.2.1)$$

Here $p(\Theta)$ is the phase function for single scattering; $N(z)$ is the number density of scattering atoms or molecules; $f$ is the oscillator strength or $f$-value, which is related to the transition probability by (1.6.5); and $\tilde{\omega}$ is the albedo for single scattering, which is

$$\tilde{\omega}(J'J_i) = \frac{A(J'J_i)}{\sum_i \{A(J'J_i) + \eta(J'J_i)[\mathrm{M}]\}} \qquad (6.2.2)$$

where $\eta$ is the rate coefficient for collisional deactivation. The phase angle $\Theta$ is related to the incident and emergent polar angles, $\cos^{-1}\mu_0$ and $\cos^{-1}\mu$, respectively, and their relative azimuth $\phi$ by (4.1.35)

Scattering in the same transition $J' \rightarrow J_0$ or frequency $h\nu_0$ as the incident solar photon constitutes *resonance scattering*. When the downward transition follows a different route, $J' \rightarrow J_1$, it is called *fluorescent scattering*.

The height integration of (6.2.1) gives, for plane parallel geometry,

$$4\pi\mathscr{I}(\mu, \phi) = g\mathscr{N}_{eq}p(\Theta)/\mu \tag{6.2.3}$$

where

$$g(J'J_i) = \pi\mathscr{F}_\nu(\infty)\frac{\pi e^2}{mc}f(J_0J')\frac{A(J'J_i)}{\sum_i A(J'J_i)} \text{ photon/sec molecule} \tag{6.2.4}$$

is the photon scattering in all directions per atom (or molecule) with no deactivation or extinction of incident flux and

$$\mathscr{N}_{eq} = \int_0^\infty \frac{N(z)e^{-\tau_\nu(z)/\mu_0}\,dz}{1 + [M(z)]\eta(J'J)/\sum_i A(J'J_i)} \tag{6.2.5}$$

is the *equivalent column-integrated density* of scattering molecules. The form

**TABLE 6.1** *Photon Scattering Coefficients (g-values) at 1 AU from the Sun*

| Atom or molecule | Line or band | $\lambda$ (Å) | $g$ (photon/sec-molecule) |
|---|---|---|---|
| [OI] | $^3P_2$–$^1D_2$ | 6300 | $4.5 \times 10^{-10}$ |
| [OI] | $^1D_2$–$^1S_0$ | 5577 | $1 \times 10^{-11}$ |
| [OI] | $^3P_1$–$^1S_0$ | 2972 | $6 \times 10^{-13}$ |
| OI | $2p^4\,^3P_{2,1,0}$–$2p^3 3s\,^3S_1^o$ | 1302–4–6 | $1 \times 10^{-4}$ |
| [NI] | $^4S^o_{3/2}$–$^2D^o_{3/2,5/2}$ | 5199 | $6 \times 10^{-11}$ |
| $O_2$ Atmospheric System | $b\,^1\Sigma_g^+ \rightarrow X\,^3\Sigma_g^-$ $(v' = 0 \rightarrow v'' = 1)$ | 8645 | $5 \times 10^{-10}$ |
| $O_2$ Infrared Atmospheric | $a\,^1\Delta_g \rightarrow X\,^3\Sigma_g^-$ $(v' = 0 \rightarrow v'' = 1)$ | 15,803 | $1.2 \times 10^{-11}$ |
| Na I "$D_1 + D_2$" | $^2S_{1/2}$–$^2P^o_{1/2,3/2}$ | 5893 | 0.80 |
| Ca II "$K$" | $^2S_{1/2}$–$^2P^o_{3/2}$ | 3933 | 0.3 |
| Li I | $^2S_{1/2}$–$^2P^o_{1/2,3/2}$ | 6708 | 16 |
| K I | $^2S_{1/2}$–$^2P^o_{1/2}$ | 7699 | 1.67 |
| $N_2^+$ First Negative System | $B\,^2\Sigma_u^+ \rightarrow X\,^2\Sigma_g^+$ $(v' = 0 \rightarrow v'' = 0)$ | 3914 | 0.05 |
| H Lyman $\alpha$ | $1s\,^2S$–$2p\,^2P^o$ | 1215 | $2.3 \times 10^{-3}$ |
| H Balmer $\alpha$ | $2s\,^2S$–$3p\,^2P^o$ | 6563 | $2.6 \times 10^{-6}$ |
| He I | $1s^2\,^1S_0$–$1s2p\,^1P_1$ | 584 | $1.7 \times 10^{-5}$ |
| He I | $2s\,^3S_1$–$2p\,^3P^o_{0,1,2}$ | 10,830 | 16.8 |
| He II | $1s\,^2S$–$2p\,^2P^o$ | 304 | $1.1 \times 10^{-4}$ |

(6.2.3) is convenient because $g$, the *photon scattering coefficient*, can be tabulated once and for all for any given transition (see Table 6.1).

When an atmosphere is viewed externally along a line of sight that is tangent to a spherical shell of radius $r$, the plane-parallel geometry is clearly inappropriate. In that case we can express the density integrated along a line of sight in terms of the local scale height $H$ and the local density of scattering molecules $N(r)$. When both solar extinction and deactivation are negligible everywhere the factor $\mathcal{N}_{eq}/\mu$ in (6.2.3) is replaced by the *tangentially integrated density*

$$\begin{aligned}\mathcal{N}_{tan}(r) &\equiv \int_{-\infty}^{\infty} N(r')\,ds \\ &= \int_{-\pi/2}^{\pi/2} N(r)e^{-(r'-r)/H} r' \sec\beta\, d\beta \\ &= 2N(r)re^{r/H}K_1(r/H) \\ &\approx N(r)(2\pi rH)^{1/2}(1 + \tfrac{3}{8}H/r + \cdots)\end{aligned} \tag{6.2.6}$$

where $K_1$ is the modified Bessel function of the third kind. The integral is evaluated by writing $r' = r\sec\beta$ and changing the independent variable to $y \equiv \sec\beta$. The leading term can be obtained most simply by expanding $\sec\beta$, but the correction term requires an asymptotic expansion of the exact solution for large arguments.

### 6.2.2 Anisotropy of Scattering, $p(\Theta)$

The anisotropy of scattering is directly related to the polarization of the scattered wave, just as Thomson scattering produces both polarization and anisotropy. In that case, if the electric vector is perpendicular to the plane of scattering, the scattered intensity is isotropic. But the scattered wave with **E** in the plane of scattering has an intensity that varies as $\cos^2\Theta$. Thus the total intensity is given by (5.2.55) (also, see Problem 5.9). The polarization produces the anisotropy and to know one is to know the other.

Thomson and Rayleigh scattering are classical effects. An electron is set in motion by the incident wave, and the direction of electron acceleration determines the allowed directions of emission of the scattered wave. Resonant (or fluorescent) scattering is a quantum effect: The electron is placed in a higher state, and then after a finite interval it cascades back to the lower state. If the molecule is in a magnetic field **B**, the absorption places the molecule in a particular Zeeman state with a particular orientation of its angular momentum relative to the external field. If there are no collisions

the molecule will re-emit from the same upper state. It can emit a $\pi$ Zeeman component, which is polarized with the electric vector of the wave parallel to the magnetic field, or a $\sigma$ component, which is polarized perpendicular to **B**. If we observe along the field, for example, we can see only the $\sigma$ component, and in general the intensity of a given Zeeman component depends on the geometry. Although it is not immediately obvious, it turns out that the total intensity of all Zeeman components in a specified direction also depends on the scattering angle.

With the polarization and consequent anisotropy dependent on the Zeeman effect (since a magnetic field provides a sense of direction in space), we might expect the polarization to disappear as $\mathbf{B} \to 0$. It does not, because the direction of radiation incidence still gives an orientation to the geometry. The polarization without an external field is the same as if there were a weak field parallel to the incident electric vector.

To understand why this is so, we recall that the Zeeman states specify an orientation in space of the angular momentum of an atom relative to a given direction. If this direction is that of a field **B**, the Larmor precession removes the energy degeneracy. But as long as an orientation is specified by an incident wave, we retain the formal distinctions of Zeeman states. The most obvious choice for orienting space might be to replace the field direction with the direction of propagation; but this arrangement gives the wrong answer, as we shall see.

Scattering that starts upward from level $FM$ to level $F'M'$ followed by cascading to the original level $F'M' \to FM$ has coherence properties. This scattering contributes to the refractive index. Cascading to a different terminal state of the lower level, $F'M' \to F''M''$, is incoherent and has no effect on the refractive index. The average contribution to the refractive index per atom must be independent of the direction of the external magnetic field if it is weak, according to Heisenberg's *principle of spectroscopic stability*. However, applying a weak magnetic field parallel to the incident electric vector does not alter the plane of polarization, whereas a field in the direction of propagation makes $\sigma$ components (circular polarization) responsible for the coherent scattering and the plane of polarization is rotated. Therefore, to treat polarization in the absence of a magnetic field, we assume a weak field parallel to the incident electric vector.

We now develop these concepts quantitatively for scattering through an angle $\Theta$. In Fig. 6.6 the $xy$ plane is the plane of scattering; "perpendicular" and "parallel" subscripts on the incident specific flux $\pi\mathscr{F}$, and the scattered radiation rate $\mathscr{E}$ will denote the direction of the electric vector relative to the scattering plane. For incident natural light, $\mathscr{F}_\perp = \mathscr{F}_{||}$; the incident flux is constant with frequency (white light). We assume that deactivating and depolarizing collisions are negligible.

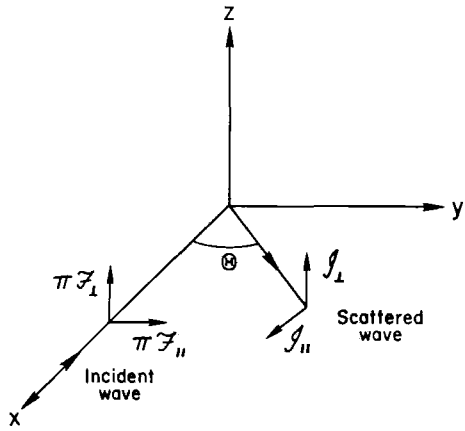



Fig. 6.6 Geometry of scattering in the $xy$ plane, with natural incident light divided into polarization components perpendicular and parallel to the plane of scattering.

The polarization and angular distribution of scattering depend critically on the level structure of the atom, including any hyperfine structure denoted by quantum number F due to nuclear spin $I$. The example of Na levels and the D lines is shown in Fig. 6.7. In addition, according to the discussion above, we assume there is an external magnetic field $B_z$ that separates the Zeeman states $M_F$. Then the scattered emission rate per unit volume and solid angle polarized with **E** along the $z$ axis is

$$\mathscr{E}_\perp = C\mathscr{F}_\perp \sum_{F'M'} (A_\pi^{F'M'})^2 \tag{6.2.7}$$

where $C$ is a constant to be obtained presently and $A_\pi^{F'M'}$ is total transition probability from state $F'M'$ to all lower Zeeman states $F''M''$ with emission of a $\pi$ Zeeman component ($=\sum_{F''}\sum_{M''} A_\pi^{F'M'F''M''}$). Similarly, the emission rate polarized in the $xy$ plane arising from scattering by $\pi\mathscr{F}_\perp$ is

$$\mathscr{E}_{||} = C\mathscr{F}_\perp \sum_{F'M'} (\tfrac{1}{2}A_\pi^{F'M'}A_\sigma^{F'M'}) \tag{6.2.8}$$

where the $\frac{1}{2}$ factor allows for the fact that perpendicular to the field $\sigma$ components radiate with only one-half the efficiency of $\pi$ components. (For example, the circularly polarized $\sigma$ components are equivalent to two linearly polarized components aligned perpendicular to one another.) Equations (6.2.7–6.2.8) give the scattered intensity as being proportional to the absorption probability [which in turn is proportional to the emission probability for Zeeman components, whose statistical weights are unity; see (1.6.5)] times the re-emission probability.

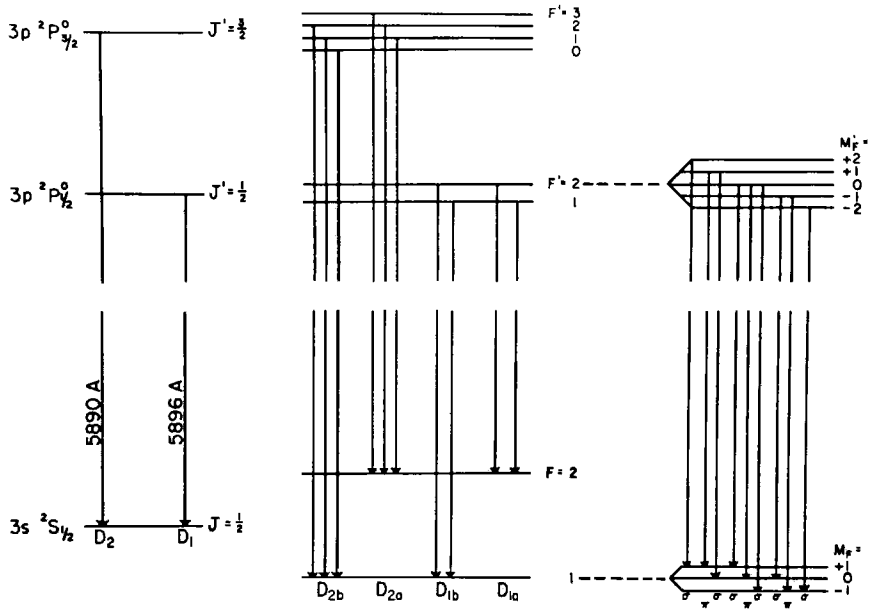



**Fig. 6.7** Energy-level diagram for the ground (3s) and first excited (3p) electron configurations of Na I. The hyperfine splitting (exaggerated in the figure) gives the hyperfine levels for a nuclear spin $I = \frac{3}{2}$. The right side shows the Zeeman states and the Zeeman line components, with their $\pi$ or $\sigma$ polarization character indicated. [After CHAMBERLAIN (1961).]

To obtain the intensity scattered from $\pi\mathscr{F}_{\parallel}$ with an external magnetic field aligned along the $z$ axis, we would proceed analogously, except that the absorption would be proportional to $A_\sigma^{F'M'}/2$. But as noted above, to treat the scattering of natural light in the limit of $\mathbf{B} \to 0$, we invoke a weak field $B_y$ parallel to the electric vector. For the upward transitions, we again have to consider only $\pi$ components. Thus to (6.2.7) and (6.2.8) we add

$$\mathscr{E}_\perp = C\mathscr{F}_{\parallel} \sum_{F'M'} [A_\pi^{F'M'}(\tfrac{1}{2}A_\sigma^{F'M'})] \tag{6.2.9}$$

and

$$\mathscr{E}_{\parallel} = C\mathscr{F}_{\parallel} \sum_{F'M'} [A_\pi^{F'M'}(A_\pi^{F'M'} \cos^2 \Theta + \tfrac{1}{2}A_\sigma^{F'M'} \sin^2 \Theta)] \tag{6.2.10}$$

The total transition probability is

$$A = A_\pi^{F'M'} + A_\sigma^{F'M'} \tag{6.2.11}$$

which is independent of $F'$ or $M'$. Summing over all $2F' + 1$ Zeeman states in the upper level, we obtain

$$\sum_{M'} (A_\pi^{F'M'} + A_\sigma^{F'M'}) = (2F' + 1)A \tag{6.2.12}$$

where $\tilde{\omega}(F') = 2F' + 1$ is the statistical weight of $F'$. For isotropic, unpolarized incident light, the scattered radiation is also unpolarized because of the spherical symmetry involved (i.e., spectroscopic stability). If the scattered radiation is observed perpendicular to a weak magnetic field, the absence of polarization requires

$$\sum_{M'} A_{\pi}^{F'M'} = \tfrac{1}{2} \sum_{M'} A_{\sigma}^{F'M'} \tag{6.2.13}$$

We now define the auxiliary quantities

$$\alpha(F') = \sum_{M'} (A_{\pi}^{F'M'})^2 + \frac{2F' + 1}{3} A^2 \tag{6.2.14}$$

and

$$\beta(F') = \tfrac{3}{2} \sum_{M'} (A_{\pi}^{F'M'})^2 - \frac{2F' + 1}{6} A^2 \tag{6.2.15}$$

Then the intensities of scattered radiation components are

$$\mathscr{E}_{\perp} = \frac{C}{4} \sum_{F'} \{[\alpha(F') + 2\beta(F')]\mathscr{F}_{\perp} + [\alpha(F') - 2\beta(F')]\mathscr{F}_{||}\} \tag{6.2.16}$$

and

$$\mathscr{E}_{||} = \frac{C}{4} \sum_{F'} \{[\alpha(F') - 2\beta(F')]\mathscr{F}_{\perp} + [\alpha(F') - 2\beta(F') + 4\beta(F')\cos^2 \Theta]\} \tag{6.2.17}$$

When the incident light is unpolarized, the polarization of radiation scattered at angle $\Theta$ is

$$\mathscr{P}(\Theta) \equiv \frac{\mathscr{E}_{\perp} - \mathscr{E}_{||}}{\mathscr{E}_{\perp} + \mathscr{E}_{||}} = \frac{\sum_{F'} \beta(F') \sin^2 \Theta}{\sum_{F'} [\alpha(F') - \beta(F') \sin^2 \Theta]} \tag{6.2.18}$$

The total volume rate of emission in direction $\Theta$ is

$$\mathscr{E}(\Theta) = \mathscr{E}_{\perp}(\Theta) + \mathscr{E}_{||}(\Theta) = \frac{C\mathscr{F}}{2} \sum_{F'} [\alpha(F') - \beta(F') \sin^2 \Theta] \tag{6.2.19}$$

where $\mathscr{F}_{||} = \mathscr{F}_{\perp} = \frac{1}{2}\mathscr{F}$, and the mean emission rate over a sphere is

$$\langle \mathscr{E} \rangle \equiv \frac{E(z)}{4\pi} = \frac{C\mathscr{F}}{2} \sum_{F'} [\alpha(F') - \tfrac{2}{3}\beta(F')] \tag{6.2.20}$$

With no collisions to deactivate or depolarize the emission (by collisionally altering the upper state $F'$) and with $f(JJ')$ related to $A(J'J)$ by (1.6.5), a comparison of the total line $(J' \to J)$ emission given by (6.2.1) and (6.2.20)

yields

$$C = \frac{9c^2 N(z)}{64\pi\nu^2 \tilde{\omega}(J) \sum_{J'} A(J'J)} \tag{6.2.21}$$

where the total statistical weight of the lower level $J$ is $\tilde{\omega}(J) = (2J + 1)(2I + 1) = \sum_F (2F + 1)$, and $I$ is the nuclear spin.

The *anisotropy* or *scattering phase function* is then

$$p(\Theta) \equiv \frac{\mathscr{E}(\Theta)}{\langle \mathscr{E} \rangle} = \frac{\sum_{F'} [\alpha(F') - \beta(F') \sin^2 \Theta]}{\sum_{F'} [\alpha(F') - \frac{2}{3}\beta(F')]} \tag{6.2.22}$$

In the notation of Section 4.1.4, we may express $p$ as

$$p(\cos\Theta) = a + b\cos^2\Theta \tag{6.2.23}$$

Then defining

$$r = \sum_{F'} \beta(F') \Big/ \sum_{F'} \alpha(F') \tag{6.2.24}$$

we have coefficients

$$a = \frac{1 - r}{1 - \frac{2}{3}r}, \qquad b = \frac{r}{1 - \frac{2}{3}r} \tag{6.2.25}$$

The relative intensities of all the Zeeman components in an unresolved multiplet must be known to compute $\sum\alpha(F')$ and $\sum\beta(F')$. These intensities are most readily obtained in three steps. First, if there are two or more lines $(J' \to J'')$ in a multiplet $(L' \to L'')$, their relative intensities are given for $LS$ coupling by well-known formulas found in texts on atomic spectra and involving the quantum numbers $S', L'$, and $J'$.

Then to find the relative intensities of hfs lines $(F' \to F'')$ within one line $(J' \to J'')$ when the nuclear spin $I \neq 0$, the same formulas apply with the quantum numbers $S', L'$, and $J'$ replaced with $I', J'$, and $F'$, respectively. Some texts give convenient tables for quickly obtaining the relative intensities. An example is given in Problem 6.4.

Finally, the relative intensities of Zeeman components $(M' \to M'')$ in an hfs line $(F' \to F'')$ follow from the intensity sum rules, which state that the sum of all transitions starting (or stopping) on any Zeeman state is the same as the sum of all transitions starting (or stopping) on any other Zeeman state in the same hfs line. From these rules we obtain the following formulas: For a transition $F' \to F'$ (i.e., for $F'' = F'$)

$$A_\pi^{F'M'F'M'} = aM'^2 \tag{6.2.26}$$

and

$$A_{\sigma}^{F'M'F'(M'\pm 1)} = \tfrac{1}{2}a(F' \pm M' + 1)(F' \mp M') \tag{6.2.27}$$

For a transition $F' \to F' + 1$,

$$A_{\pi}^{F'M'(F'+1)M'} = b(F' + M' + 1)(F' - M' + 1) \tag{6.2.28}$$

and

$$A_{\sigma}^{F'M'(F'+1)(M'\pm 1)} = \tfrac{1}{2}b(F' \pm M' + 1)(F' \pm M' + 2) \tag{6.2.29}$$

For $F' \to F' - 1$,

$$A_{\pi}^{F'M'(F'-1)M'} = c(F' + M')(F' - M') \tag{6.2.30}$$

and

$$A_{\sigma}^{F'M'(F'-1)(M'\pm 1)} = \tfrac{1}{2}c(F' - 1 \mp M')(F' \mp M') \tag{6.2.31}$$

Here $a$, $b$, and $c$ are constants that can be expressed in terms of $A$ through (6.2.11) by summing the $A_{\pi}$'s and $A_{\sigma}$'s over all lower states $F''M''$ and by considering the ratios of the strengths for the different hfs components $F' \to F''$.

## 6.3 Day Airglows of the Planets

Various reviews noted in the bibliography discuss the airglow emissions for the various planets. Our approach here will be to review airglow from the viewpoint of excitation mechanisms to see what the emissions can tell us about atmospheric structure and processes.

### 6.3.1 Hydrogen and Helium

Lyman $\alpha$ emission ($1s\,^2S \leftarrow 2p\,^2P$) was first found in the Earth's night sky from a rocket flight and initially was thought to arise from H in the interplanetary medium. Subsequent analysis has made clear that the emission arises from resonant scattering by atmospheric H at very high altitudes. In the outer atmosphere H becomes the dominant constituent because of diffusive separation, and with the high temperature, low mass, and diminished gravity, the $e$-folding scale height $kT/Mg$ reaches the order of the Earth's radius. The density distribution is then no longer given by simple consideration of hydrostatic equilibrium and the perfect-gas law (see Section 1.1) and we will develop the more complete theory in Chapter 7. For the moment we will suppose that theoretical models can be developed and compared with observations to obtain densities and temperatures.

When observed from outside the atmosphere, the resonant scattering gives the planet an extensive coronal glow. These *planetary coronas* have also been observed on Venus and Mars and the Earth's corona has been mapped from the moon.

For accurate analysis of Ly $\alpha$ it is necessary to allow for the anisotropy of the $^2S_{1/2}$–$^2P_{3/2}$ fine-structure component ($^2S_{1/2}$–$^2P_{1/2}$ emits isotropically). The H-term structure is the same as for Na (Fig 6.7) except for the important difference that the nuclear spin is $I = \frac{1}{2}$ and there is less hfs in the H spectrum. If the hfs were completely separated, the anisotropy of the $^2P_{3/2}$ component would be given by (6.2.22), with

$$r \equiv \sum_{F'} \beta(F') \Big/ \sum_{F'} \alpha(F') = 0.203 \tag{6.3.1}$$

The isotropic $^2P_{1/2}$ component contributes one-half times the mean intensity of the $^2P_{3/2}$ component.

The actual situation is a little more complicated because $^2P_{3/2}$ has a hyperfine splitting of $\Delta\nu = 0.24 \times 10^8\ \text{sec}^{-1}$, whereas the natural broadening is $\Delta\nu = A/2\pi = 1 \times 10^8\ \text{sec}^{-1}$. Thus the upper hyperfine levels are indistinct, being blurred together by the natural width. In this case the anisotropy and polarization are given by

$$r \equiv \sum_{F'} \beta(F') \Big/ \sum_{F'} \alpha(F') = 0.30 \tag{6.3.2}$$

The Ly $\alpha$ emission at 1215 Å is by far the strongest H emission ($\sim$8 kR on the Earth's day side as seen by Apollo 16 on the moon) because it is a resonant transition and because Ly $\alpha$ emission by the sun is strong. The solar Ly $\beta$ (1s–3p) emission stimulates geocoronal scattering in Ly $\beta$ at 1025 Å and Balmer $\alpha$ fluorescence ($2s\,^2S \leftarrow 3p\,^2P$) at 6563 Å. This latter emission can be detected from the ground but it is very weak, about 3 R at night.

On Mars the Mariners 6, 7, and 9 found comparable emission rates, $\sim$5–6 kR, whereas for Venus the bright dayside limb seen by Mariner 5 emitted about 40 kR. We will discuss the hydrogen planetary coronas of the terrestrial planets further in Section 7.3 after developing the theory for an outer atmosphere. The helium resonance lines 584 Å from He I and 304 Å from He II are present in the Earth's daytime spectrum, but they have not been studied intensively, as Ly $\alpha$ has.

The Pioneer 10 spacecraft carried an ultraviolet photometer that measured 5.1 R of He I ($1s^2\,^1S_0$–$1s2p\,^1P_1$), 584 Å (as well as $\sim$1 kR Ly $\alpha$) on Jupiter. With a model of the Jovian atmosphere, these measurements allow a determination of the $[He]/[H_2]$ mixing ratio. The situation is a little confused because it is not clear whether the unexpected plasma temperature ($\sim 10^{3\circ}$K) for the high Jovian ionosphere also implies high neutral temperatures

(Section 5.3.3). With a low-temperature ($\sim 150°K$) model, the ratio is $0 \lesssim [He]/[H_2] < 0.28$, but high temperatures would allow more helium. The matter is of interest to theories for the origin and evolution of the solar system; the He/H mixing ratio for the sun would give $[He]/[H_2] \approx 0.14$. Pioneer 10 also found Lyman $\alpha$ ($\sim 300$ R) in a torus associated with the orbit of Io (also see Section 6.3.4).

On Venus the Mariner 10 measured He $\lambda 584$ at about 600 R. A model analysis indicates an exospheric temperature of 375°K, and this determination does not suffer from the ambiguity of the Lyman alpha results (see Section 7.3.3). Mariner 10 also found $\lambda 584$ on the dayside of Mercury at about 70 R, indicating a surface density of 4500 $cm^{-3}$ and a scale-height temperature of 575°K.

### 6.3.2 Atomic Oxygen

Another important emission is the resonance triplet of O I at 1302–4–6 Å [$2p^4\ ^3P_{2,1,0}$–$2p^3(^4S^0)3s\ ^3S_1{}^0$]. In the Earth's atmosphere the emission arises mainly from the 190 km region and has an overhead emission rate of $4\pi\mathscr{I} \sim 7.5$ kR.

This $\lambda 1304$ triplet, being excited at high altitudes by scattering of sunlight, offers the potential of obtaining the distribution of O in the high atmosphere, which itself is important to understanding atmospheric photochemistry and the ionosphere. In practice the analysis is complicated by an extraordinarily complex problem in radiative transfer. Absorption of sunlight from one of the lower levels in the $^3P$ term can be followed by re-emission to another level. Hence the three lines are radiatively coupled. The fact that very little O is required for these resonance transitions to become optically thick means the interpretations require radiative-transfer analysis. Further, the emissions may be absorbed by such constituents as $O_2$ and $CO_2$ and produced by dissociative excitation of $O_2$ and impact of energetic photoelectrons.

Interpretations of the $\lambda 1304$ emissions from Mars and Venus have generally been approached by comparing $\lambda 1304$ with the 1356–8 Å forbidden doublet of [O I]($2p^4\ ^3P_{2,1}$–$2p^3(^4S^0)3s\ ^5S_2{}^0$) observed simultaneously. The $^5S$ and $^3S$ terms are close together and $\lambda 1356$ thus gives a means of roughly estimating the excitation due to photoelectron impact or *dissociative excitation* of $CO_2$ since $^5S$ will not be excited by radiative absorption from the ground term.

For Mars the situation seems relatively satisfactory. The 1356 Å transition has been attributed (by elimination of other mechanisms) to dissociative excitation of $CO_2$,

$$CO_2 + h\nu(\lambda < 829\ \text{Å}) \rightarrow CO + O^* \qquad (6.3.3)$$

and $\lambda 1304$ is evidently due to resonance scattering and yields $[O]/[CO_2]$ mixing ratios of $10^{-2}$ at the ionosphere peak (at 135 km). On the other hand, when the same kind of analysis is applied to Venus it gives results that are incompatible with the $\lambda 1304$ and $\lambda 1356$ intensities and especially their ratio as obtained from a rocket flight.

### 6.3.3 Molecular Excitation

The principal molecular emissions in Earth's airglow arising from resonant/fluorescent scattering are the Atmospheric ($b\,^1\Sigma_g^+ \rightarrow X\,^3\Sigma_g^-$) and IR Atmospheric ($a\,^1\Delta_g \rightarrow X\,^3\Sigma_g^-$) systems of $O_2$. Other mechanisms evidently contribute to these band systems, however. For example, $O(^1D)$ atoms, produced by Schumann–Runge dissociation of $O_3$, may react by

$$O(^1D) + O_2(^3\Sigma_g^-) \rightarrow O(^3P) + O_2(b\,^1\Sigma_g^+) \tag{6.3.4}$$

which is a principal loss mechanism (along with collisional deactivation by $N_2$) of $O(^1D)$ atoms.

The most important source of $O_2(^1\Delta_g)$ molecules is photodissociation,

$$O_3 + h\nu \rightarrow O_2(a\,^1\Delta_g) + O \tag{6.3.5}$$

in the Hartley continuum, $\lambda\lambda 2000$–$3100$. Earth-based observations have disclosed this IR Atmospheric emission on Mars by detection of rotational lines in emission against the sunlit disk with very high spectral resolution. The $O_2(a\,^1\Delta_g)$ molecules imply small amounts of ozone, about $0.2 \times 10^{-4}$ atm-cm. Ultraviolet absorption measurements performed by Mariner 9 found ozone in amounts of $50 \times 10^{-4}$ atm-cm over the winter polar caps, but could not detect amounts less than $3 \times 10^{-4}$ atm-cm. The appearance of ozone at high winter latitudes implies that it favors dry conditions, and that water-related chemistry destroys it (see Section 6.4.1).

In the Earth's nitrogen-oxygen atmosphere, the chemistry of the 70–100 km region leads to the production of nitric oxide with concentrations of $[NO] \sim 4 \times 10^7$ cm$^{-3}$, which appear in the ultraviolet $\gamma$ bands ($A\,^2\Sigma^+ \rightarrow X\,^2\Pi$) with emission rates of $\sim 10^3$ R due to resonance scattering.

A day airglow of comparable emission rate arises from the Second Positive ($C\,^3\Pi_u \rightarrow B\,^3\Pi_g$) system of $N_2$. Although the bands are permitted, this excitation occurs from the ground state by photoelectron impact from the ground state $X\,^1\Sigma_g^+$, which requires a change of electron spin. Consequently, the Second Positive system is not subject to resonant scattering and is a good indication of excitation by photoelectron impact. Other band systems of $N_2$ and $N_2^+$ are present as well.

Similarly the Mariner spacecraft have observed intense airglows from CO, O, and $CO_2^+$ on Mars. Figure 6.8 shows the spectra and Fig. 6.9 gives the limb (tangential) emission rates. The strongest of these airglow systems is the forbidden Cameron bands ($a^3\Pi \rightarrow X\,^1\Sigma$) of CO, which have a maximum emission rate when seen tangentially of 600 kR. At $\lambda$2972 appears the forbidden "transauroral" line of O I; it is apparent from Fig. 6.3 that the "auroral" transition, $\lambda$5577, must also be strong, but was out of range of the ultraviolet spectrometer. These emissions cannot be due to resonance scattering because the abundances are too small for the metastable states to become excited by absorption. Rather they arise from the direct action of sunlight

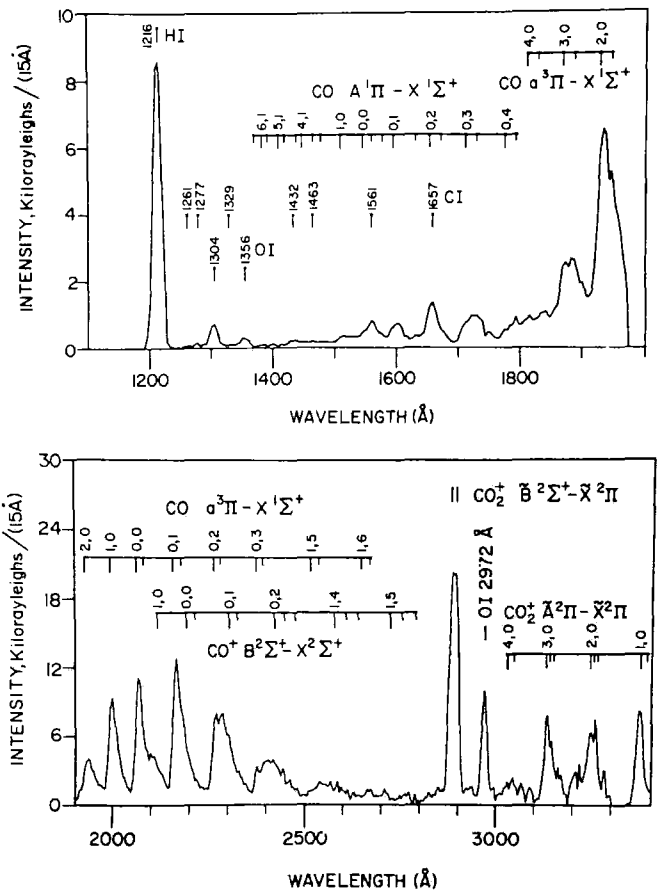



Fig. 6.8 Ultraviolet airglow of Mars obtained from Mariner 9 orbiter. The spectral tracings are averages of 120 individual limb observations with a resolution of 15 Å. [After Barth C.A.; Stewart, A. I.; Hord, C. W.; and Lane, A. L. (1972), *Icarus* **17**, 457.]

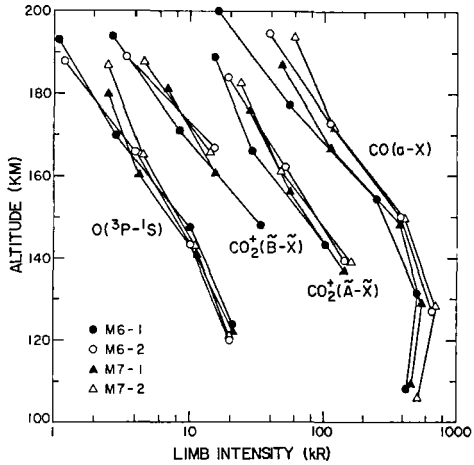



**Fig. 6.9** Height profiles of four airglow emissions on Mars. The data were obtained from four observation opportunities, as indicated: The first and second limb crossings for both the Mariner 6 and 7 fly-bys. [After STEWART (1972).]

on $CO_2$. There are three main possibilities: Dissociative excitation by photons,

$$CO_2 + h\nu \rightarrow CO^* + O^* \tag{6.3.6}$$

or by fast photoelectrons,

$$CO_2 + e \rightarrow CO^* + O^* \tag{6.3.7}$$

and photoionization followed by dissociative recombination,

$$\begin{aligned} CO_2 + h\nu &\rightarrow CO_2^+ + e \\ CO_2^+ + e &\rightarrow CO^* + O^* \end{aligned} \tag{6.3.8}$$

A physical discussion of these processes in terms of a model atmosphere can reproduce the scale height of the Cameron bands (19 $\pm$ 4.5 km) if the model has an exospheric temperature of 315 $\pm$ 75°K.

The Fourth Positive ($A\,^1\Pi \rightarrow X\,^1\Sigma^+$) system of CO appears weakly in Fig. 6.8 and this permitted system, like the resonance line $\lambda$1304 of O I, is likely due to resonance scattering. There are also two band systems from $CO_2^+$: the Fox–Duffendack–Barker ($\tilde{A}\,^2\Pi_u \rightarrow \tilde{X}\,^2\Pi_g$) and the "Doublet" ($\tilde{B}\,^2\Sigma_u^+ \rightarrow \tilde{X}\,^2\Pi_g$) systems. These emissions may arise from resonance scattering or from photoionization-excitation,

$$CO_2 + h\nu \rightarrow CO_2^+(\tilde{A}, \tilde{B}) + e \tag{6.3.9}$$

The airglow radiation from the Martian upper atmosphere constitutes a major energy loss of the extreme ultraviolet sunlight that is deposited there.

### 6.3.4 Alkali Metals

The most extensively studied alkali metal is Na I, but twilight and therefore presumably daytime, emissions are present also from Ca II, Li I, and K I. The seasonal and daily variations and the height distribution of [Na] above the peak concentration around 90 km have not been well understood. There have been problems associated with the $[Na^+]/[Na]$ ratio, the role of eddy diffusion, and the interaction between Na atoms and aerosols, all of which are suspected of being important as sources or sinks of free Na. With laser sounding techniques (*lidar*) the abundance of Na can be measured from the ground as a function of height. Observations of this nature promise to clarify the apparent peculiarities of Na distribution as well as the role of eddy transport near the mesopause.

Although not a conventional planetary airglow, the Na I emission from Io (Jupiter's inner Galilean satellite) is remarkably like airglow. The emission comes from an extensive cloud surrounding the satellite (see Fig. 6.10). The $D_2$ and $D_1$ emission rates are about 16 and 7 kR, respectively, and the excitation source seems to be resonance scattering of an optically thin, high-temperature cloud. The Pioneer 10 ultraviolet photometer measured ~ 300 R of Lyman $\alpha$ radiation from a 120° segment of a torus approximately in the orbit of Io (see Section 6.3.1). The satellite appears to have an ionosphere with peak electron densities, according to radio-occultation measures, of $N_e \sim 6 \times 10^4$ cm$^{-3}$.

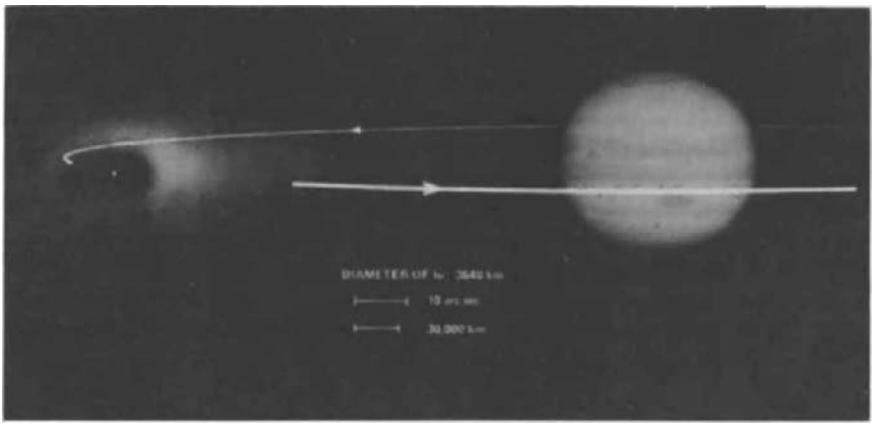


**Fig. 6.10** Image of Io's sodium cloud taken with the coude spectrograph at Table Mountain Observatory, California, on Feb. 19, 1977. The cloud was photographed in the $D_1 + D_2$ emissions. The photograph of Jupiter and the white dot indicating the position of Io have been added for orientation purposes. The dark area within the sodium cloud is produced by an occulting disk used to exclude direct continuum light from Io. [After MATSON, D. L.; GOLDBERG, B. A.; JOHNSON, T. V.; and CARLSON, R. W. (1978), *Science*, **199**, 531.]

The origin of these effects is not understood, but there is no shortage of speculation. Io passes through Jupiter's intense radiation belts and modulates the Jovian decametric radio emission. Consequently, the magnetic and/or radiation effects of Jupiter on Io are presumed to be somehow liberating vast quantities of Na and perhaps H from Io's surface. Why a body so much like Earth's moon would be packed with alkali-metal compounds is not clear. An alternative and less exhaustable source of Na ions might be Jupiter's own atmosphere.

## 6.4 Aeronomy of the Planets

The broad topic of aeronomy appears throughout this book. Chemical reactions are important at all levels—even in the Earth's troposphere where minor constituents and pollution products are concerned. The stratosphere was examined in some detail in Chapter 3 and aeronomic reactions involving ionospheres in Chapter 5. Here we look at aeronomy in the density regions of other planets that correspond roughly to the Earth's stratosphere.

### 6.4.1 Predominantly $CO_2$ Atmospheres

The problem of the low degree of association of $CO_2$ on Mars and Venus has been noted earlier (Section 5.3.1) Basically it is that $CO_2$ dissociation produces O atoms that should associate to form $O_2$ by (5.3.9) faster than $CO_2$ by (5.3.8). But the airglow and ionosphere observations indicate a near absence of $O_2$. One way out of the dilemma is a rapid downward transport of O atoms before much association into $O_2$ occurs. Then at lower altitudes, where water-related chemistry is important, a sink of CO and O is provided by the sequence

$$\begin{aligned} H + O_2 + M &\rightarrow HO_2 + M \\ HO_2 + O &\rightarrow OH + O_2 \\ CO + OH &\rightarrow CO_2 + H \\ \text{Net:}\quad CO + O &\rightarrow CO_2 \end{aligned} \tag{6.4.1}$$

On Mars the main source of $O_2$, used here as a catalyst, is from $O + OH \rightarrow O_2 + H$. With a rapid downward flow of O, its scale height is large and the eddy-transport equation (2.3.4) gives

$$[O]K/H = -\Phi_1 \tag{6.4.2}$$

where $\Phi_1$ is the O flux. Thus [O] and the production rate of $O_2$ vary as $K^{-1}$. A value of $K \sim 10^8$ cm$^2$/sec, very large by comparison with the Earth's stratosphere, seems required to fit the data.

If $K$ is somewhat smaller, the downward transport of O (and subsequent conversion back to $CO_2$) might be supplemented by catalytic conversion

of $O_2$ into $CO_2$. The sequence is

$$\begin{aligned} H + O_2 + M &\rightarrow HO_2 + M \quad (x2) \\ HO_2 + HO_2 &\rightarrow H_2O_2 + O_2 \\ H_2O_2 + h\nu &\rightarrow OH + OH \\ CO + OH &\rightarrow CO_2 + H \quad (x2) \\ \text{Net:} \quad 2CO + O_2 &\rightarrow 2CO_2 \end{aligned} \tag{6.4.3}$$

This mechanism, however, requires larger amounts of $HO_x$ than does (6.4.1) and may have problems in that regard.

Ozone destruction on Mars is tied to the $HO_x$ chemistry through (6.4.1) and through the dependence of $[O_3]$ on $[O]$ by the rapid sequence of

$$\begin{aligned} O_3 + h\nu &\rightarrow O_2 + O \\ O + O_2 + M &\rightarrow O_3 + M \end{aligned} \tag{6.4.4}$$

Thus the presence of $O_3$ on Mars mainly over the winter polar cap (see Section 6.3.3) is likely due to the freezing out of $H_2O_2$ and possibly $HO_2$, allowing $O_3$ to form there. Frozen polyoxides ($H_2O_3$, $H_2O_4$) are probably formed from $HO_2$. In addition, the Viking landers have confirmed relatively high concentrations of metal oxides in the soil and these materials may exchange oxygen with the atmosphere, the soil liberating it mainly during warm periods of dust storms.

On Venus the $O_2$ mixing ratio is $[O_2]/[CO_2] < 10^{-6}$, whereas on Mars it is $\sim 10^{-3}$. Simply this near absence of $O_2$ presents a problem, because aside from the $CO_2$ chemistry, $O_2$ should be produced as a product of $HO_x$ chemistry, involving reactions of $HO_x$ with O and $O_3$. In any case, if $O_2$ is destroyed as efficiently as it appears to be, some other catalytic reactions appear necessary, such as the scheme

$$\begin{aligned} Cl + O_2 + M &\rightarrow ClO_2 + M \\ ClO_2 + CO &\rightarrow ClO + CO_2 \\ ClO + CO &\rightarrow Cl + CO_2 \\ \text{Net:} \quad O_2 + 2CO &\rightarrow 2CO_2 \end{aligned} \tag{6.4.5}$$

### 6.4.2 Predominantly $H_2$ Atmospheres

Being farther removed from the sun, the major planets have much lower temperatures in their photochemically active middle atmospheres than does the Earth. This fact combined with the hydrogen dominated composition gives a photochemistry in which the main products are hydrocarbons—including some rather complex ones—and $NH_3$ (ammonia).

Atomic H is produced in the upper atmosphere by several processes, principally as follows: Photodissociation,

$$H_2 + h\nu(\lambda < 845\ \text{Å}) \rightarrow H + H \tag{6.4.6}$$

predissociation through absorption in the Lyman bands ($B\,^1\Sigma_u^+ \leftarrow X\,^1\Sigma_g^+$),

$$H_2(X, v = 0) + h\nu(\lambda < 1109\ \text{Å}) \rightarrow H_2(B, v') \rightarrow H_2(X, v'') \rightarrow H + H \tag{6.4.7}$$

provided $v'' > 14$; predissociation through the Werner bands ($C\,^1\Pi_u \leftarrow X\,^1\Sigma_g^+$) for $\lambda < 1009$ Å; ionization of $H_2$ leading to

$$H_2^+ + H_2 \rightarrow H_3^+ + H \tag{6.4.8}$$

which breaks an $H_2$ bond; and ionization of He, which yields

$$He^+ + H_2 \rightarrow H^+ + H + He \tag{6.4.9}$$

The H so formed diffuses downward to regions where it reacts with $CH_3$ to recycle $CH_4$ destroyed by photolysis.

The $CH_4$ (methane) molecule also undergoes predissociation. Absorption of Ly $\alpha$ leaves $CH_4$ in an excited state ($CH_4$*) with internal energy exceeding the dissociation energy (4.5 eV). Then $CH_4$* has a radiationless transition to the dissociation continuum of another state, such as

$$\begin{aligned} CH_4^* &\rightarrow CH_2(\tilde{a}\,^1A_1) + H_2 \\ CH_4^* &\rightarrow CH_2(\tilde{b}\,^1B_1) + H_2 \\ CH_4^* &\rightarrow CH(^2\Pi) + H + H_2 \\ CH_4^* &\rightarrow C(^1D) + 2H_2 \end{aligned} \tag{6.4.10}$$

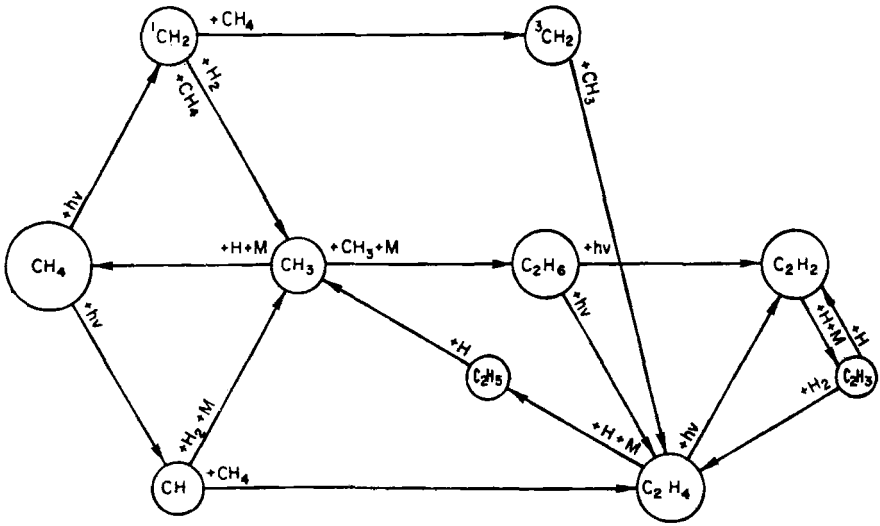



**Fig. 6.11** Principal reactions in the hydrocarbon photochemistry on Jupiter and other predominantly $H_2$ atmospheres. The notation $^1CH_2$ and $^3CH_2$ refers to radicals produced in singlet and triplet states, respectively. [After STROBEL (1975).]

From these dissociation products, $C_2H_6$ (ethane) is produced and then photodissociated into $C_2H_4$ (ethylene), $CH_3$ (methyl radical), and $CH_2$. Then $C_2H_4$* may predissociate into $C_2H_2$ (acetylene). The principal reaction routes are shown in Fig. 6.11.

Photolysis of $NH_3$ occurs primarily by

$$NH_3 + h\nu \rightarrow NH_2(\tilde{X}\,^2B_1) + H \qquad (6.4.11)$$

Other dissociation products are prevented by the screening of sunlight at $\lambda < 1600$ Å by the overhead $CH_4$.

At high pressures $NH_2$ reacts with itself to form $N_2H_4$ or with H to recycle $NH_3$. The $N_2H_4$ may be converted into $N_2$ and $2H_2$ or it may, at low temperatures, condense to form haze ("photochemical smog").

## BIBLIOGRAPHICAL NOTES

**Section 6.1 Airglow Photometry**

There are observations on record as early as 1788 of nights with an unusually large sky brightness and an increase in brightness from the zenith to the horizon. That the light of the night sky has a terrestrial component—what is now termed *airglow*—was suspected by a number of astronomers around 1900 and was definitely established through careful photometry by

YNTEMA, L. (1909), On the brightness of the sky and total amount of starlight, *Publ. Astr. Groningen* No **22**. 1–55.

The green line of [O I] had been found much earlier to be present, even when visible aurora was not, by

ÅNGSTRÖM, A. J. (1868), "Recherches sur le Spectre Solaire," W. Schultz, Upsala, pp. 41–42, and others, but the fact that the green line is permanently present at night was established by

SLIPHER, V. M. (1919), On the general auroral illumination of the sky and the wavelength of the chief aurora line, *Astrophys. J.*, **49**, 266–275.

The principal early work was done by C. FABRY and J. DUFAY in France and the fourth Lord Rayleigh (R. J. STRUTT) in England. The first measurement of the absolute intensity of the night airglow was made by

RAYLEIGH, LORD (R. J. STRUTT) (1930), Absolute intensity of the aurora line in the night sky and the number of atomic transitions, *Proc. Roy. Soc.* (*London*) **A129**, 458–467.

Rayleigh's result for the zenith was $4\pi\mathscr{I} = 1.81 \times 10^8$ quanta/cm$^2$ sec or 181 rayleighs, not far from the value (250 R) usually adopted as the average today. Incidentally, it is Rayleigh's father, the third Lord Rayleigh (J. W. Strutt), who is remembered for the theory of molecular scattering, among other things.

An extensive review of terrestrial airglow, including references to the main historical works, is included in

CHAMBERLAIN, J. W. (1961), "Physics of the Aurora and Airglow," Academic Press, New York.

More recent and excellect reviews are

HUNTEN, D. M. (1967), Spectroscopic studies of the twilight airglow, *Space Sci. Rev.* **6**, 493–573,

HUNTEN, D. M. (1971), Airglow—Introduction and review, *in* "The Radiating Atmosphere," (B. M. McCormac, ed.), pp. 3–16, Springer-Verlag, New York.

Determination of the rate coefficient of (6.1.9) in the laboratory and the resulting amount of $O(^1S)$ emission depends strongly on the rate of deactivation. See

FELDER, W. and YOUNG, R. A. (1972), Quenching of $O(^1S)$ by $O(^3P)$, *J. Chem. Phys.* **56**, 6028–6030.

In addition, there are reasons for thinking the chemical energy of association goes first into $O_2^*$ by (6.1.15) and then is transferred by a subsequent collision with $O(^3P)$ into $O(^1S)$. See

BARTH, C. A. and HILDENBRANDT, A. F. (1961), The 5577 A airglow emission mechanism, *J. Geophys. Res.* **66**, 985–986,
SLANGER, T. G. and BLACK, G. (1977), $O(^1S)$ in the lower thermosphere—Chapman vs Barth, *Planet. Space Sci.* **25**, 79–88,
DONAHUE, T. M. (1975), OGO-6 observations of 5577 Å, *in* "Atmospheres of Earth and the Planets," (B. M. McCormac, ed.), pp. 289–307, D. Reidel Publ. Co., Dordrecht, The Netherlands.

The problem of rotational temperatures of OH is discussed in

SUZUKI, K. and TOHMATSU, T. (1976), An interpretation of the rotational temperature of the airglow hydroxyl emissions, *Planet. Space Sci.* **24**, 665–671,
DICK, K. A. (1977), On the rotational temperature of hydroxyl airglow, *Planet. Space Sci.* **25**, 596–597,
KRASSOVSKI, V. I.; POTAPOV, B. P.; SEMENOV, A. I.; SOBOLEV, V. G.; SHAGAEV, M. V.; and SHEFOV, N. N. (1977), On the equilibrium nature of the rotational temperature of hydroxyl airglow, *Planet. Space Sci.* **25**, 596–597.

The history of the identification of reaction (6.1.16) as the source of the airglow continuum is given in

CHAMBERLAIN, J. W. (1961), op. cit. pp. 543–545.

Satellite photometric observations of the emission and the rate coefficient of the reaction are discussed by

DONAHUE, T. M. (1974), An upper limit to the product of NO and O densities from 105 to 120 km, *J. Geophys. Res.* **79**, 4337–4339.

The identification of the Venus nightglow as $O_2$ Herzberg II emission is due to

LAWRENCE, G. M.; BARTH, C. A.; and ARGABRIGHT, V. (1977), Excitation of the Venus night airglow, *Science* **195**, 573–574.

**Section 6.2 Resonant and Fluorescent Scattering of Sunlight**

The photon scattering coefficient, or $g$-value, is introduced in

CHAMBERLAIN, J. W. (1961), op. cit. Chapter 11, and a number of $g$-values are tabulated there (p. 425). Some additional and improved $g$-values are given by

HUNTEN, D. M. (1971), op. cit.

The tangentially integrated density (6.2.6) is due to

CHAPMAN, S. (1931), The absorption and dissociation or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, *Proc. Phys. Soc.* **43**, 483–501.

The theory of resonance scattering was developed by HEISENBERG, WEISSKOPF, DIRAC, and BREIT. The discussion here follows

BREIT, G. (1933), Quantum theory of dispersion: Parts VI and VII, *Rev. Mod. Phys.* **5**, 91–140.

**Section 6.3 Day Airglow**

Only the principal terrestrial emissions are discussed here. For additional information see reviews by

HUNTEN, D. M. (1971), op. cit.,
NOXON, J. F. (1968), Day airglow, *Space Sci. Rev.* **8**, 92–134,
LLEWELLYN, E. J. and EVANS, W. F. J. (1971), The dayglow, *in* "The Radiating Atmosphere," (B. M. McCormac, ed.), pp. 17–33, Springer-Verlag, New York,
RUNDLE, H. N. (1971), Dayglow and twilight excitation mechanisms for airglow, *in* "The Radiating Atmosphere," (B. M. McCormac, ed.), pp. 90–104, Springer-Verlag, New York,
BROADFOOT, A. L. (1971), Dayglow nitrogen band systems, *in* "The Radiating Atmosphere," (B. M. McCormac, ed.), pp. 34–44, Springer-Verlag, New York.

**Section 6.3.1 Hydrogen and Helium**

Lyman α was discovered in the night sky from a rocket flown in late 1955 by

BYRAM, E. T.; CHUBB, T. A.; FRIEDMAN, H.; and KUPPERIAN, J. (1957), Far ultraviolet radiation in the night sky, *in* "The Threshold of Space," (M. Zelikoff, ed.), pp. 203–210, Pergamon Press, London.

Observations of terrestrial Ly α and Ly β from outside the atmosphere are reported in

CARRUTHERS, G. R.; PAGE, T.; and MEIER, R. R. (1976), Apollo 16 Lyman alpha imagery of the hydrogen geocorona, *J. Geophys. Res.* **81**, 1664–1672,
MEIER, R. R.; CARRUTHERS, G. R.; PAGE, T. L.; and LEVASSEUR-REGOURD, A.-C. (1977), Geocoronal Lyman β and Balmer α emissions measured during the Apollo 16 mission, *J. Geophys. Res.* **82**, 737–739

The anisotropy of H scattering in Ly α and other transitions is discussed by

BRANDT, J. C. and CHAMBERLAIN, J. W. (1959), Interplanetary gas.I. Hydrogen radiation in the night sky, *Astrophys. J. 130*, 670–682.

Observations of H resonance lines from Venus were made by

BARTH, C. A.; WALLACE, L.; and PEARCE, J. B. (1968), Mariner 5 measurements of Lyman alpha radiation near Venus, *J. Geophys. Res.* **73**, 2541–2545,
KURT, V. G.; ROMANOVA, H. H.; SMIRNOV, A. S.; BERTAUX, J. L.; and BLAMONT, J. (1976), Hydrogen observations at Lyman alpha of the upper atmosphere of Venus with Venera-9 and Venera-10, Paper presented at COSPAR meeting, Philadelphia, July 1976,
BROADFOOT, A. L. (1976), Ultraviolet spectrometry of the inner solar system from Mariner 10, *Rev. Geophys. Space Phys.* **14**, 625–627.

The latter paper also reviews the Mariner 10 measurements of He I on Venus and of H and He I on Mercury. Observations of H from Mariners 6, 7, and 9 at Mars are summarized in

BARTH, C. A.; STEWART, A. I.; HORD, C. W.; and LANE, A. L. (1972), Mariner 9 ultraviolet spectrometer experiment: Mars airglow spectroscopy and variations in Lyman alpha, *Icarus* **17**, 457–468.

The Lyman α airglow on Jupiter may be interpreted to give the eddy diffusion coefficient for the upper atmosphere, since the H abundance there depends on the rate of downward transport to high-density regions where association of $H_2$ occurs. See

WALLACE, L. and HUNTEN, D. M. (1973), The Lyman-alpha albedo of Jupiter, *Astrophys. J.* **182**, 1013–1031.

The observations of H and He resonance radiation from Jupiter and the Lyman α torus associated with Io are described and analyzed by

CARLSON, R. W. and JUDGE, D. L. (1974), Pioneer 10 ultraviolet photometer observations at Jupiter encounter, *J. Geophys. Res.*, **79**, 3623–3633,

CARLSON, R. W. and JUDGE, D. L. (1976), Pioneer 10 ultraviolet photometer observations of Jupiter: The helium to hydrogen ratio, *in* "Jupiter," (T. Gehrels, ed.), pp. 418–440, Univ. Arizona Press, Tucson,

CARLSON, R. W. and JUDGE, D. L. (1975), Pioneer 10 ultraviolet photometer observations of the Jovian hydrogen torus: The angular distribution, *Icarus* **24**, 395–399.

**Section 6.3.2 Atomic Oxygen**

The resonance lines of O I at 1304 Å were first detected in the terrestrial daytime airglow by

CHUBB, T. A.; BYRAM, E. T.; FRIEDMAN, H.; and KUPPERIAN, J. E. (1958), The use of radiation absorption and luminescence in upper air density measurements, *Ann. Géophys.* **14**, 109–116.

The radiative transfer analysis of the terrestrial emission is due to

STRICKLAND, D. J. and DONAHUE, T. M. (1970), Excitation and radiative transport of O I 1304 Å resonance radiation, I. The dayglow, *Planet. Space Sci.* **18**, 661–689,

CHAMBERLAIN, J. W. and WALLACE, L. (1974), Formation of coupled spectral lines in a planetary atmosphere, *Astrophys. J.* **190**, 487–495,

CHAMBERLAIN, J. W. and SMITH, G. R. (1975), Formation of coupled spectral lines in a planetary atmosphere: II Numerical evaluation of approximate solutions, *Astrophys. J.* **199**, 282–288,

STRICKLAND, D. J. and ANDERSON, D. E., JR. (1977), The O I 1304-Å nadir intensity/column production rate ratio and its application to airglow studies, *J. Geophys. Res.* **82**, 1013–1016.

Day airglow spectra of Mars obtained from Mariners 6, 7, and 9 are reviewed by

BARTH, C. A. (1974), The atmosphere of Mars, *Ann. Rev. Earth Planet. Sci.* **2**, 333–368.

The analysis of Mariner Mars data on O I is given by

THOMAS, G. E. (1971), Neutral composition of the upper atmosphere of Mars as determined from the Mariner uv spectrometer experiments, *J. Atmos. Sci.* **28**, 859–868,

STRICKLAND, D. J.; THOMAS, G. E.; and SPARKS, P. R. (1972), Mariner 6 and 7 ultraviolet experiment: Analysis of the O I 1304- and 1356-Å emissions, *J. Geophys. Res.* **77**, 4052–4068.

The O I emissions (along with Ly $\alpha$) were observed on Venus with a rocket by

MOOS, H. W. and ROTTMAN, G. J. (1971), O I and H I emissions from the upper atmosphere of Venus, *Astrophys. J. Lettr.* **169**, L127–L130.

Mariner 5 and 10 Venus data are given by

BARTH, C. A.; PEARCE, J. B.; KELLY, K. K.; WALLACE, L.; and FASTIE, W. G. (1967), Ultraviolet emission observed near Venus from Mariner V, *Science* **158**, 1675–1678,

BROADFOOT, A. L.; KUMAR, S.; BELTON, M. J. S.; and MCELROY, M. B. (1974), Ultraviolet observations of Venus from Mariner 10: Preliminary results, *Science* **183**, 1315–1318.

Analysis of Venus O I emission is due to

STRICKLAND, D. J. (1973), The O I 1304- and 1356-Å emissions from the atmosphere of Venus, *J. Geophys. Res.* **78**, 2827–2836,

ANDERSON, D. E., JR. (1975), The Mariner 5 ultraviolet photometer experiment: Analysis of Rayleigh-scattered and 1304-Å radiation from Venus, *J. Geophys. Res.* **80**, 3063–3067.

**Section 6.3.3 Molecular Excitation**

The Mariner observations are reviewed by

BARTH, C. A. (1974), op. cit.

and have been analyzed by

THOMAS, G. E. (1971), op. cit.,
STEWART, A. I. (1972), Mariner 6 and 7 ultraviolet spectrometer experiment: Implications of $CO_2^+$, CO, and O airglow, *J. Geophys. Res.* **77**, 54–68.

Observations of $O_2$ Martian day airglow from Earth have been used to infer low levels of Martian $O_3$ by

NOXON, J. F.; TRAUB, W. A.; CARLETON, N. P.; and CONNES, P. (1976), Detection of $O_2$ dayglow emission from Mars and the Martian ozone abundance, *Astrophys. J.* **207**, 1025–1035.

**Section 6.3.4 Alkali Metals**

For Earth's airglow, see the reviews noted for Section 6.3, especially RUNDLE (1971). For a discussion of laser sounding of the free Na in the atmosphere see

MEGIE, G. and BLAMONT, J. E. (1977), Laser sounding of atmospheric sodium: Interpretation in terms of global atmospheric parameters, *Planet. Space Sci.* **25**, 1093–1109.

A review of the emission features of Io with models that have been put forth to account for this unique phenomenon is given by

BROWN, R. A. and YUNG, Y. L. (1976), Io, its atmosphere and optical emissions, *in* "Jupiter," (T. Gehrels, ed.), pp. 1102–1145, Univ. Arizona Press, Tucson.

The possibility that Io's atmosphere is drawn from Jupiter's ionosphere, through currents parallel to the Jovian magnetic field, is the suggestion of

CLOUTIER, P. A.; DANIELL, R. E., JR.; DESSLER, A. J.; and HILL, T. W. (1978), A cometary ionosphere model for Io, *Astrophys. Space Sci.*, in press.

**Section 6.4 Aeronomy of the Planets**

An excellent two-volume text on aeronomy, with emphasis on the terrestrial atmosphere, is

BANKS, P. M. and KOCKARTS, G. (1973), "Aeronomy," Academic Press, New York.

Aeronomy of $CO_2$ atmospheres is reviewed by

KUMAR, S. (1975), The ionosphere and upper atmosphere of Venus, *in* "Atmospheres of the Earth and Planets." (B. M. McCormac, ed.), pp. 385 399, D. Reidel Publ. Co., Dordrecht, The Netherlands,
MCCONNELL, J. C. (1973), The atmosphere of Mars, *in* "Physics and Chemistry of Upper Atmospheres," (B. M. McCormac, ed.), pp. 309–334, D. Reidel Publ. Co., Dordrecht, The Netherlands,
HUNTEN, D. M. (1974), Aeronomy of the lower atmosphere of Mars, *Rev. Geophys. Space Phys.* **12**, 529–535,
KONG, T. Y. and MCELROY, M. B. (1977a), Photochemistry of the Martian atmosphere, *Icarus* **32**, 168–189.

That $O_2$ should be a principal constituent of the Mars upper atmosphere, regardless of the rarity of $O_2$ in the lower atmosphere, was first advocated in

CHAMBERLAIN, J. W. (1962), Upper atmospheres of the planets, *Astrophys. J.* **136**, 582–293.

The model with a rapid downward flow of O is due to

MCELROY, M. B. and DONAHUE, T. M. (1972), Stability of the Martian atmosphere, *Science* **177**, 986–988.

The catalytic conversion of CO and $O_2$ into $CO_2$ by (6.60) is a proposal of

PARKINSON, T. D. and HUNTEN, D. M. (1972), Spectroscopy and aeronomy of $O_2$ on Mars, *J. Atmos. Sci.* **29**, 1380–1390.

Short-term fluctuations of Martian $O_2$ due to interchange between oxidized solids and free $O_2$ are discussed by

HUNTEN, D. M. (1974), op. cit.,

and the oxidation of the surface is discussed in

HUGUENIN, R. L. (1974), The formation of goethite and hydrated clay minerals on Mars, *J. Geophys. Res.* 3895–3905.

The ozone distribution is analyzed by

KONG, T. Y. and MCELROY, M. B. (1977b), The global distribution of $O_3$ on Mars, *Planet. Space Sci.* **25**, 839–857.

Alternate catalytic cycles that may be more appropriate for Venus are discussed in

PRINN, R. G. (1973), The upper atmosphere of Venus: A review, *in* "Physics and Chemistry of Upper Atmospheres," (B. M. McCormac, ed.), pp. 335–344, D. Reidel Publ. Co., Dordrecht, The Netherlands.

A thorough analysis of Venus' aeronomy is given by

SZE, N. D. and MCELROY, M. B. (1975), Some problems in Venus' aeronomy, *Planet. Space Sci. 23*, 763–786,
LIU, S. C. and DONAHUE, T. M. (1975), The aeronomy of the upper atmosphere of Venus. *Icarus* **24**, 148–156.

Aeronomy of the Jovian atmosphere is treated in

STROBEL, D. F. (1975), Aeronomy of the major planets: Photochemistry of ammonia and hydrocarbons, *Rev. Geophys. Space Phys.* **13**, 372–382.

The pioneering work on aeronomy of $H_2$ atmospheres was

WILDT, R. (1937), Photochemistry of planetary atmospheres, *Astrophys. J.* **86**, 321–336.

## PROBLEMS

**6.1** *Photometry for resonance scattering.* The phase function $p(\cos\Theta)$ for resonance scattering is in general a linear combination of the isotropic and Rayleigh (4.3.2) phase functions. (a) Show that in the conservative case if

$$p(\cos\Theta) = a + b\cos^2\Theta$$

then $a = 1 - b/3$, where $p$ is normalized by (1.2.4). (b) Show that when scattering is not isotropic the relation (6.1.5) between the integrated intensity and total emission rate is replaced by

$$\mu 4\pi \mathscr{J}(\mu, \phi) = p(\cos\Theta) \int_0^\infty E(z)\, dz$$

(c) If sunlight is incident from a zenith angle of 60° and the observer is also at a zenith angle of 60°, what is the observer's azimuth from the sun if $p(\cos\Theta) = 1$ when $b = 0.3$?

**6.2** *Polarization and anisotropy without hyperfine structure.* Neglect the hyperfine structure of the $D$ lines of Na and assume that there is no external magnetic field (strong enough to separate the various Zeeman components). (a) Show that $D_1$ is not polarized. (b) Find the polarization, $p(\Theta)$, and anisotropy, $\mathscr{E}(\Theta)/\langle\mathscr{E}\rangle$ for $D_2$. (Check: At $\Theta = \pi/2$, $p = \frac{3}{7}$.)

**6.3** *Polarization and anisotropy with an external magnetic field.* Repeat Problem 6.2 with an external magnetic field perpendicular to the scattering plane.

**6.4** *Polarization and anisotropy with hyperfine structure.* Repeat Problem 6.2 with the hfs shown in Fig. 6.7. Use the following tables for relative hfs intensities:

$I = \frac{3}{2}$

| $J' = \frac{3}{2}$ | $F$ \ $F'$ | $J = \frac{1}{2}$ 1 | 2 |
|---|---|---|---|
| | 0 | 14.3 | - - |
| | 1 | 35.7 | 7.1 |
| | 2 | 35.7 | 35.7 |
| | 3 | — | 100 |

| $J' = \frac{1}{2}$ | $F$ \ $F'$ | $J = \frac{1}{2}$ 1 | 2 |
|---|---|---|---|
| | 1 | 20 | 100 |
| | 2 | 100 | 100 |

**6.5** *Sum rules for Zeeman components.* For a transition $F' \to F$, write the transition probabilities from (6.2.26) and (6.2.27) for all transitions starting on $M' = F'$ and for all transitions starting on $M' = F' - 1$ and show that the sum rule is obeyed.

**6.6** *Magnetic dipole radiation.* The $\lambda$6300–64 lines of [O I] ($^3P_{2,1}$–$^1D_2$) are forbidden by electric-dipole selection rules but permitted as magnetic-dipole transitions. To calculate their scattering replace the electric vector of the radiation field by the magnetic vector and proceed as usual. Obtain the intensity of Zeeman components with (6.2.26)–(6.2.31), where $A_\pi$ and $A_\sigma$ are now partial transition probabilities for components polarized with the *magnetic* vector of the wave parallel and perpendicular, respectively, to the external magnetic field. Assume the incident fluxes at $\lambda$6300 and $\lambda$6364 are equal. The total transition probabilities are in the ratio $A_{6300}/A_{6364} = 3$ and $^{16}O$ has no nuclear spin. What is the degree of polarization of $\lambda$6300 when observed perpendicular to the direction of the sun in the absence of an external magnetic field? (Specify how you define polarization for this problem.)

**6.7** *Green-line polarization.* Many years ago it was reported that the $\lambda$5577 line of [O I] ($^1D_2$–$^1S_0$) is polarized in twilight. The conclusion drawn was that this polarization shows that the $\lambda$5577 twilight airglow has a component due to resonance scattering of sunlight. From general principles of the nature of polarization and atomic wave functions, state in one sentence why this conclusion was absurd. (Note: $\lambda$5577 arises from an electric quadrupole transition, and the dipole selection rules do not apply. However, the solution is *not* that $^1S_0$ has a long lifetime and is subject to collisional depolarization.)

**6.8** *Optical thickness of sodium on Io.* The $D_2$ line ($\lambda$5890) of Na I has an oscillator strength $f = 0.65$ and at the Earth its scattering coefficient is $g = 0.6$ photon/sec molecule. When observed at maximum elongation from Jupiter, the Na cloud around Io has an apparent emission rate in $D_2$ of 16 kR. What is the approximate minimum temperature of the cloud if it is excited by resonance scattering and its optical thickness at the center of $D_2$ is small (i.e., $\tau \lesssim 1$)? [Ignore the anisotropy of scattering for this purpose; the absorption cross section is related to $f$ by (1.6.5).]

**6.9** *Helium scattering.* The He I resonance line $\lambda$584 is observed at a zenith angle of 60° when the sun is in the local zenith. The total apparent emission rate is $4\pi\mathcal{J}(\mu = \frac{1}{2}) = 100$ R. What is the true integrated emission rate for a vertical column of atmosphere? (The $^4$He nucleus has zero spin.)

# *Chapter 7*
# STABILITY OF PLANETARY ATMOSPHERES

The stability of the Earth's atmosphere is a matter of urgent concern. We are aware of major climate changes in the past and (perhaps related in some cases) periods of selective faunal extinction. Human activities now have a major impact on the atmosphere and a potential to virtually ruin it. Hence, natural changes are not our principal concern, but an understanding of natural evolution should help our anticipation of alterations induced by society.

It may be quite fortuitous that planet Earth is habitable at all. Of the terrestrial planets, Mercury is too small and hot to hold much atmosphere. Venus, although receiving only twice the sunlight of Earth, evidently became unstable in the sense that the temperature rose until there was no surface water left (see Section 7.4.1). In spite of a low exospheric temperature on Mars, thermal escape is rapid, and a chemically induced loss of oxygen as well may have always kept a secondary atmosphere on Mars from developing fully (see Section 7.3.4).

Much of this chapter is concerned with exospheres and the escape of gases from the uppermost levels of an atmosphere. This is a subject that seemed fairly straightforward before space measurements were made of the exospheres of Earth and Venus. There were surprises in store that have considerably altered our understanding of atmospheric stability.

## 7.1 Quasi-Collisionless Exospheres

The uppermost fringes of a neutral planetary atmosphere constitute the *exosphere;* when the emphasis is on its scattered radiation (such as Lyman $\alpha$ of H), an exosphere is frequently called a *planetary corona.* The base of an exosphere (the *critical level* or, according to most current usage, the *exobase*) is defined as the height where the overhead integrated density accounts for one mean free path $l$ for a fast atom. Since an exosphere, to a first approximation, has a constant scale height $H$, the exobase is located at the radius $r_c$ from the planetary center where

$$\int_{r_c}^{\infty} N(r)Q\,dr \approx QN(r_c)H = 1 \tag{7.1.1}$$

for local density $N(r)$ and collision cross section $Q(\approx 3 \times 10^{-15}\ \text{cm}^2)$. At constant density the mean free path is $l = 1/NQ$. Hence the exobase is at the height $r_c$ where $l(r_c) = H$.

The classic exosphere is one that has a Maxwellian distribution of velocities at and below the exobase and no collisions whatever above that level. We will later modify the strict prohibition against collisions to account for a certain portion of the exospheric atoms called the *satellite population.*

### 7.1.1 Jeans Escape

Let the density of particles of mass $M$ in six-dimensional phase space be $f(\mathbf{r}, \mathbf{v})$. If the velocity distribution function is Maxwellian,

$$f(\mathbf{r}, \mathbf{v}) = N(r)\frac{\exp(-Mv_r^2/2kT - Mv_\chi^2/2kT)}{(2\pi kT/M)^{3/2}} \tag{7.1.2}$$

where $v_r$ is the radial velocity and $v_\chi$ the transverse velocity. The escape flux at the exobase is, with the element of velocity space written in cylindrical coordinates,

$$\mathcal{F}(r_c) = \int_{v_r}\int_{v_\chi}\int_0^{2\pi} v_r f(r_c, \mathbf{v}) v_\chi\,d\phi\,dv_\chi\,dv_r \tag{7.1.3}$$

where $\phi$ is the azimuthal angle for the transverse velocity component $v_\chi$. Some care is required in establishing the integration limits for $v_r$ and $v_\chi$. If the radial velocity is below the escape velocity,

$$v_{\text{esc}} = \left(\frac{2G\mathcal{M}}{r}\right)^{1/2} \tag{7.1.4}$$

(where $G$ is the gravitational constant and $\mathcal{M}$ is the mass of the planet), the atom can escape only if ${v_\chi}^2 > v_{\text{esc}}^2 - {v_r}^2 \equiv {v_2}^2$. Thus we have

$$\begin{aligned}\mathscr{F}_{\text{Jea}}(r_c) &= 2\pi \int_{v_r=0}^{v_{\text{esc}}} \int_{v_\chi=v_2}^{\infty} f(r_c, v) v_\chi v_r \, dv_\chi \, dv_r \\ &\quad + 2\pi \int_{v_r=v_{\text{esc}}}^{\infty} \int_{v_\chi=0}^{\infty} f(r_c, \mathbf{v}) v_\chi v_r \, dv_\chi \, dv_r \\ &= \frac{N(r_c)U}{2\pi^{1/2}} e^{-v_{\text{esc}}^2/U^2} \left( \frac{v_{\text{esc}}^2}{U^2} + 1 \right) \end{aligned} \tag{7.1.5}$$

where the *most probable velocity* of a Maxwellian distribution is

$$U = \left( \frac{2kT}{M} \right)^{1/2} \tag{7.1.6}$$

Equation (7.1.5) is the *Jeans formula* for the rate of escape (atom/cm$^2$ sec) by *thermal evaporation* of a gas from a planetary atmosphere. The derivation is much simpler if (7.1.3) is written with the velocity space expressed in spherical coordinates (Problem 7.1), but in the following discussion we will occasionally find it convenient to use cylindrical velocity components.

### 7.1.2 Liouville's Equation and the Density in Phase Space

The theory throughout is developed for spherical symmetry. The atmospheric structure (density, composition, and temperature) is assumed uniform over the globe; there are no latitude or day-night effects, and the atmosphere does not rotate.

We assume initially a critical level or exobase above which collisions are negligible and below which collisions maintain a complete Maxwellian distribution of velocities. In the tenuous regions where atomic collisions are rare, the spatial and momentum distribution of particles is governed by Liouville's equation. It may be derived by analogy with the ordinary equation of continuity (2.1.1). A large group of particles distributed over spatial ($q_i$) and momentum ($p_i$) coordinates in six-dimensional phase space according to some function $f(q_i, p_i)$ may then be said to have a "density" in phase space of $f(q_i, p_i)$. In (2.1.1) adapted to phase-space coordinates, the velocity component $v_i$ is replaced by $\dot{q}_i$ and $\dot{p}_i$ and Liouville's equation is

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{3} \left( \frac{\partial}{\partial q_i}(f\dot{q}_i) + \frac{\partial}{\partial p_i}(f\dot{p}_i) \right) = 0 \tag{7.1.7}$$

The particles are subject to the equations of motion (in this case the particles are under the influence of gravitation). In Hamiltonian form, we have

$$\dot{q}_i = \frac{\partial \mathscr{H}}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathscr{H}}{\partial q_i} \tag{7.1.8}$$

where $\mathscr{H} = \mathscr{H}(q_i, p_i)$ is the Hamiltonian function—the total energy expressed in terms of $q_i$ and $p_i$. Expanding the derivatives in (7.1.7) and using (7.1.8) to eliminate $\partial \dot{q}_i/\partial q_i = -\partial \dot{p}_i/\partial p_i$ gives

$$\frac{df}{dt} \equiv \frac{\partial f}{\partial t} + \sum_{i=1}^{3} \left( \frac{\partial f}{\partial q_i} \dot{q}_i + \frac{\partial f}{\partial p_i} \dot{p}_i \right) = 0 \tag{7.1.9}$$

The term on the left side is the total derivative to be taken along the path (in phase space) followed by a closed element of "volume" as it "moves" according to the equations of motion. Liouville's theorem thus states that the density in phase space remains constant along a dynamical trajectory.

Let us apply Liouville's theorem to the density and velocity distribution in a corona. We shall assume an isotropic, Maxwellian distribution of velocities for temperature $T_c$ just below the exobase, which is at radial distance $r_c$ from the center of the planet, and a total density $N(r_c) \equiv N_c$. We shall then seek the distribution function at greater heights $r$. For spherical symmetry and a steady state ($\partial f/\partial t = 0$), (7.1.9) is

$$\frac{\partial f}{\partial r}\frac{dr}{dt} = -\frac{\partial f}{\partial p_r}\frac{dp_r}{dt} \tag{7.1.10}$$

where $p_r$ is the radial component of momentum.

One may readily show from the Lagrangian of a particle in a gravitational field that the second generalized momentum is $P_\chi = Mr^2\dot{\chi}$, where $\dot{\chi}$ is the angular velocity. The Hamiltonian is thus written

$$\mathscr{H} = \frac{p_r^2}{2M} + \frac{P_\chi^2}{2Mr^2} - \frac{G\mathscr{M}M}{r} \tag{7.1.11}$$

where $P_\chi$ is the angular momentum, a constant for an atom in free flight. That this is the correct form may be verified from equations (7.1.8), the first of which recovers the radial and angular velocities, whereas the second yields the equations of motion:

$$\frac{dp_r}{dt} = \frac{P_\chi^2}{Mr^3} - \frac{G\mathscr{M}M}{r^2} \tag{7.1.12}$$

and

$$\frac{dP_\chi}{dt} = 0 \tag{7.1.13}$$

At the exobase, the Maxwellian distribution,

$$f(r_c, p_r, P_\chi) = \frac{N_c e^{-p_c^2/2MkT_c} e^{-P_\chi^2/2MkT_c r_c^2}}{(2\pi MkT_c)^{3/2}} \tag{7.1.14}$$

gives

$$\frac{\partial f}{\partial p_r} = -\frac{p_r}{MkT_c} f \tag{7.1.15}$$

Substituting (7.1.12) and (7.1.15) into (7.1.10) yields

$$\frac{d(\ln f)}{dr} = -\left(\frac{G\mathcal{M}M}{kT_c r^2} - \frac{P_\chi^2}{MkT_c r^3}\right) \tag{7.1.16}$$

which is independent of $p_r$ and can be integrated upwards from the exobase.

It is convenient to write the absolute value of the gravitational potential energy in units of $kT_c$:

$$\lambda(r) \equiv \frac{G\mathcal{M}M}{kT_c r} = \frac{v_{\text{esc}}^2}{U^2} \tag{7.1.17}$$

Then integrating (7.1.16) from $r_c$ to $r$ and using (7.1.14) for the boundary condition gives the distribution function at any height:

$$f(r, p_r, P_\chi) = \frac{N_c e^{-(\lambda_c - \lambda)} e^{-p_r^2/2MkT_c} e^{-P_\chi^2/2MkT_c r^2}}{(2\pi MkT_c)^{3/2}} \tag{7.1.18}$$

This equation is valid for all combinations of $p_r$ and $P_\chi$ allowed by the equations of motion.

If these momenta are independent and can take on all values ($-\infty < p_r < \infty$; $0 < P_\chi < \infty$), as they do below the exobase, an integration over all momentum space gives a density we shall denote as $N_b$:

$$N_b(r) \equiv \int f(q_i, p_i)\, d^3 p_i = N_c e^{-(\lambda_c - \lambda)} \tag{7.1.19}$$

where $d^3 p_i$ signifies that the integration is over all momentum coordinates. This is the generalized form (i.e., it allows for the variation with $r$ of gravitational attraction) of the (isothermal) *barometric law* (1.1.4). The derivation in Section 1.1 used hydrostatic equilibrium and the perfect gas law, which are valid so long as the gas pressure is isotropic and the mean kinetic energy per atom is a constant times the temperature. The derivation of (7.1.19) from Liouville's equation is a direct consequence of the Maxwellian distribution holding even in the region free of collisions, provided the momenta do extend over all possible values. For then the mean energy per atom is unchanged and the barometric law naturally emerges.

All this depends, however, on having the entire range of molecular momenta allowed. At substantial distances above the exobase this is not the case and the barometric law breaks down. If strictly true (7.1.19) would predict a finite density and pressure is achieved as $r \to \infty$ (and $\lambda \to 0$), and

to resolve this paradox it would be necessary to include the mass of the atmosphere within $r$ in deriving a more accurate barometric equation.

In practice, the pressure at large distances is decidedly directional and the mean kinetic energy per atom decreases. Hence the barometric law must break down. Even with the macroscopic equations of hydrodynamics, the atmosphere is not strictly in hydrostatic equilibrium. It is expanding slightly—that is, some matter is being lost, which in the kinetic theory corresponds to evaporative loss. Thus the problem is one of hydrodynamics rather than hydrostatics. But to treat the density distribution accurately it is necessary to examine the individual Keplerian orbits of particles in the corona. This we shall do in the next section.

### 7.1.3 Distribution of Density with Height

The density $N(r)$ is obtained by integrating (7.1.18) over the momentum space that is populated in accordance with the equations of motion (7.1.12–7.1.13). We shall write the density as a product of the barometric density of (7.1.19) and a *partition function* $\zeta$, which will describe how particles are partitioned in different classes of orbit. Thus we have

$$N(r) = N_b(r)\zeta(\lambda) = N_c e^{-(\lambda_c - \lambda)}\zeta(\lambda) \tag{7.1.20}$$

where

$$\zeta(\lambda) = \frac{2\pi}{r^2(2\pi MkT_c)^{3/2}} \iint e^{-p_r^2/2MkT_c - P_\chi^2/2MkT_c r^2} P_\chi \, dP_\chi \, dp_r \tag{7.1.21}$$

Introducing changes of variable to express the energies in units of $kT_c$, as with (7.1.17) for the potential, we let

$$v(r) = \frac{P_\chi^{\,2}}{2MkT_c r^2} = \frac{v_\chi^{\,2}}{U^2} \tag{7.1.22}$$

and

$$\xi(r) = \frac{p_r}{(2MkT_c)^{1/2}} = \frac{v_r}{U} \tag{7.1.23}$$

Alternatively, the total kinetic energy of a particle is

$$\psi(r) = \frac{Mv^2}{2kT_c} = \frac{v^2}{U^2} = \xi^2 + v \tag{7.1.24}$$

and the polar angle of the velocity vector is $\theta = \cos^{-1}\mu$, where

$$\mu = \frac{\xi}{(\xi^2 + v)^{1/2}} \tag{7.1.25}$$

The total energy, which is a constant for each particle, is from (7.1.11)

$$h = \mathscr{H}/kT_c = \xi^2 + v - \lambda = \psi - \lambda \tag{7.1.26}$$

With the substitution of (7.1.22–7.1.23), the partition function (7.1.21) becomes

$$\zeta(\lambda) = \frac{1}{\pi^{1/2}} \iint e^{-\xi^2 - v}\, dv\, d\xi \tag{7.1.27}$$

The particle orbits fall into three distinct categories with different integration limits. We shall consider each of these groups in turn. In general, $\zeta$ may be regarded as the fraction of the isotropic, Maxwellian distribution that is actually present; for no dynamical restrictions on the orbits, $\zeta = 1$.

**Ballistic Orbits.** The term *ballistic* will be used throughout to describe only those particles with negative energy ($h < 0$) that rise from the critical level in elliptic orbits and eventually fall back. Clearly in this case there will be symmetry in positive and negative values of $\xi$.

The orbits are restricted, then, by the conditions that $h < 0$ and that the orbits must intersect the exobase, designated by potential $\lambda_c$. From (7.1.22) and (7.1.26), remembering that $P_\chi$ is constant, we express $v$ as

$$v = h + \lambda - \xi^2 = \frac{\lambda^2}{\lambda_c^2} v_c = \frac{\lambda^2}{\lambda_c^2}(h + \lambda_c - \xi_c^2) \tag{7.1.28}$$

which yields

$$v = \frac{\lambda^2}{\lambda_c^2 - \lambda^2}(\lambda_c - \lambda - \xi_c^2 + \xi^2) \tag{7.1.29}$$

The condition that orbits intersect the exobase necessitates that $\xi_c$ be real. Hence the conservation of $P_\chi$ requires that

$$v \leq \frac{\lambda^2}{\lambda_c^2 - \lambda^2}(\lambda_c - \lambda + \xi^2) \equiv v_1(\xi) \tag{7.1.30}$$

The energy condition requires that

$$v < \lambda - \xi^2 \equiv v_2(\xi) \tag{7.1.31}$$

and for a given $\xi$ the maximum $v$ will be either $v_1$ or $v_2$, whichever is smaller. The condition that $v_1 < v_2$ imposes the requirement that

$$\xi^2 < \lambda(1 - \lambda/\lambda_c) \equiv \xi_1^2 \tag{7.1.32}$$

To review these conditions, we note that, with the energy limitation above (7.1.31) would be the only restriction between $v$ and $\xi$. However, there

exist *satellite* orbits, with $v < v_2$, which do not intersect the critical level; (7.1.30) and (7.1.32) exclude these orbits. We have, therefore, for ballistic orbits,

$$\zeta_{\text{bal}}(\lambda) = \frac{2}{\pi^{1/2}} \int_{\xi=0}^{\xi_1} \int_{v=0}^{v_1(\xi)} e^{-\xi^2 - v}\, dv\, d\xi + \frac{2}{\pi^{1/2}} \int_{\xi=\xi_1}^{\lambda^{1/2}} \int_{v=0}^{v_2(\xi)} e^{-\xi^2 - v}\, dv\, d\xi \tag{7.1.33}$$

After some manipulation, including an integration by parts and a change of variable, this becomes (Problem 7.2)

$$\begin{aligned} \zeta_{\text{bal}}(\lambda) &= \frac{2}{\pi^{1/2}} \int_0^{\lambda} \psi^{1/2} e^{-\psi} [1 - \mu_0(\psi)]\, d\psi \\ &= \frac{2}{\pi^{1/2}} \left( \gamma(\tfrac{3}{2}, \lambda) - \frac{(\lambda_c{}^2 - \lambda^2)^{1/2}}{\lambda_c} e^{-\psi_1} \gamma(\tfrac{3}{2}, \lambda - \psi_1) \right) \end{aligned} \tag{7.1.34}$$

where we have used

$$\begin{aligned} \mu_0(\psi) &= \frac{(\lambda_c{}^2 - \lambda^2)^{1/2}}{\lambda_c} \left( 1 - \frac{\psi_1}{\psi} \right)^{1/2} \\ &= \left[ 1 - \frac{\lambda^2}{\lambda_c{}^2} \left( 1 + \frac{\lambda_c - \lambda}{\psi} \right) \right]^{1/2}, \qquad \psi > \psi_1 \end{aligned} \tag{7.1.35}$$

and

$$\mu_0(\psi) = 0, \qquad \psi < \psi_1 \tag{7.1.36}$$

and where

$$\psi_1 = \frac{\lambda^2}{\lambda + \lambda_c} \tag{7.1.37}$$

In (7.1.34) the function $\gamma$ is the *incomplete* $\Gamma$ *function*,

$$\gamma(\alpha, x) \equiv \int_0^x y^{\alpha - 1} e^{-y}\, dy \tag{7.1.38}$$

For a more direct derivation of (7.1.34), let $\theta$ be the angle at which an orbit is inclined to the radius vector $\mathbf{r}$ (Fig. 7.1). The maximum angle that a ballistic orbit may possess for a specific energy $\psi$ and distance $r$ will be denoted $\theta_0(\psi)$. Ballistic orbits will thus fill a cone of half width $\theta_0$ centered on the radius vector; but outside this cone lie only satellite orbits, which do not touch the exobase. Within this cone there must be a uniform angular distribution, in accord with the orbits predicted from Liouville's equation, with particles moving in both directions along each orbit.

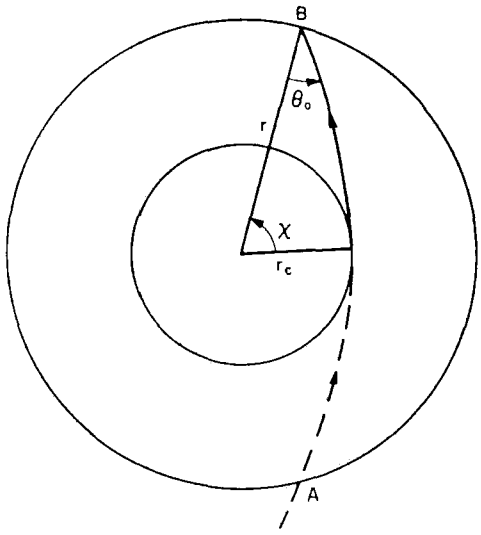



**Fig. 7.1** A hyperbolic orbit with velocity $v(r) > v_{esc}(r)$ enters sphere $r$ at $A$ and exits at $B$. The orbit is tangent to the exobase at height $r_c$ and crosses the sphere $r$ at an angle $\theta_0$.

The partition function may be integrated directly with the distribution function $f$ rewritten in polar velocity coordinates. We have

$$\zeta_{bal}(\lambda) = \frac{2}{\pi^{1/2}} \int_0^{\lambda} \int_{\theta=0}^{\theta_0(\psi)} \psi^{1/2} e^{-\psi} \sin\theta \, d\theta \, d\psi \tag{7.1.39}$$

Writing $\mu_0(\psi) = \cos\theta_0(\psi)$, this integral becomes equivalent to (7.1.34). It may be verified from the conservation of angular momentum that (7.1.35) is $\cos\theta_0(\psi)$ (cf. Problem 7.3).

If the maximum angle $\theta_0(\psi)$ is less than $\pi/2$, it designates an orbit that is tangent to the exobase (cf. Problem 7.3). A larger angle would correspond to a satellite orbit. When ballistic particles have $\theta_0(\psi) = \pi/2$, as they do for small energies $\psi$, the ballistic orbits for that and all smaller $\psi$ are isotropic [cf. (7.1.36) and (7.1.37)].

There can be no satellite orbits with energies $\psi < \psi_1$. The satellite orbit with minimum energy $\psi_1(\lambda)$ has an angle $\theta = \pi/2$ at both $r$ and $r_c$; its apogee is at $r$ and its perigee, on the opposite side of the planet, is at $r_c$.

**Satellite Orbits.** These orbits, like ballistic ones, must have negative energies, but they do not intersect the critical level. While the full range of ballistic orbits is expected, the existence of satellite orbits depends on the balance of the rare collisions that do occur within the corona with the rare destructive processes, such as photoionization and the perturbation of orbits by radiation pressure.

The production and destruction of these particles in an atmosphere is discussed in Section 7.2.1; here we will suppose that there are no loss mechanisms other than by collisions. Then even exceedingly rare coronal collisions will eventually establish an isotropic distribution for all particles that are energetically allowed, according to Boltzmann's "$H$ theorem."

Hence for all captive (ballistic plus satellite) orbits taken together, the integration limits to (7.1.27) are given by $0 < v < \lambda - \xi^2$ and $0 < \xi^2 < \lambda$. These limits lead to a form analogous to (7.1.39),

$$\zeta_{\text{bal}}(\lambda) + \zeta_{\text{sat}}(\lambda) = \frac{2}{\pi^{1/2}} \int_0^{\lambda} \psi^{1/2} e^{-\psi}\, d\psi$$

$$= \frac{2}{\pi^{1/2}} \gamma(\tfrac{3}{2}, \lambda) \tag{7.1.40}$$

This equation with (7.1.34) yields, for the maximum or full distribution of satellite orbits,

$$\zeta_{\text{sat}}(\lambda) = \frac{2}{\pi^{1/2}} \int_{\psi=\psi_1}^{\lambda} \psi^{1/2} \mu_0(\psi) e^{-\psi}\, d\psi$$

$$= \frac{2}{\pi^{1/2}} \frac{(\lambda_c{}^2 - \lambda^2)^{1/2}}{\lambda_c} e^{-\psi_1} \gamma(\tfrac{3}{2}, \lambda - \psi_1) \tag{7.1.41}$$

where $\psi_1$ is given by (7.1.37)

**Escaping Orbits.** The selection of integration limits is here similar to the procedure for ballistic orbits. There are two conditions to be satisfied: the energy is positive ($h \geq 0$) and the orbits must arise from the critical level. Note that for escaping orbits it is not sufficient to say that $\xi_c$ muct be real; it must now be positive, as particles falling to the planet from infinity must be excluded from the integration. The condition that the orbit intersects the critical level requires that $v \leq v_1(\xi)$, as with (7.1.30). The energy requirement gives $v \geq v_2(\xi)$, in analogy with (7.1.31). However, $v_2$ may turn out to be negative and in that case we merely have $v \geq 0$. This latter limit on $v$ is used, then, when $\xi > +\lambda^{1/2}$ [cf. (7.1.31)].

A further caution to be observed is that we clearly cannot have $v_2 \leq v \leq v_1$ unless $v_2 \leq v_1$. This condition necessitates, in analogy with (7.1.32), that $\xi > \xi_1$. Hence, we obtain

$$\zeta_{\text{esc}}(\lambda) = \frac{1}{\pi^{1/2}} \int_{\xi=\xi_1}^{\lambda^{1/2}} \int_{v=v_2}^{v_1} e^{-\xi^2 - v}\, dv\, d\xi$$

$$+ \frac{1}{\pi^{1/2}} \int_{\xi=\lambda^{1/2}}^{\infty} \int_{v=0}^{v_1} e^{-\xi^2 - v}\, dv\, d\xi \tag{7.1.42}$$

This expression may be transformed to

$$\zeta_{\text{esc}}(\lambda) = \frac{1}{\pi^{1/2}} \int_{\lambda}^{\infty} \psi^{1/2} e^{-\psi} [1 - \mu_0(\psi)]\, d\psi$$

$$= \frac{1}{\pi^{1/2}} \left\{ \Gamma(\tfrac{3}{2}) - \gamma(\tfrac{3}{2}, \lambda) - \frac{(\lambda_c^2 - \lambda^2)^{1/2}}{\lambda_c} e^{-\nu_1} [\Gamma(\tfrac{3}{2}) - \gamma(\tfrac{3}{2}, \lambda - \psi_1)] \right\} \tag{7.1.43}$$

where $\mu_0(\psi)$ is given by (7.1.35) and $\psi_1$ by (7.1.37); here $\Gamma(\alpha) \equiv \gamma(\alpha, \infty)$ is the *complete* $\Gamma$ *function.*

**Summary of Densities.** The total partition function to be used in computing the density by (7.1.20) is

$$\zeta = \zeta_{\text{bal}} + \zeta_{\text{sat}} + \zeta_{\text{esc}} \tag{7.1.44}$$

as given by (7.1.34), (7.1.41), and (7.1.43). These components of $\zeta$ have been tabulated versus $\lambda_c$ and $\lambda$.

**Flux of Escaping Particles.** At any height above the critical level the net number of particles moving outward across a square centimeter perpendicular to the radius vector may be expressed symbolically as

$$\mathscr{F}_{\text{esc}} = \int \frac{p_r}{M} f(q_i, p_i)\, d^3 p_i \tag{7.1.45}$$

where the integration is again over all allowed momentum coordinates. Writing $f$ in terms of polar velocity coordinates in the manner of (7.1.39), we find

$$\mathscr{F}_{\text{esc}} = N_c e^{-(\lambda_c - \lambda)} \frac{U}{2\pi^{1/2}} \int_{\lambda}^{\infty} \psi e^{-\psi} [1 - \mu_0^2(\psi)]\, d\psi \tag{7.1.46}$$

where $U = (2kT_c/M)^{1/2}$.

With $\mu_0(\psi)$ given by (7.1.35) the flux may be evaluated exactly and is

$$\mathscr{F}_{\text{esc}}(\lambda) = \frac{N_c e^{-\lambda_c} U}{2\pi^{1/2}} \frac{(\lambda_c + 1)\lambda^2}{\lambda_c^2} \tag{7.1.47}$$

This flux decreases as $r^{-2}$, as it must to satisfy continuity, and is the Jeans formula (7.1.5) at $\lambda = \lambda_c$.

### 7.1.4 Integrated Densities

The *integrated density* is the total number of atoms per cm$^2$ in a column of unit cross-sectional area above a specified height $r_1$ and along a specified

direction $\theta_1(r_1) = \cos^{-1}\mu_1(\lambda_1)$. We designate it by

$$\mathcal{N}(\mu_1, \lambda_1) \equiv \int_{r_1}^{\infty} N(r) \sec\theta(r, r_1)\, dr$$
$$= N_c e^{-\lambda_c} \lambda_1 r_1 \int_0^{\lambda_1} \frac{e^{\lambda}\zeta(\lambda)}{\lambda^2 \mu(\lambda, \lambda_1)} d\lambda \tag{7.1.48}$$

where $N(r)$ has been expressed by (7.1.20) and where $\mu(\lambda, \lambda_1)$ is defined below by (7.1.49).

The importance of the integrated density is twofold. The probability of an atom escaping from the atmosphere is governed by the total number of atoms lying above it; hence the integrated density fixes the exobase. Moreover, optical observations of coronas may yield the integrated density directly. Integrated densities depend on the density of the observed species $N_c$ and the temperature $T_c$ at the exobase and on the contribution provided by satellite orbits. In order to extract this information from the observations, we must calculate the integrated densities for various conditions.

Rocket or satellite photometers might observe $\mathcal{N}$ in any direction above the vehicle. But of special interest are observations made from outside the corona, where the line of sight passes through the corona along a chord that has its closest approach at $r_1$. This line of sight thus traverses an integrated density of $2\mathcal{N}(\mu_1 = 0, \lambda_1)$.

Except for the special case of radial integration, $\theta$ is a function of both $r_1$ and $r$. From the geometry of Fig. 7.2 and the law of sines, we have

$$\mu(\lambda, \lambda_1) = \left(1 - \frac{\lambda^2}{{\lambda_1}^2}(1 - {\mu_1}^2)\right)^{1/2} \tag{7.1.49}$$

We may then write (7.1.48) as

$$\mathcal{N}(\mu_1, \lambda_1) = N_c e^{-(\lambda_c - \lambda_1)} \zeta(\lambda_1) \frac{r_1}{\lambda_1} K(\mu_1, \lambda_1)$$
$$= N(r_1) H_1 K(\mu_1, \lambda_1) \tag{7.1.50}$$

where

$$H_1 \equiv \frac{r_1}{\lambda_1} = \frac{kT_c {r_1}^2}{G\mathcal{M}M} = \frac{kT_c}{Mg(r_1)} \tag{7.1.51}$$

and

$$K(\mu_1, \lambda_1) = \int_0^{\lambda_1} \frac{{\lambda_1}^2 \zeta(\lambda) e^{-(\lambda_1 - \lambda)}\, d\lambda}{\lambda^2 \zeta(\lambda_1)[1 - (1 - {\mu_1}^2)\lambda^2/{\lambda_1}^2]^{1/2}} \tag{7.1.52}$$

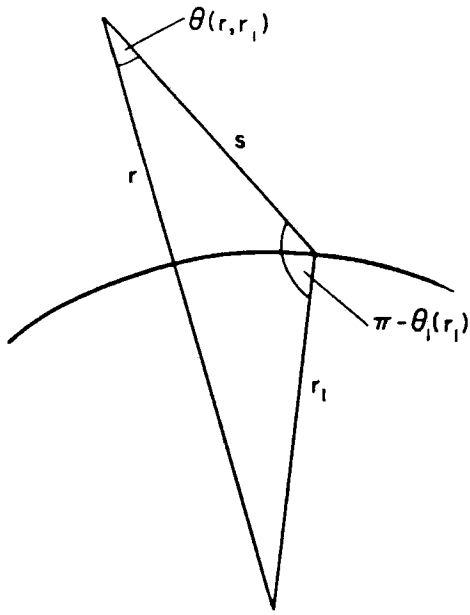



**Fig. 7.2** Geometry for integrated densities. The integration is along path $s$, extending from height $r_1$ to infinity.

In (7.1.51) the acceleration of gravity at height $r_1$ is $g(r_1) = G\mathcal{M}/r_1^2$, and $H_1$ is the conventional *scale height* at temperature $T_c$ used in plane-parallel models of atmospheric structure. When the variation of $g(r)$ is neglected above $r_1$, as is frequently done in plane-parallel models, the integrated density is $N(r_1)H_1\mu_1^{-1}$. In practice, the error involved in taking $g$ as a constant has a trivial effect on $\mathcal{N}$ only if the distance over which $g$ actually varies appreciably is very large compared with $H_1$—that is, if $r_1 \gg H_1$ or, equivalently, if $\lambda_1 \gg 1$.

In coronal theory, where we necessarily abandon the approximation of constant gravity, it is necessary to examine the situation more thoroughly, since the generalized barometric formula (7.1.9) predicts $N \to$ const as $\lambda \to 0$ and therefore $\mathcal{N} \to \infty$. We may consider, then, $\mu_1 K(\mu_1, \lambda_1)$ to be a correction factor for $\mathcal{N}(\mu_1, \lambda_1)$ to the simple, plane-parallel approximation with constant gravity. In the general case, the integral (7.1.52) is so complicated that a reasonably simple analytic expression does not exist, although tables exist of numerical calculations for various $\lambda_c$ and $\lambda_1$. For some purposes approximate expressions will suffice. First we evaluate $K$ for $\lambda_1 \gg 1$ and find to what extent $\mathcal{N}$ may depart appreciably from the usual approximation. Then we shall consider the distant regions where $\lambda_1 \ll 1$, as observations of this region will be of particular use in disclosing the contribution from satellite orbits.

**Integrated Density for Large $\lambda_1$.** When $\lambda_1 \gg 1$ the integral (7.1.52) is conveniently evaluated in two parts: the contribution from large $\lambda$ ($\lambda_2 < \lambda < \lambda_1$), which we designate $\Delta_1 K$, and the part at small $\lambda$ ($0 < \lambda < \lambda_2$), called $\Delta_2 K$. Here $\lambda_2$ is of order unity and designates the transition between large and small $\lambda$.

First consider the region of small $\lambda$. If there is a full population of particles in satellite orbits, the asymptotic density at small $\lambda$ is given by (see Problem 7.4)

$$\zeta \approx \frac{4}{3\pi^{1/2}} \lambda^{3/2} \tag{7.1.53}$$

Letting $\zeta(\lambda_1) = 1$, we have

$$\begin{aligned} \Delta_2 K(\mu_1, \lambda_1) &\leq \frac{4}{3\pi^{1/2}} \int_0^{\lambda_2} \frac{{\lambda_1}^2}{\lambda^{1/2}} e^{-(\lambda_1 - \lambda)} d\lambda \\ &= \frac{8}{3\pi^{1/2}} {\lambda_1}^2 \lambda_2^{1/2} e^{-\lambda_1} \end{aligned} \tag{7.1.54}$$

Since the approximate formula (7.1.53) actually overestimates $\zeta(\lambda)$ around $\lambda \approx 1$, it is clear that (7.1.54) gives an upper limit and that $\Delta_2 K \ll 1$. When satellite orbits are depleted, the contribution to $K$ is, of course, still smaller. This illustrates that, in contrast with the prediction of the barometric formula (7.1.19), the actual contribution to the integrated density at large distances is very small.

For the region $\lambda_2 < \lambda < \lambda_1$, we set $\zeta(\lambda) = 1$, corresponding to a full Maxwellian distribution. This is especially appropriate if the full component of satellite orbits is present and is likely to be a suitable approximation when $\lambda_1 \gg 1$, since most of the contribution to $\Delta_1 K$ arises from large $\lambda$. We thus have

$$\begin{aligned} K(\mu_1, \lambda_1) \approx \Delta_1 K(\mu_1, \lambda_1) &\approx \int_{\lambda_2}^{\lambda_1} \frac{{\lambda_1}^2 e^{-(\lambda_1 - \lambda)} d\lambda}{\lambda^2 [1 - (1 - {\mu_1}^2)\lambda^2 / {\lambda_1}^2]^{1/2}} \\ &= \int_0^{\lambda_1 - \lambda_2} \frac{e^{-x} dx}{(1 - x/\lambda_1)^2 [1 - (1 - {\mu_1}^2)(1 - x/\lambda_1)^2]^{1/2}} \end{aligned} \tag{7.1.55}$$

where $x = \lambda_1 - \lambda$.

We examine separately the two situations $\mu_1 = 1$ and $\mu_1 = 0$. In the first case we have

$$K(\mu_1 = 1, \lambda_1) = \int_0^{\lambda_1 - \lambda_2} \frac{e^{-x} dx}{(1 - x/\lambda_1)^2} \tag{7.1.56}$$

The integrand on the right side is unity at $x \equiv \lambda_1 - \lambda = 0$ (or $\lambda = \lambda_1$) and decreases to a minimum at $x = \lambda_1 - 2$ (or $\lambda = 2$). At still larger $x$ (or smaller

$\lambda$) the integrand increases and becomes infinite at $x = \lambda_1$ (i.e., $\lambda = 0$). But the latter region is excluded by the integration limit and is included more appropriately in (7.1.54). Expanding $(1 - x/\lambda_1)^{-2}$ by the binomial theorem, we may integrate term by term. These individual integrals may be expressed in closed form in terms of the *incomplete* $\Gamma$-*functions*, which can in turn be written in terms of the *truncated exponential series*. We find

$$K(\mu_1 = 1, \lambda_1) = \sum_{n=1}^{\infty} \frac{n!}{\lambda_1^{n-1}} [1 - e_{n-1}(\lambda_1 - \lambda_2) e^{-(\lambda_1 - \lambda_2)}] \tag{7.1.57}$$

where

$$e_n(\lambda_1 - \lambda_2) = 1 + (\lambda_1 - \lambda_2) + \frac{(\lambda_1 - \lambda_2)^2}{2!} + \cdots + \frac{(\lambda_1 - \lambda_2)^n}{n!} \tag{7.1.58}$$

When $\lambda_1 \gg 1$ the first few terms in the series (7.1.57) will decrease rapidly because of the factor $\lambda_1^{-(n-1)}$. At higher members of the series, $n!/\lambda_1^{n-1}$ will increase, but this occurs only after the exponential series $e_{n-1}(\lambda_1 - \lambda_2)$ begins to converge. The $n$th term in (7.1.57) will then itself be a series whose leading term is

$$e^{-(\lambda_1 - \lambda_2)} \frac{(\lambda_1 - \lambda_2)^n}{\lambda_1^{n-1}} \tag{7.1.59}$$

which is small for $\lambda_1 \gg 1$. (If $\lambda_2$ were allowed to be of order zero instead of unity, the convergence of the individual terms (7.1.59) as $n \to$ large would be very gradual, in accord with the rapid increase of the integrand in (7.1.57) as $x \to \lambda_1$.)

Hence for $\lambda_1 \gg 1$ and $\lambda_2 \sim 1$ we find the integral (7.1.57) yields

$$K(\mu_1 = 1, \lambda_1) = 1 + \frac{2}{\lambda_1} + \cdots \tag{7.1.60}$$

as the correction factor imposed by the variation of gravity in a spherical atmosphere. For $\lambda_1 = 10$, the plane-parallel model is 20 percent in error; for still smaller $\lambda_1$, the discrepancy is worse.

When $\mu_1 = 0$, (7.1.55) yields

$$K(\mu_1 = 0, \lambda_1) \approx \frac{\lambda_1^{1/2}}{2^{1/2}} \int_0^{\lambda_1 - \lambda_2} \frac{e^{-x}\, dx}{x^{1/2}(1 - x/\lambda_1)^2 (1 - x/2\lambda_1)^{1/2}} \tag{7.1.61}$$

Again expanding the denominator and integrating term by term, we have, to the same degree of approximation as (7.1.60) for $\lambda_1 \gg 1$,

$$K(\mu_1 = 0, \lambda_1) \approx \left(\frac{\lambda_1 \pi}{2}\right)^{1/2} \left(1 + \frac{9}{8\lambda_1} + \cdots\right) \tag{7.1.62}$$

From (7.1.50) the integrated density perpendicular to the radius vector is

$$\mathcal{N}(\mu_1 = 0, \lambda_1) \approx N(r_1)\left(\frac{\pi r_1 H_1}{2}\right)^{1/2}\left(1 + \frac{9}{8\lambda_1}\right) \tag{7.1.63}$$

The leading term is half the density along a complete tangential traversal through the atmosphere given in the constant-gravity approximation by (6.2.6). The correction term $\frac{9}{8}\lambda_1$ is a rather crude approximation and should not be compared literally with the correction term $\frac{3}{8}\lambda_1$ in the constant-gravity expression (6.2.6).

**Integrated Density for Small $\lambda_1$.** When $\lambda_1 \ll 1$ so that $\zeta(\lambda)$ is represented adequately by asymptotic expansions in powers of $\lambda$, the exponential in (7.1.52) is of order unity. Let us now consider a single component (ballistic, satellite, or escaping particles) and suppose that the density of that component varies as some power of $r$; that is, $\zeta_{\text{com}}(\lambda) = \text{const}\ \lambda^{n+2}$.

Equation (7.1.52) for a single component (or, more correctly, a single term in a power-law expansion of the density) is then

$$K_{\text{com}}(\mu_1, \lambda_1) = \lambda_1 \int_0^1 \frac{y^n\, dy}{[1 - (1 - \mu_1^2)y^2]^{1/2}} \tag{7.1.64}$$

Clearly we must have $n > -1$ in this discussion or else the integrated density would be infinite, violating our basic assumption that the total gravitational mass of the atmosphere may be neglected.

For an arbitrary value of $\mu_1$ the integral may be evaluated numerically, with the binomial expansion for the denominator. The convergence of terms in the integrated series is rather slow if $\mu_1 \sim 0$. However, the two special cases of most interest may be handled analytically. When $\mu_1 = 1$ (7.1.64) becomes simply $\lambda_1/(n+1)$, and the integrated density of the component is, from (7.1.50)

$$\begin{aligned}\mathcal{N}_{\text{com}}(\mu_1 = 1, \lambda_1) &= N_c e^{-\lambda_c}\zeta_{\text{com}}(\lambda_1)\frac{r_1}{n+1} \\ &= N_{\text{com}}(r_1)\frac{r_1}{n+1}\end{aligned} \tag{7.1.65}$$

As $\lambda_1 \to 0$, $n = +\frac{1}{2}, -\frac{1}{2}$, and 0 in the leading terms for ballistic, satellite, and escaping particles, respectively (see Problem 7.4). The integrated density in the radial direction at large distances $r_1$ is thus computed with an effective scale height $H_1 K(1, \lambda_1)$ the same order of magnitude as $r_1$.

For the transverse direction (7.1.64) gives

$$K_{\text{com}}(\mu_1 = 0, \lambda_1) = \frac{\lambda_1}{n+1}\frac{\Gamma[(n+3)/2]\Gamma(\frac{1}{2})}{\Gamma[(n+2)/2]} \tag{7.1.66}$$

where $\Gamma$ is the *complete* $\Gamma$-*function*. The integrated density in the transverse direction is then

$$\mathcal{N}_{\text{com}}(\mu_1 = 0, \lambda_1) = N_{\text{com}}(r_1)\frac{r_1}{n+1}\frac{\Gamma[(n+3)/2]\Gamma(\frac{1}{2})}{\Gamma[(n+2)/2]} \qquad (7.1.67)$$

The integrated density increases only moderately between the radial and transverse directions. The leading term in the ration $\mathcal{N}_{\text{com}}(\mu_1 = 0, \lambda_1)/\mathcal{N}_{\text{com}}(\mu_1 = 1, \lambda_1)$ is 1.79720 for ballistic orbits ($n = \frac{1}{2}$); 1.57080 ($=\pi/2$) for escaping orbits ($n = 0$); and 1.31102 for satellite orbits ($n = -\frac{1}{2}$).

### 7.1.5 Doppler Profiles of Spectral Lines

By the use of very sensitive detectors and high-resolution techniques, it is possible in principle to measure the line profile of radiation scattered by a planetary corona. With slightly less spectral resolution, one might measure the line width containing a specific fraction of the total intensity, that is, the integrated profile between two finite frequencies. Such measurements would be extremely useful for deriving the structure of the exosphere, especially if the intensities are measured in absolute units and at a variety of positions within the corona. In this section we derive expressions for those line profiles in terms of the basic parameters ($\lambda_c, \lambda_{cs}, U, N_c$) of the corona and the position and direction of observation.

The term *line profile* is used here in a rather special sense. Strictly speaking, we shall derive only the distribution of velocities in a specified direction in space, summing over all particles in the line of sight. This distribution is proportional to the spectral profile only when the corona is optically thin so that extinction and multiple scattering are negligible. Unfortunately, in practice these effects are likely to be troublesome. For example, outward observations from relatively near the planet might utilize the planet's shadow to obscure the atoms located below some height $r_1$. The setting or rising sun would then create an opportunity to collect data as a function of $r_1$. However, some of the incident sunlight passing through the corona is scattered before reaching the region to be observed ($r > r_1$), producing a complex situation of radiation transfer.

Without these complications, it is merely necessary to multiply the particle densities in a velocity range $dv_s$ by $g/4\pi$, where $g$ is the total number of photons scattered per atom per second, to obtain the profile in absolute radiation units per unit velocity interval (photons/cm$^2$ sec sr cm sec$^{-1}$). For the hydrogen geocorona, $g = 2.3 \times 10^{-3}$ photon/sec. For very accurate measurements it would be necessary to correct this theoretical intensity by as much as 17 per cent for the anisotropy of resonance scattering in Ly $\alpha$ (Section 6.3.1).

Consider a column of unit cross section extending above height $r_1$ at an angle $\theta_1(r_1)$ (see Fig. 7.2). This column contains an integrated density denoted in Eq (7.1.48) by $\mathcal{N}(\mu_1, \lambda_1)$. We shall let $I(p_s, \lambda_1)\, dp_s$ be the optically thin intensity of photons scattered by atoms with a momentum component parallel to the column between $p_s$ and $p_s + dp_s$, where s specifies the direction along the column. Then the spectrum is

$$I(p_s, \lambda_1) = \frac{g}{4\pi} \int_{r_1}^{\infty} \int_{p_\beta} \int_{p_\alpha} \frac{f(p_s, p_\alpha, p_\beta)\, dp_\alpha\, dp_\beta\, dr}{\mu(\lambda, \lambda_1)} \tag{7.1.68}$$

and the integrated density, expressed in terms of the spectrum, is

$$\mathcal{N}(\mu_1, \lambda_1) = \frac{4\pi}{g} \int_{-\infty}^{\infty} I(p_s, \lambda_1)\, dp_s \tag{7.1.69}$$

Here $p_\alpha$ and $p_\beta$ are momentum components in the plane perpendicular to s, $\mu$ specifies the direction of s at any level and is given by (7.1.49), and the density in momentum phase space is a modified form of (7.1.18):

$$f(p_s, p_\alpha, p_\beta) = \frac{N_c e^{-(\lambda_c - \lambda)} e^{-(p_s^2 + p_\alpha^2 + p_\beta^2)/2MkT_c}}{(2\pi MkT_c)^{3/2}} \tag{7.1.70}$$

The integration limits in (7.1.68) are subject to the dynamical restrictions on coronal particles and are quite complicated in the completely general case.

The profile in the radial direction provides the simplicity of azimuthal symmetry and this special case is much less involved. Thus with $r\lambda = r_1\lambda_1$ the radial profile is

$$I(p_r, \lambda_1) = \frac{g}{4\pi} \frac{N_c e^{-\lambda_c} r_1 \lambda_1 e^{-\xi^2}}{(2\pi MkT_c)^{1/2}} \mathcal{S}(\xi, \lambda_1) \tag{7.1.71}$$

where the *radial spectrum function* is

$$\mathcal{S}(\xi, \lambda_1) = \int_0^{\lambda_1} \int_v \frac{e^{\lambda} e^{-v}}{\lambda^2} dv\, d\lambda \tag{7.1.72}$$

If $\mathcal{S}$ were independent of $\xi$, then $I(p_r, \lambda_1)$ would have the shape of $\exp(-\xi^2)$, as does the common Doppler profile for a Maxwellian distribution, where the absorption cross section per atom is, in frequency units,

$$\alpha_\nu = \frac{\pi e^2}{mc} f \frac{c}{\nu_0 U \pi^{1/2}} \exp\left[ -\frac{c^2}{U^2} \left( \frac{\nu - \nu_0}{\nu_0} \right)^2 \right] \tag{7.1.73}$$

for an oscillator strength $f$.

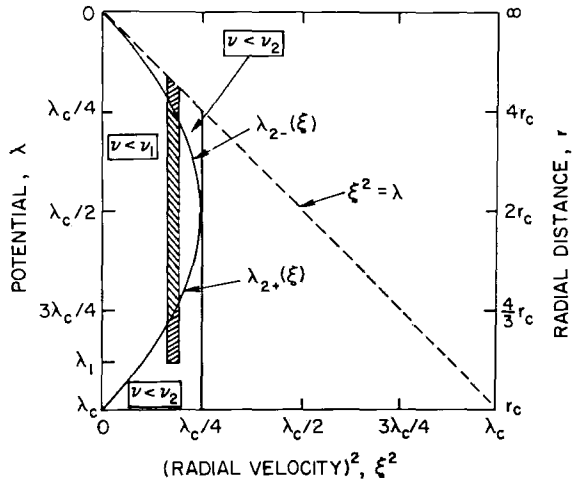



**Fig. 7.3** Integration limits in $\lambda$-space for a specified radial velocity $\xi$ and a run of transverse energies $\nu$. The diagram relates the limits on $\lambda$ to the corresponding limits on $\xi$ when the $\nu$ regime is specified. The shading shows an integration path in $\lambda$ space for a ballistic orbit with $\xi < \lambda_c/4$. The path covers three distinct regimes of $\nu$, such that $\xi^2 < \lambda < \lambda_{2-}$ (for $\nu < \nu_2$); $\lambda_{2-} < \lambda < \lambda_{2+}$ (for $\nu < \nu_1$); and $\lambda_{2+} < \lambda < \lambda_1$ (again for $\nu < \nu_2$).

The evaluation of (7.1.72) still requires assignment of limits to $\nu$ from the dynamical restrictions. The limits on $\nu$ are derived in a similar manner to that used in Section 7.1.3; however, there the second integration was over $\xi$ and here it is over $\lambda$. The relationship between the two sets of integration limits is illustrated in Fig. 7.3.

First consider ballistic orbits (with negative total energy). The limits on $\nu$ are given by (7.1.30)–(7.1.31), defining $\nu_1$ and $\nu_2$. We have $0 < \nu < \nu_1$ if condition (7.1.32) is fulfilled. Expressed as a condition on $\lambda$ instead of one on $\xi$, (7.1.32) becomes

$$\lambda \lessgtr \lambda_{2\pm} = \frac{\lambda_c}{2}\left[1 \pm \left(1 - \frac{4\xi^2}{\lambda_c}\right)^{1/2}\right] \tag{7.1.74}$$

We have to ascertain, however, which inequality is associated with the $\pm$ signs of the radical. As shown in Fig. 7.3 the integration $0 < \nu < \nu_1$ is valid when $\lambda_{2-} < \lambda < \lambda_{2+}$; this interpretation of (7.1.74) may be checked analytically by (7.1.74) and (7.1.32) (see Problem 7.6). These are the ballistic orbits whose principal restriction is that they intersect $r_c$.

In a similar fashion we find $0 < \nu < \nu_2$ when $\lambda < \lambda_{2-}$ and when $\lambda > \lambda_{2+}$. These ballistic orbits are restricted only by the condition of negative energy; no satellite orbits are possible with these $\nu$, $\lambda$ combinations because $\nu < \nu_1$ (the intersection condition) is obeyed as well. It is interesting that one such

region for any particular $\xi$, lies at low altitudes ($\lambda > \lambda_{2+}$) and the other at high ones ($\lambda < \lambda_{2-}$). In intermediate heights and small values of $\xi$, satellite orbits are allowed [compare (7.1.77)].

For $0 < \xi^2 < \lambda_c/4$ we have, therefore,

$$\mathscr{S}_{\text{bal}}(\xi, \lambda_1) = \int_{\xi^2}^{\lambda_{2-}} \frac{e^\lambda}{\lambda^2}(1 - e^{-\nu_2})\, d\lambda + \int_{\lambda_{2-}}^{\lambda_{2+}} \frac{e^\lambda}{\lambda^2}(1 - e^{-\nu_1})\, d\lambda$$

$$+ \int_{\lambda_{2+}}^{\lambda_1} \frac{e^\lambda}{\lambda^2}(1 - e^{-\nu_2})\, d\lambda \tag{7.1.75}$$

When $\xi^2 > \lambda_c/4$, the limits $\lambda_{2\pm}$ are imaginary and the three integrals in (7.1.75) coalesce into a single one. Hence for $\lambda_c/4 \leq \xi^2 < \lambda_1$ the spectrum is obtained from

$$\mathscr{S}_{\text{bal}}(\xi, \lambda_1) = \int_{\xi^2}^{\lambda_1} \frac{e^\lambda}{\lambda^2}(1 - e^{-\nu_2})\, d\lambda \tag{7.1.76}$$

For ballistic plus satellite orbits only the energy condition, $0 < v < v_2$ prevails. Hence for all $\xi^2 < \lambda_1$, the spectrum function for all captive particles is given by (7.1.76). Thus we write

$$\mathscr{S}_{\text{bal+sat}}(\xi, \lambda_1) = \int_{\xi^2}^{\lambda_1} \frac{e^\lambda}{\lambda^2}(1 - e^{-\nu_2})\, d\lambda \tag{7.1.77}$$

which may be evaluated with tabulated exponential integrals.

In practice, however, satellite orbits are never filled to capacity and one may estimate the actual contribution of ballistic plus satellite particles by the scheme used in Section 7.2.1. This involves defining the satellite critical level $\lambda_{cs}$ and using (7.1.75) with $\lambda_c$ everywhere replaced by $\lambda_{cs}$ to describe all particles of negative energy. Escaping particles (see below) are still described with $\lambda_c$.

In Section 7.1.3 we established that escaping particles in the range $v_2 < v < v_1$ are allowed when $\xi$ is in the range $[\lambda(1 - \lambda/\lambda_c)]^{1/2} < \xi < \lambda^{1/2}$. This condition means for a fixed $\xi$ that $\lambda > \xi^2$ and that either $\lambda > \lambda_{2+}$ or $\lambda < \lambda_{2-}$. The entire range $0 < v < v_1$ is allowed for particles with $\xi^2 > \lambda$.

We therefore obtain the spectrum function for escaping particles ($\xi > 0$) as

$$\mathscr{S}_{\text{esc}}(\xi, \lambda_1) = \int_0^{\xi^2} \frac{e^\lambda}{\lambda^2}(1 - e^{-\nu_1})\, d\lambda + \int_{\xi^2}^{\lambda_{2-}} \frac{e^\lambda}{\lambda^2}(e^{-\nu_2} - e^{-\nu_1})\, d\lambda$$

$$+ \int_{\lambda_{2+}}^{\lambda_1} \frac{e^\lambda}{\lambda^2}(e^{-\nu_2} - e^{-\nu_1})\, d\lambda \tag{7.1.78}$$

Figure 7.4 shows some profiles for radially outward observations above various $r_1$ with various assumed satellite critical levels (cf. Section 7.2.1).

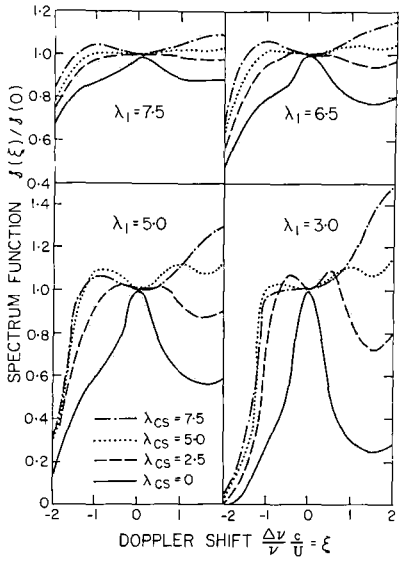



**Fig. 7.4** Spectrum functions for various shadow heights $r_1$ and satellite critical level $r_{cs}$ for zenith observations from below $r_1$. In all cases the base of the exosphere is specified by $\lambda_c = 7.5$ (e.g., appropriate for 500 km altitude on Earth with H atoms at 940°K). For $\lambda_{cs} = 7.5$ all satellite orbits are depleted; for $\lambda_{cs} = 0$ all possible satellite orbits for negative total energy are filled, so that particles at any height are isotropic for all velocities less than $\lambda^{1/2}$. The curves, when multiplied by $\exp(-\xi^2)$, give the Doppler line profiles. [After CHAMBERLAIN (1976).]

## 7.2 Collisions in Exospheres

No exosphere is strictly collisionless, of course. Here we examine (1) the effects of collisions on creating the satellite population, (2) the role of collisions in maintaining a near-Maxwellian distribution, and (3) the particular effects of charge-exchange collisions when ions as well as neutrals are present. Since these collisions can effectively introduce new unknowns into the problem, Doppler line profiles of a planetary corona (see Section 7.1.5) may be desirable to supplement the information available from the radial distribution of scattered radiation (i.e., the integrated exospheric density).

### 7.2.1 Equilibrium of Satellite Particles: Collisions and Radiation Pressure

In order to estimate the population of satellite particles, we introduce the concept of *satellite critical level* at $r_{cs}$; this level will assume a role for these particles analogous to that of the exobase $r_c$ for ballistic and escaping particles. Specifically, we suppose that below $r_{cs}$ there is, because of at least occasional collisions, a complete, isotropic distribution of the satellite particles that are energetically allowed ($h < 0$). Above $r_{cs}$ we suppose that the collisional creation of satellite orbits is negligible and that the only such orbits present are those that have perigees between $r_c$ and $r_{cs}$.

With a sharply defined satellite critical level, we may compute the density with the complete distribution of allowable satellite orbits up to $r_{cs}$; beyond, we may use the distribution for ballistic orbits alone as though these orbits arose from $r_{cs}$ rather than $r_c$. Thus (7.1.40) is appropriate for $\lambda > \lambda_{cs}$, and for $\lambda \leq \lambda_{cs}$ we may use (7.1.34) with $\lambda_c$ replaced by $\lambda_{cs}$. The ballistic or escaping components are not actually affected by this modification. It is only that the ballistic formula and the numerical tables based on it may, with this approach, be used to describe satellite plus ballistic particles.

Similarly, the integrated densities may be obtained from $K(\mu_1, \lambda_1)$ of (7.1.52) and published tables of that integral. Some caution is necessary if the point $r_1$ is below $r_{cs}$ (i.e., if $\lambda_1 > \lambda_{cs}$). The total bound ($h < 0$) population in the interval $\lambda_1 > \lambda > \lambda_{cs}$ is evaluated with the entire permitted population ($\zeta_{\text{bal}} + \zeta_{\text{sat}}$). For the interval $\lambda_{cs} > \lambda > 0$, the bound component is regarded as a ballistic component, $K_{\text{bal}}(\mu_{cs}, \lambda_{cs})$, where $\mu_{cs}$ is related to $\mu_1$, $\lambda_1$, and $\lambda_{cs}$ in the manner of (7.1.49), and where $\zeta_{\text{bal}}(\lambda)$ is found from (7.1.34) with $\lambda_c$ replaced with $\lambda_{cs}$.

The location of $\lambda_{cs}$ is specified by the condition that satellite particles are created, through collisions of ballistic particles with other particles, at the same rate that they are destroyed, principally through particle collisions, radiation pressure, photoionization, and collisions with solar-wind ions.

Let $f_{\text{sat}}(r, p)$ be the density in phase space of satellite particles with total momentum $p$, and let $f_{\text{bal}}(r, p)$ be the density of particles whose orbits intersect the critical level $r_c$. For a given $r$ and $p$ (or $\lambda$ and $\psi$), satellite particles occupy a solid angle $4\pi\mu_0$ and ballistic orbits fill the remaining portion of the sphere with solid angle $4\pi(1 - \mu_0)$, where $\mu_0$ is given by (7.1.35) and (7.1.36).

The rate of change of the population of satellite particles, with changes in total particle momentum ignored, is

$$4\pi\mu_0 \frac{d}{dt} f_{\text{sat}}(r, p) = N'Qv[4\pi(1 - \mu_0) f_{\text{bal}}(r, p)]\mu_0 - N'Qv[4\pi\mu_0 f_{\text{sat}}(r, p)](1 - \mu_0) - \frac{4\pi\mu_0}{\tau_{\text{loss}}} f_{\text{sat}}(r, p) \qquad (7.2.1)$$

Here $N'$ is the total number density of particles available for collisions, $Q$ the cross section for a collision, and $\tau_{\text{loss}}$ the lifetime for a $1/e$ probability of loss of an exospheric atom in the absence of neutral collisions. The first term on the right side is the rate of collisional injection of ballistic atoms (which fill a fraction $1 - \mu_0$ of a sphere) into a fraction $\mu_0$ of a sphere—that is, it represents the rate of creation of satellite orbits. Similarly, the second term is the rate of loss of satellite orbits by collisions that change satellite orbits into ballistic ones. The last term is the destruction rate of satellite particles through external forces.

The equilibrium ratio of densities in phase space is then

$$\frac{f_{\text{bal}}(r,p)}{f_{\text{sat}}(r,p)} = 1 + \frac{\tau_{\text{coll}}}{\tau_{\text{loss}}(1-\mu_0)} \tag{7.2.2}$$

where $\tau_{\text{coll}} = (N'Qv)^{-1}$ is the lifetime for a $1/e$ probability of a particle having a collision. The ratio depends on $p$ as well as $r$. But we may obtain an upper limit to $\lambda_{cs}$, the potential at which satellite orbits effectively cease to be created, by evaluating (7.2.2) for the maximum kinetic energy $\psi = \lambda$, that captive particles may possess. Particles with lower kinetic energy may enter satellite orbits at still smaller values of $\lambda$, so this procedure will yield the lowest height at which a significant number of bound orbits are absent.

Satellite orbits with $\psi = \lambda$ are filled to one-half capacity when $f_{\text{sat}}/f_{\text{bal}} = \frac{1}{2}$, which yields from (7.2.2)

$$N'(\lambda_{cs})\lambda_{cs}^{1/2}[1-\mu_0(\lambda_{cs})]QU\tau_{\text{loss}} = 1 \tag{7.2.3}$$

The most effective type of collision in a hydrogen corona is likely to involve charge transfer between H and $H^+$, since at low velocities the cross section may be quite large. Coulomb collisions continually work toward an isotropic distribution of proton velocities so that the distribution of proton density may approximate the barometric formula (7.1.19) (for a mean molecular weight of one-half that of hydrogen), remaining nearly constant at great distances, with the ionized gas confined by a planetary magnetic field.

For such charge-transfer collisions (7.2.1) must be replaced by an expression in which the total momentum of the neutral and ion are, in effect, interchanged. Writing $f'(r,p)$ for the density in phase space of the ions and $N$ for the total density of neutrals, we have

$$4\pi\mu_0 \frac{d}{dt} f_{\text{sat}}(r,p) = [4\pi\mu_0 f'(r,p)]QvN$$

$$- N'Qv[4\pi\mu_0 f_{\text{sat}}(r,p)] - \frac{4\pi\mu_0}{\tau_{\text{loss}}} f_{\text{sat}}(r,p) \tag{7.2.4}$$

So long as few of the neutral particles are escaping and the satellite orbits are filled almost to capacity ($f_{\text{sat}} \approx f_{\text{bal}}$) and so long as the ions and neutrals follow the same Maxwellian distribution for temperature $T_c$, we may write

$$\frac{f'}{(1-\mu_0)f_{\text{bal}} + \mu_0 f_{\text{sat}}} \approx \frac{N'}{N} \tag{7.2.5}$$

which states that the relative abundance of ions and neutrals is approximately independent of momentum. In this case (7.2.4) reduces to (7.2.1).

In practice it is difficult to evaluate (7.2.3) accurately because $N'$ and $\tau_{\text{loss}}$ vary with time and position and because $\lambda_{cs}$ varies inversely as their squares. The destruction of satellite particles occurs by photoionization with $\tau \sim$

$2.2 \times 10^6$ sec (= 25 days). However, charge exchange with fast solar-wind protons may be more efficient, especially at large distances where screening of the exosphere by the geomagnetic field is unimportant.

Radiation pressure from the scattering of solar Lyman $\alpha$ distorts the orbits in a way that usually lowers their perigees. The theory of orbital perturbations of elliptical orbits is rather complicated. There are two general methods for treating the problem. In dynamical astronomy it is common to work with the six orbital elements that completely specify a Keplerian orbit.

These elements are the semi-major axis $a$ and eccentricity $e$, which together specify the ellipse; $a$ also gives the period through Kepler's third law, $P = 2\pi a^{3/2}/(G\mathcal{M})^{1/2}$. The longitude of the ascending mode $\Omega$ measured from the vernal equinox is the point where the body crosses the plane of the ecliptic from south to north. The inclination $i$ to the plane of the ecliptic fixes with $\Omega$ the position of the orbital plane. The orientation of the ellipse in its plane is defined by $\omega$, the longitude of the perihelion point measured from the node. Finally, the time of perihelion passage $t_0$ fixes the body in its orbit. Having the perturbed elements as a function of time, one may compute altered positions and velocities if desired. Alternatively, one can consider the variations of the satellite coordinates directly.

The first approach, analysis of the changes in the orbital elements, is the more useful one when there is a steady perturbation acting over a long time. It is also mathematically elegant. One can relate four of the elements, $a$, $r$, $i$, and $\Omega$, to the total energy (per unit mass),

$$E = -\frac{G\mathcal{M}}{2a} \tag{7.2.6}$$

and the three components of the angular momentum (per unit mass),

$$\mathbf{H} = \mathbf{r} \times \frac{d\mathbf{r}}{dt} \tag{7.2.7}$$

where $\mathcal{M}$ is the mass of the planet. If we forget about the particle passing through the Earth's shadow, the sun's radiation pressure does not vary significantly over the orbit and the particle does not change its total energy or, by (7.2.6), the length of its semi-major axis. If the orbit is not circular its angular momentum does change. Imagine an elliptic orbit in a plane containing the Earth and the sun. The Earth is at the origin, perigee is on the $x$ axis, which is normal to the Earth–sun line, and the $y$ axis points away from the sun. Because the particle spends more time in the apogee half-plane than near perigee, there is a net torque on the particle about the Earth. There are two effects. The magnitude of the angular momentum ($r^2\, d\theta/dt$) is increased or decreased, depending on the direction of the particle motion in the orbit, and the position of perigee slips or *precesses*. The change in angular momentum $H$ causes a change in the eccentricity $e$; the precession of perigee

means that a torque that increases $e$ will eventually reverse and become a torque that decreases $e$. This effect is the reason why satellite hydrogen atoms cannot usefully be analyzed through their orbital elements. Radiation pressure produces an acceleration of the order of 1 percent that of gravitation at two Earth-radii. This is a large perturbation and leads to a rapid deterioration of the orbit. Consequently the balancing of angular momentum has no chance to occur and the orbits have to be examined individually.

With the geometry specified above, the vector equation of motion is

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{G\mathcal{M}\mathbf{r}}{r^3} + \mathbf{f}_y \tag{7.2.8}$$

where $\mathbf{f}_y$ is the acceleration due to radiation pressure,

$$f_y = hvg/M_H c \tag{7.2.9}$$

and $g$ is the mean number of Ly $\alpha$ photons scattered per second by an atom. With $g = 2.3 \times 10^{-3}$ photon/sec (Table 6.1), the acceleration is $f_y = 0.75$ cm/sec$^2$. (We neglect, for this illustration, the screening of sunlight by the Earth on the nightside and the anisotropy of Ly $\alpha$ scattering; cf. Section 6.3.1.) Equation (7.2.8) is equivalent to the two scalar equations

$$\begin{aligned} \frac{d}{dt}(r^2\dot{\theta}) &= f_y r\cos\theta \\ \frac{d^2r}{dt^2} - r\dot{\theta}^2 &= -\frac{G\mathcal{M}}{r^2} + f_y\sin\theta \end{aligned} \tag{7.2.10}$$

where superdots denote time derivatives. We will be concerned here with only the first or angular-momentum equation. Without the perturbing influence of radiation, the right side of that equation would be zero and the angular momentum a constant $H$, which is a statement of Kepler's law of areas. Taking that constancy as a first-order solution to (7.2.10), we obtain the integral over one revolution,

$$\begin{aligned} r^2\dot{\theta} &= H + \int_0^{2\pi} \frac{f_y r\cos\theta\, d\theta}{d\theta/dt} \\ &= H + \frac{f_y a^3(1-e^2)^3}{H}\int_0^{2\pi} \frac{\cos\theta\, d\theta}{(1+e\cos\theta)^3} \\ &= H - \frac{3\pi e f_y a^3(1-e^2)^{1/2}}{H} \end{aligned} \tag{7.2.11}$$

where we use the equation for a Keplerian ellipse,

$$r = \frac{a(1-e^2)}{1+e\cos\theta} \tag{7.2.12}$$

and regard the particle as moving directly ($\dot{\theta} > 0$).

The Newtonian solution to the two-body gravitational problem gives an angular momentum (the law of areas) of

$$H = [G\mathcal{M}a(1 - e_0{}^2)]^{1/2} \tag{7.2.13}$$

for an unperturbed eccentricity $e_0$, and the harmonic law, relating period $P$ to semi-major axis $a$,

$$P = \frac{2\pi}{\omega_0} = \frac{2\pi a^{3/2}}{(G\mathcal{M})^{1/2}} \tag{7.2.14}$$

where $\omega_0 = \langle\dot{\theta}\rangle$. The perigee distance is $r_p = a(1 - e)$ and the area of an ellipse is $\pi a^2(1 - e^2)^{1/2}$. Thus the difference in areas swept out by the radius vector in one revolution, with the perturbation and without it, can be written in terms of the change in $e$, which in turn gives the change in $r_p$. Averaged over a period (7.2.11) shows that the perigee drops at the rate

$$\left\langle \frac{dr_p}{dt} \right\rangle = -\frac{3}{2}\frac{f_y}{\omega_0}(1 - e^2)^{1/2} \tag{7.2.15}$$

This result can be readily generalized to the case where perigee is at an angle $\theta_0$ to the $x$ axis by replacing $f_y$ with $f_y \cos\theta_0$. Although this equation can give us a sense of the radiation–pressure perturbation, the orbits have to be integrated directly from (7.2.10) for actual situations. One such study indicates that H atoms with perigees at $2R_E$ will decay within times $\tau_{rad} \sim 5 \times 10^5$ sec, more or less independent of the apogee height. This lifetime is the same order as the orbital period for particles with apogees at $15R_E$ (Problem 7.5).

With this $\tau_{rad} = \tau_{loss}$ and ion densities of $N' \sim 3 \times 10^3$ cm$^{-3}$, and with $T_c = 10^{3\circ}$K, $U = 4 \times 10^5$ cm/sec, $\lambda_c = 7.04$, and $Q = 5 \times 10^{-15}$ cm$^2$, (7.2.3) yields $\lambda_{cs} = 2.1$ or $r_{cs}/r_c \approx 3.35$. At this height $\mu_0$ (for $\psi = \lambda_{cs}$) $= 0.84$. This example is little more than illustrative, however. Observations indicate a strong day–night difference with satellites much more depleted at several $R_E$ on the day side. Also the assumption above that the ion and neutral temperatures are equal is not correct, and their difference probably has profound effects on the exospheric structure (Section 7.2.3).

### 7.2.2 Departures from a Maxwellian Distribution near the Exobase

The collisionless theory invokes the Maxwellian velocity distribution at the exobase, except for particles moving downward with velocities exceeding the escape velocity $v_{esc}$. This assumption is equivalent to postulating partition

functions (7.1.21) of

$$\zeta(\lambda_c) = \begin{cases} 1, & \psi < \lambda, \quad -1 < \mu < 1 \\ 1, & \psi > \lambda, \quad 0 < \mu < 1 \\ 0, & \psi > \lambda, \quad -1 < \mu < 0 \end{cases} \tag{7.2.16}$$

With the occurrence of collisions, such sharp demarcations are clearly unrealistic. At energies $\psi \sim \lambda$, the sudden truncation in $\psi$ for $\mu < 0$ must be smoothed, and at $\mu \sim 0$, deflecting collisions will smooth the angular distribution for $\psi > \lambda$. The question is, how much will such smoothing decrease the Jeans escape flux?

A particle with $\psi > \lambda$ that is moving upward has a finite chance of having a collision in the exosphere that changes its direction or energy, or both, in ways that prevent its escape. In a complete Maxwellian distribution, these collisions that alter the momentum coordinates from state $p_j$, say, to state $p_i$ are exactly balanced by inverse collisions that alter momentum $p_i$ into $p_j$. This *detailed balancing* occurs because the Maxwellian distribution is a description of matter in thermodynamic equilibrium, wherein all processes are balanced. In the planetary geometry, compensating collisions do not generally occur, and we have to ascertain what is the equilibrium situation.

The analytic approach to this problem is through the *Boltzmann equation*, the extension of Liouville's equation (7.1.9) to include effects of collisions. With the equation of motion (7.1.12) for a particle in a gravitational field, the steady-state extension of (7.1.10) is

$$v_r \frac{\partial f(r, p_i)}{\partial r} + \left( \frac{P_\chi^2}{Mr^3} - \frac{G\mathcal{M}M}{r^2} \right) \frac{\partial f(r, p_i)}{\partial p_r} = \int [f^*(r, p_k) f(r, p_j) - f^*(r, p_k) f(r, p_i)] Q g \, d^3 p_k \tag{7.2.17}$$

Here $Q$ is the collisional cross section, $g$ is the relative scalar velocity of collision (which has the same magnitude before and after collision). We suppose that the escaping substance, with phase-space density $f$, is a minor constituent having most of its collisions with the major constituent of density $f^*$. The first term in the integral represents collisions that produce a particle with momentum $p_i$. For a given final momentum $p_i$ and initial momentum $p_k$, the conservation of energy and momentum fix the other initial momentum $p_j$ and the relative velocity $g$. The final term gives the loss rate of particles with momentum $p_i$.

It is possible to obtain a first-order solution to $f(r, p_i)$ when the departures from a Maxwellian are not great, and we shall deal with a related problem in the next section. The only way in which an accurate solution has been obtained for the flux at the exobase is by the *Monte Carlo* computing technique. Monte Carlo is a kind of numerical experiment involving a large

number of test particles whose paths are followed through the background gas (say, H atoms in a gas of O atoms) by the computer. At every collision a number of parameters—those describing the collision itself and the distance to the next collision—are specified by random numbers.

The "particles" may be initially released several mean free paths below the exobase to assure a reasonably complete Maxwellian, and followed until they either escape or return to the level of initial release. An alternative and more efficient method involves the consideration that, above the atmosphere, $\zeta = 0$ for $\psi \gtrsim \lambda_c$ and $-1 < \mu < 0$. Thus one can release these "virtual" or absent particles with a Maxwellian density $f(\psi > \lambda_c, \mu < 0)$, and see how they migrate through the gas and eventually escape. Since the real population is a Maxwellian less the virtual particles caused by the absent component from outer space, the number of virtual particles that escape gives the depletion of the real escape flux from the Jeans formula.

The loss of light particles from the top of an atmosphere also has an effect on the density distribution below the exobase. Instead of being in strict hydrostatic equilibrium, the upward flowing gas is in hydrodynamic diffusive equilibrium, and the density gradient is distorted from the barometric distribution accordingly.

Figure 7.5 shows the correction factor of $\mathscr{F}_{esc}/\mathscr{F}_{Jea}$ for a variety of temperatures for H and He in a background atmosphere of O. If the background gas is $CO_2$, the departures from the Jeans fluxes are slightly greater, with $\mathscr{F}_{esc}/\mathscr{F}_{Jea} \sim 0.5$.

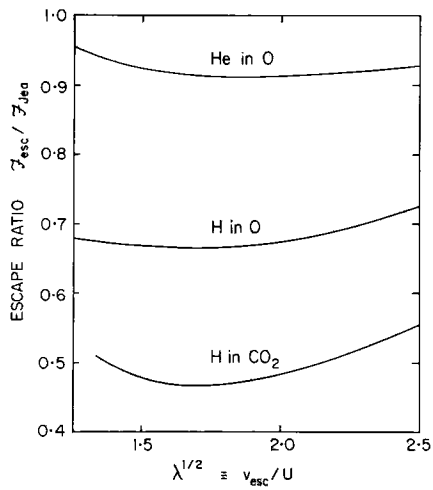



**Fig. 7.5** Ratio of actual escape flux, obtained from Monte-Carlo analysis, to the flux computed from the Jeans formula (7.1.5). The correction factor is close to unity when the escaping gas has a mass close to that of the background gas and decreases to about 0.5 for a ratio of 1/44 (H in $CO_2$). For H in the Earth's atmosphere ($T_c \approx 10^{3\circ}$K), $\lambda_c^{1/2} \approx 2.6$; for H in Mars' atmosphere ($T_c \approx 350^\circ$K), $\lambda_c^{1/2} \approx 2.0$. [Based on data of CHAMBERLAIN and SMITH (1971).]

### 7.2.3 Charge-Exchange Collisions in a Hot Plasmasphere

On a planet with a magnetic field such as Earth the direct escape of $H^+$ is prevented, except possibly near the magnetic poles where field lines tend to be open. However, if the ions have a temperature $T^* > T_c$, then charge-exchange collisions provide a source of fast H atoms. In the collision, the electron moves from a slow H atom to the fast ion forming a fast H atom.

It appears that such a mechanism can account for a discrepancy between the upward flux of total H in all chemical forms in the mesosphere and the Jeans flux of H inferred from exospheric observations of $T_c$ and the H concentration.

The basic Boltzmann equation is (7.2.17), where $f^*$ is the ion density and $f$ is the neutral density. To solve this equation we adopt a perturbation technique. Let

$$f(r, p_j) = f_0(r, p_j) + \Phi(r, p_j) f_0(r, p_j) \tag{7.2.18}$$

where $f_0$ is the known solution to Liouville's equation. The Liouville equation says simply that the density in phase space $f_0$ is constant along a dynamical trajectory. Hence, when the collision term is added, we integrate along a collisionless dynamical trajectory to find the gain and loss of density $\Phi f_0$.

For resonant charge exchange we can introduce a greatly simplifying approximation. The collision has a relatively large cross section at low velocities, which means that the collision normally occurs at large internuclear distances with little momentum transfer. Thus in (7.2.17) the only collision that produces an H atom with momentum $p_i$ is one that involves an incident $H^+$ with $p_k = p_i$. Similarly, the incident H atom can then take on any momentum, and we replace $p_j$ with $p_k$.

The cross section is only mildly velocity dependent. Theoretical cross sections have been confirmed by experiments down to 20 eV. The extrapolation to thermal energies ($\sim 0.1$ eV) is uncertain, but the theoretical $Q$ varies logarithmically with energy and increases only a factor of 2 over the extrapolation range. Within the thermal regime, $Q$ varies as $g^{-0.2}$ and may be regarded as being effectively constant at $Q_0 = 5 \times 10^{-15}$ cm$^2$.

A final simplification to (7.2.17) is that the velocities of collision, insofar as escape is concerned, are mainly ones close to the local escape velocity $v_{esc} = (2G\mathcal{M}/r)^{1/2}$. Little error should result from adopting $g = v_{esc}$.

With these approximations the Boltzmann equation appropriate to $H$—$H^+$ charge-transfer collisions is

$$\frac{p_r}{M}\frac{\partial f(r, p_i)}{\partial r} + \left(\frac{P_\chi^{\,2}}{Mr^3} - \frac{G\mathcal{M}M}{r^2}\right)\frac{\partial f(r, p_i)}{\partial p_r} = Q_0 v_{esc}(r) \int [f^*(r, p_i) f(r, p_k) - f^*(r, p_k) f(r, p_i)]\, d^3 p_k \tag{7.2.19}$$

To evaluate the collision term, we assume that the H population entering a collision is given by the Liouville equation (or Boltzmann equation without collisions). Then on the right side, $f(r, p_k)$ becomes

$$f_0(r, p_r, P_\chi) = \frac{N_c e^{-(\lambda_c - \lambda)} \exp(-p_r{}^2/2MkT_c) \exp(-P_\chi{}^2/2MkT_c r^2)}{(2\pi MkT_c)^{3/2}} \tag{7.2.20}$$

for all combinations of $p_r$ and $P_\chi$ allowed by the equations of motion.

We transform to the dimensionless notation of Section 7.1. For ions the temperature will be $T^* = \alpha T_c$. Since the electric attraction caused by the tendency for ions and electrons to separate gravitationally reduces the net downward force by a factor of 2 [for a single-component plasma, see (5.1.25)] the potential for $H^+$ is

$$\lambda^* = \frac{G\mathscr{M}M}{2\alpha kT_c r} = \frac{\lambda}{2\alpha} \tag{7.2.21}$$

With the ions following a Maxwellian distribution (out to some cut-off distance) the Boltzmann equation becomes

$$\begin{aligned} -\xi \frac{\partial f(\lambda, \xi, v)}{\partial \lambda} + \left(\frac{v}{\lambda} - \frac{1}{2}\right) \frac{\partial f(\lambda, \xi, v)}{\partial \xi} \\ = r_c \lambda_c Q_0 \lambda^{-3/2} N_c N_c{}^* \exp[-(\lambda_c - \lambda)(1 + 1/2\alpha)](2\pi MkT_c)^{-3/2} \\ \times \left[\frac{\zeta(\lambda)}{\alpha^{3/2}} \exp\left(-\frac{\xi^2 + v}{\alpha}\right) - \exp[-(\xi^2 + v)]\right] \end{aligned} \tag{7.2.22}$$

With (7.2.18) substituted on the left side, we have

$$-\xi \frac{\partial \Phi}{\partial v} + \left(\frac{v}{\lambda} - \frac{1}{2}\right) \frac{\partial \Phi}{\partial \xi} = \frac{K e^{\lambda/2\alpha}}{\lambda^{3/2}} \left\{\frac{\zeta(\lambda)}{\alpha^{3/2}} \exp\left[(\xi^2 + v)\left(\frac{\alpha - 1}{\alpha}\right)\right] - 1\right\} \tag{7.2.23}$$

where the *collision coefficient* is

$$K = r_c \lambda_c Q_0 N_c{}^* e^{-\lambda_c/2\alpha} \tag{7.2.24}$$

In writing (7.2.23) we have used the fact that the coefficient of a term in $\Phi$ on the left side vanishes identically through the Liouville equation, which governs $f_0$.

Equation (7.2.23) is to be integrated along dynamical trajectories for free ballistic flight. Along such a path, $\Phi$ then represents (in units of $f_0$) the accumulation (or loss) of particles due to collisions. The integration path is fixed by conservation of energy and angular momentum, or

$$\xi^2(\lambda) = (\xi_c{}^2 + v_c - \lambda_c) + \lambda - v \tag{7.2.25}$$

$$v/\lambda^2 = v_c/\lambda_c{}^2 \tag{7.2.26}$$

Thus (7.2.23) may be written as

$$\frac{d\Phi}{ds} = \frac{Ke^{\lambda/2\alpha}}{\lambda^{3/2}} \left\{ \frac{\zeta(\lambda)}{\alpha^{3/2}} \exp\left[ (\xi^2 + v)\left(\frac{\alpha - 1}{\alpha}\right)\right] - 1 \right\} \tag{7.2.27}$$

where the element of trajectory is

$$ds = -d\lambda \sec\theta \tag{7.2.28}$$

and

$$\tan\theta \equiv d\xi/d\lambda = (1 - 2v_c\lambda/\lambda_c{}^2)/2\xi \tag{7.2.29}$$

Integrating (7.2.27) along such a trajectory in $\lambda$, $\xi$ phase space for a specified $v_c = \text{const}$ yields a contribution to the density and flux at any height. Trajectories corresponding to positive energy terminate at $\lambda = 0$; those for negative energy turn over at $\xi = 0$ and return to the exobase.

An interesting particle-reflection or "splash" phenomenon arises when fast neutral H atoms are produced within the exosphere. We noted in Section 7.2.2 that one technique for treating the departures from a Maxwellian is to examine "virtual" incoming particles from space. Roughly 30 percent of such entrants are reflected back out according to the Monte Carlo calculations. Hence if we here follow trajectories from $\lambda = 0$ to the exobase for particles that exceed $v_{esc}$ (i.e., $\psi > \lambda$), we obtain a "splash" component of escape, which arises from collisions that produce particles initially with a downward velocity.

Some sample results of calculations of this sort are shown in Fig. 7.6.

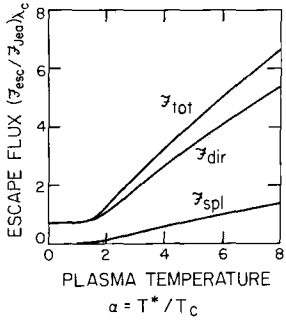



**Fig. 7.6** Escape fluxes relative to Jeans escape, for H—$H^+$ charge exchange in Earth's atmosphere as a function of the plasma temperature. Curves are computed for $\lambda_c = 7.5$ ($T_c = 940°K$), a plasmapause at $\lambda = 5.0$, and $K = 0.175 \exp(-\lambda_c/2\alpha)$. The curve for $\mathcal{F}_{dir}$ is the enhancement of the direct, upward moving flux while $\mathcal{F}_{spl}$ is the splash component, which has no Jeans analogue. [After CHAMBERLAIN (1977).]

## 7.3 Atmospheric Escape

### 7.3.1 Escape Mechanisms and the Limiting Flux

The loss of H and He from a planet may occur by several means. The Jeans "thermal evaporation" (as slightly modified in Section 7.2.2) gives a lower limit to the escape flux, but other processes may work as well. The

importance of the other mechanisms—and how they might vary over geologic periods—has not been appreciated until recent years. Consequently, much of the earlier (and some of the current) literature on atmospheric evolution should be read with caution.

On the Earth, where ions are magnetically confined, the charge-exchange mechanism (cf. Section 7.2.3) appears to be perhaps five times as important as Jeans escape for hydrogen. Another process for the Earth is direct escape of protons at high geomagnetic latitudes, where the geomagnetic field is stretched far "down stream" by the solar wind. The ions are drawn out these open magnetic field lines by the electric field of charge separation (Section 5.1.3). At high altitudes the $H^+$ ions are sufficiently accelerated by the $O^+$-$e$ field that they can escape rather than settle into diffusive equilibrium. This mechanism, called the *polar wind*, develops because $[H^+]$, when minor compared with $[O^+]$, increases with altitude.

A different kind of solar-wind interaction with atmospheric H is likely on Venus and Mars, which have been found to have virtually no planetary magnetic field (by Mariners 5 and 4, respectively). Being a plasma with an imbedded magnetic field, the solar wind can sweep up charges as it blows by an unmagnetized planet. The significance of this process depends in part on the efficiency with which atmospheric H atoms are ionized.

Still another ion-neutral reaction involves a reported auroral-type bombardment of the upper atmosphere by high-energy $O^+$ ions. (Usually the aurora is produced by $H^+$ ion and electron bombardment.) If this kind of particle bombardment is common, it could cause a loss of O atoms, which acquire energy from $O^+$-O collisions (in part by charge exchange) and take a cosmic trip.

Finally, highly exothermic reactions could produce H atoms with sufficient energy to escape provided the reactions can occur at sufficiently low density. Thus the sequence

$$\begin{aligned} X^+ + H_2 &\rightarrow XH^+ + H \\ XH^+ + e &\rightarrow X + H \end{aligned} \tag{7.3.1}$$

where $X^+$ is $O^+$ or $CO_2{}^+$, may produce H atoms energetic enough to escape.

Regardless of the manner of escape from the high atmosphere, the rate of escape is likely to be fixed in the lower portions of the atmosphere. So long as some mechanism is at work to remove atoms from the top of the atmosphere, their rate of removal in a steady state must be equivalent to the rate of supply from below. There is a common misconception that, if the exospheric temperature were to increase, the Jeans escape flux would increase dramatically because of its exponential dependence on temperature. For this to occur the atmosphere would have to be nearly in hydrostatic, diffusive equilibrium capable of quickly replenishing lost atoms. In fact, the hydrogen flux is transported mainly by mixing in the *homosphere*—that portion of

the atmosphere that is strongly mixed by vertical motions (frequently but not necessarily turbulent)—and mainly by molecular diffusion at higher levels. As this concept is important to atmospheric escape, we will develop it in detail.

The vertical flux from molecular diffusion and eddy diffusion (see Section 2.3) of a minor, light-weight constituent of density $N_1$ is

$$\Phi_1 = -DN_1\left(\frac{1}{N_1}\frac{\partial N_1}{\partial z} + \frac{M_1 g}{kT} + \frac{\alpha_T + 1}{T}\frac{\partial T}{\partial z}\right) - KN\frac{\partial(N_1/N)}{\partial z} \tag{7.3.2}$$

where $N$ is the "background" density of the atmosphere. At low densities only the molecular or $D$-term is important; at high densities $D$ is small and the eddy-diffusion or $K$-term dominates. The height where $D = K$ may be defined as the *homopause*.

We define the fractional abundance $f_1 = N_1/N$ and the actual density scale height $H^*$ by

$$\frac{1}{H^*} = -\frac{1}{N}\frac{\partial N}{\partial z} \tag{7.3.3a}$$

Then the reciprocal density scale height in equilibrium is (cf. Section 1.1, with the thermal-diffusion term incorporated)

$$\frac{1}{H_E{}^*} = \frac{Mg}{kT} + \frac{\alpha_T + 1}{T}\frac{\partial T}{\partial z} \tag{7.3.3b}$$

Thus we have

$$\Phi_1 = -b_1 f_1\left(\frac{1}{H_{1E}^*} - \frac{1}{H_1{}^*}\right) - KN\frac{\partial f_1}{\partial z} \tag{7.3.4}$$

where we have extracted the density dependence of the diffusion coefficient by $D = b_1/N$. Here $b_1$ is the *binary collision parameter*, which must be measured for each pair of colliding substances; for $H_2$ in air, $b_1 \sim 1.4 \times 10^{19}$ cm$^{-1}$ sec$^{-1}$.

The maximum rate of diffusion occurs for complete mixing, $\partial f_1/\partial z = 0$, since $\partial f_1/\partial z$ cannot be negative when $M_1 < M$. When a light-weight constituent of mass $M_1$ is completely mixed, then $H_1{}^* = H^* \equiv H_E{}^*$, and the diffusion is said to have its *limiting flow* rate of

$$\begin{aligned}\Phi_l &= b_1 f_1\left(\frac{1}{H_E{}^*} - \frac{1}{H_{1E}^*}\right) \\ &\approx \frac{b_1 f_1}{H}\left(1 - \frac{M_1}{M}\right) \approx \frac{b_1 f_1}{H}\end{aligned} \tag{7.3.5}$$

Then in general the flux from molecular and eddy diffusion combined is

$$\Phi_1 = \Phi_l + b_1 f_1 \left( \frac{1}{H_1{}^*} - \frac{1}{H^*} \right) - KN \frac{\partial f_1}{\partial z}$$

$$= \Phi_l - (D + K)N \frac{\partial f_1}{\partial z} \tag{7.3.6}$$

When the minor constituent (say, H) occurs in many forms ($H_2O$, $CH_4$, $H_2$), it is merely necessary to add the several equations of the form (7.3.5) or (7.3.6), weighted by the number of H atoms in a molecule to obtain the total H flux. Thus if $N_1$, $f_1$, $H_1$, and $\Phi_1$ pertain to H atoms in all forms and $N_a$ is the total density, the equations here are still valid, although $D$ and $b_1$ have to be regarded as weighted mean values for the various constituents.

To summarize, the maximum vertical transport occurs when a region is well mixed ($\partial f_1/\partial z = 0$) and conversely a well-mixed region fixes the limit on the rate of upward flow and hence the escape flux. In this case, no matter how hot the exosphere, the loss rate cannot be increased. The density $N_1$ has its steepest possible gradient and cannot be buoyed upward any faster.

At the other extreme, when the loss at the exobase is small, that region effectively controls the loss. In the mixing region ($K > D$), hydrogen may then not be well mixed because molecular diffusion will increase its scale height giving $H_1{}^* > H^*$. In the limit of no escape, there is a balance between mixing ($H_1{}^* = H^*$) and hydrostatic, diffusive equilibrium ($H_1{}^* = H_{1E}^*$). At the homopause this balancing yields $H_1{}^* = 2H^*$ (Problem 7.7).

With mechanisms at work in addition to Jeans escape, one cannot be certain the escape rate is known. If the escape rate at the top of the atmosphere (or in the case of helium the production rate in the lower atmosphere) is not known it is necessary to examine the upward flow by ascertaining mixing ratios at more than one height.

### 7.3.2 Earth's Loss of H and He

As an example, the Earth's atmosphere is well mixed to about 100 km. The mixing ratio of $H_2$ at 100 km is about $f_1 = 9 \times 10^{-6}$, not greatly different from the mixing ratio for H in all forms (mainly $H_2O$, $CH_4$, and $H_2$) in the stratosphere. Then (7.3.5) gives the limiting flux for H atoms in all forms as about $2 \times 10^8$ cm$^2$/sec.

Direct measurements of the exospheric temperature by satellite drag and the hydrogen abundance from the Lyman $\alpha$ emission allow the Jeans escape to be computed. Not only is the Jeans escape too low by a factor of 3 to 5, but its variations with exospheric temperature $T_c$ and density $N_c(H)$ do not give the constant flux expected. Jeans escape has an exponential relationship

with $T_c$, and hence an increase in $T_c$ should be accompanied by a drastic drop in $N_c(H)$. The actual decrease in $N_c(H)$ is not enough to compensate for the temperature excursion. However, charge-exchange loss varies as the integrated overhead density, $\mathcal{N}_c(H) \approx N_c(H)H_c$, and $N_c$ requires much less adjustment to keep the loss rate constant.

Helium loss is limited by the loss mechanism at the top of the atmosphere. For the case of $^4$He produced by $\alpha$-particle decay of radioactive elements, the crust exhales $2 \times 10^6$ atom/cm$^2$ sec. But (7.3.5) predicts a flux of $\Phi_l = 10^8$ cm$^{-2}$ sec$^{-1}$ for a mixing ratio $f = 5.24 \times 10^{-6}$. Moreover, He approximates the scale height of the atmosphere over most of the homosphere, although it does increase slightly at higher altitudes. However, it is estimated that at the exobase $N_c(^4\text{He}) \approx 3 \times 10^6$ cm$^{-3}$, and at $T_c = 10^{3\circ}$K, $\mathcal{F}_{\text{Jea}} \approx$ 2.4 atom/cm$^2$ sec. The discrepancy in fluxes is thus a factor of $10^6$.

In addition, $^3$He, thought to be produced by cosmic-ray bombardment in the atmosphere, also escapes too slowly by Jeans evaporation. If the temperature of the exobase were higher, a fit could be obtained for either $^3$He or $^4$He. For many years a satisfactory simultaneous solution to both isotopes was pursued as a means of ascertaining $T_c$. The problem is that no single $T_c$ will fit the data, and therefore the accuracy of the $^3$He data was often challenged.

The problem will not go away and it is unsolved. It appears now that He escape, like the loss of H, must occur through some process other than Jeans thermal evaporation, although much of the He loss may occur in the form of $He^+$ out the open magnetic field lines over the poles. Another possibility that holds promise is charge exchange between He or H and hot $He^+$ in the plasmasphere to produce energetic He atoms (Section 7.2.3).

### 7.3.3 Venus' Dual Exosphere and Escape

Mariner 5 presented an anomalous situation for Venus. The preliminary analysis of the observations indicated that a single exospheric temperature could not explain the data but that two components with scale heights in the ratio 2:1 or 3:1 could mimic the observations. One possibility seemed to be a large deuterium abundance on Venus, but a rocket spectrometer failed to show the predicted isotope splitting of Lyman $\alpha$.

Later analysis (Fig. 7.7) supports a two-component interpretation—either two masses or two temperatures. If the low-temperature ($T_c \approx 275^\circ$K) component represents the thermal corona, the hotter one ($T_c \approx 1020^\circ$K) may arise from a different mechanism, perhaps involving reactions such as (7.3.1) or plasma interactions not clearly understood. In such a dual exosphere, the thermal escape from the cooler component is trivial in magnitude. On the other hand, spectroscopic evidence for $H_2O$ and HCl near the cloud tops

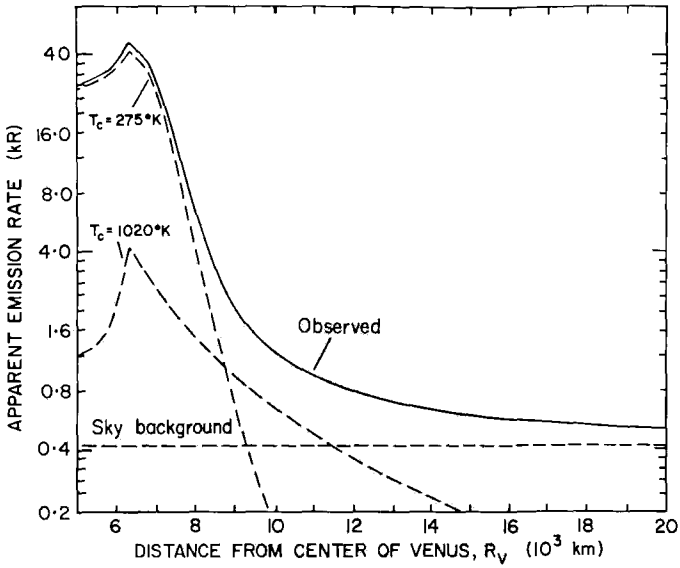



**Fig. 7.7** Dual exosphere of Venus. The Ly $\alpha$ observations of the bright limb by Mariner 5 can be reproduced by a model that includes (in addition to the sky background) two exosphere components with $T_c = 275$°K, $N_c = 2 \times 10^5$ cm$^{-3}$ and $T_c = 1020$°K, $N_c = 1.3 \times 10^3$ cm$^{-3}$ [After ANDERSON (1976).]

and photochemical theory for hydrogen compounds together indicate from (7.3.5) an escape flux about the same as for Earth, $\mathscr{F}_{esc} \approx \Phi_l \gtrsim 10^8$ cm$^{-2}$ sec$^{-1}$, provided of course that the atmosphere is well mixed from the clouds up to the homopause, which seems reasonable. There seem to be two ways to understand the situation.

First, if we take the limiting rate $\Phi_l$ at face value as an upward flowing current of hydrogen, the escaping atoms at the exosphere are evidently those we observe as the hotter component, implying a very effective nonthermal (i.e., chemical or ion-interaction) escape mechanism. The implication is that Venus loses H at about the same rate as (or faster than) Earth.

The second possibility is that the limiting-flow condition breaks down. In going from (7.3.4) to (7.3.5) we took the density scale height $H_1$* for the minor constituent equal to $H_E$*, the equilibrium value for the atmosphere. This assumption, valid for complete mixing, prevents (7.3.5) from vanishing. If we assume, however, that the escape flux is much less than $10^8$ cm$^{-2}$ sec$^{-1}$, we obtain a solution to (7.3.4) in which $\partial f_1/\partial z \neq 0$ (cf. Problem 7.7). The question concerning the validity of limiting flow is not whether $\partial f_1/\partial z$ is small but whether or not [see (7.3.6)]

$$(D + K)N \frac{\partial f_1}{\partial z} \ll \Phi_l \tag{7.3.7}$$

Other things being the same, a large $K$ reduces the upward flow of an escaping gas since it counteracts molecular diffusion. For Venus, if the escape flux is low (i.e., the bottleneck is the exobase), $K$ has to be quite large ($\gtrsim 10^8$ cm$^2$/sec) to keep the H density in the exosphere down at the observed values. With this option, the hot component presumably is still composed largely of nonthermal escapees, but the demands on the escape mechanism are reduced.

### 7.3.4 Mars' Hydrogen and Oxygen Escape

As with so many other aeronomic matters involving the two $CO_2$ atmospheres, the exosphere of Mars seemed (before the Viking landers) much easier to comprehend than did its counterpart on Venus. According to Mariner data the exobase has $T_c \approx 350°$K and there is no complication of a second exosphere. Quite possibly the bottleneck for H (and $H_2$) escape is the exobase itself. There we find $\lambda_c \approx 4.2$, and such a small value indicates rapid thermal evaporation. With $N_c(H) = 3 \times 10^4$ atom/cm$^3$, we have $\mathscr{F}_{Jea}(H) \approx 1.5 \times 10^8$ atom/cm$^2$ sec, comparable to Earth's loss rate. Nevertheless, if the limiting diffusive flow is estimated with (7.3.5), we obtain a similar value (as we would expect for a diffusion bottleneck), so it is really not clear where the limitation occurs.

In any case, if one supposes that escape varies as the density of hydrogen at the exobase, an amusing hypothesis can be developed, namely, that the thermal evaporation of hydrogen is twice the nonthermal loss of oxygen:

$$\mathscr{F}_{Jea}(H) + 2\mathscr{F}_{Jea}(H_2) = 2\mathscr{F}(O) \tag{7.3.8}$$

The nonthermal loss of O is presumed to occur by

$$O_2^+ + e \rightarrow O + O \tag{7.3.9}$$

with the O atoms having escape velocities. If the reaction rate of (7.3.9) were to change, the total abundance of $O_2$ would, after a time to achieve equilibrium, change in the opposite sense by about the same factor since the source of $O_2$ should remain constant. This source is

$$O + OH \rightarrow O_2 + H \tag{7.3.10}$$

and it remains constant if $H_2O$ (the source of OH) and $CO_2$ (the main source of O) remain constant.

Suppose the rate of (7.3.9) increases. The decreased $O_2$ in the lower atmosphere is accompanied by an $O_3$ decrease, thence an $O(^1D)$ decrease and an increase in the lifetime of $H_2$ against destruction by $O(^1D)$. Other reactions increase H, effectively taking it out of storage in $HO_2$ and OH. The overall effect is to increase the free H and $H_2$ throughout the atmosphere leading to an increase in escape flux of H + $H_2$ and Eq. (7.3.8).

The hypothesis is especially fascinating because it implies that the nonthermal escape of O controls the escape of H. The result is that $H_2O$ escapes from Mars!

The nonthermal escape of O would have a selective effect in escaping $^{16}O$ compared with $^{18}O$ (as would, of course, thermal evaporation). In view of the likelihood that large amounts of O have escaped over geologic periods, interesting questions arise regarding isotopic compositions of oxygen-containing materials on Mars, since the isotopic ratio would be related to their period of formation.

Both Viking landers obtained density distributions during descent that imply a completely unexpected temperature structure in the thermosphere. The profiles oscillate as much as 40°K over vertical distances of the order of 20 km, perhaps indicating a wave motion originating in the lower atmosphere. The average temperature between 120 and 200 km is under 200°K (compare Fig. 1.18).

## 7.4 Atmospheric Evolution and Climate

### 7.4.1 Development of Atmospheres of the Terrestrial Planets

The principal feature of the atmospheres of the terrestrial planets Venus, Earth, and Mars is that they are composed mainly of elements in the *oxygen group*—the second row of the periodic table of the elements—which account for only about 0.1 percent of the atoms of the sun or the universe. Jupiter and the other major planets, on the other hand, are mainly composed of H and He, with traces of C, N, and O that show up spectroscopically as $CH_4$ (methane), $NH_3$ (ammonia), and $H_2O$. Whether the H/He ratio coincides with the solar abundance is still uncertain (see Section 6.3.1), but the evidence does not rule out that possibility.

The Earth has retained water in its atmosphere because it has retained its oceans. Were the oceans to evaporate, $H_2O$ would become the most abundant constituent (about 300 times the present atmosphere), resembling the situation that may one day have existed on Venus (see below).

The atmospheres of the terrestrial planets must be secondary atmospheres. The primary atmospheres that may have existed after the time of planetary formation have long since disappeared; the present abundance ratios are very different from those of the sun. The composition of the crust–ocean–atmosphere system can be accounted for only by incorporating into the system the steady volcanic flow of volatiles (H, C, N, O, S, Cl) upward from the mantle. Except for the volatiles, the erosion of elements in crustal rocks (Si, Al, Fe, Ca) leads to a reasonably accurate composition of the oceans, sedimentary rocks, and the atmosphere.

The abundance of $O_2$ in the Earth's atmosphere is particularly remarkable, since the crust is suboxidized—quite the opposite of Mars. Photosynthesis of green plants is probably responsible in part for today's high abundance, but past photodissociation of $H_2O$ with subsequent escape of the H may also have been a factor.

The atmosphere of Saturn's major satellite Titan holds special interest since Titan is comparable to the Earth's moon in size and mass, but it was likely formed earlier from the same region of the protosun that formed the major planets. Spectra of Titan show weak absorptions corresponding to vibratation–rotation lines from quadrupole transitions of $H_2$. If they do indeed indicate $H_2$, they present an interpretational dilemma that has received considerable attention. The problem is that $H_2$ will readily escape a planet with such a low surface gravity (11 percent of Earth's), even at exospheric temperatures as low as 100°K ($\lambda_c = 7.8$). Thus taken at face value, the spectra indicate abundances of 5 atm-km of $H_2$ and 2 atm-km of $CH_4$ (methane). If there is no other gas present, the limiting diffusive flux (Section 7.3.1) gives an enormous escape flux of $10^{12}$ cm$^{-2}$ sec$^{-1}$. In spite of a number of imaginative proposals, no credible source has been identified for this amount of $H_2$ (1/300 of Titan's mass during the age of the solar system).

For comparison, the kind of source that would be expected is photodissociation of $CH_4$ and its photochemical offspring, $C_2H_2$ and $C_2H_6$. The yield would be only $10^{10}$ $H_2$ molecules cm$^{-2}$ sec$^{-1}$, and the mixing ratio of $[H_2]/[CH_4]$ would then be only 0.5 percent. More recent measurements have thrown doubt on the $H_2$ abundance, setting an upper limit of 1 atm-km.

Also, Mercury has an anomalous atmosphere that may indicate the rate of exhalation of elements from its crust and capture from the solar wind. Escape from Mercury should be rather rapid because of the small surface gravity and high temperatures on the sunlit side. However, there are difficulties in estimating the distribution of velocities of atoms that are thermally ejected from Mercury's surface.

The low abundance of water on Mars is evidently due to the low value of $\lambda_c$, as we have seen, which causes a rapid thermal escape of H and $H_2$. While O may also be escaping by the nonthermal mechanism (7.3.9), isotope analyses by Viking on Mars indicate, as do Venus $CO_2$ spectra, that the $^{18}O/^{16}O$ ratio is close to that on Earth. This ratio may place important constraints on the evolutionary history of Martian oxygen and other volatiles. The lack of enrichment of $^{18}O$ may imply, for example, interchange of atmospheric and subsurface oxygen on a time scale of $10^9$ years. Geochemical analyses of measurements of isotopic ratios ($^{40}Ar/^{36}Ar$) and abundances of heavy gases indicate that Mars has not outgassed as completely as Earth. In addition, it seems that Mars, like the moon, may have had a small initial endowment of volatile elements, but just why is not clear.

Compared with the massive atmosphere of Venus, that of Earth seems anomalously thin and poor in $CO_2$. The resolution of this dilemma probably lies in the large amounts of liquid water on Earth, through which $CO_2$ has formed carbonates and thence sedimentary rocks. But why Venus should have lost so much water (compared with Earth) in its early stages is an interesting problem that has led to the concept of a *runaway greenhouse effect*, a positive feedback between increasing temperature and increasing opacity.

It is apparent that, so long as there is water on the surface, the upper part of a convecting atmosphere must be saturated. Convection acting alone would tend to supersaturate the entire troposphere. The only process that prevents the atmosphere from being saturated at all heights is the descent of relatively dry air. At the tropopause, where the temperature becomes essentially constant or reverses, there is negligible descent of air from above and the air arriving there from below will be saturated.

The crucial feature of the runaway greenhouse is that, above a certain threshold of solar flux (when the infrared-active substance is condensable), an atmosphere cannot exist in radiative equilibrium. The energy discrepancy is in the sense that more solar energy is absorbed than emitted, and hence the condensed vapor (i.e., the oceans) will evaporate. The atmosphere will then heat up until the vapor is no longer condensable and the atmosphere achieves radiative equilibrium. With the water totally mixed in the atmosphere, it is photodissociated and the H can flow upward and escape.

We will examine the basic feature of this scenario—namely, the departure from radiative equilibrium—with the gray-atmosphere two-stream approximation of Section 1.2.3. From (1.2.43) we have

$$B(\tau) = B(0)(1 + \tfrac{3}{2}\tau) = \frac{\sigma}{\pi} T^4(\tau) \tag{7.4.1}$$

where $B(0)$ is the Planckian radiation characteristic of the skin temperature $T_0 \equiv T(\tau = 0)$ (cf. Fig. 1.4). In radiative equilibrium the atmosphere radiates the absorbed solar flux, which is

$$B(\tau = \tfrac{2}{3}) = 2B(0) \tag{7.4.2}$$

For a mass absorption coefficient $\kappa$ (independent of frequency in the gray approximation), the optical thickness is

$$\tau(z) = \int_z^\infty \kappa_1 \rho_1 \, dz = \frac{M_1 \kappa_1}{Mg} p_1(z) \tag{7.4.3}$$

where subscripts 1 refer to the absorbing-radiating substance and where we have used hydrostatic equilibrium (1.1.1).

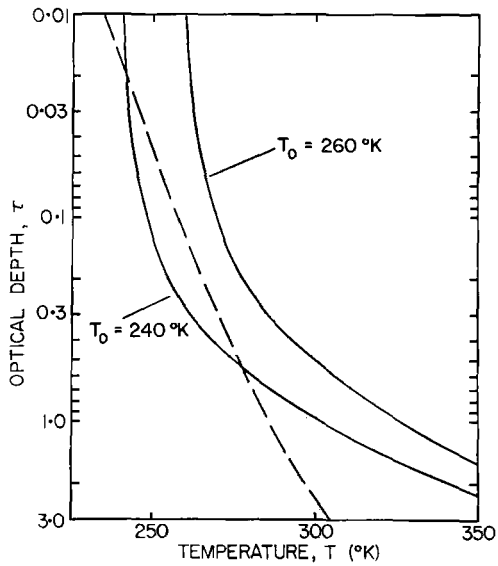



**Fig. 7.8** Model atmospheres in radiative equilibrium at two skin temperatures, $T_0$ (solid lines), and the saturation condition (7.4.4). The higher temperature model cannot exist in an equilibrium situation where there is liquid water on the surface, since there is no way for the atmosphere to convert vapor back to liquid.

If liquid exists in large quantities on the surface, then the partial pressure of the vapor at the top of the convective zone $z_c$ will be approximately the saturation vapor pressure, or $p_1(z_c) \approx p_s(T_c)$. Thus the condition imposed by saturation is

$$p_s(T_c) = \frac{Mg}{M_1 \kappa_1} \tau_c \equiv p_0 \tau_c \tag{7.4.4}$$

which defines $p_0$. For the Earth we may take $\kappa_1 = 0.1$ cm$^2$/gm (appropriate for the 8–20 $\mu$m window), $M_1 = 18$, $M = 29$, $g = 980$ cm/sec$^2$, obtaining $p_0 \approx 1.5 \times 10^{-2}$ atm. Figure 7.8 shows from (7.4.1) $T(\tau)$ for radiative equilibrium for two values of $T_0$ and from (7.4.4) the $T_c(\tau_c)$ imposed by the saturation condition. We see that if $T_0$ were to exceed about 250°K, a gray atmosphere in radiative equilibrium could not be saturated. Alternatively, if the vapor in the upper troposphere were saturated, the atmosphere could not be in radiative equilibrium, because it would be too cool to re-radiate all the energy absorbed.

There is some ambiguity in this analysis in principle, but it is not serious in practice. One does not know in advance either the $T_c$ or $\tau_c$ of the top of

the convective region; in practice we can require without serious error that the $T(\tau)$ curves defined by (7.4.1) and (7.4.4) merely intersect somewhere.

### 7.4.2 Climate Variations

*Climate* means different things to different scientists, so let us be clear that we are here concerned merely with the globally averaged temperature and its systematic change over periods of at least tens of years but usually much longer.

Changes in climate may be attributed to quasi-periodic internal changes of energy (among atmosphere, oceans, ice caps) with long response times, or the changes in climate may be forced by external modulations. Of the external modulations the most obvious hypothesis is that the amount of sunlight reaching the surface of the Earth changes. This variability might come about for several reasons:

(a) The sun may be a variable star (it does have an eleven-year cycle in the extreme ultraviolet, at least). Only one or two percent variability would be significant.

(b) The sun occasionally passes through clouds of galactic dust that increase its radiation output.

(c) The Earth's spin and orbital geometry go through secular changes, mainly in the obliquity of the rotational axis, the precession of the equinoxes, and the orbital eccentricity. The first two of these factors mainly change the seasonal and latitudinal distribution of energy input.

(d) The volcanic dust content of the air may be variable.

Simple models of the growth of polar ice caps make the radiation balance of the Earth seem precarious. If the Earth's incident sunlight decreases by a few percent, the ice caps would begin to spread to lower latitudes. The additional ice cover would (other things remaining unchanged) significantly increase the Earth's albedo, causing a further spread of ice, and so forth. A runaway polar ice cap! The principal difficulty with these models concerns the degree to which the Earth's cloudiness would change.

A second class of mechanisms deals with compositional changes in the atmosphere. The runaway greenhouse on Venus, discussed above, is an example (an extreme one) of such a mechanism. Another, if not quite as dramatic at least more alarming in an Earth-chauvinistic sense, is the accelerating increase in $CO_2$ occurring on Earth because of man-created combustion. This increase is occurring at the alarming rate of 3.8 percent per decade and the trend appears much worse if the higher derivatives are

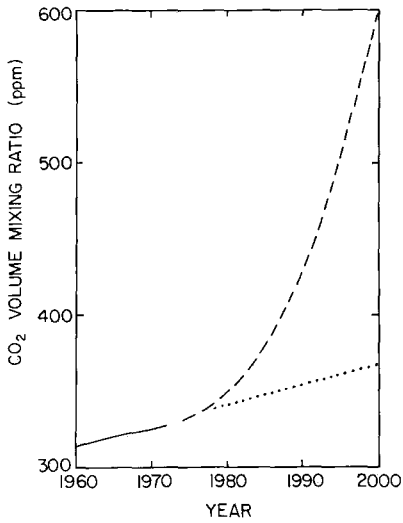



**Fig. 7.9** The growth of $CO_2$ mixing ratio (annual means in parts per million) over Mauna Loa, Hawaii. The solid line shows measured values. The dashed line gives an extrapolation with a power series fitted to the data with terms up to $t^3$, where $t = (\text{year} - 1958)$. The dotted line is a similar extrapolation but including only a linear term in $t$. [Based on data of KEELING *et al.* (1976).]

taken literally and the $CO_2$ extrapolated forward (Fig. 7.9). A $CO_2$ increase is likely to increase the greenhouse insulation of Earth.

Nevertheless the $CO_2$ released into the atmosphere since the beginning of the industrial revolution by the burning of fossil fuels is unaccounted for. Less than half of it is in the atmosphere. The excess $CO_2$ is not in the upper layers of the oceans and their deep circulation is commonly thought to be much too slow to provide the missing sink. Also, this excess $CO_2$ does not seem to be locked in the world's trees. Indeed, the harvesting of tropical forests and clearing of land for agricultural use may have led to a net transfer of $CO_2$ from the biota to the atmosphere, perhaps accounting in recent years for about 50 percent of the annual increase.

In addition to $CO_2$ a number of wastes are injected into the atmosphere through the burning of hydrocarbons. Most of the physical-chemical research on tropospheric pollutants has so far concentrated on identifying and measuring the secondary products and the complex photochemistry that is responsible for creating them from the primary automobile exhausts and power-plant discharges. Much additional work has been directed toward their toxicity on humans and their impact on ecosystems. It is too early to predict what will be the effects on climate, if any, of both the gases and aerosols, but there are indications of changes occurring in the average weather conditions many kilometers downwind from major urban centers.

The principal primary pollutants in urban areas are the hydrocarbons, the aldehydes (e.g., formaldehyde, $CH_2O$), nitrogen oxides, and carbon monoxide. The major oxidant in smog is ozone, and the question is, How is ozone formed? The central reactions start with $NO_2$, which can be dis-

sociated by sunlight reaching the ground. Thus we have

$$NO_2 + h\nu(\lambda < 3977\ \text{Å}) \rightarrow O + NO \tag{7.4.5}$$

followed by

$$O + O_2 + M \rightarrow O_3 + M \tag{7.4.6}$$

The ozone is then destroyed by

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{7.4.7}$$

completing the catalytic cycle. This sequence goes rapidly once O is formed and gives an equilibrium $O_3$ abundance of

$$[O_3] = J_5[NO_2]/k_7[NO] \tag{7.4.8}$$

where $J$ is a dissociation rate ($\text{sec}^{-1}$) and $k$ is a rate coefficient ($\text{cm}^3\,\text{sec}^{-1}$). However, (7.4.5)–(7.4.7) do not account for the observed diurnal variations of ozone and $NO_2$. There are other problems: reactive hydrocarbons are oxidized faster than they should be if the O and $O_3$ concentrations were those due only to the ozone cycle.

The resolution of both dilemmas appears to lie in the effect of $HO_x$ chemistry on the system by the cycle

$$OH + CO \rightarrow CO_2 + H$$
$$H + O_2 + M \rightarrow HO_2 + M$$

and

$$HO_2 + NO \rightarrow OH + NO_2 \tag{7.4.9}$$

giving a net

$$CO + NO + O_2 \rightarrow CO_2 + NO_2$$

Thus ozone follows the cycle of (7.4.5)–(7.4.7), but NO is converted to $NO_2$ by (7.4.9), in addition to (7.4.7).

Regarding pollution of the troposphere by industrial by-products, OH plays a kind of scavenger role whose cleansing effects may soon become overworked. The main OH reactions are

$$OH + CH_4 \rightarrow CH_3 + H_2O \tag{7.4.10}$$

and

$$OH + CO \rightarrow CO_2 + H \tag{7.4.11}$$

which are the principal sinks for $CH_4$ and CO, although oxidation of $CH_3$ provides the main natural source of CO. Probably OH also controls the concentrations of $H_2S$, $SO_2$, $CH_3Cl$, and other less soluble trace gases.

A major decrease in OH in the troposphere may have repercussions that are only beginning to become apparent. It can not only affect the natural disposal of technological products in the air, but also impact on the stratospheric chemistry. One danger is the fact that a major portion of the CO produced is technological and increasing. Another is that the stratosphere and troposphere may couple in both directions; a man-induced change in stratospheric $O_3$ could affect tropospheric OH (and hence $CH_4$), just as $CH_4$ produced in the troposphere eventually enters the stratosphere chemistry. The main source of tropospheric OH is photodissociation of tropospheric ozone,

$$O_3 + h\nu(\lambda < 4110\,\text{Å}) \rightarrow O(^1D) + O_2 \tag{7.4.12}$$

followed by

$$H_2O + O(^1D) \rightarrow 2OH \tag{7.4.13}$$

The possibilities for feedbacks of all this tropospheric chemistry on the thermal balance of the atmosphere are numerous and complex. What is certain is that the time constants are the order of decades, and pollution problems that were local a few years ago are rapidly becoming global.

Another external climate factor seems to be associated with changes in the magnetic field of the Earth. The statistical correlations between magnetic and climatic changes, as inferred from deep-sea sediment cores, indicate a cause–effect relation extending back over a million years. More recent and minor climate fluctuations have been associated with secular changes in sunspot activity, suggesting a controlling effect by the magnetic field in the solar wind (whose intensity closely follows solar activity). The problem here is to identify the physical mechanism. Some additional correlations of $^{14}C$ (which is produced by cosmic-ray impacts) both with climate and with geomagnetic or heliomagnetic fields suggest that galactic and solar cosmic rays reaching the Earth may be inducing the climatic changes. One possibility for a mechanism of cosmic-ray influence on climate is through ozone destruction by cosmic-ray produced $NO_x$ in the stratosphere (see Section 3.2.3).

The stratospheric ozone contributes directly to the tropospheric greenhouse effect by absorption in the 9.6 $\mu$m band. There could, however, be even more important interactions. Reduced ozone will lower the stratospheric heating, leading to increased condensation of stratospheric water vapor (which is close to the saturation pressure at the tropical tropopause; see Section 1.5.3). This ice could increase the Earth's albedo or it could simply reduce the $H_2O$ vapor content of the stratosphere and thereby reduce the greenhouse heating of the troposphere. Either way, the direction is the same: Any reduction of ozone promises to lower the mean temperature at the ground. Should the solar system move through a dense interstellar

cloud, its accretion by the sun would raise the solar luminosity and restrain the outflow of the solar wind, both of which could have climatological effects.

We can only touch on climate theory qualitatively here because even elementary quantitative projections necessarily involve the creation of elaborate computer models. The major difficulty in climate theory arises from the complex set of interactions and feedbacks, some of which defy quantitative modeling because the physics is inadequately understood. A simple listing of some of these interactions will make the point.

**Chemical Air-Sea Interactions.** An increase in $CO_2$ will raise the temperature of the air and surface layers of the oceans so that less $CO_2$ is soluble. The ocean thereby causes the atmospheric $CO_2$ to increase further, which constitutes a *positive feedback.*

**Heat Capacity of the Oceans.** The oceans not only absorb heat from the air, but transport it in latitude (e.g., by the Gulf Stream). The oceanic general circulation is poorly understood. Probably the main eddies are a few hundred kilometers in diameter, a decade smaller than their atmospheric analogues. Any changes in the air–sea interchange could conceivably involve changes in oceanic circulation.

**Cloudiness.** An increase of $H_2O$ in the atmosphere from increased evaporation brought on by higher temperatures might increase the cloudiness. The higher albedo of the Earth would then lower the temperature, providing a *negative feedback.* However, it may not work this way. Satellite measurements show that the northern hemisphere has a slightly higher albedo in winter than in summer.

**Polar Ice Caps.** A $CO_2$ warming (see above) would cause some recession of the polar ice. But if it also caused greater evaporation and more precipitation over the ice caps, the resultant growth of the ice mass might ultimately lead to fissures and slippage of the existing glaciers into the sea. The question is whether the net change in sea level would be up or down. Will water be transferred from the sea to the ice cap or *vice versa*? We do not know the answer because the feedbacks are complex. But if atmospheric $CO_2$ continues to increase at its present rate, it may be unwise to invest heavily in beach property.

## BIBLIOGRAPHICAL NOTES

### Section 7.1 Quasi-Collisionless Exospheres

The escape of light gases from the planets, the moon, and the sun was first discussed in a paper read to the Royal Society in 1846 by J. J. WATERSTON. The paper was mainly concerned with

matters of kinetic theory and first advanced the principle that the mean kinetic energy of each molecular species is proportional to the temperature. Thus the subject matter of this chapter is as old as the kinetic theory of gases, and older than Maxwell's law of the distribution of velocities. Waterston's manuscript was rejected, however, and was published eventually only because Lord Rayleigh, then secretary of the Royal Society, later found the manuscript in the Society's archives and recognized its historical importance. The original paper appears as

WATERSTON, J. J. (1892), On the physics of media that are composed of free and perfectly elastic molecules in a state of motion, *Phil. Trans. Roy. Soc. (London)* **A183**, 1–80.

The Jeans formula for escape by a Maxwellian velocity distribution over the upward hemisphere was developed in

JEANS, J. H. (1916), "The Dynamical Theory of Gases," Chapter 15, Cambridge University Press (4th ed., 1925).

Although out of date in some respects, a classic review of the escape problem is contained in

SPITZER, L., JR. (1949), The terrestrial atmosphere above 300 km, *in* "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), Univ. Chicago Press, Chicago, (2nd ed., 1952).

Hydrogen exospheres, their escape, and their observation (optically and otherwise) are treated in

TINSLEY, B. A. (1974), Hydrogen in the upper atmosphere, *Fundam. Cosmic Phys.* **1**, 201–300.

The mathematical analysis of the collisionless exosphere follows

CHAMBERLAIN, J. W. (1963), Planetary coronae and atmospheric evaporation, *Planet. Space Sci.* **11**, 901–960.

The latter paper gives tables of the partition functions $\zeta$ and $K$, which facilitate the calculation of densities and integrated densities.

Line profiles due to Doppler broadening by the non-Maxwellian velocity distributions of a collisionless exosphere have been computed by

CHAMBERLAIN, J. W. (1976), Spectral line profiles for a planetary corona, *J. Geophys. Res.* **81**, 1774–1776,

BERTAUX, J. L. (1978), Interpretation of OGO-5 line shape measurements of Lyman-alpha emissions from terrestrial exospheric hydrogen, *Planet. Space Sci.* **25**, in press.

Computer models of exospheres have made it possible to remove the restrictions imposed at the beginning of Section 7.1.2 in order to have an analytically simple problem. See, for example,

HARTLE, R. E. and MAYR, H. (1976), Wind-enhanced escape with application to terrestrial helium, *J. Geophys. Res.* **81**, 1207–1212,

HARTLE, R. E. (1973), Density and temperature distributions in nonuniform rotating planetary exospheres with applications to Earth, *Planet. Space Sci.* **21**, 2123–2138.

VIDAL-MADJAR, A.; CAZES, S.; and EMERICH, C. (1974), Influence of Earth's rotation and of possible perturbations on the exobase and exospheric hydrogen densities, *Planet. Space Sci.* **22**, 1375–1402.

**Section 7.2 Collisions in Exospheres**

The concept of a satellite critical level was introduced in

CHAMBERLAIN, J. W. (1963), op. cit.

Perturbation theory of Keplerian orbits subjected to a small second force (such as radiation pressure) is developed in textbooks on dynamical astronomy and in

SHAPIRO, I. I. (1962), The prediction of satellite orbits, *in* "Dynamics of Satellites," (M. Roy, ed.), pp. 257–312. Academic Press, New York.

A clear and simple presentation of the perturbations on the orbital elements by an external force is given in

BURNS, J. A. (1976), Elementary derivation of the perturbation equations of celestial mechanics, *Amer. J. Phys.* **44**, 944–949.

The importance of radiation pressure in removing H atoms in satellite orbits was demonstrated by

BERTAUX, J. L. and BLAMONT, J. E. (1973), Interpretation of Ogo 5 Lyman alpha measurements in the upper geocorona, *J. Geophys. Res.* **78**, 80–91.

The standard work on elastic-collision effects in neutral gases is

CHAPMAN, S. and COWLING, T. G. (1970), "The Mathematical Theory of Non-Uniform Gases," 3rd ed., Cambridge University Press, Cambridge.

Results of Monte Carlo calculations of the departures of thermal escape fluxes from the Jeans rate are given by

BRINKMANN, R. T. (1971), More comments on the validity of Jeans' escape rate, *Planet. Space Sci.* **19**, 791–794,

CHAMBERLAIN, J. W. and SMITH, G. R. (1971), Comments on the rate of evaporation of a non-Maxwellian atmosphere, *Planet. Space Sci.* **19**, 675–684.

The application of the Boltzmann equation to escape by charge-exchange collisions follows the treatment in

CHAMBERLAIN, J. W. (1977), Charge exchange in a planetary corona: Its effect on the distribution and escape of hydrogen, *J. Geophys. Res.* **82**, 1–9.

Observations of ion and electron temperature and concentrations have been applied to charge-exchange by

MAHER, L. J. and TINSLEY, B. A. (1977), Atomic hydrogen escape rate due to charge exchange with hot plasmaspheric ions, *J. Geophys. Res.* **82**, 689–695.

**Section 7.3.1 Escape Mechanisms and the Limiting Flux**

The suggestion that charge exchange with a hot plasma could be important to escape is due to

COLE, K. D. (1966), Theory of some quiet magnetospheric phenomena related to the geomagnetic tail, *Nature* **211**, 1385–1387.

The direct escape of ions from high geomagnetic latitudes was proposed by

NICOLET, M. (1961), Helium, an important constituent in the lower exosphere, *J. Geophys. Res.* **66**, 2263–2264.

This mechanism is the ion analog of Jeans escape, made possible by the down-wind opening of geomagnetic field lines. An additional mechanism, the *polar wind*, involves the existence of two or more positive ions and a resulting electrostatic acceleration, to supersonic speeds at high altitudes, once escape begins. See

BANKS, P. M. and KOCKARTS, G. (1973), "Aeronomy" Part B, Academic Press, New York, pp. 206–235,

for a summary of the physics and references to the original literature.

Solar wind sweeping of newly created charges from the upper region of an unmagnetized planet is treated by

MICHEL, F. C. (1971), Solar-wind induced mass loss from magnetic field-free planets, *Planet. Space Sci.* **19**, 1580–1583.

The upward splash of O atoms from the atmosphere as a result of $O^+$ auroral-type bombardment is a suggestion of

TORR, M. R.; WALKER, J. C. G.; and TORR, D. G. (1974), Escape of fast oxygen from the atmosphere during geomagnetic storms, *J. Geophys. Res.* **79**, 5267–5271.

Chemical reactions that lead to energetic atoms that may escape are listed by

SZE, N. D. and MCELROY, M. B. (1975), Some problems in Venus' aeronomy, *Planet. Space Sci.* **23**, 763–786.

The concept of a limiting flow for H in the thermosphere was implicit in the work on atmospheric diffusion being pursued actively around 1960; see

MANGE, P. (1961), Diffusion in the thermosphere, *Ann. Géophys.* **17**, 63–77 (reprinted as "Symposium d'Aéronomie," Intern. Assoc. Geomagn. Aeron. Symposium No. 1).

The concept, however, was clearly formulated and applied to a number of escape problems in

HUNTEN, D. M. (1973), The escape of light gases from planetary atmospheres, *J. Atmos. Sci.* **30**, 1481–1494.

**Section 7.3.2 Earth's Loss of H and He**

Direct determinations of Jeans escape of H from satellite data and the implied loss from other mechanisms are discussed by

BERTAUX, J. L. (1976), Observations of hydrogen in the upper atmosphere, *J. Atmos. Terr. Phys.* **38**, 821–827,

HUNTEN, D. M. and DONAHUE, T. M. (1976), Hydrogen loss from the terrestrial planets, *Ann. Rev. Earth Planet Sci.* **4**, 265–292.

BREIG, E. L.; HANSON, W. B.; and HOFFMAN, J. H. (1976), In situ measurements of hydrogen concentration and flux between 160 and 300 km in the thermosphere, *J. Geophys. Res.* **81**, 2677–2686.

The difficulties of accounting for He escape by the Jeans formula are treated in

MACDONALD, G. J. F. (1963), The escape of helium from the Earth's atmosphere, *Rev. Geophys.* **1**, 305–349,

and an update is contained in

HUNTEN, D. M. (1973), op. cit.

A more recent review, which also examines the possibility of $He^+$ charge transfer with neutral atoms to form fast He, is

CARLSON, R. W. (1978), The terrestrial helium problem: History and recent advances, *Planet. Space Sci.* **26**, in press.

**Section 7.3.3 Venus' Dual Exosphere and Escape**

Analysis of the Mariner 5 data in terms of one mass (H) and two temperatures (as well as one temperature and two masses, D or $H_2$) is due to

WALLACE, L. (1969), Analysis of the Lyman-Alpha observations of Venus made from Mariner 5, *J. Geophys. Res.* **74**, 115–131.

Of the two-mass solutions, deuterium was subsequently eliminated by a rocket-borne spectrometer flown by Wallace and his associates. And $H_2$ photodissociation (producing H with the $H_2$ scale height) is eliminated on grounds of continuity: The production rate of H would have to grossly exceed the rate at which it will diffuse to lower regions of the atmosphere and reassociate. The two-temperature model has been reanalyzed by

ANDERSON, D. E., JR. (1976), The Mariner 5 ultraviolet photometer experiment: Analysis of hydrogen Lyman alpha data, *J. Geophys. Res.* **81**, 1213–1216.

The analysis of Mariner 10 data is given in

BROADFOOT, A. L. (1976), Ultraviolet spectrometry of the inner solar system from Mariner 10, *Rev. Geophys. Space Phys.* **14**, 625–627.

Additional discussion and references are given in

HUNTEN, D. M. and DONAHUE, T. M. (1976), op. cit.

Plausible sources of nonthermal H atoms ejected from the ionosphere of Venus have been suggested by

KUMAR, S. and HUNTEN, D. M. (1974), Venus: An ionospheric model with an exospheric temperature of 350°K, *J. Geophys. Res.* **79**, 2529–2532,

SZE, N. D. and MCELROY, M. B. (1975) op. cit.

**Section 7.3.4 Mars' Hydrogen and Oxygen Escape**

A detailed analysis of data from three Mariners is given by

ANDERSON, D. E., JR. (1974), Mariner 6, 7, and 9 ultraviolet spectrometer experiment: Analysis of hydrogen Lyman alpha data, *J. Geophys. Res.* **79**, 1513–1518.

The Viking composition measurements are given by

NIER, A. O. and MCELROY, M. B. (1977), Composition and structure of Mars' upper atmosphere: Results from the neutral mass spectrometers on Viking 1 and 2, *J. Geophys Res.* **82**, 4341–4349.

The near equivalence of the escape losses of O and $H_2$ was first noted by

MCELROY, M. B. (1972), Mars: An evolving atmosphere, *Science* **175**, 443–445.

Further investigation by

LIU, S. C. and DONAHUE, T. M. (1976), The regulation of hydrogen and oxygen escape from Mars, *Icarus* **28**, 231–246,

demonstrated that the $HO_x$ photochemistry of the lower atmosphere keeps the relative abundances in the correct ratio for the oxygen and hydrogen escape to remain matched. The $^{18}O/^{16}O$ abundance ratio and its increase with time for nonthermal escape has been discussed by

MCELROY, M. B. and YUNG, Y. L. (1976), Oxygen isotopes in the Martian atmosphere: Implications for the evolution of volatiles, *Planet. Space Sci.* **24**, 1107–1113.

**Section 7.4.1 Development of Atmospheres of the Terrestrial Planets**

An older article, but still a comprehensive one, is

KUIPER, G. P. (1952), Planetary atmospheres and their origin, *in* "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.), pp. 306–405, 2nd ed., Univ. Chicago Press, Chicago.

Other major articles with emphasis on the Earth's atmosphere (and oceans) are

RUBEY, W. W. (1951), Geologic history of sea water, *Bull. Geol. Soc. Am.* **62**, 1111–1147.

BERKNER, L. V. and MARSHALL, L. C. (1965), On the origin and rise of oxygen concentration in the Earth's atmosphere, *J. Atmos. Sci.* **22**, 225–261,

BERKNER, L. V. and MARSHALL, L. C. (1966), Limitation on oxygen concentration in a primitive planetary atmosphere, *ibid.* **23**, 133–143.

BRINKMANN, R. T. (1969), The dissociation of water vapor and evolution of oxygen in the Earth's atmosphere, *J. Geophys. Res.* **74**, 5355–5368,

VAN VALEN, L. (1971), The history and stability of atmospheric oxygen, *Science* **171**, 439–443.

An analysis of the atmosphere of Saturn's massive satellite Titan is given in

HUNTEN, D. M. (1977), Titan's atmosphere and surface, *in* "Planetary Satellites," (J. A. Burns, ed.), pp. 420–437. University of Arizona Press, Tucson.

High resolution measurements of the $H_2$ absorption by Titan, with references to the earlier observations, are reported by

MÜNCH, G; TRAUGER, J. T.; and ROESLER, F. L. (1977), A search for the $H_2$ (3,0) S1 line in the spectrum of Titan, *Astrophys. J.* **216**, 963–966.

The problems associated with estimating the escape flux from the surface of Mercury or the moon are summarized in

SHEMANSKY, D. E. and BROADFOOT, A. L. (1977), Interaction of the surfaces of the moon and Mercury with their exospheric atmospheres, *Rev. Geophys. Space Phys.* **15**, 491–500.

For references to recent papers dealing with other planets see

HUNTEN, D. M. (1973), op. cit.
HUNTEN, D. M. and DONAHUE, T. M. (1976), op. cit.

The various factors contributing to the low abundance of atmospheric gases on Mars are reviewed and weighed in importance in

ANDERS, E. and OWEN, T. (1977), Mars and Earth: Origin and abundance of volatiles, *Science* **198**, 453–465.

The concept of the runaway greenhouse for Venus is due to

INGERSOLL, A. P. (1969), The runaway greenhouse: A history of water on Venus, *J. Atmos. Sci.* **26**, 1191–1198.

Additional discussion is given in

RASOOL, S. I. and DE BERGH, C. (1970), The runaway greenhouse and the accumulation of $CO_2$ in the Venus atmosphere, *Nature* **226**, 1037–1039.
POLLACK, J. B. (1971), A nongray calculation of the runaway greenhouse: Implications for Venus' past and present, *Icarus* **14**, 295–306.

The basic idea behind the runaway greenhouse—namely, the radiative-equilibrium temperature near the top of the Earth's troposphere cannot be so high that water vapor is unsaturated—was anticipated in the classic work of

SIMPSON, G. C. (1928), Further studies in terrestrial radiation, *Mem. Roy. Meteorol. Soc.* **3**, No. 21, 1–26.

**Section 7.4.2 Climate Variations**

General reviews on climate problems are

National Academy of Sciences (1975), "Understanding Climatic Change," U.S. Com. Global Atmos. Res. Pgm., N.R.C., N.A.S., Washington, D.C. 20418,
SCHNEIDER, S. H. and DICKINSON, R. E. (1974), Climate modelling, *Rev. Geophys. Space Phys.* **12**, 447–493.

The likelihood that certain periodicities in Earth's climate are due to orbital changes is developed in

HAYS, J. D.; IMBRIE, J.; and SHACKLETON, N. J. (1976), Variations in the Earth's orbit: Pacemaker of the ice ages, *Science* **194**, 1121–1132.

The energy budget of the Earth and climatological effects of a change in it are discussed by

BUDYKO, M. I. (1969), The effect of solar radiation variations on the climate of the Earth, *Tellus* **21**, 611–619,
SELLERS, W. D. (1969), A global climate model based on the energy balance of the Earth-Atmosphere system, *J. Appl. Meteorol.* **8**, 392–400,
SELLERS, W. D. (1973), A new global climate model, *J. Appl. Meteorol.* **12**, 241–254,
HELD, I. M. and SUAREZ, M. J. (1974), Simple albedo feedback models of the icecaps, *Tellus* **26**, 613–629,

NORTH, G. (1975), Analytical solution to a simple climate model with diffuse heat transport, *J. Atmos. Sci.* **32**, 1301–1307.

LINDZEN, R. S. and FARRELL, B. (1977), Some realistic modifications of simple climate models, *J. Atmos. Sci.* **34**, 1487–1501.

A survey of the effects resulting from the passage of the solar system through an interstellar cloud of gas and "dust" is given in

TALBOT, R. J., JR. and NEWMAN, M. J. (1977), Encounters between stars and dense interstellar clouds, *Astrophys. J. Suppl.* **34**, 295–308.

The $CO_2$ increases have been studied at the South Pole and at Mauna Loa; the results from the latter site are given in

KEELING, C. D. et al. (1976), Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii, *Tellus* **28**, 538–551.

The impact of changes in composition on temperatures of the lower atmosphere have been treated by radiative modeling in

MANABE; S. and WETHERALD, R. T. (1975), The effects of doubling the $CO_2$ concentration on the climate of a general circulation model, *J. Atmos. Sci.* **32**, 3–15,

WANG, W. C.; YUNG, Y. L.; LACIS, A. A.; MO, T.; and HANSEN, J. E. (1976), Greenhouse effects due to man-made perturbations of trace gases, *Science* **194**, 685–690.

An analysis of the future impact on climate of $CO_2$ released into the atmosphere by the burning of wood and fossil fuels has been published by the

National Academy of Sciences (1977), "Energy and Climate," N.A.S., Washington, D.C. 20418.

Also see the report of the 1976 Dahlem Workshop,

STUMM, W. (ed.) (1977), "Global Chemical Cycles and Their Alterations by Man," Abkon Verlagsgesellschaft, Berlin.

Analyses of the exchange of $CO_2$ between the atmosphere, oceans, and biomass are given in

WOODWELL, G. M.; WHITTAKER, R. H.; REINERS, W. A.; LIKENS, G. E.; DELWICHE, C. C.; and BOTKIN, D. B. (1978), The biota and the world carbon budget, *Science* **199**, 141–146,

STUIVER, M. (1978), Atmospheric carbon dioxide and carbon reservoir changes, *Science* **199**, 253–258.

Tropospheric pollution by automobile exhaust and industrial sources, its photochemistry and toxicity, and the associated ecological problems are thoroughly surveyed in

National Academy of Sciences (1977), "Ozone other photochemical oxidants," N.A.S., Washington, D.C. 20418.

National Academy of Sciences (1977), "Nitrogen Oxides," N.A.S., Washington D.C. 20418.

The wide-range importance of OH as a cleansing agent was noted by

LEVY, H. (1972), Photochemistry of the lower troposphere, *Planet. Space Sci.* **20**, 919–935,

and future implications have been examined in a preliminary way by

CHAMEIDES, W. L.; LIU, S. C.; and CICERONE, R. J. (1977), Possible variations in atmospheric methane, *J. Geophys. Res.* **82**, 1795–1798,

SZE, N. D. (1977), Anthropogenic CO emissions: Implications for the atmospheric CO—OH—$CH_4$ cycle, *Science* **195**, 673–675.

The possible relationship of sunspot activity with climate is developed from historical records by

EDDY, J. A. (1976), The Maunder minimum, *Science* **192**, 1189–1202,

and is developed from correlations of $^{14}C$ production with solar magnetic activity on the one hand and with climate on the other by

SUESS, H. E. (1968), Climate changes, solar activity, and the cosmic-ray production rate of natural radiocarbon, *Meteor. Monog.* **8**, 146–150.

Geomagnetic correlations with climate to 1.2 million years have been reported by

WOLLIN, G.; ERICSON, D. B.; and RYAN, W. B. F. (1971), Variations in magnetic intensity and climatic changes, *Nature* **232**, 549–551.

A speculative mechanism relating long-term magnetic activity to climate, through effects of cosmic rays on the stratosphere, was proposed by

CHAMBERLAIN, J. W. (1977), A mechanism for inducing climatic variations through the stratosphere: Screening of cosmic rays by solar and terrestrial magnetic fields, *J. Atmos. Sci.* **34**, 737–743.

## PROBLEMS

**7.1** *Jeans escape flux.* Write (7.1.3) with the velocity components expressed in spherical coordinates and derive (7.1.5).

**7.2** *Ballistic partition function.* Starting with (7.1.33), change the velocity components to $\mu$ and $\psi$ and derive (7.1.34).

**7.3** *Cone of acceptance for ballistic orbits.* From the conservation of angular momentum and energy, show that the maximum angle $\theta_0(\psi)$ in the integral (7.1.39) is given by (7.1.35), where $\mu_0 = \cos\ \theta_0$.

**7.4** *Asymptotic expansions of partition functions.* (a) From the exact integrals (7.1.34), (7.1.41), and (7.1.43), show that as $\lambda \to 0$,

$$\zeta_{\text{bal}} \to \frac{2}{\pi^{1/2}} \frac{\lambda^{5/2}}{\lambda_c}, \qquad \zeta_{\text{sat}} \to \frac{4}{3\pi^{1/2}} \lambda^{3/2} \qquad \text{and} \qquad \zeta_{\text{esc}} \to \frac{\lambda^2}{4\lambda_c^2}(1 + 2\lambda_c)$$

(b) Show that the integrated column density above $r_1$, when $\lambda_1 \ll 1$ and in the absence of satellite particles, is

$$\mathcal{N}(\mu_1 = 1, \lambda_1) = N_c e^{-\lambda_c r_c}\left(\frac{(1 + 2\lambda_c)\lambda_1}{4\lambda_c} + \frac{2\lambda_1^{3/2}}{3\pi^{1/2}}\right)$$

**7.5** *Orbital period.* (a) Show that the time spent by a ballistic particle in free flight above the exosphere is

$$\tau = \frac{2r_c\lambda_c}{U} \int_{\lambda_a}^{\lambda_c} \frac{d\lambda}{\lambda^2(h + \lambda - \nu_c\lambda^2/\lambda_c^2)^{1/2}}$$

where $\lambda_a$ is the potential at the particle's apogee. (b) Evaluate the integral for an elliptical orbit when $\lambda_c$ is at the perigee. Then show that for the special case of a circular orbit at $r_c$ the orbital period is

$$\frac{2\pi r_c}{(G\mathcal{M}/r_c)^{1/2}}$$

which is Kepler's harmonic law. (c) Compare the circular orbital period at 1.05 $R_E$ with the period of an orbit having an apogee at 15 $R_E$ and perigee at 2 $R_E$.

**7.6** *Integration limits for spectrum function.* Verify that if $v_1 < v_2$ so that $0 < v < v_1$, the $\lambda$ limits corresponding to (7.1.32) for ballistic particles are $\lambda_{2-} < \lambda < \lambda_{2+}$, where $\lambda_{2\pm}$ are given by (7.1.74).

**7.7** *Limiting flow.* Consider a planet with the radius and surface gravity of Earth. The major constituent is $CO_2$ with trace amounts of hydrogen compounds that are completely dissociated near and above the homopause. The binary collision parameter is $b_1 = 2 \times 10^{19}$ cm$^{-1}$ sec$^{-1}$. Just below the homopause the eddy diffusion coefficient is $K = 5 \times 10^7$ cm$^2$/sec, $D = K$, and the mixing ratio is $f_1(\mathrm{H}) = 10^{-4}$. Above the homopause $K = 0$. The temperature is everywhere constant at 400°. The gas-kinetic cross section is $Q = 3 \times 10^{-15}$ cm$^2$ and the only loss mechanism is Jeans escape.

(a) What is the H density at the exobase and what is the Jeans escape flux?

(b) What is the ratio of the actual H scale height to the hydrostatic, diffusive-equilibrium scale height for H above the homopause?

(c) Immediately below the homopause, what is the ratio of the actual H scale height to that of $CO_2$?

(d) Compute the limiting flow from (7.3.5), $\Phi_l \approx b_1 f_1/H$, below the homopause and explain why it differs from $\mathscr{F}_c$ found in part (a).

# *Appendix I*
# A TABLE OF PHYSICAL CONSTANTS

## Fundamental Constants

| | |
|---|---|
| Velocity of light | $c = 2.998 \times 10^{10}$ cm/sec |
| Gravitational constant | $G = 6.670 \times 10^{-8}$ dyne cm$^2$/gm$^2$ |
| Planck's constant | $h = 6.624 \times 10^{-27}$ erg sec |
| Mass of unit atomic weight | $M_0 = 1.66053 \times 10^{-24}$ gm |
| Mass of hydrogen atom | $M_H = 1.673 \times 10^{-24}$ gm |
| Mass of electron | $m = 9.1096 \times 10^{-28}$ gm |
| Electronic charge | $e = 4.8025 \times 10^{-10}$ esu |
| | $= 1.602 \times 10^{-20}$ emu |
| Avogadro's number | $1/M_0 = 6.0222 \times 10^{23}$ molecule/mole |
| Loschmidt's number (gas density at 273° K and 1 atmosphere pressure | $N_0 = 2.687 \times 10^{19}$ molecule/cm$^3$ |
| Boltzmann constant (gas constant per atom) | $k = R_0 M_0 = 1.380 \times 10^{-16}$ erg/deg |
| Gas constant per mole | $R_0 = k/M_0 = 1.987$ cal/deg mole |
| | $= 8.314 \times 10^7$ erg/deg mole |
| Gas constant for mass $M$ (molecular wt. $\mu = M/M_0$) | $R = k/M = R_0/\mu$ |
| | $= 8.314 \times 10^7/\mu$ erg/deg gm |

## Atomic and Radiation Constants

| | |
|---|---|
| Radius of first Bohr orbit | $a_0 = 0.529 \times 10^{-8}$ cm |
| Area of first Bohr orbit | $\pi a_0^2 = 0.880 \times 10^{-16}$ cm$^2$ |
| Rydberg constant for hydrogen | $R_H = 109.677.58$ cm$^{-1}$ |
| Rydberg constant for infinite mass | $R_\infty = 2\pi^2 me^4/ch^3 = 109{,}737.31$ cm$^{-1}$ |
| Energy per unit wave number | $hc = 1.986 \times 10^{-16}$ erg cm |
| Integrated absorption coefficient per atom for unit $f$-value | $\pi e^2/mc = 2.647 \times 10^{-2}$ cm$^2$/sec |

| | |
|---|---|
| Fine structure constant | $\alpha = 2\pi e^2/hc = 7.297 \times 10^{-3}$ |
| First radiation constant (specific intensity, $\lambda$ units) | $c_1 = 2hc^2 = 1.191 \times 10^{-5}$ erg cm$^2$/sec sr |
| Second radiation constant | $c_2 = hc/k = 1.439$ cm deg |
| Wein displacement-law constant | $\lambda_{max}T = 0.2898$ cm deg |
| Stefan–Boltzman constant | $\sigma = 5.67 \times 10^{-5}$ erg/cm$^2$ deg$^4$ sec |

## The Earth

| | |
|---|---|
| Radius (mean) | $R_E = 6.37 \times 10^8$ cm |
| Surface area | $4\pi R_E{}^2 = 5.10 \times 10^{18}$ cm$^2$ |
| Mass | $\mathcal{M} = 5.98 \times 10^{27}$ gm |
| Acceleration of gravity (at the surface) | $g = 9.807 \times 10^2$ cm/sec$^2$ |
| Magnetic moment of the dipole | $M = 8.1 \times 10^{25}$ G cm$^3$ |
| Velocity (mean) about the sun | $= 3.0 \times 10^6$ cm/sec |
| Distance (mean) from sun | 1 AU $= 1.496 \times 10^{13}$ cm |

## The Sun

| | |
|---|---|
| Radius (mean) | $= 6.96 \times 10^{10}$ cm |
| Surface area | $= 6.09 \times 10^{22}$ cm$^2$ |
| Mass | $= 1.99 \times 10^{33}$ gm |
| Acceleration of gravity (at the surface) | $= 2.740 \times 10^4$ cm/sec$^2$ |
| Acceleration of gravity (at 1 AU from the sun) | $= 0.593$ cm/sec$^2$ |
| Magnitude, absolute photovisual | $= +4.84$ |
| Magnitude, apparent photovisual | $= -26.73$ |
| Effective temperature (total energy radiated is equivalent to a black body with $T = T_{eff}$) | $T_{eff} = 5710$°K |
| Flux at sun's surface | $= 6.25 \times 10^{10}$ erg/cm$^2$ sec |

## Conversion Factors

| | |
|---|---|
| 1 day | $= 8.64 \times 10^4$ sec |
| 1 year | $= 3.1557 \times 10^7$ sec |
| 1 electron volt | $\varepsilon_0 = 1.602 \times 10^{-12}$ erg |
| | $\varepsilon_0/hc = 8.067 \times 10^3$ cm$^{-1}$ |
| 1 erg | $1/hc = 5.035 \times 10^{15}$ cm$^{-1}$ |
| | $1/\varepsilon_0 = 6.242 \times 10^{11}$ eV |
| 1 cm$^{-1}$ (kayser) | $hc = 1.986 \times 10^{-16}$ erg |
| | $hc/\varepsilon_0 = 1.2397 \times 10^{-4}$ eV |
| 1 eV/molecule | $= 23.053$ kcal/mole |
| Speed of 1 eV electron | $(2\varepsilon_0/m)^{1/2} = 5.93 \times 10^7$ cm/sec |
| Speed of 1 eV proton | $(2\varepsilon_0/M_H)^{1/2} = 1.38 \times 10^6$ cm/sec |
| Wavelength equivalent of 1 eV | $hc/\varepsilon_0 = 12{,}397 \times 10^{-8}$ cm |
| Temperature associated with 1 eV | $\varepsilon_0/k = 11{,}605$°K |

# *Appendix II*
# PLANETARY CHARACTERISTICS

| Planet | Mean Solar Distance (AU) [1 AU = $1.496 \times 10^{13}$ cm] | Period of Solar Revolution | Rotational Period (Sidereal, Eastward, Equatorial) | Orbital Eccentricity | Equatorial Inclination to Orbital Plane |
|---|---|---|---|---|---|
| Mercury | 0.387 | 88.0 day | 59 day | 0.206 | $<28^\circ$ |
| Venus | 0.723 | 224.7 day | −244 day | 0.007 | $3^\circ$ |
| Earth | 1.000 | 365.26 day | $23^h56^m4^s$ | 0.017 | $23^\circ27'$ |
| Mars | 1.524 | 687 day | $24^h37^m23^s$ | 0.093 | $23^\circ59'$ |
| Jupiter | 5.203 | 11.86 yr | $9^h50^m30^s$ | 0.048 | $3^\circ05'$ |
| Saturn | 9.539 | 29.46 yr | $10^h14^m$ | 0.056 | $26^\circ44'$ |
| Uranus | 19.18 | 84.01 yr | $-10^h49^m$ | 0.047 | $82^\circ05'$ |
| Neptune | 30.06 | 164.8 yr | $15^h48^m$ | 0.009 | $28^\circ48'$ |
| Pluto | 39.44 | 247.7 yr | $6^d9^h$ | 0.25 | — |

| Planet | Mass (Earth = 1) | Equatorial Radius (km) | Surface Gravity (Earth = 1) | Escape Velocity (km/sec) | Bond Albedo |
|---|---|---|---|---|---|
| Mercury | 0.0554 | 2425 | 0.37 | 4.2 | 0.056 |
| Venus | 0.815 | 6070 | 0.88 | 10.3 | 0.72 |
| Earth | 1.000 | 6378 | 1.00 | 11.2 | 0.39 |
| Mars | 0.1075 | 3395 | 0.38 | 5.0 | 0.16 |
| Jupiter | 317.8 | 71,300 | 2.64 | 61 | 0.45 |
| Saturn | 95.15 | 60,100 | 1.15 | 37 | 0.75 (?) |
| Uranus | 14.54 | 24,500 | 1.17 | 22 | 0.90 (?) |
| Neptune | 17.2 | 25,100 | 1.18 | 25 | 0.82 (?) |
| Pluto | 0.1 (?) | 3200 | — | — | 0.15 |

# *Appendix III*
# A MODEL OF EARTH'S ATMOSPHERE

The tables given in this appendix are adapted from

"U.S. STANDARD ATMOSPHERE, 1976," National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, U.S. Air Force, NOAA-S/T 76-1562, Supt. Documents, U.S. Gov. Printing Office, Washington, D.C. 20402.

In the tables the notation $a(b)$ stands for $a \times 10^b$.

**TABLE III.1** *Model Atmosphere*

| Height $z$ (km) | Geopotential Height $\Phi/g_0$ (km) | Temperature $T$ (°K) | Pressure $p$ (mb) | Number Density $N$ ($cm^{-3}$) | Mean Molecular Wt. $\langle\mu\rangle$ (gm/mole) | Pressure Scale Height $H$ (km) | Accel. of Gravity $g$ ($cm/sec^2$) |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 288 | 1.013(3) | 2.547(19) | 28.96 | 8.434 | 980.7 |
| 5 | 4.996 | 256 | 5.405(2) | 1.531(19) | 28.96 | 7.496 | 979.1 |
| 10 | 9.98 | 223 | 2.650(2) | 8.598(18) | 28.96 | 6.555 | 977.6 |
| 15 | 14.97 | 217 | 1.211(2) | 4.049(18) | 28.96 | 6.372 | 976.1 |
| 20 | 19.94 | 217 | 5.529(1) | 1.849(18) | 28.96 | 6.382 | 974.5 |
| 25 | 24.90 | 222 | 2.549(1) | 8.334(17) | 28.96 | 6.536 | 973.0 |
| 30 | 29.86 | 227 | 1.197(1) | 3.828(17) | 28.96 | 6.693 | 971.5 |
| 35 | 34.81 | 237 | 5.746(0) | 1.760(17) | 28.96 | 7.000 | 970.0 |
| 40 | 39.75 | 250 | 2.871(0) | 8.308(16) | 28.96 | 7.421 | 968.4 |
| 45 | 44.68 | 264 | 1.491(0) | 4.088(16) | 28.96 | 7.842 | 966.9 |
| 50 | 49.61 | 271 | 7.978(−1) | 2.135(16) | 28.96 | 8.047 | 965.4 |
| 55 | 54.53 | 261 | 4.253(−1) | 1.181(16) | 28.96 | 7.766 | 963.9 |
| 60 | 59.44 | 247 | 2.196(−1) | 6.439(15) | 28.96 | 7.368 | 962.4 |
| 65 | 64.34 | 233 | 1.093(−1) | 3.393(15) | 28.96 | 6.969 | 960.9 |
| 70 | 69.24 | 220 | 5.221(−2) | 1.722(15) | 28.96 | 6.570 | 959.4 |
| 75 | 74.13 | 208 | 2.388(−2) | 8.300(14) | 28.96 | 6.245 | 957.9 |
| 80 | 79.00 | 198 | 1.052(−2) | 3.838(14) | 28.96 | 5.962 | 956.4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 85 | 83.89 | 189 | 4.457(−3) | 1.709(14) | 28.96 | 5.678 | 955.0 |
| 86 | 84.85 | 187 | 3.734(−3) | 1.447(14) | 28.95 | 5.621 | 954.7 |
| 90 | 88.74 | 187 | 1.836(−3) | 7.12(13) | 28.91 | 5.64 | 953 |
| 95 | 93.60 | 189 | 7.597(−4) | 2.92(13) | 28.73 | 5.73 | 952 |
| 100 | 98.45 | 195 | 3.201(−4) | 1.19(13) | 28.40 | 6.01 | 951 |
| 110 | 108.13 | 240 | 7.104(−5) | 2.14(12) | 27.27 | 7.72 | 948 |
| 120 | 117.78 | 360 | 2.538(−5) | 5.11(11) | 26.20 | 12.09 | 945 |
| 130 | 127.40 | 469 | 1.250(−5) | 1.93(11) | 25.44 | 16.29 | 942 |
| 140 | 136.98 | 560 | 5.403(−6) | 9.32(10) | 24.75 | 20.03 | 939 |
| 150 | 146.54 | 634 | 4.542(−6) | 5.19(10) | 24.10 | 23.38 | 936 |
| 160 | 156.07 | 696 | 3.040(−6) | 3.16(10) | 23.49 | 26.41 | 933 |
| 180 | 175.04 | 790 | 1.527(−6) | 1.40(10) | 22.34 | 31.70 | 927 |
| 200 | 193.90 | 855 | 8.474(−7) | 7.189(9) | 21.30 | 36.18 | 922 |
| 220 | 212.64 | 899 | 5.015(−7) | 4.049(9) | 20.37 | 40.04 | 916 |
| 240 | 231.27 | 930 | 3.106(−7) | 2.429(9) | 19.56 | 43.41 | 911 |
| 260 | 249.78 | 951 | 1.989(−7) | 1.529(9) | 18.85 | 46.35 | 905 |
| 280 | 268.19 | 966 | 1.308(−7) | 9.818(8) | 18.24 | 48.93 | 900 |
| 300 | 286.48 | 976 | 8.770(−8) | 6.518(8) | 17.73 | 51.19 | 894 |
| 350 | 331.74 | 990 | 3.450(−8) | 2.528(8) | 16.64 | 55.83 | 881 |
| 400 | 376.32 | 996 | 1.452(−8) | 1.068(8) | 15.98 | 59.68 | 868 |
| 450 | 420.25 | 998 | 6.248(−9) | 4.687(7) | 15.25 | 63.64 | 855 |
| 500 | 463.54 | 999 | 3.024(−9) | 2.197(7) | 14.33 | 68.79 | 843 |
| 750 | 670.85 | 1000 | 2.260(−10) | 1.646(6) | 6.58 | 161.1 | 785 |
| 1000 | 864.07 | 1000 | 7.514(−11) | 5.445(5) | 3.94 | 288.2 | 732 |

**TABLE III.2** *Composition of Dry Atmosphere*[a] *(By Volume)*

| Species | Fractional Volume | Molecular Wt. |
|---|---|---|
| $N_2$ | 0.780,840 | 28 |
| $O_2$ | 0.209,476 | 32 |
| A | 9.34(−3) | 40 |
| $CO_2$ | 3.14(−4) | 44 |
| Ne | 1.818(−5) | 20.2 |
| He | 5.24(−6) | 4 |
| $CH_4$ | 2(−6) | 16 |
| Kr | 1.14(−6) | 83.8 |
| $H_2$ | 5(−7) | 2 |
| $O_3$ | 4(−7) | 48 |
| $N_2O$ | 2.7(−7) | 44 |
| CO | 2(−7) | 28 |
| Xe | 8.7(−8) | 131.3 |
| $NH_3$ | 4(−9) | 17 |
| $SO_2$ | 1(−9) | 64 |
| $NO_2$ | 1(−9) | 46 |
| NO | 5(−10) | 30 |
| $CCl_4$ | 1.2(−10) | 154 |
| $H_2S$ | 5(−11) | 34 |
| HBr, BrO | ~1(−11) | 81, 96 |

[a] Total thickness of the dry atmosphere is $\xi = 8.9 \times 10^5$ atm-cm or $\mathcal{N} = 2.15 \times 10^{25}$ molecule/cm$^2$.

**TABLE III.3** *Mid-Latitude Ozone Model*

| Height $z$ (km) | Number Density $[O_3]$ (cm$^{-3}$) | Local Thickness $\Delta\xi = 10^5[O_3]/N_0$ atm-cm/km | Volume Mixing Ratio $[O_3]/N$ |
|---|---|---|---|
| 2 | 6.8(11) | 2.5(−3) | 3.2(−8) |
| 6 | 5.7(11) | 2.1(−3) | 4.2(−8) |
| 10 | 1.13(12) | 4.2(−3) | 1.3(−7) |
| 15 | 2.65(12) | 9.9(−3) | 6.5(−7) |
| 20 | 4.77(12) | 1.77(−2) | 2.6(−6) |
| 25 | 4.28(12) | 1.59(−2) | 5.1(−6) |
| 30 | 2.52(12) | 9.38(−3) | 6.5(−6) |
| 35 | 1.40(12) | 5.21(−3) | 7.9(−6) |
| 40 | 6.07(11) | 2.26(−3) | 7.3(−6) |
| 45 | 2.22(11) | 8.26(−4) | 5.4(−6) |
| 50 | 6.64(10) | 2.47(−4) | 3.1(−6) |
| 60 | 7.33(9) | 2.73(−5) | 1.1(−6) |
| 70 | 5.4(8) | 2.0(−6) | 3.1(−7) |

**TABLE III.4** *Thermospheric Composition*

| Height $z$ (km) | Concentrations ($cm^{-3}$) | | | | | |
|---|---|---|---|---|---|---|
| | $[N_2]$ | $[O]$ | $[O_2]$ | $[A]$ | $[He]$ | $[H]$ |
| 86 | 1.13(14) | 8.6(10) | 3.0(13) | 1.4(12) | 7.6(8) | — |
| 90 | 5.5(13) | 2.4(11) | 1.5(13) | 6.5(11) | 4.0(8) | — |
| 100 | 9.2(12) | 4.3(11) | 2.2(12) | 9.5(10) | 1.1(8) | — |
| 150 | 3.1(10) | 1.8(10) | 2.8(9) | 5.0(7) | 2.1(7) | 3.7(5) |
| 200 | 2.9(9) | 4.1(9) | 1.9(8) | 1.9(6) | 1.3(7) | 1.6(5) |
| 250 | 4.8(8) | 1.4(9) | 2.5(7) | 1.5(5) | 9.7(6) | 1.2(5) |
| 300 | 9.6(7) | 5.4(8) | 3.9(6) | 1.6(4) | 7.6(6) | 1.0(5) |
| 400 | 4.7(6) | 9.6(7) | 1.3(5) | 2.1(2) | 4.9(6) | 9.0(4) |
| 500 | 2.6(5) | 1.8(7) | 4.6(3) | 3.4(0) | 3.2(6) | 8.0(4) |
| 750 | 2.7(2) | 3.7(5) | 1.8(0) | 2.0(−4) | 1.2(6) | 6.2(4) |
| 1000 | 4.6(−1) | 9.6(3) | 1.3(−3) | 2.2(−8) | 4.9(5) | 5.0(4) |

# *Appendix IV*
# PLANETARY SPACECRAFT MISSIONS

| Planet | Spacecraft | Launch | Mission Summary | Notes |
|---|---|---|---|---|
| Venus | Mariner 2 | 27 Aug. 62 | Fly-by 14 Dec. 62 | |
| Venus | Venera 3 | 16 Nov. 65 | Venus impact 1 Mar. 66 | |
| Venus | Venera 4 | 12 June 67 | Soft atmospheric entry 18 Oct. 67 | 1, 2 |
| Venus | Mariner 5 | 14 June 67 | Fly-by 19 Oct. 67 | 1 |
| Venus | Venera 5 | 5 Jan. 69 | Soft atmospheric entry 16 May 69 | 2 |
| Venus | Venera 6 | 10 Jan. 69 | Soft atmospheric entry 17 May 69 | 2 |
| Venus | Venera 7 | 17 Aug. 70 | Soft surface landing 15 Dec. 70 | 2 |
| Venus | Venera 8 | 27 Mar. 72 | Soft surface landing 22 July 72 | 2 |
| Venus/Mercury | Mariner 10 | 3 Nov. 73 | In solar orbit: Venus fly-by, 5 Feb, 74; Mercury fly-by, 29 Mar. 74; additional Mercury encounters in Sept. 74 and in Mar. 75 | 3 |
| Venus | Venera 9 | 8 June 75 | Orbiter and surface lander 22 Oct. 75 | 4 |
| Venus | Venera 10 | 14 June 75 | Orbiter and surface lander 25 Oct. 75 | 4 |
| Venus | Pioneer Venus 1 | 20 May 78 | Orbiter; Dec. 78 | 5 |
| Venus (planned) | Pioneer Venus 2 | Aug. 78 | Multiprobes; enter atmosphere Dec. 78 week after arrival of Orbiter | 5 |
| Mars | Mariner 4 | 28 Nov. 64 | Fly-by 14 July 65 | |
| Mars | Mariner 6 | 25 Feb. 69 | Fly-by 31 July 69 | |
| Mars | Mariner 7 | 27 Mar. 69 | Fly-by 5 Aug. 69 | |
| Mars | Mars 2 | 19 May 71 | Orbiter; hard landing probe 27 Nov. 71 | |
| Mars | Mars 3 | 28 May 71 | Orbiter; soft landing probe 2 Dec. 71 | 6 |
| Mars | Mariner 9 | 30 May 71 | Entered Mars orbit 13 Nov. 71 | 6 |
| Mars | Viking 1 | 20 Aug. 75 | Orbiter; soft landing probe 20 July 76 | 7 |
| Mars | Viking 2 | 9 Sept. 75 | Orbiter; soft landing probe 3 Sept. 76 | 7 |
| Jupiter | Pioneer 10 | 3 Mar. 72 | Fly-by 4 Dec. 73 | 8 |

| Planet | Spacecraft | Launch | Mission Summary | Notes |
|---|---|---|---|---|
| Jupiter/Saturn | Pioneer 11 | 6 Apr. 73 | Fly-by Jupiter 3 Dec. 74; expected to fly by Saturn Sept. 79 | 8 |
| Jupiter/Saturn | Voyager 1 | 5 Sept. 77 | Jupiter fly-by Mar. 79; Saturn fly-by Nov. 80 | 9 |
| Jupiter/Saturn/ Uranus (?)/ Neptune (?) | Voyager 2 | 20 Aug. 77 | Jupiter fly-by July 79; Saturn fly-by Aug. 81; possibly Uranus Jan. 86, Neptune Sept. 89 | 9 |
| Jupiter (planned) | Jupiter Orbiter/ Probe (Galileo) | Dec. 81 | Orbiter and atmospheric entry probe; expected arrival Nov. 84 | |

*Notes on publication of scientific results.*

**1.** Preliminary results of the American Mariner 5 fly-by of Venus are given in several brief articles by the participating investigators in

*Science* **158**, No. 3809, 29 Dec. (1967), pp. 1665–1688.

In March 1968, Soviet and American investigators met in Tucson, Arizona to discuss results of the Venera 4 and Mariner 5. The collected papers of this "Second Arizona Conference on Planetary Atmospheres" are published in

*J. Atmos. Sci.* **25**, No. 4, July (1968), pp. 533–671.

**2.** Summary articles on the Soviet series of Venus entry probes, Venera 4–8, are

VAKHNIN, V. M. (1968), A review of the Venera 4 flight and its scientific program, *J. Atmos. Sci.* **25**, pp. 533–534;
AVDUEVSKY, V. S.; MAROV, M. Ya.; and ROZHDESTVENSKY, M. K. (1970), A tentative model of the Venus atmosphere based on the measurements of Veneras 5 and 6, *J. Atmos. Sci.* **27**, pp. 561–579;
AVDUEVSKY, V. S.; MAROV, M. YA.; ROZHDESTVENSKY, M. K.; BORODIN, N. F.; and KERZHANOVICH, V. V. (1971), Soft landing of Venera 7 on the Venus surface and preliminary results of the investigations of the Venus atmosphere, *J. Atmos. Sci.* **28**, pp. 263–264;
MAROV, M. YA. *et. al.* (1973a), Preliminary results on the Venus atmosphere from the Venera 8 descent module, *Icarus* **20**, 407–421;
MAROV, M. YA. (1973b), Venera 8: Measurements of temperature, pressure, and wind velocity on the illuminated side of Venus, *J. Atmos. Sci.* **30**, pp. 1210–1214;
AVDUEVSKY, V. S.; MAROV, M. YA.; MOSHKIN, B. E.; and EKONOMOV, A. P. (1973), Venera 8: Measurements of solar illumination through the atmosphere of Venus, *J. Atmos. Sci.* **30**, 1215–1218.

**3.** Preliminary analyses of the U.S. Mariner 10 Venus fly-by are given in

*Science* **183**, No. 413; 29 Mar. (1974), pp. 1289–1321.

Preliminary reports of the first Mercury encounter of M 10 were published in

*Science* **185**, No. 4146, 12 July (1974), pp. 141–180.

**4.** Results of the Venera 9 and 10 orbiter/lander missions are reported in a series of papers, available in English translation from the Russian (*Kosmicheskie Issledovaniya*) in

*Cosmic Res.* **14**, No. 5, Sep.–Oct. (1976), pp. 573–701.

A summary article is

KELDYSH, M. V. (1977), Venus exploration with the Venera 9 and 10 spacecraft, *Icarus* **30**, 605–625.

**5.** The exploration of Venus, with particular emphasis on the expected role of the Pioneer Venus program, is reviewed in 11 papers in two issues of

*Space Sci. Rev.* **20**, Nos. 3–4 (1977), pp. 249–525.

**6.** Preliminary results from the Mariner 9 orbiter of Mars are given in

*Science* **175**, No. 4019, 21 Jan. (1972), pp. 293–323.

A number of papers on results of Mariner 9 and the Soviet Mars 3 are collected in

*Icarus* **17**, No. 2, Oct. (1972), pp. 289–327.

**7.** A series of three issues of *Science* were largely devoted to preliminary reports on Viking 1 and 2. They are

*Science* **193**, No. 4255, 27 Aug. (1976), pp. 759–815.
*Ibid.* **194**, No. 4260, 1 Oct. (1976), pp. 57–109.
*Ibid.* **194**, No. 4271, 17 Dec (1976), pp. 1274–1353.

Later discussion of the results by the experimenters appears in a special issue of

*J. Geophys. Res.* **82**, No. 28, 30 Sept. (1977), pp. 3959–4681, "Scientific Results of the Viking Project."

**8.** The Pioneer 10 and 11 preliminary reports appear, respectively, in

*Science* **183**, No. 4122, 25 Jan. (1974), pp. 301–324.
*Ibid.* **188**, No. 4187, 2 May. (1975), pp. 445–477.

Detailed analyses have been collected, following a conference on the Pioneer results, in

"Jupiter" (1976), (T. Gehrels, ed.), Univ. Arizona, Tucson, 1254 pp.

**9.** Twelve papers on various aspects of the Voyager missions are published together in

*Space Sci. Rev.* **21**, Nos. 2–3 (1977), pp. 75–376.

# *Appendix V*
# SUPPLEMENTARY READING

In order to develop the physics and chemistry of planetary atmospheres in some mathematical detail qualitative descriptions of atmospheric phenomena have been kept at a minimum. Several excellent sources of such material are available, and some students may wish to use one or more of them in conjunction with this book. Two texts that are mainly oriented toward physical and dynamical meteorology are

WALLACE, J. M. and HOBBS, P. V. (1977), "Atmospheric Science: An Introductory Survey," Academic Press, New York, 467 pp.

HOUGHTON, J. T. (1977), "The Physics of Atmospheres," Cambridge Univ. Press, Cambridge, 203 pp.

Meteorology is treated in a very clear and entertaining manner, with many illustrations, in

ANTHES, R. A.; PANOFSKY, H. A.; CAHIR, J. J.; and RANGO, A. (1975), "The Atmosphere," Chas. Merrill Publ. Co., Columbus, Ohio, 339 pp.

Atmospheric chemistry is examined in detail in

MCEWAN, M. J. and PHILLIPS, L. F. (1975), "Chemistry of the Atmosphere," Wiley, New York, 301 pp.

MCELROY, M. B. (1976), Chemical processes in the solar system: A kinetic perspective, *Chemical Kinetics, Ser. 2* **9**, pp. 127-211.

A very good descriptive book that concentrates on atmospheres of the planets is

GOODY, R. M. and WALKER, J. G. (1972), "Atmospheres," Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 150 pp.

A text oriented toward the solar system, somewhat out of date but still useful for fundamentals, is

BRANDT, J. C. and HODGE, P. W. (1964), "Solar System Astrophysics," McGraw-Hill, New York, 457 pp.,

which is written for advanced undergraduates.

A more elementary text, beautifully illustrated, is

HARTMANN, W. K. (1972), "Moons and Planets: An Introduction to Planetary Science," Wadsworth Publ. Co., Belmont, California, 404 pp.

At the popular level, the solar system is the subject matter of an issue of

*Scientific American* **233**, No. 3, Sept. (1975). [Reprinted as "The Solar System," W. R. Freeman & Co., San Francisco.]

From time to time the U.S. National Research Council (NRC) issues reports on some aspect of atmospheric physics and chemistry. The NRC is the operational arm of the National Academy of Sciences and National Academy of Engineering and as such undertakes these studies at the request of government agencies. Among the advantages of these reports are that they are written by committees of (unpaid) experts and are very authoritative and timely when issued; they are published rapidly and inexpensively, once completed and reviewed internally; they are addressed to nonspecialists and hence generally develop the current state of knowledge in a field before assessing the particular problem area that may have inspired the report; and they contain excellent bibliographies. Among the disadvantages are that they are not widely advertised; not widely available in libraries; and usually do not contain subject indices. To obtain a current list of titles one may write

Printing and Publishing Office
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

A few titles of recent years are

"Atmospheric Chemistry: Problems and Scope" (1975), Committee on Atmospheric Sciences, 130 pp.
"The Atmospheric Sciences: Problems and Applications" (1977), Committee on Atmospheric Sciences, 124 pp.
"Halocarbons: Effects on Stratospheric Ozone" (1976), Panel on Atmospheric Chemistry, Committee on Impacts of Stratospheric Change, 352 pp.;

and a companion report

"Halocarbons: Environmental Effects of Chlorofluoromethane Release" (1976), Committee on Impacts of Stratospheric Change, 126 pp.;
"Environmental Impact of Stratospheric Flight" (1975), Climatic Impact Committee, 348 pp.
"Understanding Climatic Change: A Program for Action" (1975), U.S. Committee for the Global Atmospheric Research Program, 239 pp;
"Ozone and Other Photochemical Oxidants" (1977), Committee on Medical and Biologic Effects of Environmental Pollutants, 719 pp;
"Nitrogen Oxides" (1977), Committee on Medical and Biologic Effects of Environmental Pollutants, 333 pp;
"Energy and Climate" (1977), Geophysics Study Committee, Geophysics Research Board, 281 pp;
"Climate, Climatic Change, and Water Supply" (1977), Geophysics Study Committee, Geophysics Research Board, 132 pp;
"Long-Term Worldwide Effects of Multiple Nuclear-Weapons Detonations" (1975), Assembly of Mathematical and Physical Sciences, 213 pp.

# AUTHOR INDEX

## C



## D



## E



## F



## G



## H

## I



## J



## K



## L

## M



## N



## O



## P



## R

## S



## T



## U



## V

## W



## Y



## Z

# SUBJECT INDEX

Boldface pages refer to definitions of terms or defining equations. Atoms and molecules are listed under their chemical formulae; thus methane appears under $CH_4$. Band systems and continua are listed under the molecule.

## F



## G



## H

## I



## J



## K



## L

## M



## N

## O

## P

## Q



## R

## S



## T

## U



## V

## W



## Z



A
B
C 8
D 9
E 0
F 1
G 2
H 3
I 4
J 5