Atmospheric Radiation

Theoretical Basis

SECOND EDITION

R. M. GOODY
and
Y. L. YUNG

Atmospheric Radiation

This page intentionally left blank

Atmospheric Radiation

Theoretical Basis

Second Edition

R. M. GOODY
Harvard University

and

Y. L. YUNG
California Institute of Technology

New York Oxford
OXFORD UNIVERSITY PRESS

Oxford University Press

Oxford New York
Athens Auckland Bangkok Bombay
Calcutta Cape Town Dar es Salaam Delhi
Florence Hong Kong Istanbul Karachi
Kuala Lumpur Madras Madrid Melbourne
Mexico City Nairobi Paris Singapore
Taipei Tokyo Toronto

and associated companies in
Berlin Ibadan

First published in 1989 by Oxford University Press, Inc.,
198 Madison Avenue, New York, New York 10016

First issued as an Oxford University Press paperback, 1995

Library of Congress Cataloging-in-Publication Data
Goody, Richard M.
Atmospheric radiation.
Includes bibliographies and index.
1. Atmospheric radiation. I. Yung, Y. L.
(Yuk Ling) II. Title
QC912.3.G66 1989 551.5'273 88-1442
ISBN 0-19-505134-3
ISBN 0-19-510291-6 (Pbk.)

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid free paper

PREFACE

The twenty-five years that have passed since *Atmospheric Radiation, Theoretical Basis,* was published have seen many altered perceptions, usually direct results of changing technologies. Fortunately, there have been fewer advances in fundamental ideas, which are the concern of this book. The second edition has been extensively revised to reflect current knowledge.

We have preserved the organization of the first edition to the extent possible. The task of reconciling manuscripts by two authors writing a continent apart was greatly eased by working to an established pattern.

Two new sections have been added. Whereas remote sensing (§ 6.5) was in an early stage when the first edition was prepared, and at that time offered few interesting and novel ideas, it is now a major discipline with extensive relationships to many other fields.

Solutions to scattering problems were mentioned only in summary form in the earlier work because Chandrasekhar's landmark book, *Radiative Transfer,* appeared then to be definitive, leaving little room for improvement. (That was before the advent of large computers.) Now the symbiosis between transfer theory and numerical methods has given rise to another important and sometimes elegant discipline (Chapter 8).

We wish to express our gratitute to a number of our colleagues who assisted us in different ways: John Shaw helped us to come to grips with modern theories of molecular spectra; Peter Gierasch contributed a thoughful review of Chapter 10 at a time when it threatened to turn into a monograph; Stephen Fels put much effort into reviewing some of the chapters and gave us the benefit of his deep insights into many topics; we are indebted to B. Crofton Farmer and his colleagues at the Jet Propulsion Laboratory for access to their computer, their programs, and their data, that has allowed us to prepare the illustrative spectra in Chapter 3; Andrew Ingersoll, Donald Hunten, and Darrell Strobel made available to use their lists of typographic errors in the first edition; and we are particularly grateful to David Crisp who reviewed the entire manuscript and offered invaluable comments on both style and content.

Finally we would not have attempted this revision but for kind and encouraging comments by many individuals to the effect that they or their students had derived benefit from the first edition; and we would not

have been able to complete the task but for the generous hospitality of the Jet Propulsion Laboratory that provided us with a meeting place.

Cambridge, Massachusetts — R. M. G.
Pasadena, California — Y. L. Y.
January, 1989

CONTENTS

This page intentionally left blank

Atmospheric Radiation

This page intentionally left blank

1

INTRODUCTION

1.1. The nature of the problem

Earth, like the other inner planets, receives virtually all of its energy from space in the form of solar electromagnetic radiation. Its total heat content does not vary significantly with time, indicating a close overall balance between absorbed solar radiation and the diffuse stream of low-temperature, thermal radiation emitted by the planet. The transformation of the incident solar radiation into scattered and thermal radiation, and the thermodynamic consequences for the earth's gaseous envelope, are the subjects of this book.

The scope must be narrowed, however, because in its broadest interpretation our title could include atmospheric photochemistry and many other topics usually treated in books dealing with the upper atmosphere. By restricting attention to the thermodynamic aspects, this problem of selection usually resolves itself. For example, the absorption of energy accompanying photodissociation or photoionization will be considered if the energy involved is comparable to that of other sources or sinks, but not otherwise. Similarly, the oxygen airglow has some thermodynamic consequences in the upper atmosphere, but the important topic of the airglow will be mentioned only in this limited context.

The irradiance[1] at mean solar distance—the *solar constant*—is slightly less than $1400\ \mathrm{W\,m^{-2}}$, giving an average flux of solar energy per unit area of the earth's surface equal to $350\ \mathrm{W\,m^{-2}}$ (the factor 4 is the ratio of surface area to cross section for a sphere). Of this energy, approximately 31% is scattered back into space, 43% is absorbed at the earth's surface, and 26% is absorbed by the atmosphere. The ratio of outward to inward flux of solar radiation is known as the *albedo*. We may speak of the albedo of the entire earth or of individual surfaces with reference either to monochromatic radiation or to a weighted average over the solar spectrum. In this last sense the albedo of the earth as a whole is about 0.31, and an average of $224\ \mathrm{W\,m^{-2}}$ is available for heating, directly and indirectly, the earth and its atmosphere.

The redistribution of this absorbed solar energy by dynamic and radiative processes and its ultimate return to space as low-temperature

[1] A brief account of the sun as a source of radiation is given in Appendix 9.

planetary or *terrestrial radiation* are the most important topics of this book. They are related mainly to conditions in the troposphere and lower stratosphere, where most of the atmospheric mass resides, but there are also other problems of interest concerned with the small amounts of ultraviolet radiation that can be absorbed high in the atmosphere where densities are low and the resulting thermal effects may be large. This raises problems, unfamiliar in the lower atmosphere, such as nonequilibrium conditions, chemical energy, etc.

Assuming that the earth radiates as a black body in the infrared spectrum, we may compute the general level of terrestrial temperatures. The rate at which energy is absorbed by earth is

$$\text{rate of absorption} = f\pi r^2(1-a), \tag{1.1}$$

where f = solar constant ($\sim 1400\ \text{W m}^{-2}$),

r = earth's radius, and

a = albedo for solar radiation (~ 0.31).

The rate at which energy is emitted by earth is

$$\text{rate of emission} = 4\pi r^2 \sigma \theta_e^4, \tag{1.2}$$

where σ = the Stefan–Boltzmann constant and

θ_e = effective emission temperature of earth.

If the planet is in a steady state, we may equate (1.1) to (1.2) to give 255.5 K for the effective emission temperature. This is lower than the average temperature of the earth's surface but approximately equal to the average temperature of the atmosphere, indicating that much of the radiation to space must come from the atmosphere rather than from the surface, a conclusion that is confirmed from a cursory examination of the atmospheric absorption spectrum.

At a temperature of 255.5 K emission of thermal radiation is negligible for wavelengths less than 4 μm (micron)[2] and, since solar radiation carries little energy at longer wavelengths, it is possible and convenient to treat the solar and terrestrial fluxes independently. First consider the terrestrial component. The principal gaseous constituents of the atmosphere (O_2, N_2, A) are almost transparent to wavelengths longer than 4 μm, but minor polyatomic constituents such as H_2O, CO_2, O_3, N_2O, CO, and CH_4 have intense and complex absorption spectra and are present in sufficient quantities to absorb a considerable proportion of the

[2] The micron (μm) and other spectrographic units are discussed and defined in Appendix 2.

terrestrial radiation. Dust, haze, and, perhaps most importantly, clouds absorb and scatter light strongly throughout the entire spectrum. Since clouds, ground, and atmosphere do not differ greatly in temperature, it follows from Kirchhoff's laws that emission and absorption are approximately equal to each other. Terrestrial radiation is therefore passed from layer to layer in the atmosphere, creating a transfer problem of great intricacy. The situation differs in the upper atmosphere because Kirchhoff's laws are not obeyed if the pressure is very low.

Absorption by polyatomic gases is complicated, each band consisting of many lines whose individual shapes can affect the radiative transfer. If the data are used in their entirety, even a relatively simple problem may be impractical because of the requirements for computer time. Methods are required that treat correctly only the essential statistics of the problem. The development of such methods has led to a fairly complete understanding of the transfer problem for constant-pressure paths. Absorption along an atmospheric path, for which temperature, pressure, and chemical composition all vary together, presents new problems, not all of which have been solved.

Figure 1.1 gives a general picture of the importance of different absorptions in the lower atmosphere for mid-latitudes. An indication of the energy absorbed by the stratosphere or troposphere can be found by multiplying (a) by (c) or by (b) − (c), respectively.

Most of the solar absorption in the stratosphere is by the ultraviolet Hartley and Huggins bands of ozone. At much higher levels, the small amount of solar radiation with wavelengths less than 0.2 μm is absorbed, mainly by molecular oxygen. In the troposphere, depletion of sunlight is caused by a group of near infrared bands of water vapor. For terrestrial radiation, water vapor is the most important single constituent of the lower atmosphere, although carbon dioxide is always significant. In the stratosphere, water vapor, carbon dioxide, and ozone are of comparable importance, while in the mesosphere carbon dioxide dominates.

In addition to the absorptions shown in Fig. 1.1, both solar and terrestrial streams of radiation are absorbed or scattered by dust, haze, molecules, and clouds. The theory of scattering by molecules and by water drops is well developed, but dust and haze cannot be treated so precisely and their amounts are variable and difficult to relate to other physical or meteorological phenomena.

The motions, temperatures, chemical composition, and amounts and types of cloud are all related. Ideally, nothing should be taken as given; all atmospheric properties, including the radiation field, should result from model calculations, given only the fixed boundary conditions and the incident flux of solar radiation. Fully interactive models are now possible using the largest computers, but their very complexity presents troublesome questions of verification and raises serious barriers to understanding. Advances have, in fact, been quite limited and it is

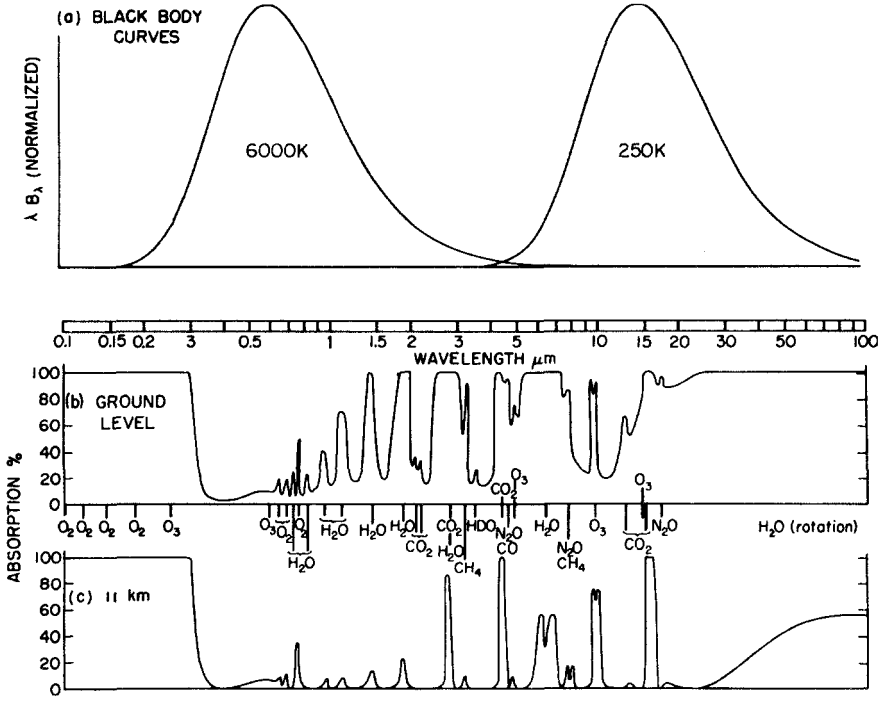


FIG. 1.1 Atmospheric absorptions. (a) Black-body curves for 6000 K and 250 K. (b) Atmospheric absorption spectrum for a solar beam reaching ground level. (c) The same for a beam reaching the temperate tropopause. The axes are chosen so that areas in (a) are proportional to radiant energy. Integrated over the earth's surface and over all solid angles, the solar and terrestrial fluxes are equal to each other; consequently, the two black-body curves are drawn with equal areas. Conditions are typical of mid-latitudes and for a solar elevation of 40° or for a diffuse stream of terrestrial radiation.

possible that the promise of very large computers is illusory if the purpose is understanding as opposed to elaborate bookeeping.

Historically, before large computers became available, two classes of study developed, one synthetic but limited in scope and the other descriptive but aiming at completeness. The former, typified by the early work of Gold and Emden, approaches from the standpoint of local radiative equilibrium in an atmosphere of known cloudiness and chemical composition, and inquires how far such models can account for the observed atmosphere. Modifications, such as adding a convective troposphere, can make models more realistic. This kind of a priori approach was also important in the early days of planetary exploration because almost no direct observations were available. The second class of study, for which Simpson's early work is the paradigm, accepts the observed thermal structure and attempts to calculate the accompanying radiation field. The two methods are complementary; each throws some

light on the nature of the hydrodynamic processes, the former from the ways in which the models differ from observations and the latter by computing the diabatic heating to which the motions must respond. These classical methods, rather than the theory and results from large numerical models, are emphasized in this book.

The theoretical basis of atmospheric radiation involves a wide range of physical concepts and mathematical methods. It may help the reader to indicate how the different chapters fit into the scheme of ideas that has been outlined.

The remainder of Chapter 1 provides background information about the observed thermal structure and chemical composition of the atmosphere. Chapter 2 outlines the formal mathematical theory required to handle radiative transfer problems. Chapter 3 describes the physics of molecular absorption and Chapter 4 discusses methods that isolate the essential statistics from the mass of spectral details. Chapter 5 makes some reference to the spectrographic data available for application to atmospheric problems. Having provided in the last three chapters the information required about atmospheric absorption, Chapter 6 is concerned with calculating atmospheric radiation in an atmosphere of arbitrary structure. The final section in Chapter 6 discusses the inverse problem, namely the retrieval of atmospheric parameters from observed satellite data. To this point dust and haze have not been considered and clouds have been considered only in idealized form. Chapter 7 outlines the theory of scattering by small particles and droplets and Chapter 8 discusses the mathematical methods required to use these data. Chapters 9 and 10 deal with some simple investigations involving atmospheric radiation. Chapter 9 discusses atmospheres in radiative and radiative–convective equilibrium, and what we can learn from them. Chapter 10 outlines formal approaches to the problem of greatest concern to meteorology, the interaction between fields of radiation and of motion.

1.2. The thermal structure of the atmosphere

The average thermal structure of the atmosphere differs little from year to year, and it is useful to think in terms of climatological mean conditions with small variations superimposed. Figure 1.2 shows longitudinally averaged temperatures for January and July in the Northern Hemisphere. For the purposes of this book, these cross sections represent conditions at all longitudes and in both hemispheres, although the distribution of land and sea does influence climatological means.

The following features of Fig. 1.2 may be noted. The structure of the lowest 2 or 3 km is complicated, with inversions at some latitudes. Above 3 km, however, there are some regular features. The equally spaced isotherms indicate a constant lapse rate of about 6.5 K km^{-1}, independent of both season and latitude. At the tropopause a sudden change to

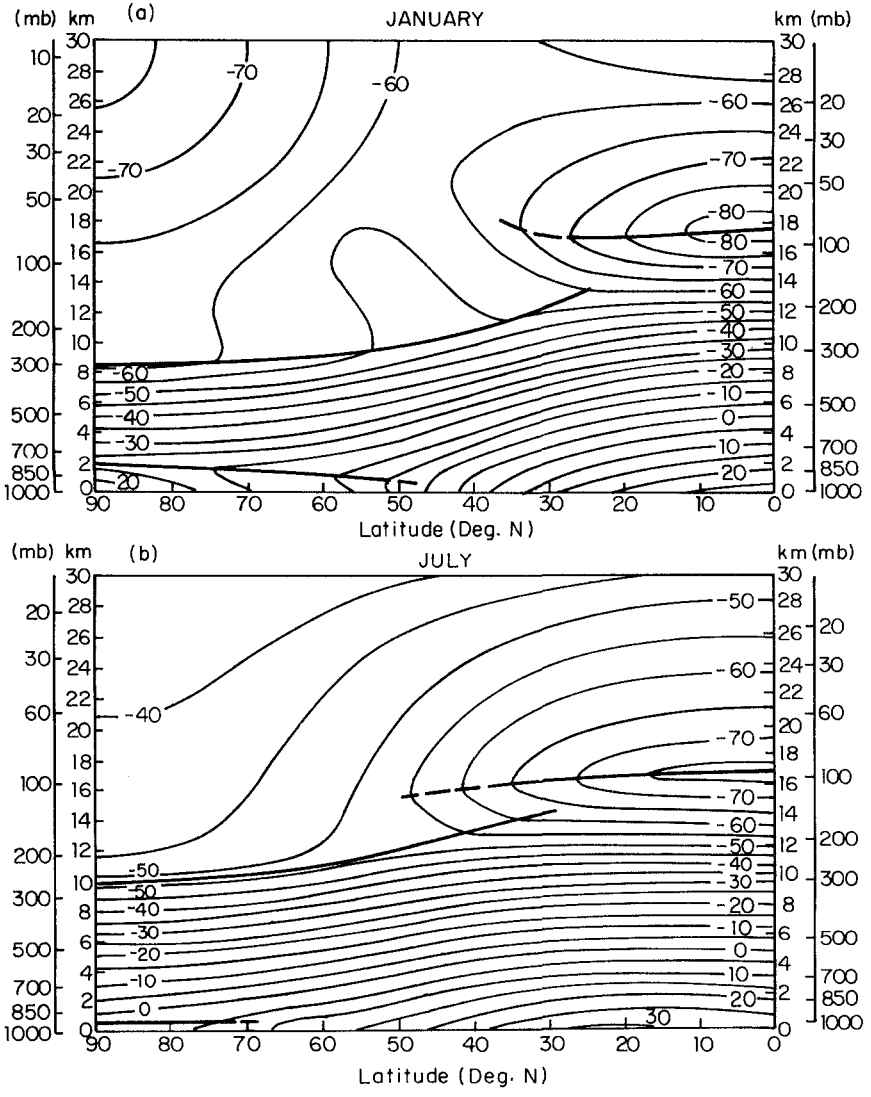


FIG. 1.2. Meridional cross sections of longitudinally averaged temperatures. Temperatures are in Centrigrade. Northern Hemisphere data: (a) January, (b) July. Heavy lines indicate the tropopause and the arctic inversion. After Holton (1979).

isothermal or inversion conditions takes place. The tropopause is usually multiple between latitudes 30° and 50°, where the high tropical tropopause overlaps the low arctic tropopause. In the stratosphere, temperatures are, curiously, lower in tropical regions than in the arctic. In Appendix 3 atmospheric parameters are given for a *model atmosphere* in the range 0–30 km.

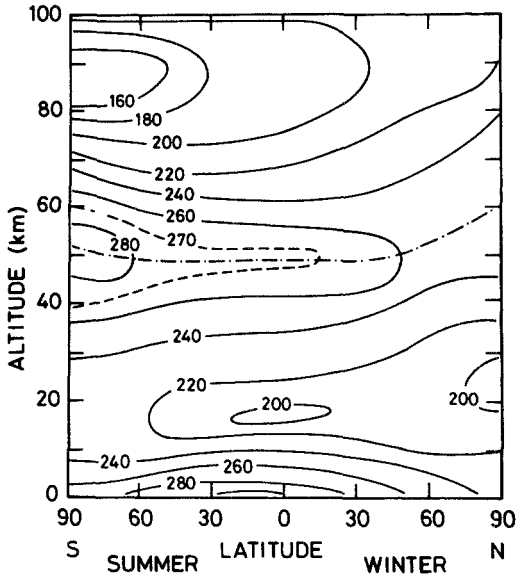


FIG. 1.3. Temperature of the middle atmosphere at solstice. After Brasseur and Solomon (1984).

Above 30 km thermal data are more sparse. The main feature of Fig. 1.3 is a temperature maximum close to 50 km at all latitudes. The temperature minimum at 80 km may be deeper (down to 130 K) than is indicated in the figure. Above this minimum, the rise of temperature into the thermosphere is a dominating feature of the upper atmosphere. Numerical details are given in Appendix 3.

Above 100 km the physical and chemical state of the atmosphere is variable. Temperature always increases with height up to about 250 km, above which level the thermal conductivity of air is so large as to create a near-isothermal state. Whether this constant temperature is as low as 1000 K or as high as 2000 K depends on the level of solar activity; the diurnal variation is also large. Figure 1.4 shows two temperature profiles up to 500 km, and Appendix 3 gives data on a model atmosphere up to 1000 km.

Atmospheric nomenclature is generally derived from the thermal state. We have already used the terms *troposphere, stratosphere, mesosphere,* and *thermosphere.* These designations suffer from being too evocative and there is almost no physical distinction to be drawn between the stratosphere and the mesosphere. Better terms are *lower atmosphere* (troposphere), *middle atmosphere* (stratosphere and mesosphere), and *upper atmosphere* (above 80 km).

The lowest 1 or 2 km of the atmosphere differs from the remainder

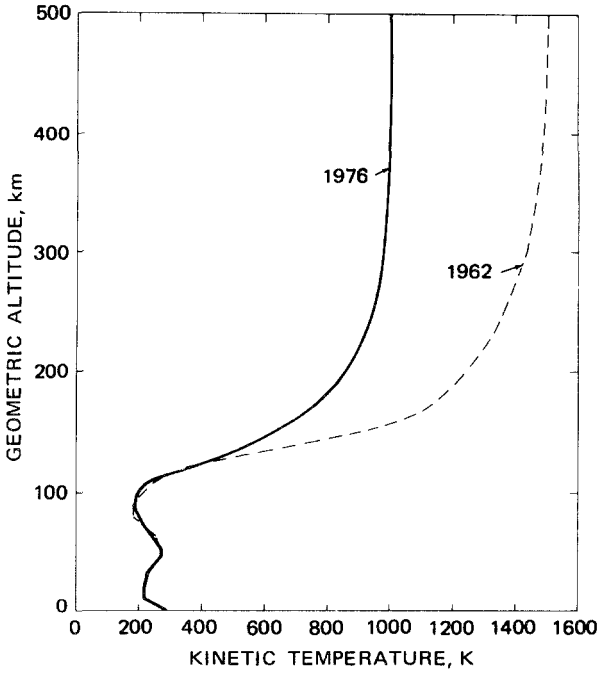


FIG. 1.4. Kinetic temperature up to 500 km. From the *U.S. Standard Atmosphere 1976.*

of the troposphere. Interactions with the surface are strong and diurnal variations are large. The lowest 100 m or so has been intensively studied and is referred to as the *planetary boundary layer.*

1.3. The chemical composition of the atmosphere

A summary of the available data on the composition of dry air is given in Table 1.1. Relative concentrations for important isotopes are given in Table 1.2. From these figures the totals in Table 1.1 can be broken down into isotopic components, if required.

Concentrations of atmospheric gases are subject to chemical and photochemical alteration. Molecular oxygen is decomposed into atoms above 90 km. Methane and nitrous oxide are unstable in the stratosphere. Figure 1.5 shows representative vertical profiles of the species listed in Table 1.1, for average, mid-latitude conditions.

The concentration of water vapor is influenced by condensation, and the gas is not evenly mixed with other constitutents below 15 km. In the troposphere, the average relative humidity is close to 50% and the vapor pressure varies over a very wide range. For a ground temperature of 278 K the vapor pressure is close to 8 mb, on the average; above 15 km

Table 1.1. The composition of dry air

Molecule	Fraction by volume in the troposphere	Comments
N_2	7.8084×10^{-1}	Photochemical dissociation high in the ionosphere; mixed at lower levels
O_2	2.0946×10^{-1}	Photochemical dissociation above 95 km; mixed at lower levels
A	9.34×10^{-3}	Mixed up to 110 km; diffusive separation above
CO_2	3.45×10^{-4}	Slightly variable; mixed up to 100 km; dissociated above
CH_4	1.6×10^{-6}	Mixed in troposphere; dissociated in mesosphere
N_2O	3.5×10^{-7}	Slightly variable at surface; dissociated in stratosphere and mesosphere
CO	7×10^{-8}	Variable photochemical and combustion product
O_3	$\sim 10^{-8}$	Highly variable; photochemical origin
$CFCl_3$ and CF_2Cl_2	$1\text{–}2 \times 10^{-10}$	Industrial origin; mixed in troposphere, dissociated in stratosphere

Note: Many other trace gases are present in the atmosphere, some of which (e.g., H_2, NO, SO_2, NH_3, Ne, He, Rn, Kr, Xe) have been extensively studied, but none influences the radiation fluxes to a significant extent. In addition to these gases, the atmosphere contains solid matter in suspension, whose concentration and composition are highly variable. Water vapor is discussed in the text; the above figures apply to dry air.

the relative humidity falls to 1% and the vapor pressure averages 4×10^{-4} mb. Measured frost-point temperatures are shown in Fig. 1.6. These may be interpreted in terms of vapor pressures with the aid of the vapor–pressure data in Appendix 4. The highest available frost-point measurements from aircraft indicate a frost-point temperature of 189 K at 15 km at all times and over a wide range of latitudes. Other observations suggest that, above this level, oxidation of methane provides an additional source of water. Photochemical decomposition of water starts around 80 km, reducing the mixing ratio above this level.

Table 1.2. Isotopic abundances in nature

Isotope	Percentage relative abundance
^{1}H	99.9851
^{2}D	0.0149
^{12}C	98.892
^{13}C	1.108
^{16}O	99.758
^{17}O	0.0373
^{18}O	0.2039
^{14}N	99.631
^{15}N	0.369

Note: Almost all terrestrial hydrogen is combined in the form of water. Since HHO and HDO have different vapor pressures, the relative concentrations of ^{2}D to ^{1}H can vary from phase to phase by as much as 10%. Small differences in the concentrations of oxygen isotopes also occur.

The vertical distribution of ozone differs from that of other atmospheric gases, having a maximum number density near to 25 km (Fig. 1.7). Ozone is formed photochemically from oxygen: rapidly above 30 km so that equilibrium is obtained during daylight hours and slowly below this level so that ozone concentration depends on mixing and transport processes, and is highly variable. This variability is reflected in the

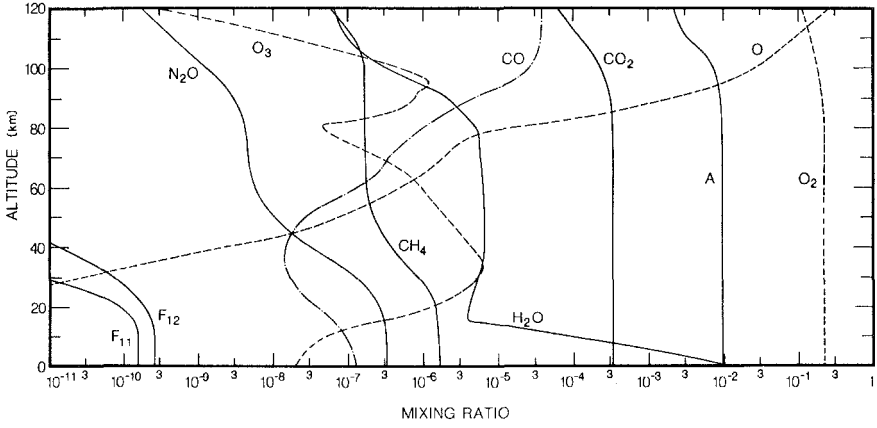


FIG. 1.5. Vertical profiles of mixing ratio of selected species at equinox. After Allen et al. (1981, 1984).

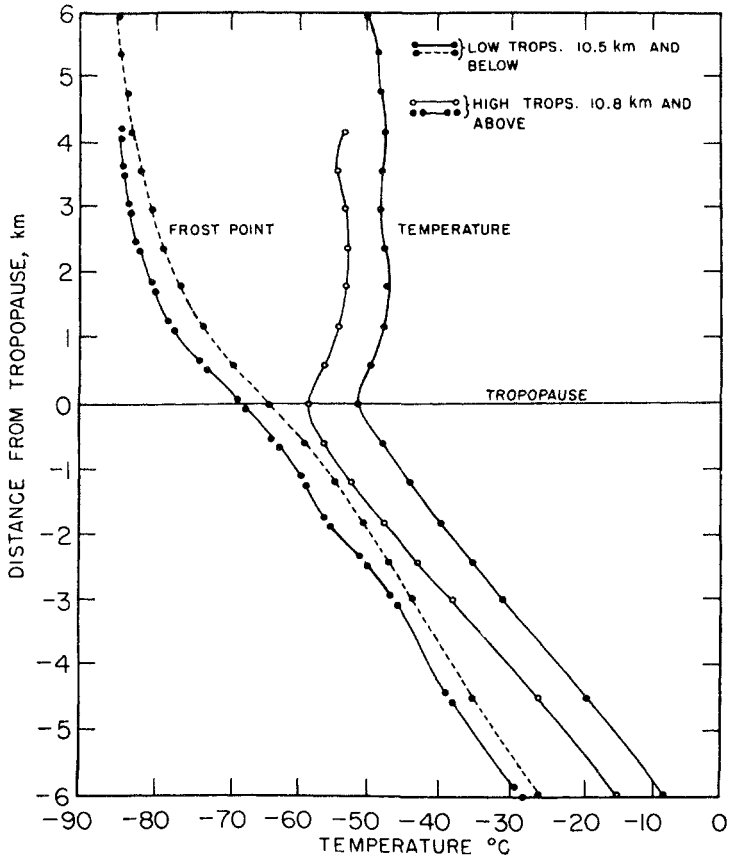


FIG. 1.6. Mean temperature and frost-point over southern England. After Murgatroyd et al. (1955).

seasonal and latitudinal changes shown in Fig. 1.8. The maximum ozone amount occurs in the polar night, where none is formed. Day-to-day ozone changes are related to the passage of weather systems.

Carbon dioxide strongly influences the radiation field at all levels below 100 km. It is chemically unreactive and has its main sources and sinks in industrial and biological processes at the earth's surface. In the planetary boundary layer, its concentration is variable but, at higher levels, its mixing ratio is essentially constant below the dissociation level of molecular oxygen; above this level carbon dioxide dissociates. The total amount of carbon dioxide in the atmosphere is slowly increasing with time (Fig. 1.9) because of industrial and agricultural activity.

Methane and nitrous oxide are of biospheric origin. Both are destroyed by photochemical processes in the middle atmosphere. Observations show that their mixing ratios are constant in the lower atmosphere and slowly increase with time.

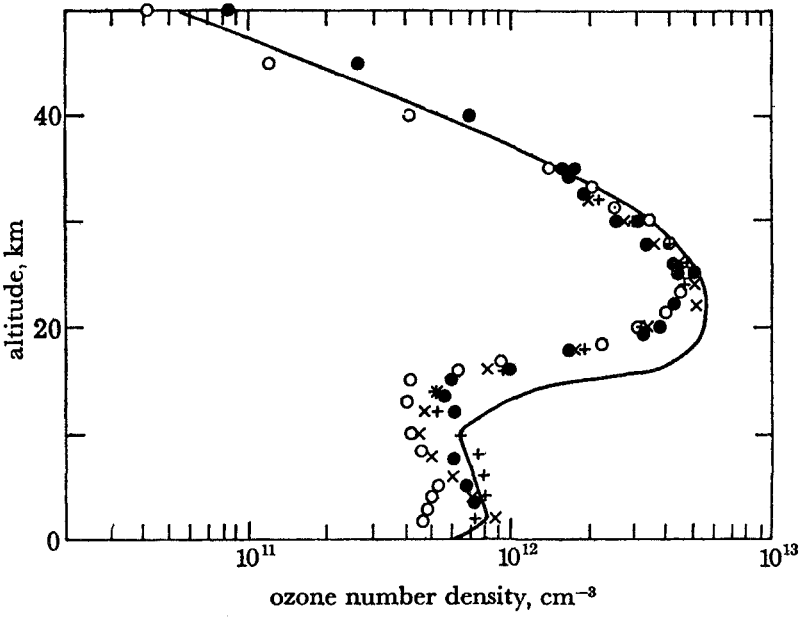


FIG. 1.7. The vertical distribution of ozone. The solid line is the result of theoretical calculations by Logan et al. (1978) for middle latitudes, and the points are measurements by a number of different authors.

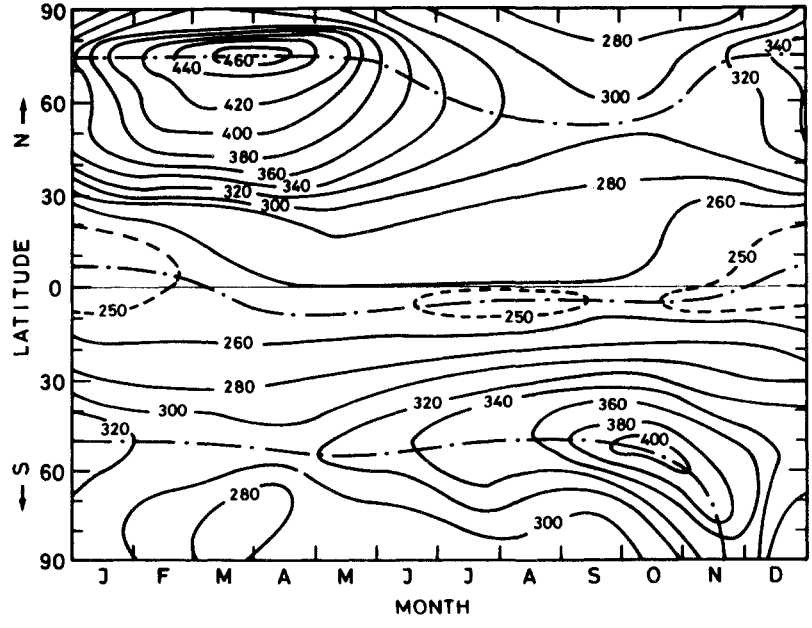


FIG. 1.8. Latitude–season cross section of total ozone. The total ozone in a vertical column is measured in units of 10^{-3} cm of the gas reduced to s.t.p. (Dobson units). After Brasseur and Solomon (1984).

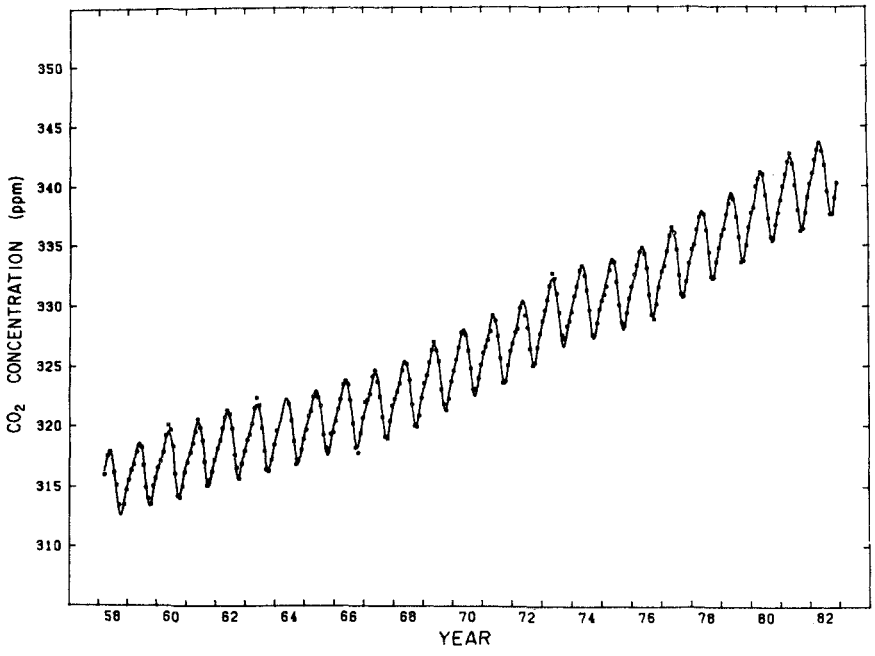


FIG. 1.9. Molecular concentration of CO_2 at the Mauna Loa observatory, Hawaii. Both long-term and seasonal changes are shown. After Bacastow et al. (1985).

Tropospheric carbon monoxide is a photochemical and combustion product, more variable in composition than are methane and nitrous oxide. The molecular mixing ratio varies from 5×10^{-8} in the Southern Hemisphere to 2×10^{-7} in the Northern Hemisphere. The vertical distribution is irregular, reflecting the production and loss processes.

Dust and haze are so variable in amount and chemical composition that few generalizations can be made, particularly about the vertical distribution. Visibility measurements reflect the aerosol concentration at ground level. The visual range can vary from a meter or two to 200 km, depending upon atmospheric conditions. Generally speaking, dust and haze concentrations decrease rapidly with height in the troposphere; a scale height of 1 km is typical. Visible tops are formed at inversions. Haze particles are frequently hygroscopic and their size depends upon the relative humidity. In the stratosphere the concentration of aerosols smaller than 0.1 μm continues to fall off with height, but larger particles (0.1 μm–1.0 μm) increase to a maximum near 20 km. Aerosols are usually considered important for their effects on solar radiation, but they may also modify the terrestrial radiation, particularly close to the ground.

Clouds have very great influence on both the solar and the terrestrial radiation. They are properly described as aerosols but are always treated as an independent phenomenon. It is not possible to give a useful, brief

description of cloud climatology, cloud physics, and the associated dynamic processes.

BIBLIOGRAPHY

1.1. The nature of the problem
Two books that also deal with the fundamentals of atmospheric radiation are

Kondratyev, K. Ya., 1969, *Radiation in the atmosphere.* New York: Academic Press.

Liou, K.-N., 1980, *An introduction to atmospheric radiation.* New York: Academic Press.

1.2. The thermal structure of the atmosphere
The data in Figs. 1.2, 1.3, and 1.4 are from

Holton, J. R., 1979, *An introduction to dynamic meteorology.* New York: Academic Press.

Brasseur, G., and Solomon, S., 1984, *Aeronomy of the middle atmosphere: Chemistry and physics of the stratosphere and mesosphere.* Dordrecht: Reidel.

U.S. Standard Atmosphere 1976. Washington D.C.: U.S. Government Printing Office.

The data in Fig. 1.3 and in Fig. 1.8 were compiled by J. London but Brasseur and Solomon (1984) is the only source in the open literature.

A recent book on the thermal structure and aeronomy in the upper atmosphere is

Chamberlain, J. W., and Hunten, D. M., 1987, *Theory of planetary atmospheres: An introduction to their physics and chemistry.* New York: Academic Press.

Details of the temperature of the lower atmosphere are to be found in any textbook on meteorology. The earth's boundary layer is discussed by

Priestley, C. H. B., 1959, *Turbulent transfer in the lower atmosphere.* Chicago: University of Chicago Press.

1.3. The chemical composition of the atmosphere
A recent review of tropospheric chemistry is by

Logan, J. A., Prather, M. J., Wofsy, S. C., and McElroy, M. B., 1981, "Tropospheric chemistry: A global perspective," *J. Geophys. Res.* **86,** 7210.

The data in Fig. 1.5 are from

Allen, M., Yung, Y. L., and Waters, J. W., 1981, "Vertical transport and photochemistry in the terrestrial mesosphere and lower thermosphere (50–120 km)," *J. Geophys. Res.* **86,** 3617.

Allen, M., Lunine, J. I., and Yung, Y. L., 1984, "The vertical distribution of ozone in the mesosphere and lower thermosphere," *J. Geophys. Res.* **89,** 4841.

The water vapor measurements shown in Fig. 1.6 are very old but still useful, see

Murgatroyd, R. J., Goldsmith, P., and Hollings, W. E. H., 1955, "Some recent measurements of the humidity from aircraft up to heights of about 50,000 ft over southern England," *Quart. J. R. Meteorol. Soc.* **81,** 533.

The ozone calculations shown in Fig. 1.7 are by

Logan, J. A., Prather, M. J., Wofsy, S. C., and McElroy, M. B., 1978, "Atmospheric chemistry: Response to human influence," *Phil. Trans. Roy. Soc. London Ser. A* **290,** 187.

The Hawaiian carbon dioxide measurements are reported by

Bacastow, R. B., Keeling, C. D., and Whorf, T. P., 1985, "Seasonal amplitude increase in the atmospheric CO_2 concentration at Mauna Loa, Hawaii, 1959–1982," *J. Geophys. Res.* **90,** 10529.

Isotopic ratios are given by

Evans, R. D., 1955, *The atomic nucleus.* New York: McGraw-Hill.

Data on atmospheric aerosol and clouds are reviewed by

Mason, B. J., 1957, *The physics of clouds.* Oxford: The Clarendon Press.

Junge, C. E., and Manson, J. E., 1961, "Stratospheric aerosol studies," *J. Geophys. Res.* **66,** 2163.

Pruppacher, H. R., and Klett, J. D., 1978, *Microphysics of clouds and precipitation.* Dordrecht: Reidel.

2

THEORY OF RADIATIVE TRANSFER

2.1. Definitions

2.1.1. *Intensity, flux, energy density*

In common with astrophysical usage the word *intensity*[1] will denote *specific intensity of radiation*, i.e., the flux of energy in a given direction per second per unit frequency (or wavelength) range per unit solid angle per unit area perpendicular to the given direction. In Fig. 2.1 the point P is surrounded by a small element of area $d\pi_s$, perpendicular to the direction of the unit vector $\mathbf{s}$. From each point on $d\pi_s$ a cone of solid angle $d\omega_s$ is drawn about the $\mathbf{s}$ vector. The bundle of rays, originating on $d\pi_s$, and contained within $d\omega_s$, transports in time dt and in the frequency range ν to $\nu + d\nu$, the energy

$$E_\nu = I_\nu(P, \mathbf{s})\, d\pi_s\, d\omega_s\, d\nu\, dt, \tag{2.1}$$

where $I_\nu(P, \mathbf{s})$ is the specific intensity at the point P in the $\mathbf{s}$-direction. If I_ν is not a function of direction the intensity field is said to be *isotropic*; if I_ν is not a function of position the field is said to be *homogeneous*. If it is more convenient to use wavelength than frequency we have the alternative definition of intensity,

$$I_\lambda\, d\lambda = I_\nu\, d\nu,$$

or

$$I_\lambda = \frac{\nu^2}{c} I_\nu, \tag{2.2}$$

where c is the velocity of light.

The component of *flux* in the $\mathbf{d}$ direction, $F_{\nu,d}(P)$, is defined as the total energy flowing across unit area perpendicular to $\mathbf{d}$, per unit frequency interval. An infinitesimal area $d\pi_d$ has a projected area in the $\mathbf{s}$

[1] Much of the early work on radiative transfer is to be found in the astrophysical literature. SI nomenclature differs from astrophysical nomenclature and is more suited to situations involving finite sources. Equivalents are *specific intensity* (this book) and *radiance* (SI) and *flux* (this book) and *exitance* (SI). This book does not use an equivalent to *radiant intensity* (SI) and the SI system does not use an equivalent to *irradiance* (this book).

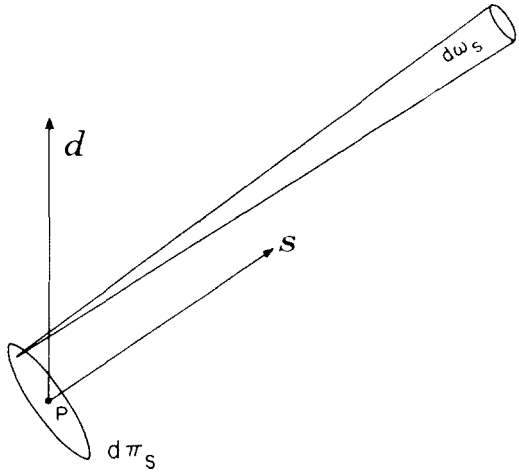


FIG. 2.1. Specific intensity of radiation.

direction,

$$d\pi_s = d\pi_d \cos(\mathbf{d}, \mathbf{s}),$$

where $(\mathbf{d}, \mathbf{s})$ denotes the angle between the two vectors. The energy flux across $d\pi_d$, integrated over all $\mathbf{s}$ directions is, from the definitions of $I_\nu(P, \mathbf{s})$ and $F_{\nu,d}(P)$,

$$F_{\nu,d}(P)\, d\nu\, d\pi_d = \int_{\omega_s} I_\nu(P, \mathbf{s}) \cos(\mathbf{d}, \mathbf{s})\, d\nu\, d\pi_d\, d\omega_s,$$

or

$$F_{\nu,d}(P) = \int_{\omega_s} I_\nu(P, \mathbf{s}) \cos(\mathbf{d}, \mathbf{s})\, d\omega_s, \tag{2.3}$$

where the integral extends over all solid angles.

If $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ are three orthogonal unit vectors, we have the trigonometric identity

$$\cos(\mathbf{d}, \mathbf{s}) \equiv \cos(\mathbf{d}, \mathbf{x}) \cos(\mathbf{s}, \mathbf{x}) + \cos(\mathbf{d}, \mathbf{y}) \cos(\mathbf{s}, \mathbf{y}) + \cos(\mathbf{d}, \mathbf{z}) \cos(\mathbf{s}, \mathbf{z}).$$

Substituting in (2.3),

$$F_{\nu,d}(P) = F_{\nu,x}(P) \cos(\mathbf{d}, \mathbf{x}) + F_{\nu,y}(P) \cos(\mathbf{d}, \mathbf{y}) + F_{\nu,z}(P) \cos(\mathbf{d}, \mathbf{z}). \tag{2.4}$$

Equation (2.4) is the transformation law for the components of the vector

$$\mathbf{F}_\nu(P) = \mathbf{x} F_{\nu,x}(P) + \mathbf{y} F_{\nu,y}(P) + \mathbf{z} F_{\nu,z}(P). \tag{2.5}$$

The sun's disk subtends, on the average, an angle of 32′ at the earth's surface; for most practical purposes sunlight can be regarded as a parallel beam of radiation. The above definition of intensity is unnecessarily general for this case. Let us suppose that the sun's direction is that of the vector $\mathbf{s}_\odot$, and let its disk subtend a solid angle $d\omega_\odot$ at the earth. For unmodified solar radiation, $I_\nu(P, \mathbf{s})$ is only nonzero if $\mathbf{s}$ is very close to $\mathbf{s}_\odot$. In most circumstances, therefore, we may replace the angle $(\mathbf{d}, \mathbf{s})$ by $(\mathbf{d}, \mathbf{s}_\odot)$. From (2.3),

$$\begin{aligned} F_{\nu,d}(P) &\simeq \int_{\omega_s} I_\nu(P, \mathbf{s}) \cos(\mathbf{d}, \mathbf{s}_\odot)\, d\omega_s, \\ &= \cos(\mathbf{d}, \mathbf{s}_\odot) \bar{I}_\nu(P)\, d\omega_\odot, \end{aligned} \tag{2.6}$$

where $\bar{I}_\nu$ is the mean value of the intensity, averaged over the sun's disk.

If we write

$$F_{\nu,d}(P) = f_\nu \cos(\mathbf{d}, \mathbf{s}_\odot),$$

then

$$f_\nu = \bar{I}_\nu\, d\omega_\odot \tag{2.7}$$

is the solar *irradiance,* a positive quantity that is a function of solar distance only.

We can now evaluate the *energy density* (u_ν) of a radiation field. Consider a cylinder, parallel to the $\mathbf{s}$ direction, of length ds and cross section $d\pi_s$ (Fig. 2.2). A photon traveling in the $\mathbf{s}$ direction will spend a time $dt = ds/c$ in the cylinder, where c is the velocity of light. The total amount of energy in the cylinder made up of photons traveling in the $\mathbf{s}$ direction within the solid angle $d\omega_s$ is that which crosses $d\pi_s$ in time dt:

$$I_\nu(P, \mathbf{s})\, d\pi_s\, d\nu\, d\omega_s\, dt = \frac{I_\nu(P, \mathbf{s})}{c} d\pi_s\, d\nu\, d\omega_s\, ds.$$

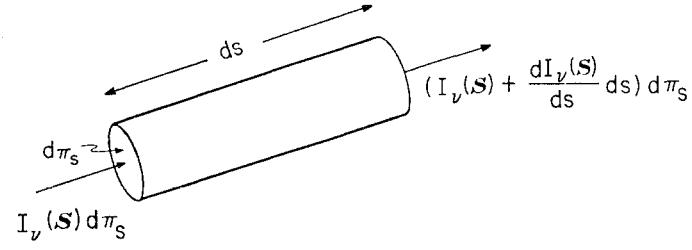


FIG. 2.2. Radiative heating.

The volume of the cylinder is $ds\, d\pi_s$ and these quanta contribute to the energy density an amount

$$\frac{I_\nu(P, \mathbf{s})}{c} d\nu\, d\omega_s.$$

Integrating over all directions we find for the energy density in the frequency range ν to $\nu + d\nu$:

$$u_\nu(P)\, d\nu = \frac{d\nu}{c} \int I_\nu(P, \mathbf{s})\, d\omega_s.$$

Hence,

$$u_\nu(P) = \frac{4\pi}{c} \bar{I}_\nu(P), \tag{2.8}$$

where

$$\bar{I}_\nu(P) = \frac{1}{4\pi} \int I_\nu(P, \mathbf{s})\, d\omega_s \tag{2.9}$$

is the *mean intensity* of the radiation field.

The divergence of the energy flux equals the rate at which energy is *added* to the field per unit volume, i.e., the rate at which energy is *lost* by the matter. Let h_ν be the rate per unit volume at which heat is gained by matter from radiation in unit frequency range

$$h_\nu = -\nabla \cdot \mathbf{F}_\nu. \tag{2.10}$$

Expanding the right-hand side of (2.10) and using the definition of flux (2.3), we find

$$\begin{aligned} \nabla \cdot \mathbf{F}_\nu &= \frac{\partial F_{\nu,x}}{\partial x} + \frac{\partial F_{\nu,y}}{\partial y} + \frac{\partial F_{\nu,z}}{\partial z}, \\ &= \int d\omega_s \left[\frac{\partial I_\nu}{\partial x} \cos(\mathbf{s}, \mathbf{x}) + \frac{\partial I_\nu}{\partial y} \cos(\mathbf{s}, \mathbf{y}) + \frac{\partial I_\nu}{\partial z} \cos(\mathbf{s}, \mathbf{z}) \right] \\ &= \int d\omega_s \frac{dI_\nu}{ds}. \end{aligned} \tag{2.11}$$

The meaning of (2.11) is made clear by Fig. 2.2. The energy *lost* by matter in the small cylinder is $[dI_\nu(\mathbf{s})/ds]\, ds\, d\pi_s\, d\omega_s$ per second per unit frequency range. Since $ds\, d\pi_s$ is the volume of the cylinder, the heat loss per unit volume from radiation traveling in the $\mathbf{s}$ direction is $[dI_\nu(\mathbf{s})/ds]\, d\omega_s$, and (2.11) follows by integrating over all ω_s.

In later sections of this chapter the operator d/ds occurring in (2.11) will be written in different forms that we will state now so that they may be conveniently compared:

$$\frac{d}{ds} \equiv \cos(\mathbf{s}, \mathbf{x}) \frac{\partial}{\partial x} + \cos(\mathbf{s}, \mathbf{y}) \frac{\partial}{\partial y} + \cos(\mathbf{s}, \mathbf{z}) \frac{\partial}{\partial z}$$

$$\equiv \xi_x \frac{\partial}{\partial x} + \xi_y \frac{\partial}{\partial y} + \xi_z \frac{\partial}{\partial z}$$

(where $\xi_{x,y,z}$ is the cosine of the angle between the direction of the intensity and the **x**, **y**, and **z** directions);

$$\equiv \xi_i \frac{\partial}{\partial x_i} \qquad \text{(sum rule for repeated indices)}$$

$$\equiv \mathbf{s} \cdot \nabla. \tag{2.12}$$

In later chapters we will frequently make use of the concept of *radiative equilibrium,* whereby there is no net energy exchange between matter and the radiation field,

$$h = \int_0^\infty h_\nu \, d\nu = 0. \tag{2.13}$$

A special case of radiative equilibrium is *monochromatic radiative equilibrium* where

$$h_\nu = 0. \tag{2.14}$$

In defining radiative equilibrium, and in other problems concerning atmospheric thermodynamics, the relevant quantities are those that are integrated over the entire spectrum. This operation will be indicated by omitting the suffix ν from frequency-dependent quantities. Thus,

$$I = \int_0^\infty I_\nu \, d\nu,$$

$$F = \int_0^\infty F_\nu \, d\nu,$$

$$u = \int_0^\infty u_\nu \, d\nu, \text{ etc.}$$

There are circumstances in which limited frequency ranges have particular physical significance, e.g., the frequency range embracing an

isolated absorption band. Integrals over such ranges will be designated by a suffix other than ν. Thus,

$$I_i = \int_i I_\nu \, d\nu,$$

$$I = \sum_i I_i, \text{ etc.}$$

2.1.2. *Extinction and emission*

We now require formal definitions for the interactions between matter and a radiation field. For most atmospheric problems a name can be assigned to a process without difficulty, and careful definitions may seem pedantic. However, if we wish to examine the whole range of atmospheric phenomena, without having to modify definitions, it pays to examine the situation closely.

The totality of interactions between radiation and matter will be classed as either *extinction* or *emission*. The two processes are distinguished by the sign of the change of radiant intensity as a result of the interaction. If the intensity *decreases* then we have *extinction*; if the intensity *increases* we have *emission*. No interaction at all can be pictured as the simultaneous extinction and emission of identical quanta, or, alternatively, as a case of vanishingly small interaction coefficients. This prosaic distinction between extinction and emission is the only one that applies to all phenomena described by the two terms.

The fundamental law of extinction is that of *Lambert*.[2] It states that the extinction process is linear, independently in the intensity of radiation and in the amount of matter, provided that the physical state (i.e., temperature, pressure, composition) is held constant. The possible processes nonlinear in the light intensity have been fully explored. The scattering of light by light is a tractable theoretical problem and the nonlinear scattering from bound electrons can be observed in the laboratory with the aid of coherent light amplifiers. The photon densities required to exhibit nonlinear effects are greatly in excess of those occurring in planetary atmospheres, and deviations from Lambert's law on this account are completely negligible. On the other hand, the optical properties of individual molecules are strongly influenced by the proximity of other molecules; the problem of pressure broadening of spectral lines, which will be discussed in Chapter 3, is one example. However, if we postulate that the matter is always in the same physical state, then the intermolecular forces are fixed. Under this condition, only a linear dependence of extinction on amount of matter is possible.

[2] The name of *Bouguet* is more commonly used in European literature.

From Lambert's law, the change of intensity along a path ds is proportional to the amount of matter in the path,

$$dI_\nu \text{ (extinction)} = -e_{\nu,\nu} I_\nu \, ds, \tag{2.15}$$

where $e_{\nu,\nu}$ is the *volume extinction coefficient.* To satisfy Lambert's law, $e_{\nu,\nu}$ must be proportional to the local density of absorbing matter.

The argument that the extinction process is linear in the amount of matter applies with equal force to the emission process. As a formal statement, we may write

$$dI_\nu \text{ (emission)} = e_{\nu,\nu} J_\nu \, ds, \tag{2.16}$$

defining the *source function* J_ν, the form of which will be the subject of later discussion.

The dependence of $e_{\nu,\nu}$ on the density of matter gives rise to a number of different expressions: $e_n = e_\nu / n$, where n is the number density of absorbing molecules, defines the *molecular extinction coefficient*; $e_m = e_\nu / \rho$, where ρ is the density of absorbing matter, gives the *mass extinction coefficient*; $e_s = e_\nu n / n_s$, where n_s is Loschmidt's number, yields the *extinction coefficient per cm s.t.p.*, alternatively stated, per *cm-Amagat.* The interrelations between these coefficients are discussed in Appendix 2.

Our earlier statement that all interactions can be classed as extinction or emission can now be summed up in the statement that any change in intensity resulting from the interaction of matter and radiation must be the sum of (2.15) and (2.16),

$$\begin{aligned} dI_\nu(P, \mathbf{s}) &= dI_\nu \text{ (extinction)} + dI_\nu \text{ (emission)} \\ &= -e_{\nu,\nu} \, ds [I_\nu(P, \mathbf{s}) - J_\nu(P, \mathbf{s})], \end{aligned}$$

and hence

$$-\frac{1}{e_{\nu,\nu}} \frac{dI_\nu(P, \mathbf{s})}{ds} = I_\nu(P, \mathbf{s}) - J_\nu(P, \mathbf{s}). \tag{2.17}$$

Equation (2.17) is known as the *equation of transfer,* and was first given in this form by Schwarzschild. While it sets the pattern of the formalism used in transfer problems, its physical content is very slight. The physics is mainly contained in the definitions of the extinction coefficient and the source function. Equation (2.17) can be written in any of the other notations (2.12); in particular the vector notation is

$$-\frac{1}{e_{\nu,\nu}} \mathbf{s} \cdot \nabla I_\nu(P, \mathbf{s}) = I_\nu(P, \mathbf{s}) - J_\nu(P, \mathbf{s}),$$

where the del operator operates on the position vector of P.

Combining (2.10), (2.11), and (2.17), we can now express the heating rate in terms of the source function and extinction coefficient,

$$h_\nu = -\int d\omega_s \frac{dI_\nu}{ds} = 4\pi e_{\nu,\nu}(\bar{I}_\nu - \bar{J}_\nu), \tag{2.18}$$

where

$$(\bar{I}_\nu, \bar{J}_\nu) = \int \frac{d\omega_s}{4\pi}(I_\nu, J_\nu). \tag{2.19}$$

Equation (2.18) expresses the heating rate as the difference between the absorption from the mean radiation field and the mean emission. The condition for monochromatic radiative equilibrium (2.14) is now

$$\bar{I}_\nu = \bar{J}_\nu. \tag{2.20}$$

We will find it convenient for some purposes to use only quantities with the dimensions of intensity. The *heating function* H_i has such dimensions, but is defined for a finite frequency interval only,

$$H_i = \frac{\int_i h_\nu \, d\nu}{2\pi \int_i e_{\nu,\nu} \, d\nu}. \tag{2.21}$$

Further subdivision of interactions between matter and the radiation field depends on the physical process involved, and particularly on changes of internal energy of the matter. If radiation interacts with matter whose only mode of internal energy is translational, then the interaction coefficients are very small in all circumstances discussed in this book. For example, if a 0.1 μm photon interacts with a free electron (Compton scattering), only 5×10^{-5} of its energy is transformed into kinetic energy, and this is the most efficient conversion of any relevance to the atmosphere. If there is no change of any form of internal energy the interaction is a *simple scattering process*. There is a close approach to *simple scattering* when the matter has only narrow (quantized) states of internal energy and the incident photon has a frequency far from that of a possible transition.

All matter with which we will be concerned has electronic, vibrational, and rotational internal energy, and some small interaction at least will take place between these energy states and incident radiation. The process may still be classified as *scattering*, however. An interacting photon will cause a transition to a higher, excited state. The excited state has a limited lifetime, and if the absorbed photon is reemitted with

negligible conversion to translational energy, the process is one of *scattering*. If the transition to the lower state takes place in one step, then the emitted photons will have frequencies identical to those of the absorbed photons, and we refer to a process of *coherent scattering*. The molecule may, however, revert to the ground state by a cascade process through intermediate levels. The emitted photons now differ in frequency from those incident, but the total energies of the extinction and emission processes are the same; this is *incoherent scattering*.

It often happens, however, that before the matter can reemit, molecular collisions occur, during which nonradiating transitions (*deactivation*) can take place. The energy then ends up in other forms of internal energy. In the case of complete *thermodynamic equilibrium* the energy is shared equally among all the accessible degrees of freedom. Where energy is transferred to kinetic energy, the process is called *absorption*; the reverse process will be called *thermal emission*. Thermal emission and scattering are not mutually exclusive, and frequently occur simultaneously. We shall see that circumstances occur when there is essentially no distinction, other than semantic, between them. Since all processes are linear, the extinction coefficient can be expressed as the sum of an *absorption coefficient* (k_ν) and a *scattering coefficient* (s_ν):

$$e_\nu = k_\nu + s_\nu. \tag{2.22}$$

The most general problem in atmospheric radiation therefore has a source function consisting of two parts,

$$e_\nu J_\nu = k_\nu J_\nu \text{ (thermal)} + s_\nu J_\nu \text{ (scattering)}. \tag{2.23}$$

One physical observation can usefully be introduced at this stage; namely, that the atmosphere is effectively *isotropic*. Exceptions exist, such as raindrops falling under gravity and the ionospheric propagation of radio waves in the earth's magnetic field, but these are of a relatively obvious character. Assuming isotropy, and accepting the random nature of molecular agitation, there is no sufficient reason for absorption and thermal emission to be anything but isotropic.

Simple scattering, on the other hand, involves a very direct connection between incident and emitted radiation. Since light has vector properties (see § 2.1.3), isotropy is no longer expected, and is not observed for scattering either by molecules or small particles. The more complex scattering process, involving transitions between quantized states, may or may not preserve some memory of the vector properties of the incident photon. Thus, although it must be with circumspection, it is possible to think of absorption and scattering processes as differing in the symmetry of the source function, although this is neither as fundamental nor as useful a distinction as that adopted in this section.

If we consider carefully the meaning of the word "heating", it is synonymous with an increase in the translational energy of matter. Our definitions of scattering and radiative equilibrium therefore correspond, and there is, for example, no formal mathematical distinction between absorption in radiative equilibrium and isotropic scattering.

2.1.3. *Simple scattering*

In the previous paragraph the vector properties of electromagnetic radiation were first mentioned. We now discuss how to write the equation of transfer in a suitable matrix form.

According to the Maxwell equations, the most general propagating electromagnetic wave consists of vibrations of the electric vector in the plane transverse to the direction of propagation. To specify the phase and polarization of the radiation field requires a pair of complex numbers. However, in the formulation of the equation of radiative transfer, it is more convenient to use an equivalent representation employing four real quantities known as *Stokes parameters,* each with the physical dimensions of an intensity. We will denote these by $I_\nu^{(i)}$ $(i = 1, 2, 3, 4)$.

The most general polarization from a single source is elliptical. The polarization ellipse can be defined in terms of the intensities of two components polarized at right angles to each other, and the direction in space of the major or minor axis of the ellipse. Let **l** and **r** be two unit vectors forming an orthogonal set with **s**, the direction of propagation. Let $I_\nu^{(l)}$ and $I_\nu^{(r)}$ be the intensities of the two polarized components of the beam. The total intensity is

$$I_\nu = I_\nu^{(l)} + I_\nu^{(r)}, \tag{2.24}$$

and both component intensities are determined if we also know

$$Q_\nu = I_\nu^{(l)} - I_\nu^{(r)}. \tag{2.25}$$

Let $\tan\beta$ equal the ratio of the axes of the polarization ellipse, and let χ be the angle between **l** and the major axis. We define

$$U_\nu = Q_\nu \tan 2\chi, \tag{2.26}$$

and

$$V_\nu = Q_\nu \tan 2\beta \sec 2\chi. \tag{2.27}$$

We can now make a variety of choices for the Stokes parameters: $(I_\nu, Q_\nu, U_\nu, V_\nu)$ is one possibility; $(I_\nu^{(l)}, I_\nu^{(r)}, U_\nu, V_\nu)$ is another. Four parameters are more than are necessary to specify an elliptically

polarized beam, for which the following relation exists,

$$I_\nu^2 = Q_\nu^2 + U_\nu^2 + V_\nu^2. \tag{2.28}$$

A more general field may contain many independent, polarized components. If, as in nature, these possess no systematic phase relations, then the Stokes parameters are additive, but (2.28) no longer holds. Then all four parameters are required to define the field. *Natural* (or unpolarized) *light* has $I_\nu^{(l)} = I_\nu^{(r)}$, and therefore $Q_\nu = U_\nu = V_\nu = 0$.

If we now consider the equation of transfer (2.17), we need to make no modifications except to substitute $I_\nu^{(i)}$ and $J_\nu^{(i)}$ in place of I_ν and J_ν. In particular, we do not need to modify our definition of the extinction coefficient, which is the same for all Stokes parameters if the medium is isotropic. The problem is to write the source function in a form that can be related to the *phase matrix* P_{ij}, which in turn can be calculated from electromagnetic theory.

When an incident bundle of radiation, characterized by the fluxes $I_\nu^{(i)}(\mathbf{s})\, d\omega_s$ $(i = 1, 2, 3, 4)$, is scattered by an infinitesimally small quantity of matter, we observe a new radiation field characterized by the vector $dI_\nu^{(j)}(\mathbf{d})$ (emission) $(j = 1, 2, 3, 4)$. According to Lambert's law, these quantities must be linearly related. Stating this relationship in its most general form

$$dI_\nu^{(j)}(\mathbf{d}) \text{ (emission)} = W_{ij}(\mathbf{s}, \mathbf{d}) I_\nu^{(i)}(\mathbf{s})\, d\omega_s, \tag{2.29}$$

where, following the sum rule for repeated indices, the right-hand side is summed over all i. $W_{ij}(\mathbf{s}, \mathbf{d})$ is an intensity transformation matrix, whose form will be discussed in Chapter 7, and which, according to Lambert's law, must be proportional to the amount of matter. As a matter of definition, which is obviously pertinent, we may assume W_{ij} to be proportional to $s_{\nu,\nu}\, ds/4\pi$, where $s_{\nu,\nu}$ is the volume scattering coefficient. Thus we can write

$$W_{ij}(\mathbf{s}, \mathbf{d}) = s_{\nu,\nu} \frac{ds}{4\pi} P_{ij}(\mathbf{s}, \mathbf{d}), \tag{2.30}$$

where the constant of proportionality P_{ij} is the phase matrix.

Substituting (2.30) in (2.29), and integrating over all angles of incidence, we find

$$dI_\nu^{(j)}(\mathbf{d}) \text{ (emission)} = s_{\nu,\nu}\, ds \int_{\omega_s} \frac{d\omega_s}{4\pi} P_{ij}(\mathbf{s}, \mathbf{d}) I_\nu^{(i)}(\mathbf{s}), \tag{2.31}$$

and hence, from (2.16), the source function is

$$J_\nu^{(j)}(\mathbf{d}) = \frac{s_{\nu,\nu}}{e_{\nu,\nu}} \int_{\omega_s} P_{ij}(\mathbf{s}, \mathbf{d}) I_\nu^{(i)}(\mathbf{s}) \frac{d\omega_s}{4\pi}. \tag{2.32}$$

The components of the phase matrix are related to each other by Maxwell's equations, but one component must be normalized if energy is to be conserved in a strictly scattering or *conservative process*. This normalization may be expressed by

$$\int_{\omega_s} P(\mathbf{s}, \mathbf{d}) \frac{d\omega_s}{4\pi} = \int_{\omega_d} P(\mathbf{s}, \mathbf{d}) \frac{d\omega_d}{4\pi} = 1, \tag{2.33}$$

where $P = P_{11}$ is the matrix component for transformation of intensity to intensity. For most purposes in this book, we need only consider this one component.

A special case, for which there is an extensive literature, is that of *isotropic scattering*, $P = 1$. Atmospheric aerosols do not scatter isotopically, but for molecules and very small particles this can be a useful approximation.

One consequence of (2.33) is that there is no formal distinction between calculations for conservative processes on the one hand and for radiative equilibrium on the other. For conservative scattering (2.32) reads

$$J_\nu(\mathbf{d}) = \int_{\omega_s} P(\mathbf{s}, \mathbf{d}) I_\nu(\mathbf{s}) \frac{d\omega_s}{4\pi}.$$

From (2.33),

$$\begin{aligned} \bar{J}_\nu &= \int_{\omega_d} J_\nu(\mathbf{d}) \frac{d\omega_d}{4\pi} \\ &= \int_{\omega_s} I_\nu(\mathbf{s}) \frac{d\omega_s}{4\pi} \int_{\omega_d} P(\mathbf{s}, \mathbf{d}) \frac{d\omega_d}{4\pi} \\ &= \int_{\omega_s} I_\nu(\mathbf{s}) \frac{d\omega_s}{4\pi} = \bar{I}_\nu, \end{aligned} \tag{2.34}$$

which, from (2.20), is the condition for radiative equilibrium.

2.2. Thermal emission

2.2.1. *Thermodynamic equilibrium*

Thermodynamic equilibrium describes the state of matter and radiation inside a constant-temperature enclosure. The equilibrium is complete for, according to the second law of thermodynamics, no thermal changes can be produced by any mechanism without the intervention of external work. The implications are far reaching and were first described by Kirchhoff in 1882. Radiation inside the enclosure is known as *black-body*

or *enclosed* radiation. The latter is the better name, but the former is more commonly used.

Kirchhoff's deductions were as follows.

(*a*) In each separate, homogeneous medium within the enclosure, the radiation is *homogeneous, unpolarized,* and *isotropic.*

(*b*) The source function is equal to the intensity,

$$I_\nu = J_\nu.$$

(*c*) $c'^2 I_\nu$ is the same in all media in the enclosure, where c' is the velocity of light in the medium concerned.

(*d*) As a direct consequence of (*c*), $c'^2 I_\nu$ must be a universal function of temperature (θ) only, and will be written $\mathbf{c}^2 B_\nu$ where $\mathbf{c}$ is the velocity of light *in vacuo.*

A further consequence of the existence of strict thermodynamic equilibrium is that B_ν must have the form

$$B_\nu = \frac{\nu^3}{\mathbf{c}^2} F\left(\frac{\theta}{\nu}\right), \tag{2.35}$$

where F is an unknown function. The proof of this statement is to be found in standard textbooks on thermodynamics. Differentiating (2.35) with respect to ν, we can show that the turning points of the function B_ν are determined by the ratio θ/ν, or the product $\theta\lambda$, only. This is *Wien's displacement law.*

Integrating (2.35) over all frequencies and making the substitution $x = \theta/\nu$, there results

$$\int_0^\infty B_\nu \, d\nu = \int_0^\infty \frac{\theta^4}{\mathbf{c}^2 x^5} F(x) \, dx = \frac{\boldsymbol{\sigma}}{\pi} \theta^4. \tag{2.36}$$

This is the *Stefan–Boltzmann radiation law,* and $\boldsymbol{\sigma}$ is the *Stefan–Boltzmann constant.*

Investigation of the function F was the historical reason for the development of quantum theory. Planck's theory leads to the expression

$$B_\nu(\theta) = \frac{2\mathbf{h}\nu^3}{\mathbf{c}^2(e^{\mathbf{h}\nu/\mathbf{k}\theta} - 1)}, \tag{2.37}$$

where $\mathbf{h}$ = Planck's constant, and

$\mathbf{k}$ = Boltzmann's constant.

Using equation (2.2), we may write alternatively,

$$B_\lambda = \frac{\nu^2}{\mathbf{c}} B_\nu = \frac{C_1 \lambda^{-5}}{\pi(e^{C_2/\lambda\theta} - 1)}, \tag{2.38}$$

where

$$C_1 = 2\pi\mathbf{hc}^2 \tag{2.39}$$

and

$$C_2 = \mathbf{hc/k} \tag{2.40}$$

are known as the *first* and *second radiation constants,* respectively.

Both forms of the Planck function (2.37) and (2.38) are shown in Fig. 2.3. Tabulated values are given in Appendix 5. B_ν has a single maximum at

$$\frac{\theta \mathbf{c}}{\nu} = 0.50995 \text{ cm K},$$

while the maximum of B_λ is at

$$\theta\lambda = 0.28978 \text{ cm K}.$$

The curve with ordinate λB_λ in Fig. 2.3 illustrates an important practical point in atmospheric computations. A black body at 6000 K has

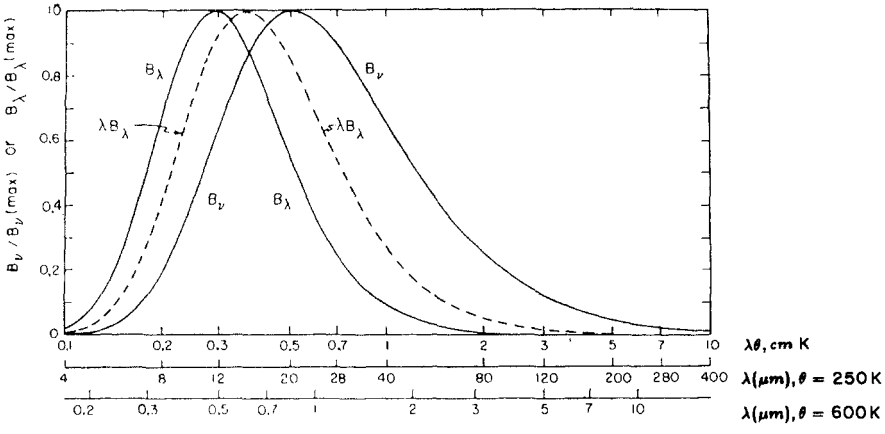


FIG. 2.3. The Planck function. The curve with ordinate λB_λ gives areas proportional to energies.

only 0.4% of its energy at wavelengths longer than 5μm; a 250 K black body, on the other hand, has only 0.4% of its energy at wavelengths shorter than 5 μm. For most practical purposes, therefore, the solar and terrestrial radiation streams can be treated independently.

The Planck function behaves very differently in its two wings. As $\lambda \to \infty$ or $\nu \to 0$,

$$B_\nu \to 2\mathbf{k}\theta\nu^2/\mathbf{c}^2, \tag{2.41}$$

$$B_\lambda \to 2\mathbf{k}\theta\mathbf{c}/\lambda^4. \tag{2.42}$$

(2.41) and (2.42) are known as the *Rayleigh–Jeans distribution.* As $\lambda \to 0$ or $\nu \to \infty$,

$$B_\nu \to \frac{2\mathbf{h}\nu^3}{\mathbf{c}^2} e^{-\mathbf{h}\nu/\mathbf{k}\theta}, \tag{2.43}$$

$$B_\lambda \to \frac{2\mathbf{h}\mathbf{c}^2}{\lambda^5} e^{-\mathbf{h}\mathbf{c}/\lambda\mathbf{k}\theta}, \tag{2.44}$$

which is the *Wien distribution.*

2.2.2. *Breakdown of thermodynamic equilibrium*

In strict thermodynamic equilibrium, the source function depends only upon temperature, frequency, and the velocity of light. This so greatly simplifies the problem of radiative transfer for thermal radiation that it is important to understand the conditions under which Planck's source function may be an adequate approximation. Before Milne's analysis of the problem in 1930, it was usually assumed that the Planck function was correct for thermal radiation, for reasons that appear to have been based on the following argument. Consider a small element of matter inside a constant-temperature enclosure, where it absorbs and emits according to Planck's and Kirchhoff's laws. If we extract this element from the enclosure without altering its physical state, the only change to which it can react is to the incident radiation, which is not the same outside as inside the enclosure. Since it is reasonable to suppose that emission is a property of matter alone, the source function should not be affected by this change, and should continue to be the Planck function.

This argument is fallacious, however, because, as first pointed out by Einstein, emission is also influenced by the incident radiation field (*induced emission*). The important practical question is the extent to which the absorption coefficient and the source function can be changed from their equilibrium values by the action of the incident radiation.

Since the discussion now involves disequilibrium states, we can no longer use thermodynamic arguments but, instead, we must employ a

microscopic statistical model. In this section, we shall follow the consequences of this model to the point of understanding the conditions under which Kirchhoff's and Planck's laws fail. A more difficult problem is to derive an accurate source function under these conditions. That question will be considered in § 6.3.3.

Thermal equilibrium can be defined in terms of Boltzmann's law for the distribution of molecules between two states (1 and 2),

$$\frac{n(1)}{n(2)} = \frac{g_1}{g_2} \exp\left(-\frac{E_1 - E_2}{\mathbf{k}\theta}\right)$$
$$= \frac{g_1}{g_2} \exp\left(\frac{-\mathbf{h}\nu_{1,2}}{\mathbf{k}\theta}\right), \tag{2.45}$$

where the n's are state populations, the g's are statistical weights, and the E's are energies; $\nu_{1,2}$ is, by Planck's quantum relationship, the frequency of a photon absorbed or emitted during a radiative transition between the two states, and $\mathbf{h}$ and $\mathbf{k}$ are Planck's and Boltzmann's constants. As applied to translational modes, (2.45) leads to *Maxwell's distribution* of molecular velocities. For complete equilibrium, (2.45) is obeyed for all energy states, regardless of their nature and throughout the medium.

Complete equilibrium is an idealization, appropriate to the inside of a constant-temperature enclosure in which nothing changes and to which we have no access. We wish to treat systems not in equilibrium but to which Boltzmann's law applies in a limited sense, namely, those for which (2.45) applies to certain groups of levels, but not to all levels. We say that the relevant levels are in a state of *local thermodynamic equilibrium* (LTE) at a given temperature. To give a concrete example, consider resonant fluorescence of sodium vapor at room temperature. The translational energies of the sodium atoms will be close to local thermodynamic equilibrium, while the glowing gas emits as if it were at a temperature of thousands of kelvin.

Einstein demonstrated that Planck's source function results if (2.45) is obeyed for the levels under consideration. We may, therefore, regard Planck's and Boltzmann's laws as interchangeable; conditions leading to one lead to the other, and vice versa. We shall go further and demonstrate that a one-to-one relationship exists between the source function and the state populations and, therefore, that our task is simply to calculate the state populations. Now, it is known that collisions acting alone will bring about a Boltzmann distribution and, consequently, a Planck source function. But, in a collisionless medium, radiation can bring about almost any population of energy levels and Planck's law need not be obeyed.

The population of energy levels and the resulting source function will be the result of a conflict between collisional and radiative effects. The

rate of collisional adjustment of state populations is determined by a relaxation time, η, that is inversely proportional to the collisional rate (or proportional to the pressure). Radiative adjustment is determined by the natural lifetime of the excited states with respect to radiative transitions, ϕ, a molecular constant, independent of the state of the medium. When $\eta/\phi \ll 1$, LTE should occur and Planck's law will be valid; when $\eta/\phi \gg 1$, on the other hand, a different source function will probably be required. Because of the rapid variation of pressure with height and the dependence of η on the pressure, there will be a fairly sharply defined *relaxation level* in the atmosphere below which Planck's law is valid for transitions between a particular pair of energy levels but above which another source function will be required.

We shall, primarily, be concerned with interactions between vibrational, rotational, and translational levels. If all these levels were in disequilibrium it would be difficult, if not impossible, to derive a useful source function. Fortunately, we may treat the case of disequilibrium among vibrational states while the translational and rotational levels are in local thermodynamic equilibrium. We shall justify this statement for the rotational levels in § 2.2.4, but we may discuss the translational states now.

The temperature defined by (2.45), if applied to the populations of the translational states, is the *kinetic temperature.* Every collision makes some adjustment to the translational energy states and, from this, we shall show that there is a unique kinetic temperature wherever there is a fluid atmosphere. Consider levels in the thermosphere where diffusion is the most important heat transfer process. Disturbance of the equilibrium among translational states will be caused by the arrival of molecules from levels with differing kinetic temperatures. A relevant dimensionless parameter is,

$$A = \frac{l}{\theta}\frac{d\theta}{dz}, \tag{2.46}$$

where l is the molecular mean free path and z is the altitude. If $A \ll 1$ we have, in effect, an unlimited isothermal collisional region and we expect to obtain Maxwell's distribution. If $A \gg 1$, on the other hand, a unique Maxwell distribution is impossible. A gradient of $1\,\mathrm{K\,km^{-1}}$ at a temperature of 1000 K leads to $A \simeq l \times 10^{-8}$. If l approaches one scale height (about 50 km at these levels) molecules may leave the atmosphere without performing collisions on the way. This condition defines the lower boundary to the *exosphere,* which we may conveniently take to be the upper limit to the fluid atmosphere. At this level $A \simeq 0.05$ and equilibrium may be expected between the translational modes. The kinetic temperature and the Maxwell distribution are therefore valuable concepts at all levels considered in this book.

The model developed in the following sections treats rotation and vibration as independent forms of energy. As will be discussed in Chapter 3, this is not strictly correct for polyatomic molecules, but the question is unimportant in the present context.

2.2.3. *The interaction between matter and radiation*

An upper and a lower vibrational level (u and l) are connected by collisional and radiative transitions (Fig. 2.4). Collisional transitions (b coefficients in Fig. 2.4) involve a radiating and a colliding molecule and can raise or lower the vibrational energy. Similarly, absorption and induced emission (the C coefficients) involve a radiating molecule and a photon and cause transitions in both directions. Radiative transitions can also take place spontaneously, without the presence of a photon or a colliding molecule (A coefficients), but only in a downward direction. For completeness, we include in Fig. 2.4 transitions involving other levels of the same molecule or of different molecules (designated by k).

Radiative transitions are accompanied by the appearance or disappearance of a photon. We assume that we may distinguish photons associated with $u \leftrightarrow l$ transitions from all others. This may not always be an easy thing to do because ground-state and upper-state transitions (Chapter 3) may be very close. Given unlimited spectral resolution, the distinction can always be made, however, and we shall assume that there is no ambiguity.

Each vibrational state has a rotational–translational fine structure that gives rise to the characteristic appearance of a vibration–rotation

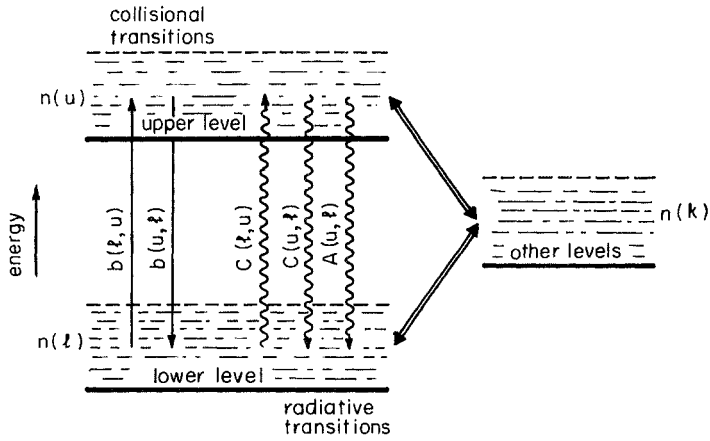


FIG. 2.4. Transitions involving two vibrational levels. $n(u)$, $n(l)$, and $n(k)$ are number densities. $A(u, l)$, $C(u, l)$, and $C(l, u)$ are Einstein coefficients. $b(u, l)$ and $b(l, u)$ are collisional transition coefficients. The heavy horizontal lines are the vibrational energy levels; the shaded areas show, schematically, the rotational fine structure.

band (Chapter 3). According to our assumptions, the states forming this fine structure are in thermodynamic equilibrium and their relative populations may be incorporated into a fixed structure factor, $f(\nu - \nu_0)$, where ν is the frequency of the absorbed or emitted photon, and ν_0 is the frequency of the vibrational transition alone. f is normalized,

$$\int_{-\infty}^{+\infty} f(\nu - \nu_0)\, d(\nu - \nu_0) = 1. \tag{2.47}$$

(i) The Einstein relations. Consider an element of matter with state populations $n(u)$ and $n(l)$, interacting with a radiation field containing $u_\nu\, d\nu$ photons per unit volume with frequencies between ν and $\nu + d\nu$. For the remainder of this section we shall express all radiation parameters in terms of the energy unit $\mathbf{h}\nu$, so that all previous expressions for I_ν, B_ν, and u_ν must be divided by $\mathbf{h}\nu$.

The radiative transitions $(u \leftrightarrow l)$ per unit volume and per unit frequency range can now be specified in terms of the structure factors and three Einstein coefficients.

Rate of spontaneous emission

$$= n(u)A(u, l)f_s(\nu - \nu_0); \tag{2.48}$$

rate of induced emission

$$= n(u)u_\nu C(u, l)f_i(\nu - \nu_0); \tag{2.49}$$

and rate of absorption

$$= n(l)u_\nu C(l, u)f_a(\nu - \nu_0). \tag{2.50}$$

We may eliminate two out of the three coefficients by requiring that, when Boltzmann's law is obeyed, the steady-state solution for u_ν is $4\pi B_\nu/\mathbf{c}$;[3] see (2.8) and (2.37), divided by $\mathbf{h}\nu$.

Boltzmann's law for the vibrational populations is from (2.45)

$$\frac{\bar{n}(u)}{\bar{n}(l)} = \frac{g_u}{g_l} \exp\left(\frac{-\mathbf{h}\nu_0}{\mathbf{k}\theta}\right), \tag{2.51}$$

where the overbar indicates thermodynamic equilibrium. Substituting (2.51) into (2.48), (2.49), and (2.50), forming a balance equation, and

[3] Here, and in the remainder of this section, we use vacuum conditions and the velocity of light in vacuo ($\mathbf{c}$). The refractive index of air is very close to unity, and no correction for it is applied in atmospheric thermal calculations. If required, the refractive index, or the local velocity of light (c'), can be systematically introduced into the discussion.

comparing u_ν to the Planck function (2.37), leads to the Einstein relations:

$$\frac{A(u, l)f_s}{C(u, l)f_i} = \frac{8\pi\nu^2}{\mathbf{c}^3}, \tag{2.52}$$

$$\frac{g_l C(l, u)f_a}{g_u C(u, l)f_i} = \exp\left[\frac{\mathbf{h}(\nu - \nu_0)}{\mathbf{k}\theta}\right]. \tag{2.53}$$

(ii) The equation of transfer. We now consider the situation represented in Fig. 2.2. The energy density is $I_\nu\, d\omega/\mathbf{c}$ and the rate of gain of photons by the radiation field is $[dI_\nu(s)/ds]\, d\omega$. From the definitions (2.48), (2.49), and (2.50) and the Einstein relations (2.52) and (2.53), we may write

$$\begin{aligned}\frac{dI_\nu}{ds}\, d\omega = &-n(l)\frac{I_\nu\, d\omega}{\mathbf{c}} C(l, u)f_a \\ &\text{absorption} \\ &+ n(u)\frac{I_\nu\, d\omega}{\mathbf{c}}\frac{g_l}{g_u} C(l, u)f_a \exp\left[\frac{-\mathbf{h}(\nu - \nu_0)}{\mathbf{k}\theta}\right] \\ &\text{induced emission} \\ &+ n(u)\frac{g_l}{g_u} C(l, u)f_a \exp\left[\frac{-\mathbf{h}(\nu - \nu_0)}{\mathbf{k}\theta}\right]\frac{8\pi\nu^2}{\mathbf{c}^3}\frac{d\omega}{4\pi}. \\ &\text{spontaneous emission}\end{aligned} \tag{2.54}$$

The factor $d\omega/4\pi$ in the third term on the right-hand side (2.54) is because spontaneous emission is isotropic. Induced emission, on the other hand, is identical in all respects to the incident radiation and all induced photons fall within the solid angle $d\omega$.

If we compare (2.54) to (2.17) we find, as matters of definition, that the absorption coefficient and source function are

$$k_{\nu,\nu} = \frac{n(l)C(l, u)f_a}{\mathbf{c}}\left[1 - \frac{n(u)}{\bar{n}(u)}\frac{\bar{n}(l)}{n(l)}\exp\left(\frac{-\mathbf{h}\nu}{\mathbf{k}\theta}\right)\right], \tag{2.55}$$

$$J_\nu = B_\nu \frac{n(u)\bar{k}_{\nu,\nu}}{\bar{n}(u)k_{\nu,\nu}}, \tag{2.56}$$

where the overbars again indicate an equilibrium state. The absorption coefficient in thermodynamic equilibrium is the quantity usually measured in the laboratory and is

$$\bar{k}_{\nu,\nu} = \frac{\bar{n}(l)C(l, u)f_a}{\mathbf{c}}\left[1 - \exp\left(\frac{-\mathbf{h}\nu}{\mathbf{k}\theta}\right)\right]. \tag{2.57}$$

Equations (2.55), (2.56), and (2.57) are rarely used without approximations. One approximation is acceptable for all atmospheric bands except for the water vapor rotation band. It is assumed that each (u, l) band is so narrow that we may replace ν by ν_0 in (2.55) and (2.57). We may then integrate both equations over all frequencies, using (2.47), and introduce the *band intensity*

$$S_\nu(l, u) = \int k_{\nu,\nu}\, d\nu$$

$$= \frac{n(l)C(l, u)}{\boldsymbol{c}} \left[1 - \frac{n(u)}{\bar{n}(u)} \frac{\bar{n}(l)}{n(l)} \exp\left(\frac{-\mathbf{h}\nu_0}{\mathbf{k}\theta}\right)\right]. \tag{2.58}$$

$$\bar{S}_\nu(l, u) = \frac{n(l)C(l, u)}{\boldsymbol{c}} \left[1 - exp\left(\frac{-\mathbf{h}\nu_0}{\mathbf{k}\theta}\right)\right]. \tag{2.59}$$

The exponential terms in (2.58) and (2.59) owe their existence to induced emission and it is often argued that they may be neglected. For most atmospheric bands the exponential is very small. For the 667 cm^{-1} band of CO_2, for example, $\exp(-\mathbf{h}\nu_0/\mathbf{k}\theta)$ is 0.0082 at 200 K and 0.041 at 300 K. Apart from the statistical weights, the same factor determines the ratio of number densities in the upper and lower states [see (2.51)]. Almost all of the molecules are, therefore, in the lower state for thermodynamic equilibrium. If the system is not very far from thermodynamic equilibrium, the factor multiplying the exponential in (2.58) is close to unity, and we may consistently write

$$S_\nu(l, u) \simeq n(l)C(l, u)/\mathbf{c} \qquad \text{and} \qquad \bar{S}_\nu(l, u) \simeq \bar{n}(l)C(l, u)/\mathbf{c}. \tag{2.60}$$

To the same degree of approximation,

$$J_\nu = B_\nu \frac{n(u)}{\bar{n}(u)} \frac{\bar{S}_\nu(l, u)}{S_\nu(l, u)} \simeq B_\nu \frac{n(u)}{\bar{n}(u)} \frac{\bar{n}(l)}{n(l)}. \tag{2.61}$$

Equation (2.61) shows that J_ν has the same angular and frequency dependence as B_ν. These two quantities differ by a scaling factor dependent upon the state populations. According to our earlier assumption of narrow bands, we may replace J_ν and B_ν by $J(l, u)$ *and* $B(l, u)$, quantities characteristic of the entire band. Since $\bar{n}(u)$ and $\bar{n}(l)$ can be calculated, given the temperature, the problem of determining the absorption coefficient and the source function comes down to determining $n(u)$ and $n(l)$.

(iii) State populations. These are the result of a steady state between collision-induced and radiative transitions. First, consider the radiative transitions.

The heating rate (2.18), when expressed in terms of photon energies, is equal to the number of photons lost to the radiation field or the net number gained by the upper state. If we integrate (2.18) over all frequencies, therefore, we obtain the radiative contribution to the time rates of change of the populations of the u and l states:

$$\left[\frac{\partial n(u)}{\partial t}\right]_{\text{rad}} = -\left[\frac{\partial n(l)}{\partial t}\right]_{\text{rad}} = h(l, u), \tag{2.62}$$

where

$$h(l, u) = \int dv \int d\omega k_{v,v}(I_v - J_v). \tag{2.63}$$

There are many possible collision partners that can lead to collisional excitation or deexcitation of the radiating molecule. Here we use the simplest prescription, collisions with "air" molecules whose number density is n_A. In terms of the binary coefficients indicated in Fig. 2.4,

$$\begin{aligned}\left[\frac{\partial n(u)}{\partial t}\right]_{\text{coll}} &= -\left[\frac{\partial n(l)}{\partial t}\right]_{\text{coll}}, \\ &= n_A n(l) b(l, u) - n_A n(u) b(u, l).\end{aligned} \tag{2.64}$$

For thermodynamic equilibrium, we must have a steady state and it follows from (2.64) that

$$\bar{n}(l) b(l, u) = \bar{n}(u) b(u, l),$$

so that

$$\begin{aligned}\left[\frac{\partial n(u)}{\partial t}\right]_{\text{coll}} &= -\left[\frac{\partial n(l)}{\partial t}\right]_{\text{coll}} \\ &= b(u, l) n_A \bar{n}(u)\left[\frac{n(l)}{\bar{n}(l)} - \frac{n(u)}{\bar{n}(u)}\right].\end{aligned} \tag{2.65}$$

In order to show the nature of the solution, we shall first consider a *two-level model,* for which the total number of molecules in states u and l is conserved:

$$n = n(l) + n(u) = \bar{n}(l) + \bar{n}(u). \tag{2.66}$$

Note that (2.62) and (2.64) referred to $l \leftrightarrow u$ transitions *only*; we now go further and require explicitly that there are no other transitions to or from u or l.

We now eliminate $n(l)$ and $\bar{n}(l)$ from (2.51) and (2.65),

$$\left[\frac{\partial n(u)}{\partial t}\right]_{\text{coll}} = -\left[\frac{\partial n(l)}{\partial t}\right]_{\text{coll}}$$

$$= b(u, l)n_{\text{A}}[\bar{n}(u) - n(u)]\left[1 + \frac{g_u}{g_l}\exp\left(\frac{-\mathbf{h}\nu_0}{\mathbf{k}\theta}\right)\right]. \quad (2.67)$$

The effect of collisions is to drive $n(u)$ toward its equilibrium value with a *collisional relaxation time*, $\eta(u, l)$,

$$\eta(u, l)^{-1} = b(u, l)n_{\text{A}}\left[1 + \frac{g_u}{g_l}\exp\left(\frac{-\mathbf{h}\nu_0}{\mathbf{k}\theta}\right)\right]. \quad (2.68)$$

The relaxation time is inversely proportional to the pressure through the number density, n_{A}.

For a steady state, with transitions caused by both radiation and collisions,

$$\left[\frac{\partial n(u)}{\partial t}\right]_{\text{coll}} + \left[\frac{\partial n(u)}{\partial t}\right]_{\text{rad}} = 0,$$

and we find, after some manipulation,

$$J(l, u) = B(l, u)\frac{\bar{S}_\nu(l, u)}{S_\nu(l, u)}\left[1 + \frac{h(l, u)}{b(l, u)}\right],$$

or

$$J(l, u) = \frac{\bar{S}_\nu(l, u)}{S_\nu(l, u)}\left[B(l, u) + \left(\frac{\eta}{2\phi}\right)H(l, u)\right], \quad (2.69)$$

where, from (2.21),

$$H(l, u) = \frac{h(l, u)}{2\pi S_\nu(l, u)}. \quad (2.70)$$

The radiative relaxation time for spontaneous emission is

$$\phi(l, u) = A(u, l)^{-1}. \quad (2.71)$$

ϕ may be calculated from (2.52), (2.53), and (2.58) (with $\nu = \nu_0$),

but it is common practice to evoke the approximations (2.60) and to write

$$\phi(l, u)^{-1} \simeq S_n(l, u) \frac{g_l}{g_u} \frac{8\pi\nu_0^2}{\mathbf{c}^2}, \tag{2.72}$$

$$\eta(l, u)^{-1} \simeq b(u, l) n_A, \tag{2.73}$$

and

$$J(l, u) \simeq B(l, u) + [\eta(l, u)/2\phi(l, u)]H(l, u). \tag{2.74}$$

S_n is the band intensity per molecule in the ground state ($=S_\nu/n(l)$).

To the same degree of approximation, we may eliminate $H(l, u)$ with the help of (2.62) and (2.70) and write

$$J(l, u) = B(l, u)\left[\frac{1}{\eta(l, u)} + \frac{1}{\phi(l, u)}\right]^{-1} \times \left[\frac{1}{\eta(l, u)} + \left(\frac{\int d\nu \int d\omega\, k_{\nu,\nu} I_\nu}{\int d\nu \int d\omega\, k_{\nu,\nu} B_\nu}\right)\frac{1}{\phi(l, u)}\right]. \tag{2.75}$$

In (2.75), it is possible to write J_ν and B_ν in place of $J(l, u)$ and $B(l, u)$ because, according to (2.61), the ratio of these two quantities is independent of the frequency. This statement does not apply to equations, such as (2.74), that involve more than this ratio.

2.2.4. *Discussion of the source function*

(i) In a constant-temperature enclosure. $I_\nu = B_\nu$, by definition, and, from (2.61) and (2.75),

$$B_\nu = J_\nu = I_\nu, \tag{2.76}$$

regardless of the collisions. This is Kirchhoff's second law, showing that our treatment is appropriately consistent with classical thermodynamics.

It also follows from (2.63) that

$$h(u, l) \equiv 0, \tag{2.77}$$

an uninterestingly restrictive condition for atmospheric calculations.

(ii) Rapid collisions. At high pressures, the collisional rate is high, the

vibrational relaxation time is short, and $\eta/\phi \ll 1$. From (2.74),

$$J(l, u) = B(l, u), \tag{2.78}$$

but there is no restriction on the heating rate. From (2.56), the levels are in local thermodynamic equilibrium. This is the situation occurring in the lower atmosphere, which greatly simplifies meteorological calculations.

(iii) Slow collisions. If $\eta/\phi \gg 1$, (2.75) gives

$$J(l, u) \simeq S_v(l, u)^{-1} \int dv \int d\omega \, k_{v,v} I_v, \tag{2.79}$$

while (2.74) yields

$$H(l, u) \simeq 0. \tag{2.80}$$

Equation (2.79) is an incoherent scattering relation, that is, one in which the frequencies of the incident and scattered photons differ. The equation may be obtained from (2.32) with an isotropic phase function, after integration over all frequencies in the (l, u) band. The state populations have arranged themselves in such a way that there is no net heating. They are not in thermodynamic equilibrium and collisions are so slow that there is effectively no communication between the radiation field and the translational modes.

The magnitude of η/ϕ at 1 bar is a critical parameter. The most important single band in the mesosphere is the 15 μm band of CO_2, with a radiative lifetime of 0.74 s. An estimate of the collisional relaxation time at 1 bar and 180 K is 25 μs. η/ϕ is approximately unity at a pressure of 34 dyne cm^{-2}, occurring 76 km above the earth's surface. This is the most important relaxation level in atmospheric calculations; we shall discuss others in § 6.3.3.

We are now in a position to discuss our earlier assumption that the rotational levels are in thermal equilibrium when vibrational relaxation takes place, i.e., that rotational relaxation takes place at the lower pressure. Before doing so, however, we briefly consider electronic transitions.

Permitted electronic transitions have a radiative lifetime of the order of 10^{-8} s. Energy levels are widely spaced and collisional relaxation times may be expected to be longer than those for vibrational transitions ($\eta \gg 10^{-6}$ s at s.t.p.); even at 1 bar, $\eta/\phi \gg 10^2$, and only resonant scattering can be expected.

On the other hand, the forbidden transition between the ground states of atomic oxygen has a radiative lifetime of about 10^4 s. The transition is important in the mesosphere because the energy gap is so

small that almost all collisions can induce transitions. The time between collisions is 10^2 s at 500 km, and Planck's source function is valid for this transition at all levels below the exosphere.

We now consider the rotational transitions. First, we consider the situation at a pressure low enough for both rotational and vibrational relaxation to have occurred. The foregoing discussion is valid but for the fact that the two levels under consideration will now be individual vibration–rotation energy levels. There will be no communication between the vibration–rotation levels and the translational levels; (2.79) will be valid, but the frequency integral will be over a single rotation line rather than a band. The difference will show in the rotational level populations but not in the heating rate, which is zero in either case. Intermediary cases, when some of the rotational levels are in equilibrium and some are not, are complicated, but of little importance.

Finally, we must justify the assertion that rotational relaxation occurs at lower pressures than does vibrational relaxation. Nearly all rotational levels have longer radiation lifetimes than any vibrational levels. The radiation lifetime of a combined transition should be equal to the shorter of the two, that of the vibrational levels. On the other hand, collisional relaxation times for rotational levels are much shorter than those of vibrational levels. According to Anderson's theory of line broadening (§ 3.5.3), rotational relaxation times are related to spectral line widths, from which we may conclude that a typical relaxation time for a rotational level at 1 bar is 10^{-10} s, about 10^5 times smaller than for vibrational levels. This justifies our assertion that rotational relaxation takes place at a lower pressure than does vibrational relaxation.

The rotation band of water vapor must be considered on a different basis. The rotational relaxation time should be the same as above, but radiative lifetimes vary between 0.1 and 10 s (the latter for small J values). Thermodynamic equilibrium should prevail for all J levels at pressures above 10^{-3} dyne cm^{-2} and for none at pressures below 10^{-5} dyne cm^{-2}. For atmospheric pressures between these two values, the water vapor rotation band would pose a formidable problem were it not for photochemical decomposition. A pressure of 10^{-3} dyne cm^{-2} corresponds to a level of about 150 km and few water vapor molecules exist above this level. Planck's source function is a good approximation for all levels at which water vapor is important.

2.2.5. *Transitions between more than two levels*

The approximate equations of §§ 2.2.3 and 2.2.4 are often used in atmospheric calculations. However, the level populations $n(u)$ and $n(l)$ can be influenced by radiative and collisional transitions to and from other states, and involving other molecules; these possibilities were generally indicated by the level k in Fig. 2.4. If u, l, and k are all in

thermodynamic equilibrium there will be no net collisional transfer between them, but if one of the k levels is highly overpopulated because of a solar absorption or a chemical reaction, as examples, and if transitions between k and u or l are possible, the state populations and source functions may be greatly changed from those for the two-level model.

The equations that we have used are readily generalized if we allow for the fact that u may now be either the upper or the lower level for a $u \leftrightarrow k$ transition. From (2.62) and (2.65) we may write a balance equation,

$$\sum_{k<u} \left\{ h(k, u) + b(u, k) n_A \bar{n}(u) \left[\frac{n(k)}{\bar{n}(k)} - \frac{n(u)}{\bar{n}(u)} \right] \right\} = \sum_{k>u} \left\{ h(k, u) + b(u, k) n_A \bar{n}(u) \left[\frac{n(k)}{\bar{n}(k)} - \frac{n(u)}{\bar{n}(u)} \right] \right\}. \quad (2.81)$$

A similar equation can be written for the lower level, l.

To solve these equations we require the collisional coefficients $b(u, k)$. If radiative transitions take place between k and u, the heating rate can be calculated from (2.63) and (2.55) (omitting induced emission),

$$h(k, u) = \int d\nu \int d\omega\, n(k) C(k, u) f_a(\nu) I_\nu / \mathbf{c} - 4\pi J(k, u) S_\nu(k, u), \quad (2.82)$$

where, from (2.61),

$$J(k, u) = B(k, u) \left[\frac{n(u)}{\bar{n}(u)} \frac{\bar{n}(k)}{n(k)} \right]. \quad (2.83)$$

In (2.82), the incident radiation I_ν must include solar radiation where solar absorption can cause transitions between coupled levels or species. Both of the foregoing equations are written in the form appropriate to $u > k$.

The equations can be closed by writing them for every k level between which communication is possible. For this closed set the sum of the population densities will be conserved. Although complicated, these calculations pose no difficulty for a modern computer provided that all of the coefficients are known, and therein lies the main difficulty of this work.

A comment on the heating rate $h(k, u)$ is in order. It does not follow that the "heating" for each band ends up as kinetic energy. That is so for a two-level model but, for multiple levels, heating may be positive for one pair of levels and negative for another pair that are strongly coupled

by collisions to the first. Only when the h's are summed over all interrelated levels and species is it possible to assess the net transfer to kinetic energy.

2.3. The integral equations

2.3.1. *Introduction*

For the rest of this chapter, we shall be concerned with formal mathematical manipulation of the equation of transfer needed for subsequent chapters. No new physics is introduced and the reader may treat the remaining paragraphs of this chapter as appendixes, to be consulted when the equations are required. There are, of course, physical grounds for choosing the particular models treated. In §§ 2.3.3 and 2.3.4 the model of a plane-parallel atmosphere is introduced, first for an isotropic thermal source function with uniform isotropic boundary conditions, and second for a scattering source function with a nearly parallel incident flux as an upper boundary condition. These two models are used to treat terrestrial and solar radiation, respectively, and owe their usefulness to the effective independence of the two radiation fields. Discussion of the detailed correspondence between the models and the actual atmosphere will, however, be left for Chapter 6.

As far as actual solutions to the equations are concerned, we discuss one general solution in § 2.3.2 that can be implemented by means of a numerical quadrature. This is the only solution that has been shown to be useful for the practical meteorological task of calculating diabatic heating of the atmosphere. For the formally equivalent problems of scattering and radiative equilibrium complete solutions have been developed, some of considerable mathematical elegance. We defer consideration of these methods to Chapter 8.

2.3.2. *The general solution*

The *optical path* along a ray trajectory from point 1 to point 2 in the direction **s** is

$$\tilde{\tau}(1, 2) = \int_1^2 e_\nu \, ds. \tag{2.84}$$

Note that we have defined $\tilde{\tau}$ to be *positive definite.*

Consider the path of integration shown in Fig. 2.5. The equation of transfer at P' is

$$\frac{dI_\nu(P', \mathbf{s})}{d\tilde{\tau}_\nu} = I_\nu(P', \mathbf{s}) - J_\nu(P', \mathbf{s}). \tag{2.85}$$

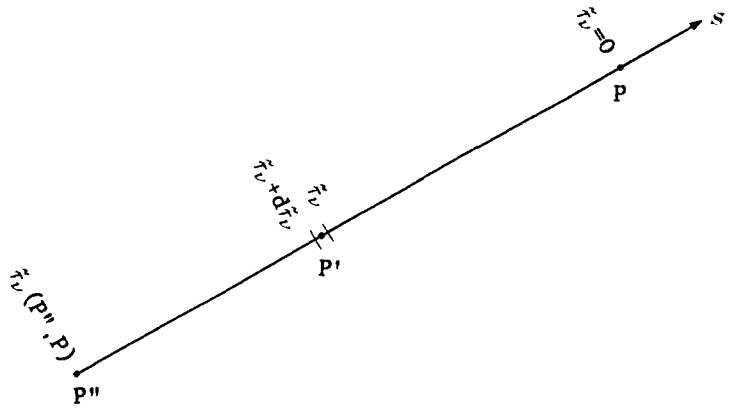


FIG. 2.5. Path of integration.

The sign of the left-hand side depends upon the sign of $d\tilde{\tau}_\nu/ds$ and upon the choice of P as the origin for $\tilde{\tau}_\nu$.

Multiplying both sides of (2.85) by $e^{-\tilde{\tau}_\nu}$ we find

$$\frac{d[I_\nu(P', \mathbf{s})e^{-\tilde{\tau}_\nu}]}{d\tilde{\tau}_\nu} = -e^{-\tilde{\tau}_\nu}J_\nu(P', \mathbf{s}).$$

Integrating from P to P'',

$$I_\nu(P, \mathbf{s}) = I_\nu(P'', \mathbf{s})e^{-\tilde{\tau}_\nu(P'', P)} + \int_0^{\tilde{\tau}_\nu(P'', P)} J_\nu(P', \mathbf{s})e^{-\tilde{\tau}_\nu(P', P)}\, d\tilde{\tau}_\nu. \quad (2.86)$$

Equation (2.86) can be evaluated if the source function between P and P'' is known, and also the incident intensity at P''. Atmospheric problems are frequently posed in such a form, and this solution is widely used. It states that the intensity at P is made up of the intensity imposed at P'', attenuated according to Lambert's law, together with contributions from each intervening, radiating element, attenuated by the appropriate optical path.

We can simplify (2.86) by observing that the equation of transfer must also be obeyed inside the boundaries, which we may take to be infinitely thick, optically. Consequently, if P'' tends to infinity, so will $\tilde{\tau}_\nu(P'', P)$ and (2.86) becomes

$$I_\nu(P, \mathbf{s}) = \int_0^\infty J_\nu(P', \mathbf{s})e^{-\tilde{\tau}_\nu}\, d\tilde{\tau}_\nu. \quad (2.87)$$

We can recover (2.86) if we write

$$J_\nu(P', \mathbf{s}) = I_\nu(P'', \mathbf{s}) \qquad \text{for } \bar{\tau}_\nu \geqslant \bar{\tau}_\nu(P'', P). \tag{2.88}$$

Thus, a boundary condition on the intensity can be introduced into the integral equation by assuming a constant source function from the boundary to infinity. This procedure is widely used in atmospheric computations.

For some special problems in Chapter 8, we shall require a volume integral for $\bar{I}_\nu(P)$. First we redefine $\mathbf{s}$ to be the *radius vector* from the point P; it is therefore the reverse of the vector in Fig. 2.5, and (2.87) becomes

$$I_\nu(P, -\mathbf{s}) = \int_0^\infty J_\nu(P', -\mathbf{s}) e^{-\bar{\tau}_\nu} \, d\bar{\tau}_\nu,$$

where

$$d\bar{\tau}_\nu = e_{\nu,\nu}(P') \, ds.$$

From the definition of $\bar{I}_\nu$ (2.9) we have

$$\bar{I}_\nu(P) = \frac{1}{4\pi} \int_0^\infty e_{\nu,\nu}(P') \, ds \int d\omega_s \, J_\nu(P', -\mathbf{s}) e^{-\bar{\tau}_\nu}.$$

A volume element at the point P' (distance s from P) is (Fig. 2.6)

$$dV(P') = s^2 \, ds \, d\omega_s, \tag{2.89}$$

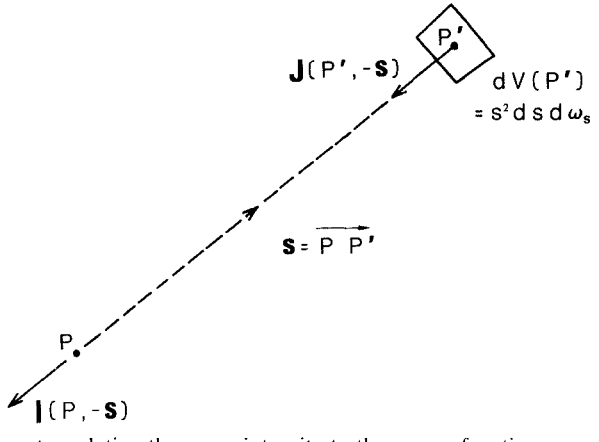


FIG. 2.6. Geometry relating the mean intensity to the source function.

and, therefore,

$$\bar{I}_\nu(P) = \int \frac{e_{\nu,\nu}(P')J_\nu(P', -\mathbf{s})e^{-\bar{\tau}_\nu(P, P')}}{4\pi s^2} dV(P'), \tag{2.90}$$

where the integral extends over all space.

2.3.3. *Thermal radiation in a stratified atmosphere*

We now consider an isotropic source function in an atmosphere for which absorption coefficient and temperature are functions of the vertical coordinate (z) alone (a *stratified atmosphere*, see Fig. 2.7).

The appropriate equation of radiative transfer in this case is, from (2.17) and Fig. 2.7,

$$-\frac{\xi}{e_{\nu,\nu}} \frac{dI_\nu(z, \xi)}{dz} = I_\nu(z, \xi) - J_\nu(z, \xi), \tag{2.91}$$

where $\xi = \cos \zeta$ is the cosine of the zenith angle of the vector $\mathbf{s}$ (2.12). Introducing the *optical depth*,

$$\tau_\nu(z, \infty) = \int_z^\infty e_{\nu,\nu}(z')\, dz', \tag{2.92}$$

we have

$$\xi \frac{dI_\nu(\tau_\nu, \xi)}{d\tau_\nu} = I_\nu(\tau_\nu, \xi) - J_\nu(\tau_\nu, \xi). \tag{2.93}$$

The change of sign between (2.91) and (2.93) follows from the definition (2.92).

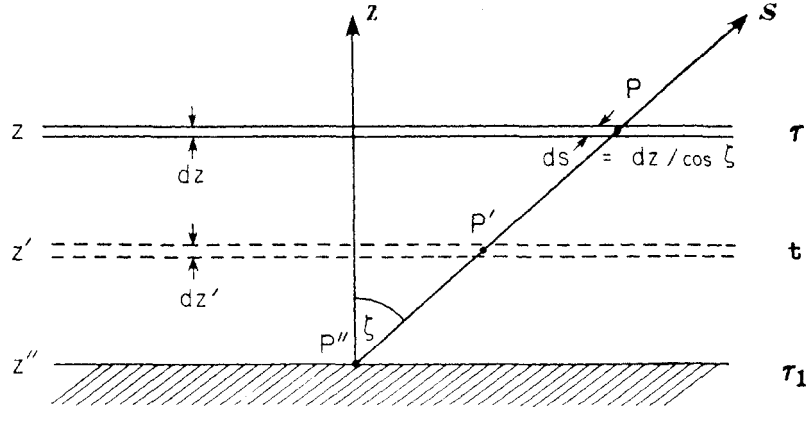


FIG. 2.7. A stratified atmosphere.

In the following treatment we must bear in mind that the solutions to (2.93) will differ for upward traveling beams ($1 \geq \xi > 0$) and downward traveling beams ($-1 \leq \xi < 0$), because the boundary conditions differ in the two cases.

For an upward traveling stream of radiation, when the lower boundary ($z = z''$) is a black-body with temperature θ^*, we have (see Fig. 2.7)

$$I_\nu(\tau_1, \xi) = B_\nu(\theta^*) \qquad 1 \geq \xi > 0. \tag{2.94}$$

For a downward traveling stream, on the other hand, the appropriate boundary condition for a planetary atmosphere is that no thermal radiation is incident upon the atmosphere from outer space. Then we have

$$I_\nu(0, \xi) = 0 \qquad -1 \leq \xi < 0. \tag{2.95}$$

We now return to the formal solution (2.86), but with the optical depth replacing the optical path. To do this we first simplify the notation and write (see Fig. 2.7)

$$\begin{aligned} \tau_\nu(z', \infty) &= t, \\ \tau_\nu(z, \infty) &= \tau, \\ \tau_\nu(z'', \infty) &= \tau_1. \end{aligned} \tag{2.96}$$

If we compare the optical paths $\tilde{\tau}_\nu(P', P)$ and $\tilde{\tau}_\nu(P'', P)$ in Figs. 2.5 and 2.7, we can see that the appropriate substitution is

$$\tilde{\tau}_\nu = \frac{t - \tau}{\xi}, \tag{2.97}$$

for both upward and downward directed beams: in both cases $\tilde{\tau}_\nu$ is positive, as required, because the numerator and the denominator in (2.97) always have the same sign. Direct substitution of (2.97) in (2.86) and use of the boundary conditions (2.94) and (2.95) yields

$$I_\nu(\tau, \xi) = B_\nu(\theta^*)e^{-(\tau_1 - \tau)/\xi} + \int_\tau^{\tau_1} J_\nu(t, \xi)e^{-(t-\tau)/\xi}\frac{dt}{\xi}, \tag{2.98}$$

for $1 \geq \xi > 0$, and

$$I_\nu(\tau, \xi) = -\int_0^\tau J_\nu(t, \xi)e^{-(t-\tau)/\xi}\frac{dt}{\xi}, \tag{2.99}$$

for $-1 \leq \xi < 0$. Expressions (2.98) and (2.99) are valid for $0 \leq \tau \leq \tau_1$.

For numerical solutions at a given zenith angle, (2.98) and (2.99) can be employed in the form given. Further simplification is possible, however, if we wish to calculate integrated quantities, such as the mean intensity (2.9) or the flux (2.3), and if the source function is isotropic.

Isotropy of $J_\nu(t)$ enables us to perform angular integrations in terms of exponential integrals of order n (see Appendix 6),

$$E_n(x) = \int_1^\infty \frac{e^{-wx}}{w^n}\, dw. \tag{2.100}$$

From Fig. 2.8, an element of solid angle is,

$$d\omega = \frac{2\pi s \sin\zeta \cdot s\, d\zeta}{s^2} = -2\pi\, d\xi, \tag{2.101}$$

and the angular integration runs from $\xi = +1$ to $\xi = 0$ in the positive hemisphere and $\xi = 0$ to $\xi = -1$ in the negative hemisphere.

The mean intensity (2.9) can now be written

$$\begin{aligned} \bar{I}_\nu(\tau) &= \frac{1}{4\pi}\int I_\nu(\tau, \xi)\, d\omega \\ &= \frac{1}{2}\int_0^1 I_\nu(\tau, \xi)\, d\xi + \frac{1}{2}\int_1^0 I_\nu(\tau, \xi)\, d\xi. \end{aligned} \tag{2.102}$$

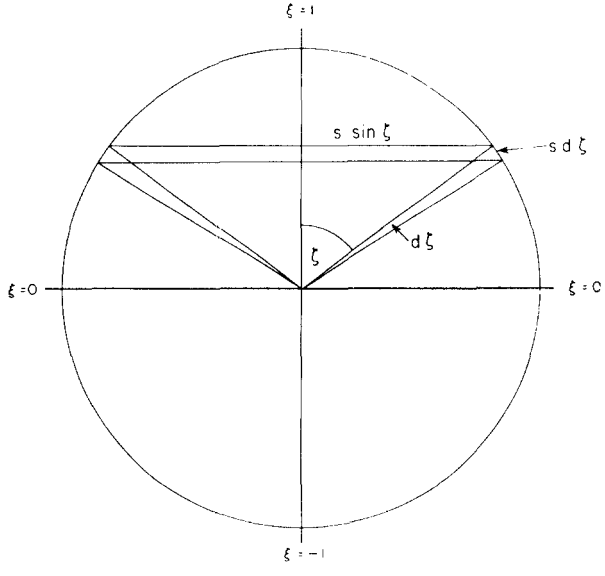


FIG. 2.8. Evaluation of the flux.

Substituting from (2.98), (2.99), and (2.100),

$$\bar{I}_\nu(\tau) = \tfrac{1}{2}B_\nu(\theta^*)E_2(\tau_1 - \tau) + \frac{1}{2}\int_\tau^{\tau_1} J_\nu(t)E_1(t-\tau)\,dt + \frac{1}{2}\int_0^\tau J_\nu(t)E_1(\tau - t)\,dt, \tag{2.103}$$

or, using a property of E_1 and E_2 that is described in Appendix 6,

$$\bar{I}_\nu(\tau) = \tfrac{1}{2}B_\nu(\theta^*)E_2(\tau_1 - \tau) + \frac{1}{2}\int_{E_2(\tau_1-\tau)}^1 J_\nu(t)\,dE_2(t-\tau) + \frac{1}{2}\int_{E_2(\tau)}^1 J_\nu(t)\,dE_2(\tau - t). \tag{2.104}$$

Similarly we find the vertical flux (2.3) (horizontal fluxes are zero in a stratified atmosphere),

$$\begin{aligned} F_\nu(\tau) &= \int I_\nu(\tau, \xi)\xi\,d\omega \\ &= 2\pi\int_0^1 I_\nu(\tau, \xi)\xi\,d\xi + 2\pi\int_{-1}^0 I_\nu(\tau, \xi)\xi\,d\xi \\ &= 2\pi B_\nu(\theta^*)E_3(\tau_1 - \tau) + 2\pi\int_\tau^{\tau_1} J_\nu(t)E_2(t-\tau)\,dt \\ &\quad - 2\pi\int_0^\tau J_\nu(t)E_2(\tau - t)\,dt. \end{aligned} \tag{2.105}$$

From a property of E_2 and E_3, given in Appendix 6,

$$F_\nu(\tau) = 2\pi B_\nu(\theta^*)E_3(\tau_1 - \tau) + 2\pi\int_{E_3(\tau_1-\tau)}^{1/2} J_\nu(t)\,dE_3(t-\tau) - 2\pi\int_{E_3(\tau)}^{1/2} J_\nu(t)\,dE_3(\tau - t). \tag{2.106}$$

For numerical computations the first two terms and the third term in (2.106) are evaluated independently since they involve two different relationships between J_ν and t (appropriate to the atmosphere below or above z, respectively). In the meteorological literature positive definite

quantities F^+ and F^- are usually defined in the following way:

$$F_\nu^+(\tau) = 2\pi B_\nu(\theta^*)E_3(\tau_1 - \tau) + 2\pi \int_{E_3(\tau_1-\tau)}^{1/2} J_\nu(t)\, dE_3(t-\tau), \tag{2.107}$$

$$F_\nu^-(\tau) = 2\pi \int_{E_3(\tau)}^{1/2} J_\nu(t)\, dE_3(\tau - t), \tag{2.108}$$

and (2.106) becomes

$$F_\nu(\tau) = F_\nu^+(\tau) - F_\nu^-(\tau). \tag{2.109}$$

Finally, from (2.18), (2.21), (2.103) and (2.104), we can express the heating function in the form

$$\begin{aligned} H_i(\tau) &= \left(\int_i e_{\nu,\nu}\, d\nu\right)^{-1} \int_i e_{\nu,\nu}\Big[-2J_\nu(\tau) + B_\nu(\theta^*)E_2(\tau_1 - \tau) \\ &\quad + \int_\tau^{\tau_1} J_\nu(t)E_1(t-\tau)\, dt + \int_0^\tau J_\nu(t)E_1(\tau - t)\, dt\Big]\, d\nu \qquad (2.110) \\ &= \left(\int_i e_{\nu,\nu}\, d\nu\right)^{-1} \int_i e_{\nu,\nu}\Big[-2J_\nu(\tau) + B_\nu(\theta^*)E_2(\tau_1 - \tau) \\ &\quad + \int_{E_2(\tau_1-\tau)}^1 J_\nu(t)\, dE_2(t-\tau) + \int_{E_2(\tau)}^1 J_\nu(t)\, dE_2(\tau - t)\Big]\, d\nu. \qquad (2.111) \end{aligned}$$

2.3.4. *Solar radiation in a stratified atmosphere*

Equation (2.87) is a general solution to all radiative transfer problems, given the appropriate source function and boundary conditions. If we treat the solar and thermal radiation fields as separable, the solar source function involves only scattering processes and, in practice, attention is usually concentrated on simple scattering by molecules and aerosol particles. The appropriate source function is therefore given by (2.32). The lower boundary condition will express a reflectivity condition involving, in general, a reflection matrix.

This is too complex for real geophysical situations, however, and a more commonly used boundary condition assumes that the surface has uniform brightness and that the albedo is

$$a_\nu = \frac{F_\nu^+(z'')}{F_\nu^-(z'')}. \tag{2.112}$$

The upper boundary condition expresses the fact that an unpolarized

and nearly parallel beam of solar radiation strikes the atmosphere at an angle defined by $(\xi_\odot, \varphi_\odot)$, where φ is an azimuth angle. Some instruments are able to discriminate in favor of the incident beam by recording radiation only in the neighborhood of the sun. Consequently, it is of value to state the equation of transfer in two parts, one valid for angles close to $(\xi_\odot, \varphi_\odot)$ and one for all other angles.

The appropriate equation of transfer is obtained by substituting the source function (2.32) into (2.93). In addition, we change the notation from **d**, **s**, $d\omega_d$, and $d\omega_d'$ to (ξ, φ), (ξ', φ'), $d\omega$, and $d\omega'$:

$$\xi \frac{dI_\nu^{(i)}}{d\tau_\nu}(\tau_\nu; \xi, \varphi) = I_\nu^{(i)}(\tau_\nu; \xi, \varphi) - \frac{s_{\nu,\nu}}{e_{\nu,\nu}} \int P_{ij}(\xi, \varphi; \xi', \varphi') I_\nu^{(j)}(\tau_\nu; \xi', \varphi') \frac{d\omega'}{4\pi}. \quad (2.113)$$

Now consider a small solid angle $d\omega_\odot$ surrounding the direction $(\xi_\odot, \varphi_\odot)$. Integrate both sides of (2.113) over $d\omega_\odot$ and let $d\omega_\odot$ become very small. The second term on the right-hand side will tend to zero, as will contributions from the other two terms, provided that no source of "parallel" radiation is involved. Such a source carries a finite irradiance $f_\nu^{(i)}(\tau_\nu)$ in an infinitesimally small solid angle, so that from (2.7)

$$\lim_{d\omega_\odot \to 0} \int_{d\omega_\odot} I_\nu^{(i)}(\tau_\nu)\, d\omega = f_\nu^{(i)}(\tau_\nu) \neq 0.$$

The differential equation governing the "parallel" irradiance is

$$\xi_\odot \frac{df_\nu^{(i)}(\tau_\nu)}{d\tau_\nu} = f_\nu^{(i)}(\tau_\nu), \quad (2.114)$$

or, integrating from 0 to τ_ν, bearing in mind that $\xi_\odot$ is negative for the solar beam,

$$f_\nu^{(i)}(\tau_\nu) = f_\nu^{(i)}(0) \exp(\tau_\nu / \xi_\odot), \quad (2.115)$$

where $f_\nu^{(i)}(0)$ is the solar irradiance outside the earth's atmosphere (2.7). Some texts choose to replace $\xi_\odot$ by $-|\xi_\odot|$, since the solar beam is always downward and $\xi_\odot$ is always a negative number.

It is appropriate to point out here that the laboratory experimenter attempts to reproduce the conditions of (2.115) by placing an absorption tube in a collimated beam formed from a small source, which is ultimately focused on the slit of a spectrophotometer. A quantity proportional to the irradiance is recorded, and by alternately filling and

emptying the absorption tube the optical path and hence the absorption coefficient can be determined.

Returning now to (2.113) we redefine $I_\nu^{(i)}$ so that it is continuous near $(\xi_\odot, \varphi_\odot)$ and does not include the direct, "parallel" beam. Equation (2.113) is still satisfactory except for that part of the integral over ω' near to $(\xi_\odot, \varphi_\odot)$. This contributes a term

$$\frac{s_{\nu,\nu}}{e_{\nu,\nu}} \int_{d\omega_\odot} P_{ij}(\xi, \varphi; \xi', \varphi') I_\nu^{(j)}(\tau_\nu; \xi', \varphi') \frac{d\omega'}{4\pi}$$

$$= \frac{s_{\nu,\nu}}{e_{\nu,\nu}} P_{ij}(\xi, \varphi; \xi_\odot, \varphi_\odot) \frac{f_\nu^{(j)}(\tau_\nu)}{4\pi}.$$

According to our redefinition of the intensity to exclude direct radiation, this term would be missed; it must, therefore, be added explicitly to the equation of transfer, which now becomes

$$\xi \frac{dI_\nu^{(i)}(\tau_\nu; \xi, \varphi)}{d\tau_\nu} = I_\nu^{(i)}(\tau_\nu; \xi, \varphi) - \frac{s_{\nu,\nu}}{e_{\nu,\nu}} \int P_{ij}(\xi, \varphi; \xi', \varphi') I_\nu^{(j)}(\tau_\nu; \xi', \varphi') \frac{d\omega'}{4\pi}$$

$$- \frac{s_{\nu,\nu}}{e_{\nu,\nu}} P_{ij}(\xi, \varphi; \xi_\odot, \varphi_\odot) \frac{f_\nu^{(j)}(0)}{4\pi} \exp\left(\frac{\tau_\nu}{\xi_\odot}\right). \tag{2.116}$$

The upper boundary condition on the scattered intensity is now the same as for thermal radiation, i.e., that there is no scattered radiation from outer space,

$$I_\nu^{(i)}(0; \xi, \varphi) = 0 \qquad \text{for } -1 \leq \xi < 0. \tag{2.117}$$

2.4. Approximate methods for thermal radiation

2.4.1. *The atmospheric problem*

The intricacies of atmospheric radiation calculations, taken together with the ready availability of large digital computers, have led to an emphasis upon the development of numerical radiation algorithms. These algorithms can be coupled to algorithms for hydrodynamic processes and interactions may be handled by iteration. If the end result is to couple algorithms for scattering and radiative heating with algorithms for atmospheric and ocean dynamics, serious questions may arise as to the significance of the results. Recent history has demonstrated that such complex numerical calculations may be flawed; they may yield unphysical results and equally competent investigators can disagree. An outsider can make no judgment. Even if complete documentation were available it would be impractical to check on the results, and documentation is often missing.

Such questions are beyond the scope of this book, the limited purpose of which is to describe the fundamental processes that lie behind the computations. Although numerical methods may be essential for accurate numbers, a valuable level of understanding of atmospheric problems can also be achieved with approximate equations; the remainder of this chapter will be concerned with an examination of some of the most important.

In this chapter we limit considerations to thermal radiation, for which the source function is isotropic. This permits approximations to angular integrals that may be less appropriate for anisotropic, scattering phase functions. Nevertheless, the methods have important extensions also to scattering phase functions. Dicussion will be delayed until Chapter 8, after examining scattering phase functions in Chapter 7.

The value of approximate methods is often underrated. For many purposes they may provide all the accuracy that is required; they should always be considered as a first step in any investigation of atmospheric radiation.

2.4.2. *Transparent and opaque approximations*

The radiation mean free path is, from (2.15), equal to $(e_{v,v})^{-1}$. It varies with frequency and, as we shall see in subsequent chapters, this variation may be very large. First, consider a single frequency.

A typical atmospheric problem may involve coupling between the radiation field and some other physical phenomenon, such as atmospheric motion, that may impose its own characteristic scales on the problem, for example, the wavelength of a periodic disturbance or the vertical grid scale in a numerical model. We may now ask whether the transfer equation may be simplified if this imposed scale (l) is, respectively, much smaller than or much larger than the radiative scale, $(e_{v,v})^{-1}$. To do so, we make some reasonable assumptions as to the conditions imposed on the problem by the scale l:

1. The average over a volume of any fluctuating quantity is assumed to tend to zero as the linear dimensions of the volume of integration become larger than l.
2. To order of magnitude we may replace d^n/ds^n by $(1/l)\, d^{n-1}/ds^{n-1}$. (In boundary-layer problems of fluid flow (2) defines a *smooth function.*)

We restrict our attention to thermodynamic equilibrium ($J_v = B_v$) and consider (2.18), expressing the heating rate as the balance between the angle-averaged absorption and emission. If $e_{v,v}l \ll 1$, the medium is transparent over many scale lengths. Contributions to the first term in the parentheses on the right-hand side of (2.18) originate largely from distances approximately equal to the mean free path of the radiation, and

the angle average in (2.19) can be approximated by an integral over a sphere of radius equal to many scale lengths. According to assumption (1), the first term in parentheses should not reflect the local fluctuations in the physical state of the medium that are, however, intrinsic to the second term because it is a function of the local temperature. Consequently, if δ indicates a small spatial variation, and provided that I_ν and B_ν are of the same magnitude,

$$\delta(h_\nu/e_{\nu,\nu}) \simeq -4\pi\delta B_\nu. \tag{2.118}$$

If we further assume that the composition is not perturbed,[4] then $\delta e_{\nu,\nu} = 0$ and (2.118) becomes

$$\delta h_\nu \simeq -4\pi e_{\nu,\nu}\delta B_\nu. \tag{2.119}$$

Equations (2.118) and (2.119) are known as the *transparent approximation.* For small temperature variations, (2.119) can be written in the form

$$\delta h_\nu \simeq -4\pi e_{\nu,\nu}(d\bar{B}_\nu/d\theta)\delta\theta, \tag{2.120}$$

where the overbar denotes a local mean; this is a statement of *Newton's law of cooling.*

The other limit that we may consider is that of the *opaque approximation* ($e_{\nu,\nu} l \gg 1$). Major contributions to the local intensity now come from distances small compared to l. According to condition (2), it is now possible to make a Taylor expansion of B_ν, provided that it and $\tilde{\tau}_\nu$ are continuous functions of the space variable. These conditions may not be satisfied near a boundary where sharp gradients of temperature, even temperature discontinuities, can exist. We must, therefore, assume that we are dealing with events that are unaffected by the (distant) boundaries.

The Taylor expansion of $B_\nu(\tilde{\tau}_\nu)$ is

$$B_\nu(\tilde{\tau}_\nu) = B_\nu(0) + \tilde{\tau}_\nu\left(\frac{dB_\nu}{d\tilde{\tau}_\nu}\right)_0 + \frac{\tilde{\tau}_\nu^2}{2!}\left(\frac{d^2B_\nu}{d\tilde{\tau}_\nu^2}\right)_0 + \frac{\tilde{\tau}_\nu^3}{3!}\left(\frac{d^3B_\nu}{d\tilde{\tau}_\nu^3}\right)_0 + \cdots. \tag{2.121}$$

Substituting (2.121) in (2.87) and integrating over all space,

$$I_\nu(0) = B_\nu(0) - \left(\frac{dB_\nu}{d\tilde{\tau}_\nu}\right)_0 + \left(\frac{d^2B_\nu}{d\tilde{\tau}_\nu^2}\right)_0 - \left(\frac{d^3B_\nu}{d\tilde{\tau}_\nu^3}\right)_0 + \cdots. \tag{2.122}$$

[4] We shall reconsider this restriction in § 10.6.

Multiplying by $d\omega$ and integrating over all solid angles,

$$4\pi\bar{I}_\nu(0) = \int B_\nu(0)\, d\omega - \int \left(\frac{dB_\nu}{d\tilde{\tau}_\nu}\right)_0 d\omega + \int \left(\frac{d^2B_\nu}{d\tilde{\tau}_\nu^2}\right)_0 d\omega - \int \left(\frac{d^3B_\nu}{d\tilde{\tau}_\nu^3}\right)_0 d\omega + \cdots. \tag{2.123}$$

For time-independent problems, (2.12) enables us to write the first differential operator in (2.123),

$$\frac{d}{d\tilde{\tau}_\nu} = \xi_x \frac{\partial}{\partial \tilde{\tau}_{\nu,x}} + \xi_y \frac{\partial}{\partial \tilde{\tau}_{\nu,y}} + \xi_z \frac{\partial}{\partial \tilde{\tau}_{\nu,z}}, \tag{2.124}$$

where ξ_x, ξ_y, and ξ_z are the three direction cosines. Since B_ν is isotropic, the angular dependence of each term is that of the direction cosines, and because

$$\int \xi_x\, d\omega = \int \xi_y\, d\omega = \int \xi_z\, d\omega = 0,$$

the second term on the right-hand side of (2.123) is zero. The third term involves quantities such as $\int \xi_x \xi_y\, d\omega$, which are zero, and quantities such as $\int \xi_x^2\, d\omega$, which equal $4\pi/3$. Applying the operator (2.124) twice,

$$\int \left(\frac{d^2B_\nu}{d\tilde{\tau}_\nu^2}\right) d\omega = \frac{4\pi}{3}\left(\frac{\partial^2 B_\nu}{\partial \tilde{\tau}_{\nu,x}^2} + \frac{\partial^2 B_\nu}{\partial \tilde{\tau}_{\nu,y}^2} + \frac{\partial^2 B_\nu}{\partial \tilde{\tau}_{\nu,z}^2}\right). \tag{2.125}$$

By similar arguments, it can be shown that all even derivatives on the right-hand side of (2.123) are zero. According to property (2) of the scale length, l, we may write,

$$\frac{d^4B_\nu}{d\tilde{\tau}_\nu^4} \simeq (e_{\nu,\nu} l)^{-2} \frac{d^2B_\nu}{d\tilde{\tau}_\nu^2} \ll \frac{d^2B_\nu}{d\tilde{\tau}_\nu^2},$$

and neglect all terms in (2.123) after the third.

From (2.18) and (2.123), we now find

$$h_\nu = (4\pi/3)\nabla \cdot e_{\nu,\nu}^{-1} \nabla B_\nu, \tag{2.126}$$

or, if $e_{\nu,\nu}$ is constant,

$$h_\nu = (4\pi/3) e_{\nu,\nu}^{-1} \nabla^2 B_\nu. \tag{2.127}$$

2.4.3. *Approximate forms for the absorption coefficient*

Stellar spectra exhibit strong absorption continua and it was a logical step in the development of astrophysical theory to assume the absorption coefficient to be independent of frequency (*grey absorption*). This assumption carried over into the early meteorological literature with the difference that two independent coefficients were assigned to the solar and terrestrial radiation streams (*semigrey absorption*).

In fact, the earth's atmosphere differs so greatly from a grey absorber that grey calculations have little meaning except in the limits of the transparent and opaque approximations, when effective mean absorption coefficients can be defined. Equations (2.119) and (2.127) can be used to express the frequency-integrated heating (the quantity of thermodynamic significance) in terms of the frequency-integrated source function if the mean absorption coefficients,

$$\bar{k}_{\mathrm{P}} = \frac{\int_0^\infty k_\nu \delta B_\nu \, d\nu}{\int_0^\infty \delta B_\nu \, d\nu}, \tag{2.128}$$

or

$$\bar{k}_{\mathrm{R}}^{-1} = \frac{\int_0^\infty k_\nu^{-1} \nabla^2 B_\nu \, d\nu}{\int_0^\infty \nabla^2 B_\nu \, d\nu}, \tag{2.129}$$

respectively, are employed.

Below 120 km in the earth's atmosphere, the temperature varies less than 20% from its mean value and we may write approximately

$$\delta B_\nu \simeq \frac{d\bar{B}_\nu}{d\theta} \delta\theta,$$

and

$$\nabla^2 B_\nu \simeq \frac{d\bar{B}_\nu}{d\theta} \nabla^2\theta,$$

so that

$$\bar{k}_{\mathrm{P}} = \frac{\int_0^\infty k_\nu (d\bar{B}_\nu/d\theta) \, d\nu}{\int_0^\infty (d\bar{B}_\nu/d\theta) \, d\nu}, \tag{2.130}$$

and

$$\bar{k}_R^{-1} = \frac{\int_0^\infty k_\nu^{-1}(d\bar{B}_\nu/d\theta)\, d\nu}{\int_0^\infty (d\bar{B}_\nu/d\theta)\, d\nu}. \tag{2.131}$$

Equation (2.131) is commonly used in astrophysics under the name of the *Rosseland mean.* Equation (2.130) is not in common use in astrophysics, although a similar mean with the weighting function B_ν in place of $dB_\nu/d\theta$ is called the *Planck mean.* Since the shapes of the two weighting functions do not greatly differ, these two means are similar, and we shall use the same name for (2.130). Since k_ν varies over a very wide range in atmospheric spectra, the Planck and Rosseland means may differ by very large factors.

2.4.4. *The method of moments in three dimensions*

From (2.12), the equation of transfer (2.85) can be written

$$\xi_i\, \partial I_\nu(\mathbf{s})/\partial \tilde{\tau}_{\nu,i} = I_\nu(\mathbf{s}) - B_\nu(\mathbf{s}), \tag{2.132}$$

where $i = (x, y, z)$, ξ_i is the direction cosine between the $\mathbf{s}$ direction and the i axis, and the sum rule for repeated indices is employed.

$I_\nu(\mathbf{s})$ is a function of frequency, position, and direction, and it is this multiple dependency that creates many of the difficulties of radiative transfer problems. If we wish to calculate angle-averaged quantities such as fluxes and heating rates we do not anticipate that details of the angular variation of the intensity should be very important, and it is common practice to approximate this aspect of the solution. We shall discuss one particular method, known as the *method of moments,* because it can be applied to three-dimensional geometry, enabling a comparison to be made with the opaque and transparent approximations. Other techniques have been restricted to stratified atmospheres but, unlike the present derivation, most can be raised to higher degrees of approximation.

The method of moments is based on repeated application of the operator $\int \xi^n\, d\omega$, where n is a positive integer, to the equation of transfer. Applied to equation (2.132), this operator leads to a relationship between a moment of order n and a moment of order $n+1$. If the precise angular variation of $I_\nu(\mathbf{s})$ is not important, we may anticipate that it will be possible to approximate a high-order moment in terms of moments of lower order, in which case the hierarchy of moment equations can be closed.

Multiply (2.132) by $d\omega$ and integrate over all angles to obtain the zero-order moment equation. From (2.3) and (2.9), we have, omitting ν

suffices for convenience,

$$\partial F_i/\partial \tilde{\tau}_i = 4\pi I - 4\pi J. \tag{2.133}$$

The first moment equation is obtained by multiplying (2.132) by ξ_j and integrating over all angles, again using (2.3),

$$\int \xi_i \xi_j (\partial I/\partial \tilde{\tau}_i)\, d\omega = F_j. \tag{2.134}$$

Equations (2.133) and (2.134) are correct for any angular distribution of the radiation field. In order to close the equations, however, we shall calculate the second moment equation on the assumption that the radiation field is approximately isotropic in each hemisphere independently, although the two hemispheres may differ. Since the cosine changes sign in opposite hemispheres, the flux integral (2.3) involves contributions with different signs and the result may be a small residuum between two large quantities. It is, therefore, important that the intensity not be approximated when evaluating the flux. The second moment of $I(\mathbf{s})$ should, however, be less sensitive to the angular distribution because no sign changes are involved in the angular integration. If (2.134) is evaluated on the basis of an isotropic radiation field, there results [see the derivation of (2.125)]

$$\int \xi_i \xi_j \frac{\partial I}{\partial \tilde{\tau}_i}\, d\omega = \frac{4\pi}{3} \frac{\partial \bar{I}}{\partial \tilde{\tau}_j}. \tag{2.135}$$

If we take (2.135) to be approximately valid under all conditions, the required approximation to the equation of transfer can be stated in either one of two forms:

$$\frac{\partial^2 F_i}{\partial \tilde{\tau}_i\, \partial \tilde{\tau}_j} \simeq 3F_j - 4\pi \frac{\partial B}{\partial \tilde{\tau}_j}, \tag{2.136}$$

or

$$\frac{1}{3} \frac{\partial^2 \bar{I}}{\partial \tilde{\tau}_j^2} \simeq \bar{I} - B. \tag{2.137}$$

The boundary conditions for these two equations are, in general, awkward. If the flux vector were fully defined at each point on the boundaries, (2.136) could be solved, but atmospheric problems are not posed in this form. More typically, the given boundary condition is the inward intensity at the boundary, from which neither F nor $\bar{I}$ can be

obtained. An appropriate treatment will be given for a stratified atmosphere in the next section.

We can now discuss the relationship between (2.136) and the transparent and opaque approximations. If both F and J vary rapidly over unit optical path the conditions are analogous to those for the transparent approximation. Under these conditions, the two differentials in (2.136) should be much larger than the first term on the right-hand side and if we set $\partial/\partial\tilde{\tau}_j = \delta$ and make use of (2.10) we rederive the transparent approximation (2.118).

In the opaque case, quantities vary only slowly over unit optical path and the balance in (2.136) should be between the two terms on the right-hand side. Making use again of (2.10), we obtain the opaque approximation (2.127).

The important conclusion from this discussion is that, despite the approximations made, (2.136) is exact in the transparent and opaque limits and can be in error only for intermediate opacities. In practice, the worst errors are not large and dependence upon physical parameters is faithfully represented. Unfortunately, as might have been guessed from the differences between the Rosseland and Planck means, equation (2.136) cannot be integrated over frequency to give useful relationships between frequency-integrated quantities, except, of course, for the case of grey absorption.

A final application of the three-dimensional equations is to derive the equation for radiative equilibrium; from (2.10) and (2.14) we have

$$\nabla \cdot \mathbf{F} = \frac{\partial F_i}{\partial x_i} = -e_v \frac{\partial F_i}{\partial \tilde{\tau}_i} = 0,$$

so that (2.136) becomes

$$3F_j = 4\pi \frac{\partial B}{\partial \tilde{\tau}_j} = -\frac{4\pi}{e_v}\frac{\partial B}{\partial x_j}. \tag{2.138}$$

In vector form,

$$\mathbf{F} = -\frac{4\pi}{3e_v}\nabla B. \tag{2.139}$$

2.4.5. *Approximations for a stratified atmosphere*

For a stratified atmosphere, we set $\partial/\partial\tilde{\tau}_x = \partial/\partial\tilde{\tau}_y = 0$ and $\tilde{\tau}_z = \tau$ in (2.136) to give

$$\frac{d^2F}{d\tau^2} = 3F - 4\pi\frac{dJ}{d\tau}, \tag{2.140}$$

and, in (2.138), for radiative equilibrium

$$F = -\frac{4\pi}{3}\frac{dJ}{d\tau} = \text{constant}. \tag{2.141}$$

In this simplified geometry, we may derive the appropriate boundary conditions, but to do so we must approximate the flux integral, so that the attempt to preserve the exact form for the flux in (2.134) loses some significance. When boundary conditions are applied, it will no longer necessarily be the case that the result will be exact in the transparent and opaque limits.

The flux and the mean intensity at the boundary must be evaluated from the given boundary conditions on the inward intensity and an approximation to the angular distribution. For this example, we choose hemispheric isotropy, for which I^+ and I^- are intensities independent of the zenith angle in their appropriate hemispheres ($\xi > 0$ and $\xi < 0$ respectively). This is consistent with, but more restrictive than, the approximation (2.135).

From (2.3) and (2.9),

$$F = \pi(I^+ - I^-), \tag{2.142}$$

$$\bar{I} = \tfrac{1}{2}(I^+ + I^-). \tag{2.143}$$

Substitute (2.142) and (2.143) in (2.133) and eliminate $\bar{I}$ and either I^+ or I^-. There results at $\tau = \tau_1$ (lower boundary)

$$I^+(\tau_1) = F(\tau_1)/2\pi + B(\tau_1) + (1/4\pi)(dF/d\tau)_{\tau=\tau_1}, \tag{2.144}$$

and at $\tau = 0$ (upper boundary)

$$I^-(0) = -F(0)/2\pi + B(0) + (1/4\pi)(dF/d\tau)_{\tau=0}. \tag{2.145}$$

As an illustration, consider the case of radiative equilibrium with black bodies emitting $B^*(0)$ or $B^*(\tau_1)$ at the two boundaries. The third terms on the right-hand side of (2.144) and (2.145) are now zero and

$$F/2\pi = B(0) - B^*(0) = B^*(\tau_1) - B(\tau_1). \tag{2.146}$$

Equation (2.146) requires a discontinuity in the Planck function, implying a discontinuity of temperature, at the boundary.

The class of approximation of which (2.140) is representative is extensive and a large number of different names and terms are used to describe members of the class: the *Schwarzschild–Schuster* approximation, the *Eddington* approximations, *Chandrasekhar's first*

approximation, and a variety of *two-stream approximations.* In addition, the *exponential-kernel* approximation to the integral equations is comparable in intent. All employ a similar degree of approximation and the differences between one method and another come down to coefficients of order unity. These coefficients can be adjusted to give better results for one particular application, but general statements that one technique is better than another are difficult to justify.

In order to give a limited but consistent comparison between the various approximations we shall apply each to the closure between second- and zero-order moments as performed in (2.135). Eddington introduced the idea of hemispheric isotropy, as expressed by (2.142) and (2.143). Since the derivation of (2.135) was based upon the idea of approximate isotropy, independently in each hemisphere, it follows that Eddington's approximation leads to (2.140) for a stratified atmosphere. The treatment in § 2.4.4 is more general than Eddington's, because it is valid in three dimensions.

Eddington introduced another approximation that is more commonly associated with his name,

$$I(\xi) = a + b\xi. \tag{2.147}$$

This is a flexible approach because the approximation can be raised to higher order by adding higher powers of ξ, although the method is rarely used except in the first approximation, (2.147), which yields the closure (2.135), without modification. This follows from the fact that the second term in (2.147) vanishes in all even moments. There may be a change in the approximate boundary conditions, however.

Chandrasekhar introduced Gaussian quadrature to evaluate the integrals in (2.3) and (2.9). A typical angular integral is replaced by a sum over $2m$ ($j = \pm 1, \pm 2, \ldots \pm m$) ordinates,

$$\int_{-1}^{+1} f(\xi)\, d\xi \simeq \sum_j a_j f(\xi_j), \tag{2.148}$$

where the a_j are weighting factors that can be calculated when the ξ_j are known, and

$$\sum_j a_j = 2.$$

If $f(\xi)$ is a polynomial of degree less than m, any method of numerical quadrature can give exact results, but Gauss' method, for which the ξ_i are zeros of the Legendre polynomial $P_m(\xi)$, gives exact results for polynomials of degree $\leqslant 2m$.

For Chandrasekhar's first approximation, we replace the radiation

field by two components I^+ and I^- with values of ξ equal to $+3^{-1/2}$ and $-3^{-1/2}$, respectively, and with weights equal to unity. It follows that

$$\bar{I} = (I^+ + I^-)/2,$$

and, from (2.134),

$$F = \int \xi^2 \frac{\partial I}{\partial \tau} d\omega = \frac{2\pi}{3} \frac{\partial}{\partial \tau}(I^+ + I^-) = \frac{4\pi}{3} \frac{\partial \bar{I}}{\partial \tau},$$

identical to (2.135). Again, the stratified approximation (2.140) will be unaltered, but the boundary conditions need not be the same as (2.144) and (2.145).

In the first paper to introduce an equation of transfer, Schuster attempted to represent the intensity by two antiparallel streams, an approximation also used by Schwarzschild before he gave the correct equation. We can state this approximation in the form

$$\begin{aligned} I(\xi) &= I^+ \delta(\xi - 1) \qquad \xi > 0, \\ I(\xi) &= I^- \delta(\xi + 1) \qquad \xi < 0, \end{aligned} \tag{2.149}$$

where

$$\int \delta(\xi - 1) \frac{d\omega}{4\pi} = \int \delta(\xi + 1) \frac{d\omega}{4\pi} = \frac{1}{2}.$$

From (2.149) we obtain

$$\bar{I} = \int I \frac{d\omega}{4\pi} = \tfrac{1}{2}(I^+ + I^-),$$

and, from (2.134),

$$F = 2\pi(\partial/\partial\tau)(I^+ + I^-).$$

These relations are inconsistent with (2.140), although the discrepancies can be reduced by redefining τ, e.g., the substitution $\tau' = 3^{1/2}\tau$ would correctly introduce the factor 3 into the first term on the right-hand side of (2.140) but, at the same time, would introduce an error in the second term. The Schwarzschild–Schuster approximation is now of historical interest only.

Closely connected to approximations to the equation of transfer for stratified atmospheres is the *exponential-kernel* approximation to the integral equations, known in the meteorological literature (see § 6.1.2) as

the *diffusivity approximation.* In (2.106) we approximate

$$E_3(x) \simeq \tfrac{1}{2} \exp(-rx), \tag{2.150}$$

where r is an adjustable parameter.

The analogy to (2.140) can be demonstrated by differentiating twice the approximate integral equation resulting from the substitution of (2.150) in (2.106),

$$\frac{d^2F}{d\tau^2} = r^2F - 2\pi r \frac{dJ}{d\tau}. \tag{2.151}$$

This equation cannot satisfy both the transparent and the opaque limits: the former involves a balance between the two differentials in (2.151) and (2.140) and requires $r = 2$; the latter involves a balance between the two terms on the right-hand side and requires $r = 3/2$. In practice, a value of r between these two limits is chosen (e.g., $r = 1.66$).

A final comment on approximations to the angular distribution of the radiation field is that we may readily improve an approximation, without a more elaborate differential equation. Suppose that we use the stratified equations to solve for the source function B. This solution can be substituted into the exact integral equation (2.87) to give an improved solution for the intensity at any level in the atmosphere. Eddington used this method to improve the solution for the outward intensity leaving a stellar atmosphere, for which reason it is referred to as *Eddington's second approximation.*

BIBLIOGRAPHY

2.1. Definitions

The formal theory of this book is based, wherever possible, on

Chandrasekhar, S., 1950, *Radiative transfer.* Oxford: Clarendon Press. Reprinted 1960 by Dover Publications, Inc.

This book was written with astrophysical problems in mind. So are

Chandrasekhar, S., 1939, *An introduction to the study of stellar structure.* Chicago: Chicago University Press. Reprinted 1957 by Dover Publications Inc.

Wooley, R. v. d. R., and Stibbs, D. W. N., 1953, *The outer layers of a star.* Oxford: Clarendon Press.

A more general approach is adopted by

Kourganoff, V., 1952, *Basic methods in transfer problems.* Oxford: Clarendon Press.

The mathematical aspects of the problem are treated by

Busbridge, I. W., 1960, *The mathematics of radiative transfer.* London: Cambridge University Press.

Sobolev, V. V., 1975, *Light scattering in planetary atmospheres.* Oxford: Pergamon Press.

The relation between the equation of transfer and the electromagnetic wave equation is discussed by

Rozenberg, G. V., 1977, "The light ray (contribution to the theory of the light field)," *Sov. Phys. Usp.* **20,** 55.

Wolf, E., 1978, "Coherence and radiometry," *J. Opt. Soc. Am.* **68,** 6.

2.2. Thermal emission

The treatment of Kirchhoff's laws is nowhere more explicit and readable than in

Planck, M., 1913, *Vorlesungen über die Theorie der Wärmestrahlung.* Leipzig: Barth. English edition 1959 by Dover Publications, Inc.

Numerical values for Planck's function are given by

Allen, C. W., 1973, *Astrophysical Quantities (third edition).* London: Athlone Press.

Pivovonsky, M., and Nagel, M. R., 1961, *Tables of blackbody radiation functions.* New York: Macmillan.

Kaye, G. W. C., and Laby, T. H., 1973, *Tables of physical and chemical constants and some mathematical functions (14th edition).* London: Longman.

The breakdown of Kirchhoff's laws for a two-level model was first treated by

Milne, E. A., 1930, *Handbuch der Astrophysik,* Vol. 3(i). Berlin: Springer.

The first application to mesospheric source functions was by

Curtis, A. R., and Goody, R. M., 1956, "Thermal radiation in the upper atmosphere," *Proc. Roy. Soc. London Ser. A* **236,** 193.

A good introduction to more recent work is to be found in an unpublished thesis

Williams, A. P., 1971, *Radiative transfer in the mesosphere.* D. Phil. Thesis, Oxford University.

A selection of other publications is

Kuhn, W. R., and London, J., 1969, "Infrared radiative cooling in the middle atmosphere (30–110 km)," *J. Atmos. Sci.* **26,** 189.

Dickinson, R. E., 1984, "Infrared radiative cooling in the mesosphere and lower thermosphere," *J. Atmos. Terr. Phys.* **46,** 995.

Lopez-Puertas, M., Rodrigo, R., Molina, A., and Taylor, F. W., 1986a, "A non-LTE radiative transfer model for infrared bands in the middle atmosphere. I. Theoretical basis and application to CO_2 15 μm bands," *J. Atmos. Terr. Phys.* **48,** 729.

The last paper contains a review of collisional exchange coefficients.

2.3. The integral equations

Alternative derivations of these equations are given by Chandrasekhar (1950) and Kourganoff (1952), § 2.1, and by

Pierls, R., 1939, "Critical conditions in neutron multiplication," *Proc. Camb. Phil. Soc.* **35,** 610.

2.4. Approximate methods for thermal radiation

Eddington emphasized the importance of the opaque approximation for stellar interiors in some of the earliest work on the subject, see

Eddington, A. S., 1926, *The internal constitution of the stars.* London: Cambridge University Press, Reprinted 1959 by Dover Publications, Inc.

A systematic treatment of the use of the Planck and Rosseland means is given by

Pomraning, G. C., 1971, "Grey radiative transfer," *J. Quant. Spectrosc. Radiat. Transfer* **11,** 597.

The three-dimensional treatment of § 2.4.4 follows

Giovanelli, R. G., 1959, "Radiative transfer in a non-uniform medium," *Aust. J. Phys.* **12,** 164.

The method of moments, upon which Giovanelli's paper is based, is of much older lineage. For a stratified atmosphere, it has been systematically extended to higher moments than the second. Such methods are, however, largely superseded by those described in Chapter 8.

As is pointed out in the text, the literature on low-order approximations in a stratified atmosphere is extensive and confusing, despite the fact that there is little to choose between one approach and another. For the interested reader, a good place to start is the early literature, and we are fortunate that there is an instructive collection by

Menzel, D. H. (Ed.), 1966, *Selected papers on the transfer of radiation.* New York: Dover Publications.

This book contains the following fundamental papers, including English translations of the two papers by Schwarzschild:

Schuster, A., 1905, "Radiation through a foggy atmosphere," *Astrophys. J.* **21,** 1.

Schwarzschild, K., 1906, "On the equilibrium of the sun's atmosphere," *Nach. K. Gesell. Wiss. Göttingen* **195,** 41.

Schwarzschild, K., 1914, "Diffusion and absorption in the sun's atmosphere," *Sitz. K. Preuss. Akad. Wiss.* 1183.

Eddington, A. S., 1916, "On the radiative equilibrium of the stars," *Mon. N. R. Astronom. Soc.* **77,** 16.

Rosseland, S., 1924, "Note on the absorption of radiation within a star," *Mon. N. R. Astronom. Soc.* **84,** 525.

Milne, E. A., 1930, "Thermodynamics of the stars," *Handbuch Astrophys.* **3**(1), 65.

The paper by Schuster gave the first equation of transfer, but in a parallel-stream approximation. The correct equation was given in Schwarzschild's second paper. Eddington's influence on the subject cannot be understated; see also Eddington (1926). The paper by Milne can be looked upon as the start of the "modern" era.

The following general references contain some discussion of approximate methods: Liou (1980, § 1.1).

Irvine, W. M., 1975, "Multiple scattering in planetary atmospheres," *Icarus* **25,** 175.

Van de Hulst, H. C., 1980, *Multiple light scattering: Tables formulas and applications, Volumes I and II.* New York: Academic Press.

3

VIBRATION–ROTATION SPECTRA OF GASEOUS MOLECULES

3.1. Introduction

In this chapter we discuss the characteristics of absorption by gaseous constituents of the earth's atmosphere. This is a complex topic and atmospheric investigators may be disturbed by the idea that weather and climate might be affected by details of the kind we shall discuss. But, as yet, we lack criteria as to what is important and what is not, leaving little alternative to developing a general understanding of the field.

A full description of the atmospheric absorption spectrum involves the intensities, state dependence, and detailed line profiles of 10^5 to 10^6 lines of 20 or more different chemical species. Given the capabilities of modern computers, it is possible to store, retrieve, and manipulate such data and this is the method of choice for purposes such as the identification of lines in high-resolution spectra. One of a number of current attempts to assemble an up-to-date archive of molecular data is the Air Force Geophysics Laboratory (AFGL) magnetic tape. Not only does this tape provide an economical means of access to the best data from a vast literature, but it also provides a convenient international standard atmosphere. Two numerical climate models, both using the AFGL data, cannot attribute their differences to the radiation data employed.

We shall, therefore, address the subject of molecular spectroscopy in the general context of the AFGL tape and many of our illustrations are composed from the tape in preference to seeking out observed spectra. As will be apparent by the end of this chapter, it may sometimes take an expert to distinguish between the two.

Figure 3.1 offers an overview of the atmospheric absorption spectrum. The six gases considered are the most important radiators, although climate studies often involve more and rarer species. All six gases are minor species (and therefore in dilute mixtures with nitrogen and oxygen) and are very simple molecules (methane is the most complex).

Figure 3.1 shows no visible or ultraviolet spectra. The missing features are mainly electronic bands of oxygen and ozone; they will not

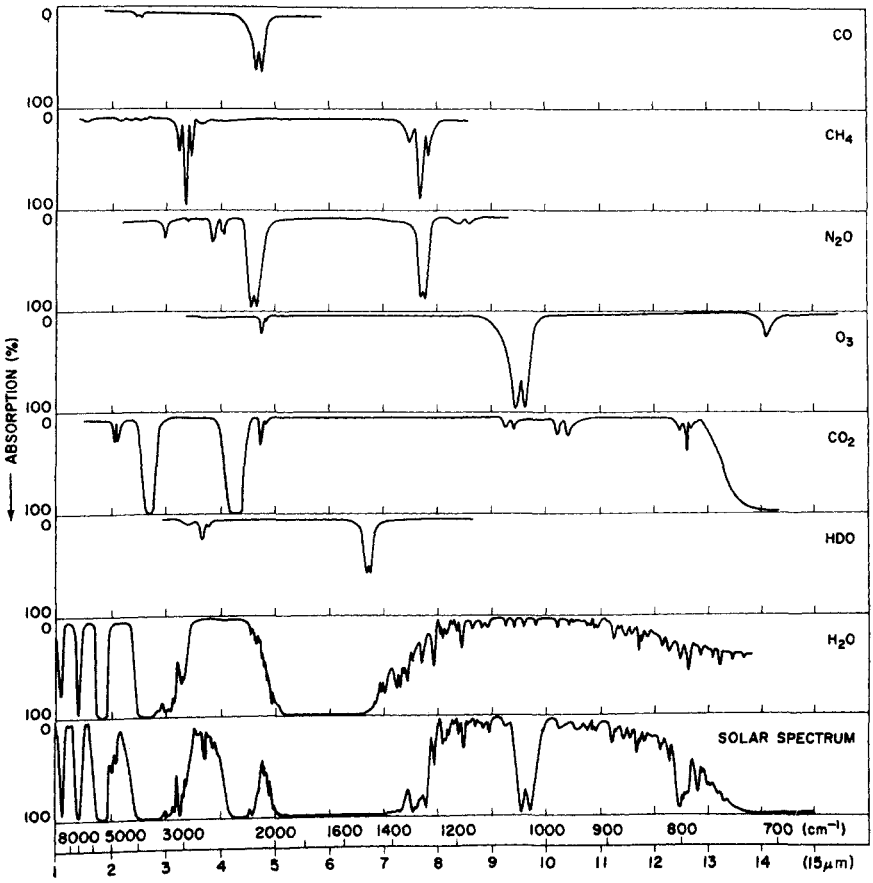


FIG. 3.1. Low-resolution absorption spectrum of the atmosphere. The top six panels are the absorption spectra of important atmospheric species. The bottom panel is a simulated absorption spectrum of the atmosphere. After Valley (1965).

be treated in this chapter since they are more complex theoretically but easier to handle empirically than the bands shown in Fig. 3.1. Data for electronic bands that are suitable for empirical calculations are discussed in Chapter 5.

The absorptions shown in Fig. 3.1 take the form of discrete bands of differing widths. Apart from the rotation band of water vapor, stretching from 16 μm to the microwave spectrum, all involve a change in the vibrational energy of the molecule. The differing widths of the bands are the result of simultaneous changes in the rotational energy, and the features are referred to as *vibration–rotation bands.*

The structures of selected vibration–rotation bands at much higher spectral resolution are illustrated by Figs. 3.2, 3.3, 3.4, 3.5, 3.6, and 3.7.

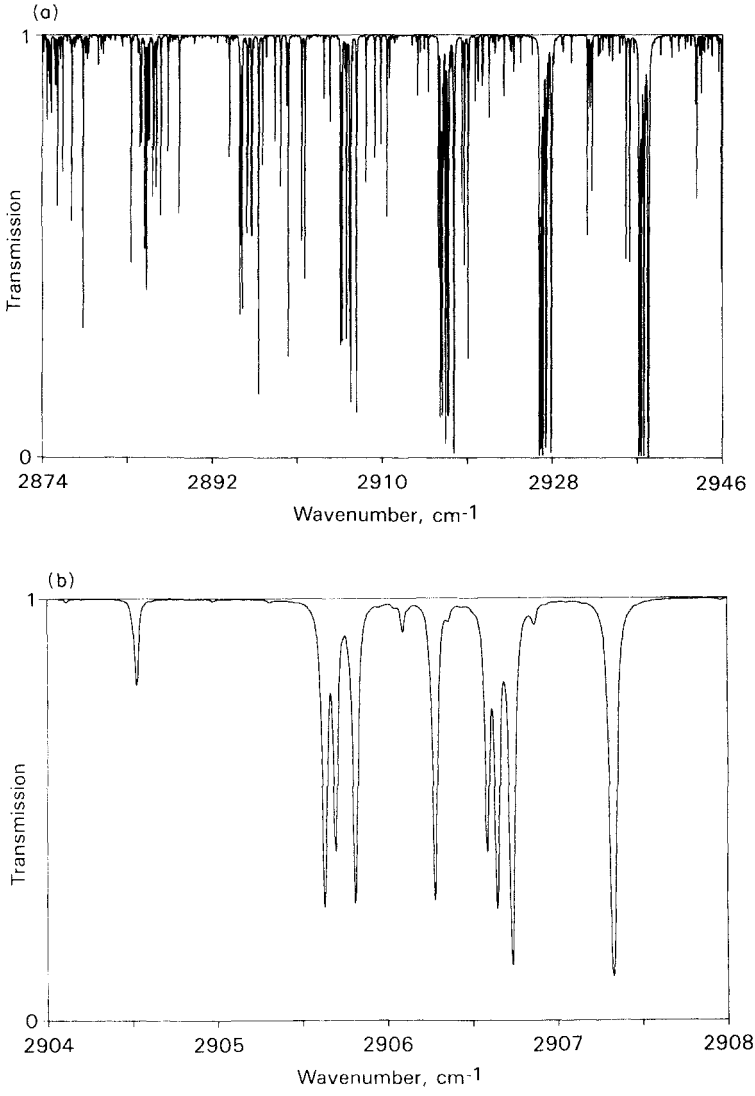


FIG. 3.2. Synthetic spectrum of CH_4 near 3.44 μm. Spectral range: (a) 2874–2946 cm^{-1}; (b) 2904–2908 cm^{-1}. Level of observation: 10 km. Zenith angle of observation: 30°. Terrestrial concentration ×1.

Each has been constructed by a computer for a single atmospheric constituent. The absorption path reaches from outside the atmosphere down to the level of observation at the given zenith angle of observation. Terrestrial gas concentrations are employed or a given multiple of them, if this makes for a clearer illustration.

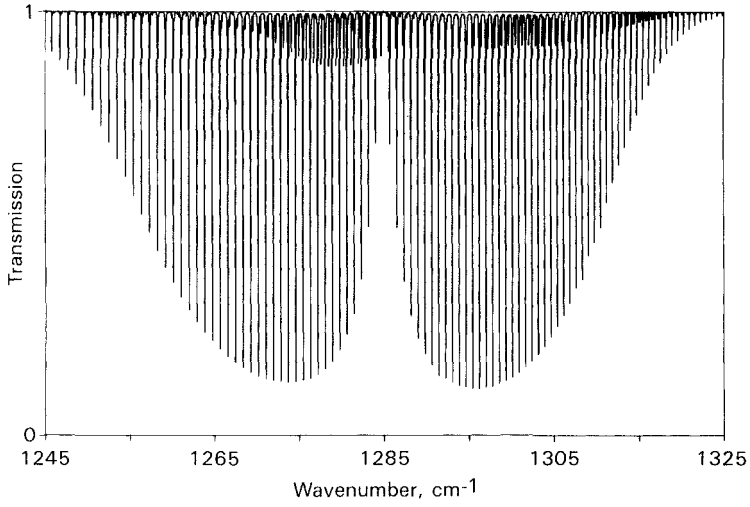


FIG. 3.3. Synthetic spectrum of N_2O near 7.78 μm. Spectral range: 1245–1325 cm^{-1}. Altitude of observation: 15 km. Zenith angle of observation: 30°. Terrestrial concentration ×1.

Figure 3.2a and Fig. 3.2b are both centered on a region in the wing of the strong 3.31 μm band of methane. Figure 3.2a shows eight groups of lines (*manifolds*) while Fig. 3.2b shows details of one manifold. At the high dispersion of spectrum (b), each line is seen to have a finite width.

Figure 3.3 shows the central region of the strongest band of nitrous

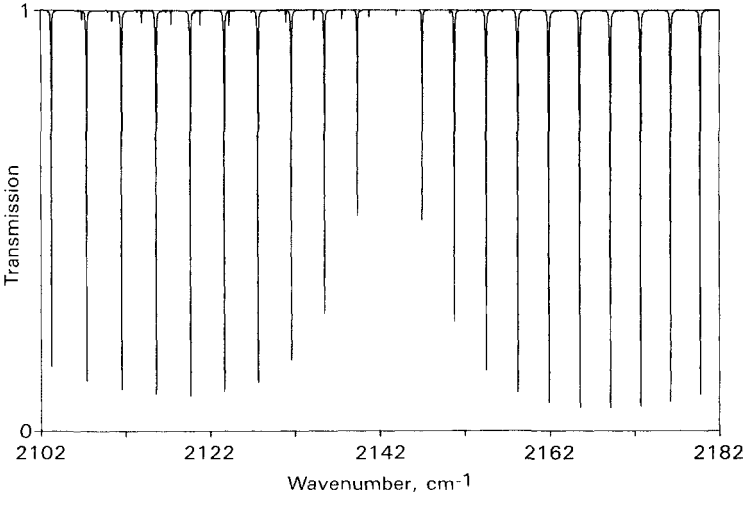


FIG. 3.4. Synthetic spectrum of CO near 4.67 μm. Spectral range: 2102–2182 cm^{-1}. Altitude of observation: 10 km. Zenith angle of observation: 30°. Terrestrial concentration ×1.

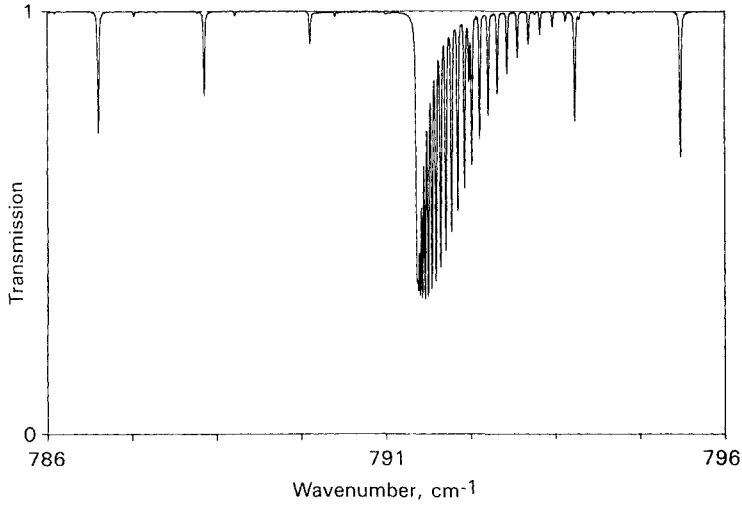


FIG. 3.5. Synthetic spectrum of CO_2 near 12.64 μm. Spectral range: 786–796 cm^{-1}. Altitude of observation: 15 km. Zenith angle of observation: 30°. Terrestrial concentration ×10.

oxide. The two very regular groups of lines are the *P- and R-branches* (to the left and the right, respectively) and they are separated by a gap caused by a missing line at the center of the band. A second, weaker band is superimposed (an *upper state band*) with a slightly different band center. Mixed in here, and also in most of the other spectra shown in this

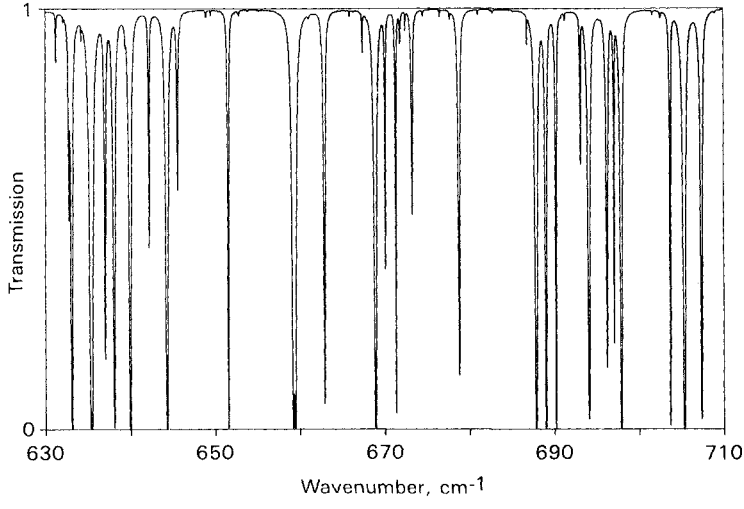


FIG. 3.6. Synthetic spectrum of H_2O near 14.9 μm. Spectral range: 630–710 cm^{-1}. Altitude of observation: 0 km. Zenith angle of observation: 30°. Terrestrial concentration ×0.03.

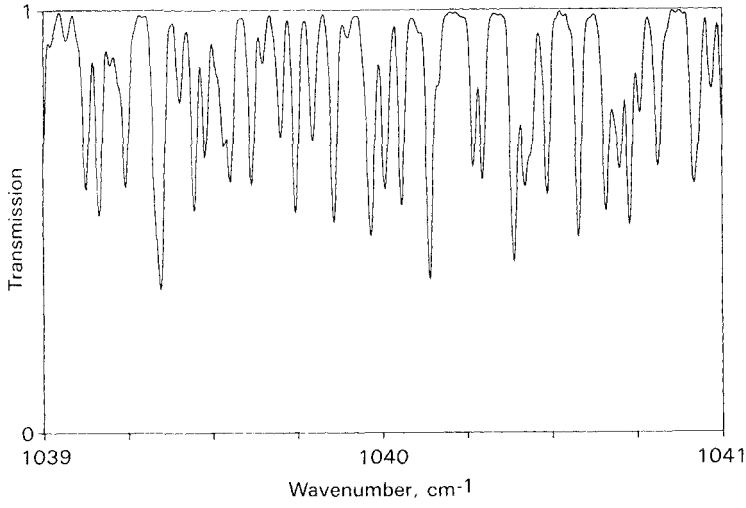


FIG. 3.7. Synthetic spectrum of O_3 near 9.61 μm. Spectral range: 1039–1041 cm^{-1}. Altitude of observation: 30 km. Zenith angle of observation: 30°. Terrestrial concentration ×1. Unlike most of the other synthetic spectra shown in this chapter, individual rotation lines are not resolved.

chapter, are weaker isotopic lines. In this spectrum, a few of the weak lines are from $^{14}N^{15}N^{16}O$.

Carbon monoxide, like nitrous oxide, is a linear molecule and the center of the 4.67 μm band shows a simple band structure (Fig. 3.4) similar to that in Fig. 3.3. The lines are more widely spaced than for nitrous oxide because the carbon monoxide molecule has a smaller moment of inertia. Again there is a gap in the band center (characteristic of a *parallel band*); a weak band of $^{13}C^{16}O$ is superimposed.

A high-resolution spectrum of a *perpendicular band* of a linear molecule is shown in Fig. 3.5. This is a weak band on the wing of the 15 μm band of carbon dioxide. It is the result of a transition involving two vibrational levels in close resonance (*Fermi resonance*), a circumstance that can greatly complicate the interpretation of a molecular spectrum. The P- and R-branches of this band are represented by the five, widely spaced, isolated lines. The gap between the bands is now filled with a partially resolved *Q-branch* near 790 cm^{-1}. Some weak isotopic lines of $^{16}O^{12}C^{17}O$ and $^{16}O^{12}C^{18}O$ are also present.

Figures 3.3, 3.4, and 3.5 illustrate arrays of lines containing some obvious order, but a glance at most regions of the atmospheric spectrum is more suggestive of a completely disordered situation. Figure 3.6 shows a region in the wing of the water vapor rotation band (this region is usually dominated by a very strong carbon dioxide band but the computer permits us to isolate species). The spectrum appears to be completely disordered, both as regards line positions and line intensities.

Figure 3.7 shows a section of the 9.6 μm band of ozone at high spectral resolution. The altitude of observation is 30 km. Little ozone lies above this level and the pressure is low, so that lines are very narrow (see § 3.3.1); nevertheless, the ozone lines overlap strongly and are not resolved. Each feature in Fig. 3.7 is a manifold of many unresolved rotation lines.

3.2. Vibration–rotation spectra

3.2.1. *The Hamiltonian for a semirigid molecule*

The quantum-mechanical Hamiltonian operator, H, is obtained by replacing variables by operators in the classical expression for the energy, E, of a system, consisting, in this case, of the atomic nuclei and electrons that make up a molecule. If ψ is the wave function, Schroedinger's equation is

$$\frac{i\mathbf{h}}{2\pi}\frac{\partial\psi}{\partial t} = H\psi. \tag{3.1}$$

We may separate the Hamiltonian into time-dependent and time-independent terms,

$$H = H_1(q) + H_2(q, t) \tag{3.2}$$

(q represents the particle coordinates). We can deduce the *stationary states* of the molecule from the time-independent term

$$H_1\psi = E\psi. \tag{3.3}$$

This equation has discrete eigenvalues (*energy levels*), E_n, and eigenfunctions, ψ_n. Transitions between energy levels result in the absorption and emission of photons (frequency, ν) following Planck's relation,

$$\Delta E = \mathbf{h}\nu. \tag{3.4}$$

Provided that the time-dependent term in (3.2) is small, it may be treated as a perturbation, from which the rate of change of the probability that a stationary state is occupied can be calculated.

It is common practice to distinguish between different forms of molecular energy, electronic, vibrational, rotational, translational, and nuclear spin interactions, each with its own Hamiltonian, wave functions, and energy levels. The total energy is then the sum of the different forms of energy, the wave function the product of the individual wave functions, and emitted or absorbed photons have energies equal to the

sum of those for individual transitions. This view is not always correct, but it is useful for a discussion of orders of magnitude.

With this caveat in mind it is useful to order the different forms of energy. Nuclear spin energy is very small indeed and need not be considered further at this stage (but see § 3.3.3). Electronic levels are commonly separated by a few electron volts (or a few $\times 10^4\,\mathrm{cm}^{-1}$, if we use the wavenumber as an energy unit, see Appendix 2 for discussion), vibrational levels by 5×10^2 to $10^4\,\mathrm{cm}^{-1}$, while rotational levels are closer, $1\text{–}10^2\,\mathrm{cm}^{-1}$.

Transitions between levels lead to electronic band systems in the ultraviolet, visible, and near-infrared spectrum ($10^5\text{–}10^4\,\mathrm{cm}^{-1}$ or $10^{-1}\text{–}1\,\mu\mathrm{m}$), vibrational bands in the near- to far-infrared ($10^4\text{–}10^2\,\mathrm{cm}^{-1}$ or $1\text{–}10^2\,\mu\mathrm{m}$), and rotational bands extending from the far-infrared to the microwave region ($1\text{–}10\,\mathrm{cm}^{-1}$ or $10^2\text{–}10^4\,\mu\mathrm{m}$). From the relative magnitudes of the energies involved, we may anticipate that vibrational and rotational changes will accompany an electronic change, but not necessarily vice versa and, similarly, rotational changes will accompany a vibrational transition. These magnitudes account, in a general way, for the appearance of the bands in Fig. 3.1; the bands themselves are vibrational, while the finite width is caused by many simultaneous, unresolved rotational transitions.

Translational energy does not have stationary states in an unlimited spatial domain, but it is involved in establishing the equilibrium populations of energy levels. It is, therefore, important to note that a typical translational energy ($\mathbf{k}\theta$ with $\theta = 300$ K) is $400\,\mathrm{cm}^{-1}$, much greater than most rotational energies, smaller than most vibrational energies, and very much smaller than electronic energies. Kinetic collisions can, therefore, influence rotational levels strongly, vibrational levels slightly, and electronic levels scarcely at all.

Before considering stationary states further we need to discuss the time-dependent term in the Hamiltonian, $H_2(q, t)$. There are two different kinds of term. One results from the interaction of two molecules during a collision. At atmospheric pressures, the time spent during collisions is very small compared to the time spent between collisions. The theory applicable to the small fraction of the time domain occupied by collisions is outlined in § 3.3. For the majority of the time, molecules may be treated as isolated and the important consideration is the time-dependent interaction between a molecule and the ambient electromagnetic field. The strongest interactions occur if the molecule has an electric dipole moment, $\mathbf{M}(q)$,

$$H_2(q, t) = -\mathbf{M}(q) \cdot \mathbf{E} \sin(2\pi\nu t), \tag{3.5}$$

where $\mathbf{E}$ is the electric field vector. Electric dipole interactions give rise

to the strongest bands in the terrestrial spectrum. Such bands are called *permitted,* whereas other bands are loosely referred to as *forbidden.*

We shall mainly discuss permitted transitions. Magnetic dipole and electric quadrupole interactions are, respectively, 10^5 and 10^8 times weaker than electric dipole interactions and, although such transitions may be detected, they are of little importance for energy transfer.

The probability of a dipole transition is proportional to the square of the absolute value of the matrix element of the dipole moment

$$\mathbf{R}_{ij} = \int \psi_i^* \mathbf{M} \psi_j \, dV, \tag{3.6}$$

where dV is a volume element in configuration space, and the integral is over all space. Subscripts i and j represent two quantum states and the asterisk denotes complex conjugation. Since wave functions are orthogonal,

$$\int \psi_i^* \psi_j \, dV = 0. \tag{3.7}$$

If $\mathbf{M}$ is not a function of the configuration coordinates of the atoms forming the molecule, it may be taken outside the integral in (3.6) and

$$\mathbf{R}_{ij} = \mathbf{M} \int \psi_i^* \psi_j \, dV = 0. \tag{3.8}$$

For a dipole transition to be permitted, the dipole moment must change with configuration coordinates and be different in the initial and the final states. Since $\mathbf{M}$ transforms as a vector, it is usual to find that (3.6) is identically zero for certain combinations of i and j. The rules that define which pairs of states have nonzero matrix elements are called *selection rules.* When selection rules depend upon the precise symmetry of the wave function they can be readily violated if perturbations affect that symmetry.

The matrix elements (3.6) and the Einstein coefficients (§ 2.2.3) are related by

$$A_{ij} = \frac{64\pi^4 \nu_{ij}^3}{3\mathbf{h}\mathbf{c}^3} |\mathbf{R}_{ij}|^2 \tag{3.9}$$

for nondegenerate levels. Orders of magnitude for A_{ij} are $10^8\,\mathrm{s}^{-1}$ for permitted electronic transitions, about $10\,\mathrm{s}^{-1}$ for vibrational transitions, and $1\,\mathrm{s}^{-1}$ for a pure rotational transition.

The relationship between equilibrium line intensity and Einstein coefficient given by (2.59) is valid for any single transition. For nondegenerate states, (2.59) may be written in matrix notation,

$$S_n(i,j) = \frac{n_j}{n}\frac{8\pi^3\nu_{ij}\,|\mathbf{R}_{ij}|^2}{3\mathbf{hc}}\left[1-\exp\left(\frac{-\mathbf{h}\nu_{ij}}{\mathbf{k}\theta}\right)\right]. \tag{3.10}$$

Equation (3.10) provides the connection between the quantum-mechanical formulation and the line and band intensities used elsewhere in this book.

We now return to the time-independent part of the Hamiltonian operator, H_1. The nuclear spin energy has already been set aside, and we may also separate both the electronic energy and the translational energy of the center of mass. The latter is rigorously separable and the former to a high degree of approximation (the *Born–Oppenheimer approximation*). We are, of course, interested in electronic states, but they may be regarded as stationary while a vibration–rotation transition takes place. This leaves us to consider, independently, the Hamiltonian for motions with respect to the center of mass.

The Hamiltonian for these motions can be separated into kinetic (T) and potential (V) energy operators,

$$H = T + V. \tag{3.11}$$

It is intuitively obvious that, given the aggregation of nuclei into a stable molecule, it will be convenient to state the problem in terms of vibrations of the nuclei with respect to each other and rotation of the entire molecule. A subtle difficulty is that a separation cannot always be made for the kinetic energy term. The potential energy, on the other hand, is only a function of nuclear separations and can be expressed in terms of *normal coordinates* (Q_k), i.e., amplitudes of the independent *normal modes* of vibration (classical frequencies, ν_k). A nonlinear molecule with N atoms ($N>2$) has three rotational degrees of freedom and $3N-6$ normal modes of vibration; a linear molecule has one less rotational degree of freedom and $3N-5$ normal modes.

For a molecule to have a stable equilibrium configuration (indicated by the subscript or superscript, e), there must be a minimum in the potential energy curve (Fig. 3.8). Near the position of equilibrium, the most important term in the potential energy will be quadratic in the normal coordinates,

$$V = \tfrac{1}{2}\sum_k \lambda_k Q_k^2 + \text{higher order terms in } Q_k, \tag{3.12}$$

where λ_k is a force constant. The first term on the right-hand side of

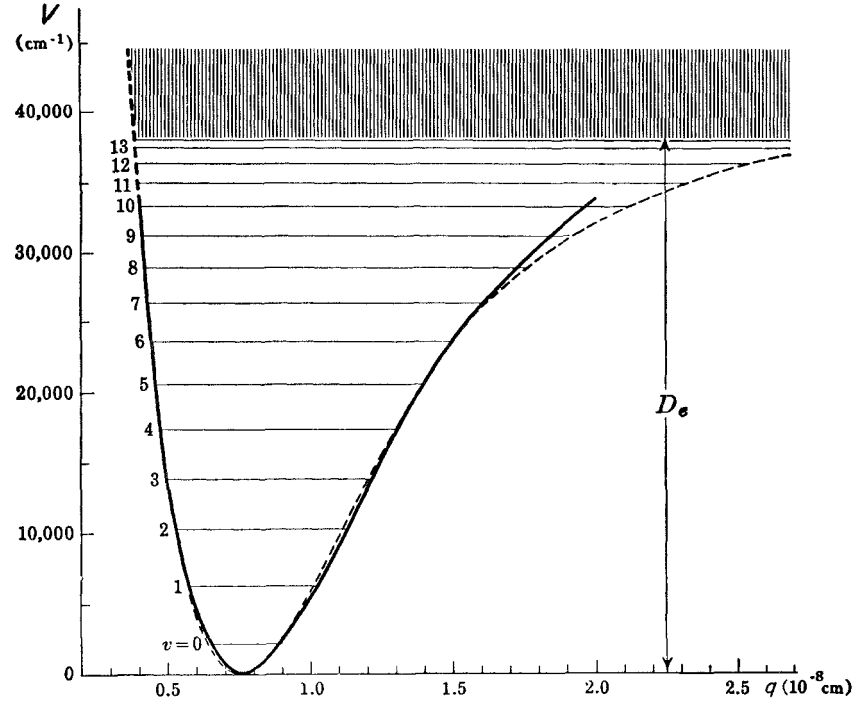


FIG. 3.8. Potential curve of the H_2 molecule in its ground electronic state. The heavy, solid line is the measured potential curve. The broken line is a commonly used approximation to the measurements. The horizontal lines represent the vibrational energy levels. The hatched area is the dissociation continuum and D_e is the dissociation energy. After Herzberg (1950).

(3.12) is the potential energy of a *harmonic oscillator*, while the higher order terms are *anharmonic*.

When we express the vibrational energy in terms of normal modes, we are working in a noninertial frame, rotating with the molecule. Transformation to a rotating frame automatically involves the appearance of two fictitious forces, *centrifugal force* and *Coriolis force*, both of which can affect the molecular vibrations. There is, therefore, a fundamental coupling between vibration and rotation that cannot be completely resolved.

The classical expression for the kinetic energy in terms of rotation and the normal modes of vibration is

$$T = \tfrac{1}{2} \sum_{\alpha,\beta} (J_\alpha - \pi_\alpha)\mu_{\alpha\beta}(J_\beta - \pi_\beta) + \tfrac{1}{2} \sum_k P_k^2. \tag{3.13}$$

Here α and β refer to two orthogonal axes, fixed with respect to the

molecule, $J_{\alpha,\beta}$ is a component of the total angular momentum, and P_k is the momentum coordinate conjugate to Q_k. $\pi_{\alpha,\beta}$ is the *vibrational angular momentum.* This is the essential coupling term between vibration and rotation. For a linear molecule it may be pictured as angular momentum associated with the rotary motion around the symmetry axis resulting from the coupling of two orthogonal oscillations out of phase with each other (§ 3.2.4). This term is zero for oscillations along the axis of a linear molecule and it is therefore zero for diatomic molecules, whose only vibration is of this type. It depends upon the amplitudes of the normal vibrations involved, as does the second, vibrational term on the left-hand side of (3.13). $\mu_{\alpha\beta}$ are components of a modified reciprocal inertia tensor. To second order, these coefficients depend upon the displacements of the atoms from their equilibrium positions and each can be written in the form

$$\mu_{\alpha\beta} = \mu^{e}_{\alpha\beta} + \text{terms involving } Q_k. \tag{3.14}$$

With a suitable choice of axes,

$$\mu^{e}_{\alpha\beta} = \delta_{\alpha\beta}/I^{e}_{\alpha},$$

where I^{e}_{α} is a principal moment of inertia of the molecule in its equilibrium configuration and $\delta_{\alpha\beta}$ is the Kronecker delta function.

The vibrational angular momenta are proportional to the amplitudes of the normal modes of vibration, so that $\pi \to 0$ as all $Q_k \to 0$. If we set Q_k equal to zero in (3.13) in all terms except P_k, we obtain the Hamiltonian for the *harmonic-oscillator, rigid-rotator* model:

$$H^0 = \tfrac{1}{2}\sum_{\alpha}\frac{J^2_{\alpha}}{I^{e}_{\alpha}} + \tfrac{1}{2}\sum_{k}(\lambda_k Q^2_k + P^2_k). \tag{3.15}$$

This Hamiltonian is the sum of independent rotational and vibrational Hamiltonians, without interacting terms. In §§ 3.2.2 and 3.2.3 we look at this model further. It has played a central role in the development of molecular spectroscopy and the nomenclature of the subject is derived from it. It gives valuable qualitative insights into the behavior of real molecules and it is a true first-order model for diatomic molecules (for which the π's are exactly zero). It is, however, not a correct first-order model for polyatomic molecules for the following reasons. First, we have neglected terms in the normal coordinates in the rotational term and have included terms of the same order in the vibrational terms. Second, while H^0 can yield a valuable approximation to the energy levels, the presence of the vibrational angular momentum in (3.13) can fundamentally alter the symmetry of the wave functions and affect both state populations and selection rules.

Although (3.15) is flawed as a first-order approximation for polyatomic molecules it is still a useful mathematical form that can yield basis functions for perturbation expansions. This is the approach of most numerical methods. Perturbations have been carried out on such basis functions up to fourth order. The higher order terms in the Hamiltonian (there are many of them) are classed generically as *interactions* (§ 3.2.4). Molecules for which such perturbation schemes converge rapidly, including all of the principal atmospheric radiators, are termed *semirigid.*

3.2.2. *The states of the harmonic-oscillator, rigid-rotator model*

(i) Vibrational states. The Hamiltonian in (3.15) is separable and we may deal with vibration and rotation independently, subsequently multiplying wave functions or adding energies for a combined state.

The energy levels for a harmonic vibrator are

$$E_v = \sum_k \mathbf{h}\nu_k(v_k + \tfrac{1}{2}), \tag{3.16}$$

where v_k is the vibrational quantum number (an integer) while k denotes the normal modes.

Diatomic molecules have only one normal mode, the stretching ν_1 frequency, shown in Fig. 3.9b. Carbon monoxide and nitric oxide fall into this class of molecule. So does hydrogen but, for homonuclear diatomic molecules, symmetry forbids the existence of any dipole moment and vibration–rotation spectra are absent. The spacing of vibrational levels for hydrogen is illustrated in Fig. 3.8. The spacing between the levels $v = 0$ and $v = 1$ is approximately twice that between $v = 9$ and $v = 10$. According to (3.16) the levels should be equally spaced. This reminds us that the harmonic-oscillator model is an idealization.

A number of atmospheric molecules are triatomic. For nonlinear molecules there are three normal modes, while for linear molecules there are four but two orthogonal bending modes (ν_{2a} and ν_{2b}) are degenerate (Fig. 3.9). Energy levels for triatomic molecules are specified by listing the quantum numbers in the form $(v_1 v_2 v_3)$.

Methane, with five nuclei, has nine normal modes, but when the spherical symmetry of the molecule is taken into account, only four are independent. Vibrational energy states for methane are specified by listing the four quantum numbers.

The existence of isotopic lines was mentioned in our overview of spectra in § 3.1. The frequency of a classical oscillator is given by

$$\nu = (1/2\pi)(\lambda/\mu)^{1/2}, \tag{3.17}$$

where λ is a force constant and μ is the reduced mass of the molecule.

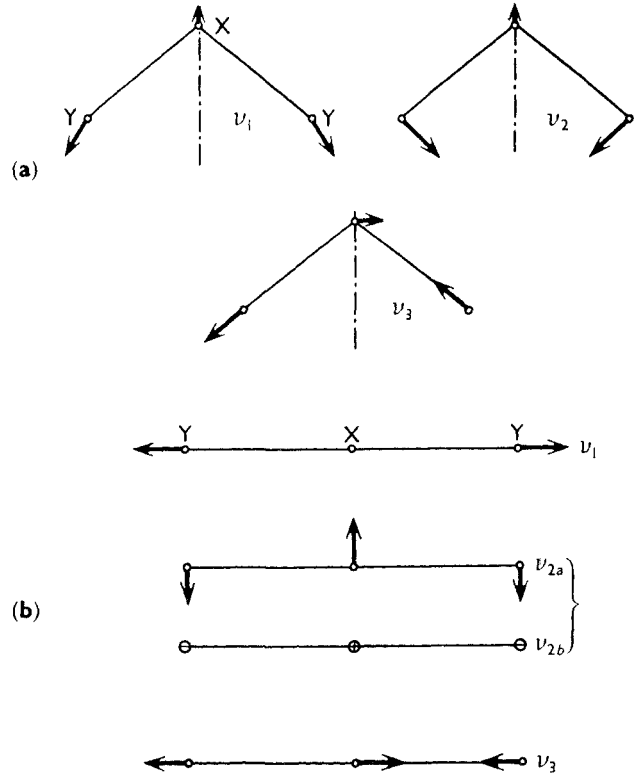


FIG. 3.9. Normal vibrations of a triatomic molecule. (a) H_2O is an example of a nonlinear triatomic molecule. (b) CO_2 is an example of a linear triatomic molecule. After Herzberg (1945).

The restoring force is caused by electronic forces that are not affected by the isotopic species of the nuclei, so that λ is the same for all species. If the superscripts i and j refer to two isotopic species,

$$\frac{\nu^i}{\nu^j} = \left(\frac{\mu^j}{\mu^i}\right)^{1/2}. \tag{3.18}$$

As a case in point, this calculation leads to a frequency shift for the $^{13}C^{16}O$ band shown in Fig. 3.4 of 47.7 cm^{-1} with respect to the normal species.

(ii) Rotational states. Turning now to the energy levels of a rigid rotator, the nomenclature is based on the principal moments of inertia and is given in Table 3.1. The angular momentum **J** is the sum from all sources, electronic, vibrational, rotational, and is quantized. The quan-

Table 3.1. Nomenclature based upon moments of inertia

Moments of inertia	Class	Examples
$I_A = 0, I_B = I_C \neq 0$	Linear	CO_2, N_2O, CO, NO
$I_A \neq 0, I_B = I_C \neq 0$	Symmetric top	$CFCl_3$
$I_A = I_B = I_C$	Spherical top	CH_4
$I_A \neq I_B \neq I_C$	Asymmetric top	H_2O, O_3

tization is

$$|\mathbf{J}| = \frac{\mathbf{h}}{2\pi}[J(J+1)]^{1/2}, \tag{3.19}$$

where J is the rotational quantum number (an integer). For the harmonic-oscillator, rigid-rotator model this quantization also applies to the rotational angular momentum alone.

There are two constants of the rotational motion: the component of the angular momentum along a unique symmetry axis (if there is one), and the component in a fixed direction in space. The former, **K**, is quantized

$$|\mathbf{K}| = \frac{K\mathbf{h}}{2\pi} \qquad \text{with } K < J. \tag{3.20}$$

The quantization (3.20) is important only for symmetric tops. The quantum number K has only positive values, so that for each value of $K \neq 0$ there are two vector directions of **J** that can give rise to the same value. These levels are, therefore, doubly degenerate.

The second directional quantization is for the component of momentum in a fixed spatial direction, but it gives rise to no energy interactions except with polarized radiation, which we shall not consider. There are $2J + 1$ values of this momentum component.

The energy levels for a rigid, symmetric top are

$$E_r = \mathbf{hc}F(J, K), \tag{3.21}$$

where

$$F(J, K) = BJ(J+1) + (A - B)K^2, \tag{3.22}$$

and the *rotational constants* are defined by

$$B, A = \mathbf{h}/8\pi^2\mathbf{c}I_{B,A}. \tag{3.23}$$

With these definitions, F, A, and B are all expressed in cm^{-1}.

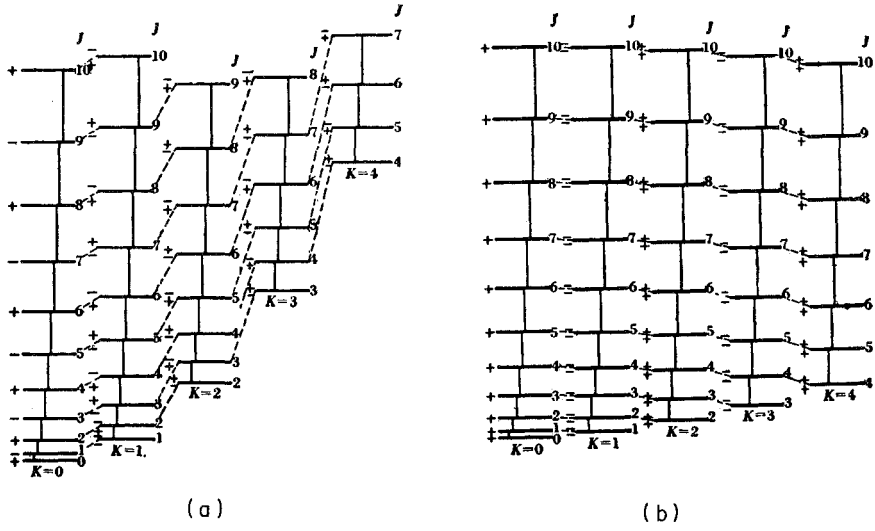


FIG. 3.10. Energy levels of a symmetric top. (a) Prolate; (b) oblate. The + and − signs indicate parity. After Herzberg (1945).

The energy levels (3.22) differ in general appearance according to whether the top is *prolate* ($A > B$) or *oblate* ($A < B$). Examples of both are shown in Fig. 3.10.

For spherical tops $A = B$, while for linear molecules $K = 0$. In both cases, (3.22) reduces to

$$F(J, K) = BJ(J + 1). \tag{3.24}$$

These energy levels appear in Fig. 3.10 in the left-hand columns, for which $K = 0$. This apparent similarity between linear molecules and spherical tops is, however, illusory. The spherical top is highly degenerate and, when the degeneracies are resolved, individual J levels turn into multiplets.

This brings us to the complicated case of the asymmetric top. It cannot be described in any simple terms but can be pictured as an intermediate stage between a prolate and an oblate symmetric top (Fig. 3.11). Each J level of a symmetric top is split into $J + 1$ sublevels with different K. For an oblate top, the energy decreases as K increases while for a prolate top the order is reversed (see Fig. 3.10a and b). Each level with $K \neq 0$ is doubly degenerate. The first deviation from a symmetric top gives rise to a splitting of these levels and the levels for an asymmetric top can be found, to a first approximation, by interpolating between the prolate and oblate symmetric tops, as illustrated in Fig. 3.11. K now plays an ambiguous role and is no longer a useful quantum number.

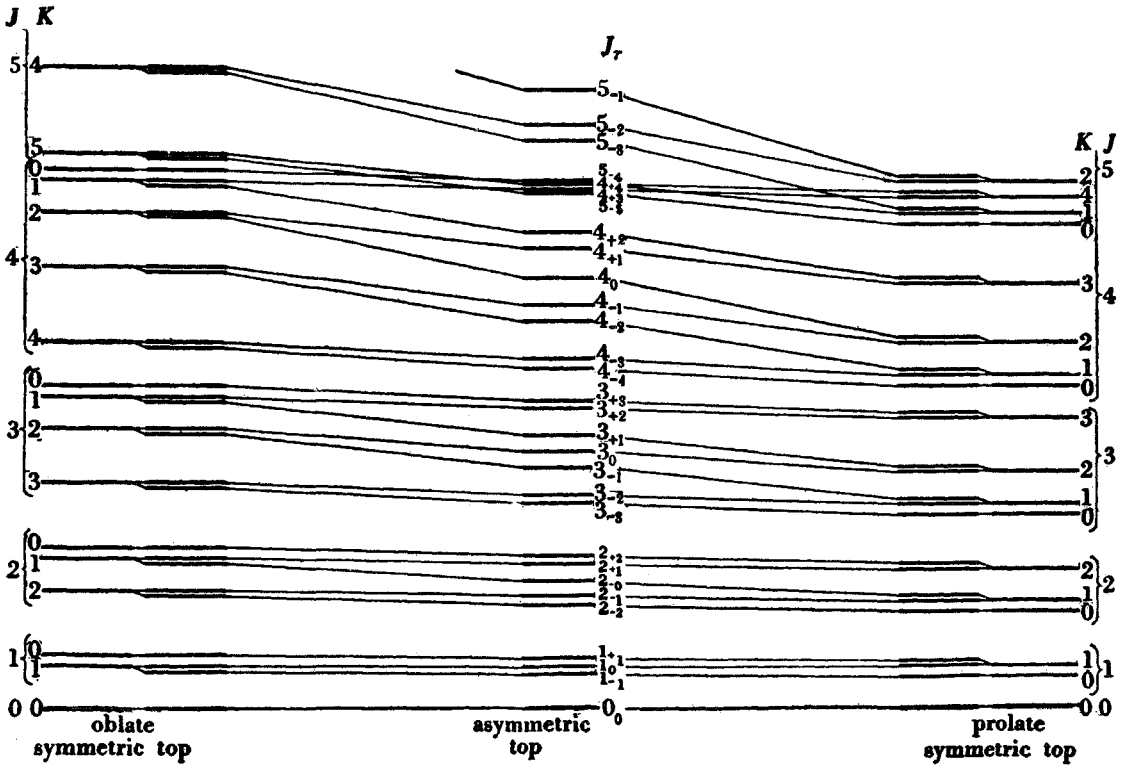


FIG. 3.11. Energy levels of an asymmetric top. The asymmetric top levels are correlated with those of prolate and oblate symmetric tops. After Herzberg (1945).

There are $2J + 1$ levels of the asymmetric top for each value of J. The levels are often labeled arbitrarily in the order of their energies, with an integer, τ, ranging from $-J$ to $+J$. This integer is set as a subscript to the J but is not a quantum number. Reference to Fig. 3.11 shows that $\tau = K_A - K_C$ where K_A and K_C are the related K values for the prolate and oblate symmetric tops. Some writers set these two values of K as

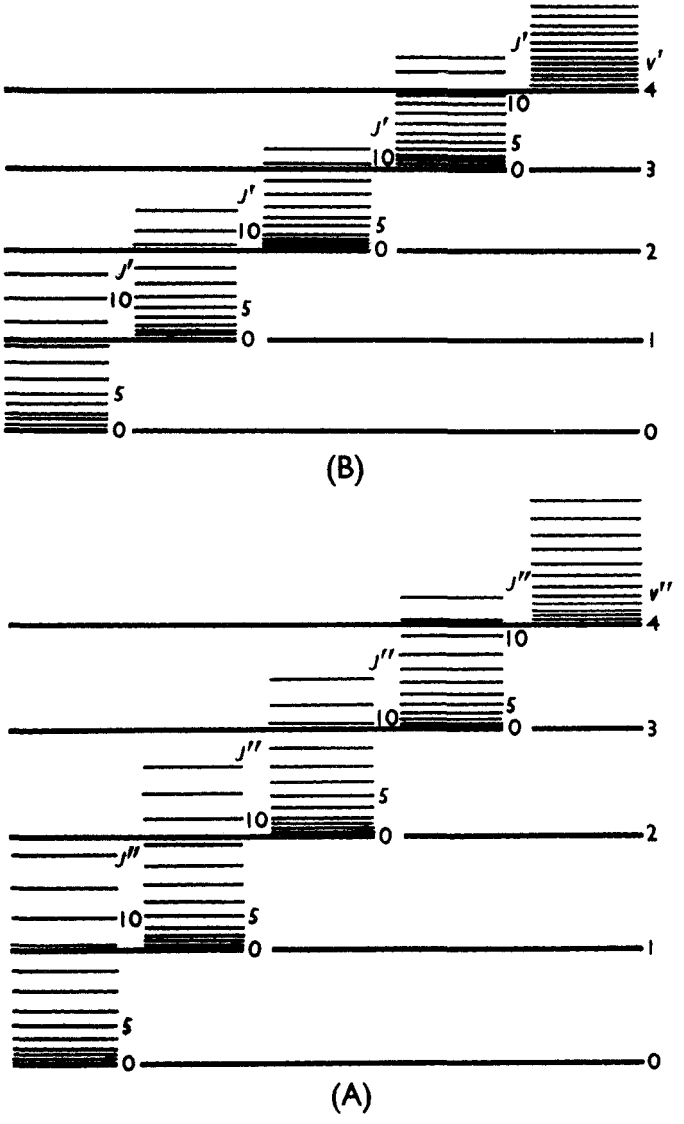


FIG. 3.12. Energy levels of an idealized diatomic molecule. (A) and (B) represent two electronic states of the molecule. After Herzberg (1950).

subscripts to J, as a level designation, in preference to the arbitrary τ enumeration.

The energy levels for an asymmetric top are given by (3.21) with

$$F(J) = \tfrac{1}{2}(A + C)J(J + 1) + \tfrac{1}{2}(A - C)E(J, \tau, \kappa), \tag{3.25}$$

where

$$\kappa = \frac{2[B - (A + C)/2]}{A - C}.$$

Tables of $E(J, \tau, \kappa)$ are available in the literature.

(iii) Vibrational–rotational states. Figure 3.12 shows how vibrational and rotational energy levels can combine in the case of a diatomic molecule. Two electronic levels, A and B, are indicated. A general absorption involves a transition from A, v'', J'' to B, v', J'. This simple addition of energy levels is, of course, correct for the harmonic-oscillator, rigid-rotator model, but does not carry over to polyatomic molecules. Interactions will be discussed in § 3.2.4.

3.2.3. *Selection rules and line intensities*

To calculate the line intensity from (3.10) we must know the population, n_j, of the lower state and the square of the matrix element of the dipole moment of the transition, $|\mathbf{R}_{ij}|^2$. From (2.45), n_j is proportional to the statistical weight of the state, g_j. This number is made up of enumerations of the degeneracies of the electronic, vibrational, rotational, and, we must now add, the nuclear spin states of the molecule. At first, we consider the rotational and the vibrational states only.

The vibrational states of a molecule have equal statistical weights if we treat each degenerate state independently. The rotational levels have a fundamental degeneracy associated with the component of the angular momentum in a fixed direction in space. There are $2J + 1$ orientations and they cannot be distinguished.

Other angular degeneracies are strict only for the harmonic-oscillator, rigid-rotator model and are resolved in real molecules. Each K level of a symmetric top, except for $K = 0$, is double. This degeneracy is usually resolved, as is an additional factor of $2J + 1$ in the degeneracy for a J level of a spherical top. This latter factor owes its existence to the indeterminacy of the direction of the angular momentum with respect to the spherical molecule. When resolved, such levels appear as $2J + 1$ manifolds; see the spectrum of methane in Fig. 3.2a.

To illustrate the calculation of vibrational and rotational state populations, we use the example of a linear molecule. The fraction of

molecules in the (v, J) level is the product of the fractions in the v level and the J level separately. From (2.45), (3.16), and (3.24)

$$\frac{n(v)}{n} = Q_v^{-1} \exp\left(\frac{-v\mathbf{h}\nu}{\mathbf{k}\theta}\right), \tag{3.26}$$

where

$$Q_v = \sum_{v=0}^{\infty} \exp\left(\frac{-v\mathbf{h}\nu}{\mathbf{k}\theta}\right) = \left[1 - \exp\left(\frac{-\mathbf{h}\nu}{\mathbf{k}\theta}\right)\right]^{-1} \tag{3.27}$$

is the *vibrational partition function,* and

$$\frac{n(J)}{n} = Q_r^{-1}(2J+1) \exp\left[\frac{-B\mathbf{hc}J(J+1)}{\mathbf{k}\theta}\right], \tag{3.28}$$

where Q_r is the *rotational partition function.* Under normal atmospheric conditions, $B\mathbf{hc}/\mathbf{k}\theta \ll 1$ and

$$Q_r = \mathbf{k}\theta/\mathbf{hc}B. \tag{3.29}$$

Equations (3.16) and (3.18) are plotted as continuous functions of v or J in Figs. 3.13 and 3.14, and Table 3.2 gives populations of certain lowest excited vibrational states at two temperatures. The iodine molecule, illustrated in Fig. 3.13, has an unusually low vibrational frequency.

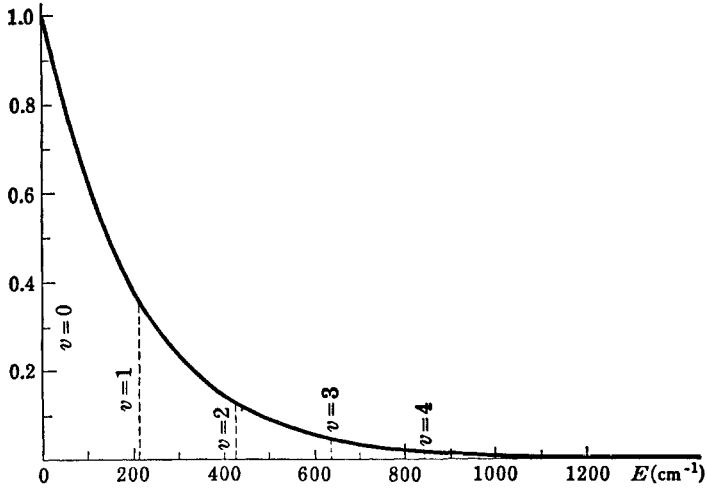


FIG. 3.13. Thermal distribution of vibrational levels. The data correspond to the iodine molecule, with $\nu_1 = 213.2\,\text{cm}^{-1}$, and for a temperature of 300 K. After Herzberg (1950).

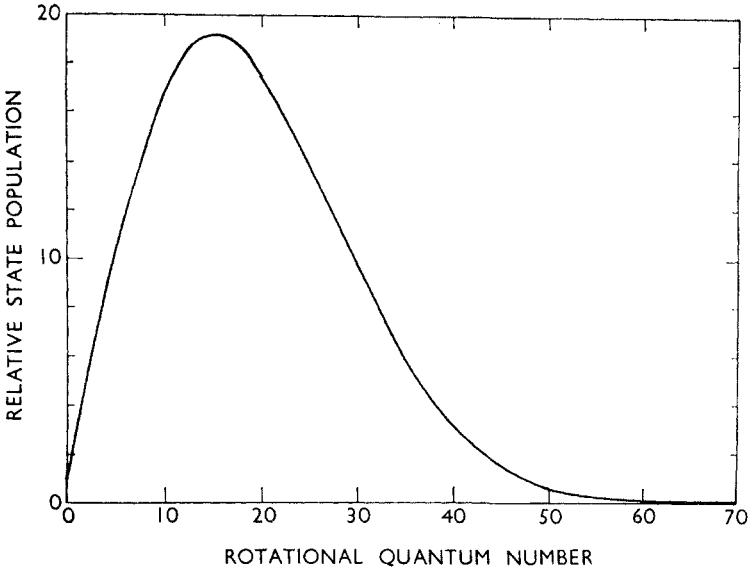


FIG. 3.14. Thermal distribution of rotational levels. The nuclear weights are assumed to be the same for all J. $B = 0.418\ \text{cm}^{-1}$ (nitrous oxide) and the temperature is 300 K.

For most molecules at atmospheric temperatures, the population of the first vibrational level is very small and varies rapidly with temperature. The rotational states have a maximum population for a J value approximately equal to

$$J_{\max} = \left(\frac{\mathbf{k}\theta}{\mathbf{hc}B}\right)^{1/2} - \tfrac{1}{2}. \tag{3.30}$$

We have not considered nuclear spins up to this point. They

Table 3.2. Ratio of the populations of the lowest two vibrational states for some molecules at 300 K and 1000 K

Gas	ν_1 (cm^{-1})	$\exp(-h\nu_1/k\theta)$ 300 K	$\exp(-h\nu_1/k\theta)$ 1000 K
H_2	4160.2	2.16×10^{-9}	2.51×10^{-3}
HCl	2885.9	9.77×10^{-7}	1.57×10^{-2}
N_2	2330.7	1.40×10^{-5}	3.50×10^{-2}
CO	2143.2	3.43×10^{-5}	4.58×10^{-2}
O_2	1556.4	5.74×10^{-4}	1.07×10^{-1}
Cl_2	556.9	6.92×10^{-2}	4.49×10^{-1}
I_2	213.2	3.60×10^{-1}	7.63×10^{-1}

Source: After Herzberg (1950).

contribute a *hyperfine structure* to the energy levels that can be detected in the microwave spectrum but does not concern us, except that these levels contribute to the statistical weights. The nuclear spin weights for linear molecules alternate from even to odd J creating an alternation in rotational state populations.

A molecule must possess a symmetry axis before nuclear spins are important, e.g., the two-fold axes of $^{16}OC^{16}O$ and $^{1}HO^{1}H$. Consider the example of carbon dioxide, and rotate the molecule 180° about this symmetry axis. There is no detectable change in the physical situation and solutions to the wave equation should be the same. However, these solutions involve only the square of the wave function, $|\psi|^2$, and it is possible for the wave function itself to remain constant (a *symmetric* wave function) or change sign (an *antisymmetric* wave function). The vibrational wave function is always symmetric, but the symmetry of rotational levels alternates and even J levels are antisymmetric while odd J levels are symmetric. Nuclear spin levels can also be either symmetric or antisymmetric; if I is the spin of each of the two identical nuclei that are interchanged by the rotation, the number of symmetric wave functions is $(2I+1)(I+1)$ while the number of antisymmetric wavefunctions is $(2I+1)I$. For our example of ^{16}O, I is zero so that the antisymmetric wave functions are missing.

The rotational and the nuclear wave functions are not independent for the reason that the symmetry of the combined wave function is known. The rotation is equivalent to the exchange of two identical nuclei and, depending upon whether they follow Fermi–Dirac or Bose–Einstein statistics, the overall wave function must be antisymmetric or symmetric, respectively, with respect to the interchange. Oxygen nuclei follow Bose–Einstein statistics and therefore symmetric (or antisymmetric) nuclear wave functions combine with symmetric (or antisymmetric) rotational wave functions, the net result of which is that odd-numbered J levels of the carbon dioxide molecule are not populated. This strong restriction will not apply to species such as $^{16}OC^{18}O$ that have no two-fold symmetry axis. Asymmetric isotopic species of carbon dioxide do not have alternating populations of rotational levels.

Nuclear weights can also be calculated for nonlinear molecules, but the calculation can be quite complicated and we shall not discuss the matter further, but turn instead to the question of transition probabilities.

According to (3.8), the electric dipole moment must be a function of the configuration coordinates if the transition probability is to be nonzero. To evaluate the vibrational matrix elements we need to express this variation of the electric moment in terms of the normal coordinates. We may expand each component of the molecule-fixed dipole moment, M_α ($\alpha = 1, 2, 3$) in a Taylor series,

$$M_\alpha = M_\alpha^e + \sum_k \mu_k^\alpha Q_k + \tfrac{1}{2} \sum_{k,l} \mu_{k,l}^\alpha Q_k Q_l + \cdots, \tag{3.31}$$

where M_α^e is the equilibrium value of the component of the dipole moment and the μ's are expansion coefficients. It is an additional assumption of the harmonic-oscillator, rigid-rotator model that only the first two terms on the right-hand side of (3.31) need to be considered. Of these, the first does not change in a vibrational transition. The matrix element for the linear term in (3.31) is zero unless one vibrational quantum number changes and that by unity

$$\Delta v_k = \pm 1. \tag{3.32}$$

This is the selection rule for a harmonic oscillator. Combined with the known energy levels, (3.16), and Planck's relation, (3.4), it means that allowed vibrational transitions give the same frequencies ν_1, ν_2, ν_3, etc. as the classical normal modes. For real molecules these are not the only transitions that occur, but they usually give rise to the strongest bands; these are the *fundamentals*.

Not all of the normal modes may be active, however. There is still a question as to the size of the expansion coefficients in (3.31). It is possible for the molecular symmetry to be such that the linear coefficient is identically zero, leading to an *inactive* fundamental mode. In the case of methane, for example, there are nine normal modes but only four are distinct (ν_1, nondegenerate; ν_2, doubly degenerate; ν_3 and ν_4 both triply degenerate). Of these only ν_3 and ν_4 are active in the infrared spectrum.

The dipole moment is a vector quantity and dependence of resolved components upon angle follows from this. Apart from a multiplicative constant, (3.6) is readily evaluated and gives a strict selection rule,

$$\Delta J = 0, \pm 1, \qquad \text{but not } J = 0 \text{ to } J = 0. \tag{3.33}$$

For symmetric tops there is a strict rule related to the conservation of angular momentum,

$$\Delta K = 0. \tag{3.34}$$

Finally, there is a *parity rule* whereby positive levels combine only with negative levels and vice versa,

$$+ \rightarrow - \text{ and } - \rightarrow + \text{ only.} \tag{3.35}$$

The parity rule refers to sign changes of the wave function for reflection through the origin of coordinates. For a symmetric linear molecule there is no distinction between reflection through the origin and rotation of 180° about the symmetry axis. Even and odd J levels, therefore, have opposite parity, with the important consequence that, for such molecules, $\Delta J = 0$ is not permitted. These restrictions apply to the

total angular momentum and are not necessarily true for the rotational momentum alone.

For the harmonic-oscillator, rigid-rotator model these selection rules apply to the rotational states and

$$\Delta J = \pm 1. \tag{3.36}$$

For transitions $(v_{k+1}, J+1) \leftarrow (v_k, J)$ we find, from (3.24), (3.36), (3.16), and (3.32),

$$\nu = \nu_k \pm 2B\mathfrak{c}(J+1). \tag{3.37}$$

For a pure rotation band, set $\nu_k = 0$; but, in this case, only the positive sign in (3.37) is admissible.

Equation (3.37) represents two branches of lines equally spaced about the vibrational frequency. The high-frequency branch ($\Delta J = +1$) is the R-branch and the low-frequency branch ($\Delta J = -1$) is the P-branch. Figures 3.3 and 3.4 show examples for nitrous oxide and carbon monoxide, respectively. P- and R-branch lines are labeled with the J value of the initial state in parentheses, e.g., R(0), R(1), R(2), . . . , P(1), P(2), . . . etc.

Figure 3.5 has two R-branch and three P-branch lines of carbon dioxide, but in the gap between the two branches there is a very strong, partially resolved feature; this is the $\Delta J = 0$ transition, a *Q-branch,* not allowed in the harmonic-oscillator, rigid-rotator model for a linear molecule. The explanation for the occurrence of this branch must wait for a discussion of interactions; but, even for our simple model, a Q-branch is possible for a prolate symmetric top because the two degenerate K levels ($K = 0$ excepted) have opposite parities and the parity rule no longer forbids $\Delta J = 0$.

For an asymmetric rotator, the parity rule has no simple consequences and all three branches intermingle with such a variety of line spacings as to present an almost random appearance (Figs. 3.6 and 3.7).

Selection rules tell us whether or not a matrix element is zero. If it is not, we need to be able to calculate its magnitude. For vibrational transitions this calculation involves the expansion coefficients in (3.31), but these must usually be obtained empirically. One of the important applications of molecular spectroscopy is to establish the coefficients in (3.31) and the potential curve, Fig. 3.8.

For a pure rotational band, or for the relative rotational intensities in a vibration-rotation band, the situation is different. We only have to deal with geometric factors and the magnitude of the permanent dipole moment. Matrix elements can be calculated quite simply in the rigid-

rotator approximation. For a linear molecule

$$\begin{aligned} |\mathbf{R}_{ij}|^2 &= |\mathbf{M}|^2(J+1)/(2J+1) \qquad \text{for } (J+1) \leftarrow J \\ &= |\mathbf{M}|^2(J+1)/(2J+3) \qquad \text{for } (J+1) \rightarrow J. \end{aligned} \tag{3.38}$$

If $J \gg 1$, both expressions become independent of J and relative line intensities are, from (3.10), proportional to the populations of the lower states. This is the quantity, with the rotational constant for nitrous oxide, which is shown in Fig. 3.14. We now see that this is also what one branch of a simple vibration–rotation band will look like. To obtain the full appearance of a band, the P- and R-branches must be generated by reflection in the vertical axis of Fig. 3.14 and the origin placed at the vibrational frequency. This gives a remarkably good representation of the band shown in Fig. 3.3.

3.2.4. *Interactions*

To calculate energy levels to the precision required by modern spectroscopic techniques requires fourth-order perturbation expansions using basis functions that are usually provided by the harmonic-oscillator, rigid-rotator model. Very many perturbation terms are required, involving higher order coefficients in the expansions of the potential energy, (3.12), and the inertia matrix, (3.14), combined in complex products with the vibrational angular momenta. No simple description of these terms is possible except in very restricted circumstances. However, the history of the subject has been one of attempts to identify successive corrections to a noninteracting model with physical "explanations" of each different interaction. The names of these interactions now form part of the language structure of the subject and we shall introduce them in the simplest possible contexts. The reader should be warned, however, that this approach is fraught with confusing issues, and that the small space allotted to the subject in this section is useful mainly as a glossary.

A starting point is provided by an expression for the energy levels of a diatomic molecule, allowing a single anharmonic term in the potential function, (3.12). If the rotational energy is small compared to the dissociation energy (Fig. 3.8), the result is

$$\begin{aligned} E_{v,J}/\mathbf{hc} = (\nu_e/\mathbf{c})(v+\tfrac{1}{2}) - (x_e\nu_e/\mathbf{c})(v+\tfrac{1}{2})^2 + B_eJ(J+1) \\ - D_eJ^2(J+1)^2 - \alpha_e(v+\tfrac{1}{2})J(J+1). \end{aligned} \tag{3.39}$$

The subscript e indicates the equilibrium configuration. ν_e is the frequency of the normal mode, B_e is the equilibrium rotational constant, and v and J are the vibrational and rotational quantum numbers. x_e is an expansion coefficient characterizing the first anharmonic term. In

addition,

$$D_e = \frac{4\mathbf{c}^2 B_e^3}{\nu_e^2}, \tag{3.40}$$

$$\alpha_e = 6\left(\frac{x_e \mathbf{c} B_e^3}{\nu_e}\right)^{1/2} - \frac{6\mathbf{c}B_e^2}{\nu_e}. \tag{3.41}$$

The first and third terms on the right-hand side of (3.39) are the familiar energy levels of the harmonic-oscillator, rigid-rotator. The second term on the right-hand side can be interpreted as an *anharmonic correction* to the frequency of the normal mode. It is a function of the vibrational quantum number, so that the frequencies of permitted vibrational transitions with $\Delta v = 1$ now depend on the quantum number of the lower state. For example,

$$\text{for } v = 1 \leftarrow 0, \qquad \frac{\Delta E_{v,J}}{\mathbf{hc}} = \frac{\nu_e}{\mathbf{c}}(1 - 2x_e),$$

$$\text{for } v = 2 \leftarrow 1, \qquad \frac{\Delta E_{v,J}}{\mathbf{hc}} = \frac{\nu_e}{\mathbf{c}}(1 - 4x_e).$$

The second example, $2 \leftarrow 1$, is an *upper-state band.* It no longer has the same frequency as the *ground-state band,* $1 \leftarrow 0$. An example is shown in Fig. 3.3; the weak, superposed band has the first ν_2 level as its lower state.

Anharmonic corrections must be made to all vibrational levels. Corrections to some of the energy levels of water vapor are shown in Fig. 3.15.

Anharmonicity also changes the selection rules from those for a harmonic oscillator, (3.32); all integral changes of the quantum numbers are now allowed. $\Delta v = 2$ gives the first *overtone band,* with twice the frequency of the normal mode. Simultaneous changes in two different vibrational quantum numbers give rise to *combination* and *difference bands,* with frequencies that are sums or differences of the normal-mode frequencies. This change in selection rules together with the unequal spacings of the vibrational levels can lead to a rich variety of vibrational bands from a term scheme such as that in Fig. 3.15. The apparent preponderance of ground-state bands in the atmospheric spectrum is due partly to low atmospheric temperatures and small upper-state populations (Table 3.2), and partly to the fact that overtone and combination bands normally have smaller transition probabilities than fundamentals.

The fourth term on the right-hand side of (3.39) can be rationalized as a *centrifugal stretching* term modifying the rotational constant. It is not an anharmonic term because it does not involve the anharmonic

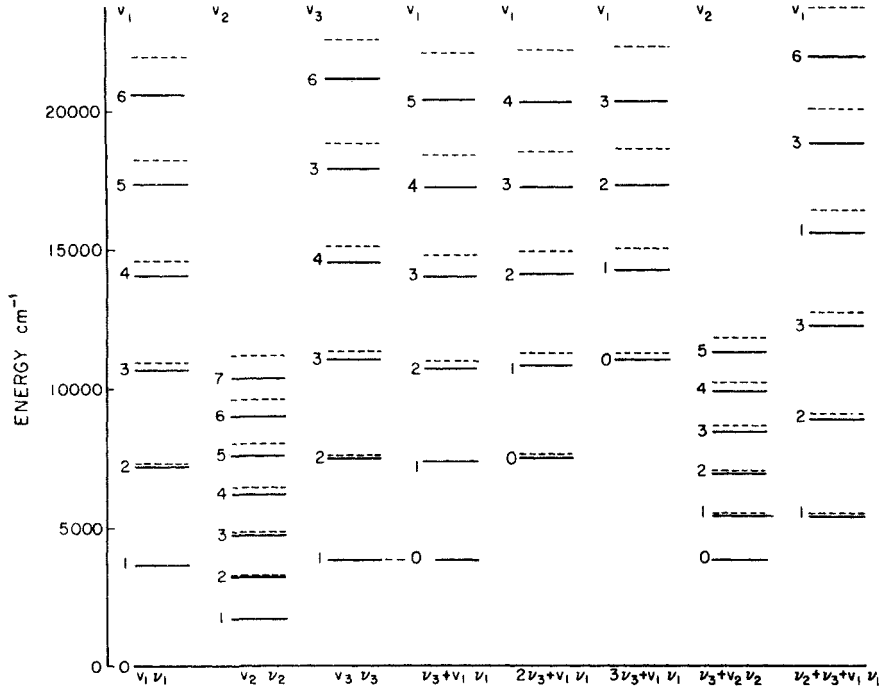


FIG. 3.15. Vibrational energy levels of water vapor. The solid lines represent the measured energy levels while the broken lines indicate the levels anticipated in the absence of anharmonicities. Because anharmonic terms do not combine in any simple way, a separate term scheme is required for each combination of modes. Not all possibilities are shown. After Herzberg (1945).

coefficient x_e. It should have also appeared in the harmonic-oscillator, rigid-rotator model if that model were fully consistent. It has the effect of modifying the simple sequence of equally spaced P- and R-branch lines for a linear molecule. In place of (3.37) we now obtain

$$\nu = \nu_k \pm [2B\mathbf{c}(J+1) - 4D\mathbf{c}(J+1)^3]. \tag{3.42}$$

The final term on the right-hand side of (3.39) can be combined with the third term to give an effective rotational constant, $B_e - \alpha_e(v + \frac{1}{2})$. This treats the term as an anharmonic adjustment to the rotational constant, an interpretation that is consistent for the first term on the right-hand side of (3.41), because this term involves the anharmonic constant, x_e. But this is not consistent with the second term on the right-hand side of (3.41), which is harmonic. This second term is better described as the result of *Coriolis interactions*. One important consequence of its presence is that it resolves the structure of a Q-branch, if

that branch is permitted. The third and fourth terms on the right-hand side of (3.39) are functions of J alone and they contribute nothing to the energy change if J does not change (as is the case for a Q-branch). The Coriolis term, on the other hand, contributes a term,

$$\Delta\nu = -\alpha_e \Delta v J(J+1), \tag{3.43}$$

splitting the Q-branch into an array with a spacing increasing with the J value. A carbon dioxide Q-branch in Fig. 3.5 shows this behavior.

The nature and importance of one aspect of Coriolis interactions can be seen from Fig. 3.9b. The Coriolis force acts at right angles to the motion. Applied to the ν_{2a} motions, this produces a pattern of forces that reinforces the motions of ν_3. Each of these normal modes can therefore force a response in the other.

We have seen that anharmonic terms and Coriolis interactions can both give interactions between levels. Given an example, such as the levels of water vapor in Fig. 3.15, and also considering the rotational levels superposed on all of them, it is clear that many levels must be very close to each other and that there is much opportunity for close *resonances*. When resonances occur the idea of independent states of a molecule breaks down and the resonating states, assuming that they can combine, will form a single state with strongly perturbed energies. Calculations of perturbed states present no difficulties for an approach based upon perturbation expansions; but the questions can be very difficult to understand in a framework of ideas based upon independent levels, such that as outlined in §§ 3.2.2 and 3.2.3.

The best-known case of *Fermi resonance*, as this resonance is named, is found in the carbon dioxide molecule. The $2\nu_2$ level lies very close to the ν_1 level. Fermi resonance between these levels leads to two combined levels, shifted from both ν_1 and $2\nu_2$ [see the (100) and the (02^00) levels in Fig. 3.16]. At the same time, the rotational constants of the two levels are perturbed, but in such a way that the sum over the two levels is unchanged.

The $2\nu_2$ level of the carbon dioxide molecule is doubly degenerate because of the fundamental degeneracy of the ν_2 vibration (Fig. 3.9) but, because of selection rules on the vibrational angular momentum, only one of these can combine with ν_1, leaving the other, (02^20) in Fig. 3.16, unperturbed.

We now return to the question of the vibrational angular momentum, the existence of which is fundamental to many issues in the interactions of vibration and rotation. It appeared in the expression for the classical kinetic energy, (3.13), and was subsequently omitted partly because of the conceptual difficulties that it introduces. Its nature can be illustrated from Fig. 3.9b. If the ν_{2a} and ν_{2b} vibrations combine with a $\pi/2$ phase difference between them, the central atom will perform a

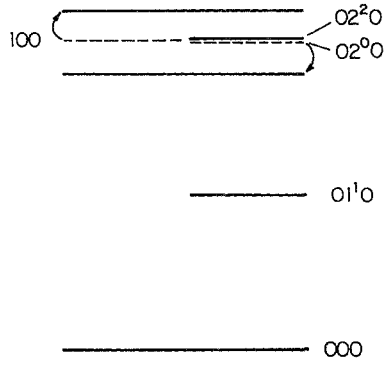


FIG. 3.16. Fermi resonance in CO_2. The broken lines go over to the solid lines indicated by the arrows under the influence of Fermi resonance. After Herzberg (1945).

rotary motion about the symmetry axis and give rise to a component of angular momentum along this axis. This momentum is quantized with a quantum number l (positive integer) that is less than or equal to v_2 and is closely analogous to the quantum number K for the symmetric top. Since the effect of the vibration is to transform a linear molecule into a symmetric top, this is not a surprising result. The quantum number l is sometimes added as a superscript to v_2 in the designation of the vibrational level, e.g., (01^10), (02^00), and (02^20) in Fig. 3.16.

The energy levels of a linear molecule (omitting centrifugal and anharmonic terms) can be obtained from (3.22) by taking the limit $A \ll B$ (with K replaced by l),

$$F(J, l) = BJ(J+1) - Bl^2. \tag{3.44}$$

Since (3.44) depends only upon l^2, each level for $l \neq 0$ is doubly degenerate, as we also saw to be the case for the quantum number, K. This degeneracy is resolved by Coriolis interactions even in a harmonic-vibrator model. For $l = 1$, the level shift has been given approximately by

$$\Delta F = (\mathbf{c}B_e^2/\nu)(v_2+1)J(J+1). \tag{3.45}$$

This leads to a doubling of lines known as *l-type doubling.*

l-type doubling is illustrated in Fig. 3.17. The letters C and D denote transitions involving the $+l$ and $-l$ levels, respectively. The spacing between C, D pairs follows from (3.45) and the midpoint of each pair from (3.43); both aspects of the Coriolis interaction are required to explain the interesting structure shown in Fig. 3.17.

The vibrational angular momentum also plays a part in selection

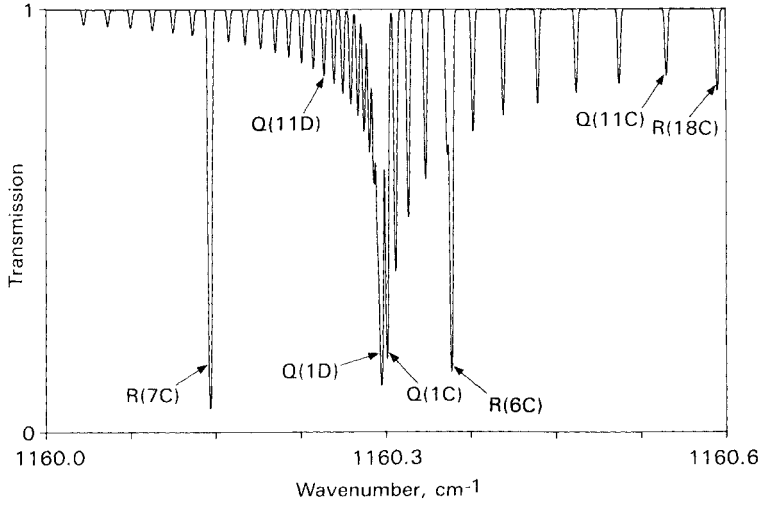


FIG. 3.17. *l*-type doubling in a Q-branch of N_2O near 8.62 μm. This is a synthetic spectrum. All strong lines except for the three R-branch lines indicated belong to the Q-branch of $(03^10) \leftarrow (01^10)$. Four Q-branch lines are also labeled to assist with the identification of pairs of lines caused by *l*-type doubling, such as Q(1C) and Q(1D), Q(11C) and Q(11D). This spectrum is based on physical conditions very different from those occurring in the earth's atmosphere.

rules and, for linear molecules in particular, it strongly influences the appearance of a band. There is a strict selection rule,

$$\Delta l = 0, \pm 1. \tag{3.46}$$

Transitions with $\Delta l = \pm 1$ give rise to *perpendicular bands.* Since the parities of odd and even l levels differ, a change in l can allow $\Delta J = 0$. Perpendicular bands can therefore have strong Q-branches. The CO_2 band shown in Fig. 3.5 results from the transition $(11^10) \leftarrow (02^00)$, and is a perpendicular band.

Bands for which $\Delta l = 0$ are known as *parallel bands.* If the transition is from $l = 0$ to $l = 0$ there is no relief from the parity rule and Q-branches are forbidden. However, for $l \neq 0$ the doubly-degenerate levels have opposite parities and weak Q-branches may be allowed. Figure 3.17 is an example of a Q-branch in a parallel band.

3.3. The shape of a spectral line

3.3.1. *Introduction*

An excited molecule will decay spontaneously to a lower state with the emission of a photon (§ 2.2.3), even it it is completely undisturbed. Each

state (apart from the ground state) has a finite lifetime and, according to the uncertainty principle, each must possess a finite but narrow range of energies. When transitions take place to or from an excited state a narrow band of frequencies results, constituting a spectral line of finite width. If spontaneous emission were the only cause of line broadening the width would be the *natural line width,* and the shape the *natural line shape.*

Natural line widths do not concern us in practice, because they are invariably small compared to the line widths of atmospheric lines. In practice, lines are broadened for one of two other reasons, either because of the Doppler shifts resulting from molecular thermal motions (*Doppler broadening*), or because of interactions between pairs of molecules (*pressure broadening* or *collision broadening*). Doppler broadening and collision broadening are usually treated as independent phenomena although, when we come to consider *Dicke's theory* (§ 3.3.6), we shall find that this is not strictly justifiable. Nevertheless, we shall follow this course. Doppler effects, taken independently of collisional effects, are comparatively easy to deal with. Collision broadening, on the other hand, is a topic of extraordinary complexity.

Fortunately, collision broadening has some simple aspects when the time spent in close interaction is very small compared to the time spent between collisions (*impact theories*). Impact conditions are always found in the earth's atmosphere. The paradigm of impact theories, is the *Michelson–Lorentz theory* (§ 3.3.2); it is oversimplified according to modern ideas, and we introduce it solely to illustrate principles.

When two molecules approach one another, a time-dependent term appears in the Hamiltonian (3.2). Under the assumptions of impact theory and a number of other simplifications, it is possible to calculate the Michelson–Lorentz parameters from quantum theory. This is the *Anderson–Tsao–Curnutte* theory, which is widely used to calculated line widths that have not been directly measured.

If the interaction Hamiltonian is large the situation may be very complicated. We may have to consider composite molecules (*dimers*); energy levels may be distorted and symmetries may be destroyed, allowing *collision-induced* transitions, not discussed in § 3.2 (see § 3.4). Under less extreme conditions, the time spent during collisions and the size of the interaction at that time may combine severely to distort the wings of a line. Line-wing shapes may sometimes be treated by *statistical theories* (§ 3.3.5) in which the interactions between stationary molecules are considered.

None of the theories mentioned so far embraces the full range of interactions in the time domain, and it is difficult for the nonspecialist to understand how the different partial theories come together. There is, however, one complete impact theory for collisions that involves no transitions (*adiabatic interactions*); we shall discuss the *Lindholm theory*

in § 3.3.3 because the connections between line centers and line wings are displayed explicitly. The conditions under which this theory might be applicable are, however, not encountered in atmospheric work.

We shall describe line profiles in terms of a line shape factor, $f(\nu - \nu_0)$, defined by

$$k_\nu = Sf(\nu - \nu_0), \tag{3.47}$$

where k_ν is the absorption coefficient,

$$S = \int k_\nu \, d\nu$$

is the line intensity, ν_0 is the unperturbed frequency of the transition, and the integral is over a single line, i.e., it includes contributions only from a single vibrational–rotational transition.

3.3.2. *The Michelson–Lorentz theory*

The natural line shape for spontaneous decay can be obtained from a solution to Schroedinger's equation; it is

$$f(\nu - \nu_0) = \frac{\alpha}{\pi[(\nu - \nu_0)^2 + \alpha^2]}, \tag{3.48}$$

where

$$\alpha = \frac{1}{2\pi\tau} \tag{3.49}$$

is half the frequency width measured to half the maximum absorption coefficient or, simply, the *line width*. τ is the average lifetime of the excited state. For spontaneous decay, it is equal to the reciprocal of the appropriate Einstein coefficient, (2.71); as has already been pointed out, the natural lifetime is usually too long, and the natural line width too narrow to be important for atmospheric calculations.

Under conditions of spontaneous decay, transitions take place at random time intervals about a mean value, τ; that is to say, time intervals (t) follow a Poisson distribution with the probability of occurrence

$$p(t) = (1/\tau) \exp(-t/\tau). \tag{3.50}$$

Almost identical circumstances exist for collisions under the combined conditions (impact, diabatic, small interactions) that characterize the Michelson–Lorentz theory. The only effect of collisions is to cause a transition, otherwise the molecules are unchanged; and kinetic collisions follow the Poisson distribution, (3.50), if the molecules possess a single

velocity. The same line shape, (3.48), must result, with the only difference that τ is now the mean time between collisions.

Collisions are here defined in terms of their ability to cause transitions (*optical collisions*). Since rotational energy changes are, generally, smaller than translational energies, optical collisions are usually more frequent than kinetic collisions, and τ may be expected to be smaller than the average time between kinetic collisions (as obtained from viscosity measurements, for example).

We return to examine a point that was glossed over. How is it possible to have collisionally induced transitions (diabatic conditions) while, at the same time, requiring that the energy interaction is negligible? This is an insuperable difficulty in classical mechanics but is resolved in quantum mechanics. During a collision the two molecules form a joint mechanical system in which transitions may take place simultaneously in both molecules. One molecule may make an upward transition while the other makes a downward transition. If the collision is between identical molecules, exact resonance can occur; even for dissimilar molecules, so many joint transitions are possible that some will be close to resonance and require only a very small interaction Hamiltonian. Conditions under which the Michelson–Lorentz theory is approximately obeyed are, therefore, common. Atmospheric calculations make frequent use of the *Lorentz line shape,* (3.48), and the *Lorentz line width,* (α_L). From kinetic theory, the Lorentz width may be written

$$\alpha_L = \sum_i n_i^2 \left[\left(\frac{2\mathbf{k}\theta}{\pi} \right) \left(\frac{1}{m} + \frac{1}{m_i} \right) \right]^{1/2}, \tag{3.51}$$

where n_i is the number density of the ith species of perturber, m_i is its mass, m is the mass of the absorber, σ_i is the optical collision diameter, and θ is the temperature.

If the composition is held constant, all of the n_i are proportional to the total pressure and (3.51) gives the important result, common to all impact theories, that the line width is proportional to the pressure. The temperature variation indicated by (3.51) is dependent upon the details of the collision, and is uncertain, but this is rarely a crucial factor in thermal calculations.

The time between optical collisions for a typical atmospheric gas at s.t.p. is approximately 10^{-10} s. For this value we find $\alpha_L = 0.05\ \text{cm}^{-1}$ at s.t.p. Figure 3.18 shows three Lorentz profiles for this line width, at pressures of 1, 0.5, and 0.25 bar.

Before leaving the Michelson–Lorentz theory, two peripheral matters should be mentioned. One is that the distribution of collision times (3.50) is correct only for a gas with a single molecular velocity. We should take an average over a Maxwellian distribution of velocities. The

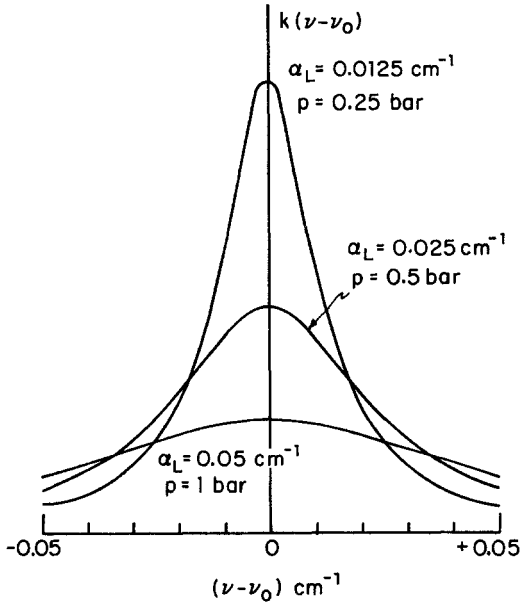


FIG. 3.18. Lorentz profiles for three pressures. A line width of $0.05\ \text{cm}^{-1}$ at a pressure of 1 bar is typical for vibration–rotation bands.

result depends upon the collisional model. For a billiard-ball model, it results in deviations from the Lorentz profile amounting to a few percent.

The *van Vleck–Weisskopf* line shape is a correction to the Lorentz line shape appropriate to the microwave region of the spectrum. This spectral region is of little importance for thermal calculations, but the theory may also be used for the very far wings of lines in the infrared spectrum. We shall see, however, that other factors become important in the far wings of lines and the significance of this particular matter is moot.

Michelson's derivation of (3.51) implicitly assumes that conditions are randomized after the collision. They should, however, be consistent with Boltzmann statistics. When this is done correctly, we obtain the van Vleck–Weisskopf profile,

$$f(\nu-\nu_0)=\frac{1}{\pi}\left(\frac{\nu}{\nu_0}\right)^2\alpha_L\left[\frac{1}{(\nu-\nu_0)^2+\alpha_L^2}+\frac{1}{(\nu+\nu_0)^2+\alpha_L^2}\right]. \qquad (3.52)$$

For $(\nu-\nu_0)\ll\nu_0$, (3.52) reduces to the Lorentz profile. For $\nu=0$ it agrees with Debye's theory of static polarizability. When microwave measurements can distinguish between this and the Lorentz profile it appears to be an improvement.

3.3.3. *An adiabatic model*

We have shown that vibrational–rotational transitions take place very readily during kinetic collisions and, therefore, that adiabatic models (models without transitions) have little direct relevance to atmospheric calculations. The discussion of this section has a different purpose: it is to illustrate a number of important concepts in a comparatively simple framework.

Lindholm's theory is a complete, adiabatic impact theory valid for large as well as for small energy interactions. This gives a consistent model of a line for all displacements from the line center (the larger the interaction energy the larger the frequency displacement).

Since we do not consider transitions, we may use a classical approach, although a quantal analogue to the discussion exists. Use is made of the *phase-shift approximation,* by which the encounter between molecules is assumed to affect the phase of the vibration but not the amplitude. This approximation can be shown to be satisfactory for all circumstances applicable to the atmosphere. We may then write the time-dependent oscillation in the form

$$a(t) = a \exp[2\pi i \nu_0 t + 2\pi i \eta(t)], \tag{3.53}$$

where $\eta(t)$ is the phase, the time variation of which is caused by the interaction; according to the phase-shift approximation, a, the amplitude, is constant.

The power in the Fourier components of this expression gives the required line shape. After some manipulation, and making use of the impact approximation, this power spectrum can be written

$$f(\nu - \nu_0) \propto \mathscr{R} \int_0^\infty C(t) \exp[-2\pi i(\nu - \nu_0)t]\, dt, \tag{3.54}$$

where $\mathscr{R}$ denotes the real part, and

$$C(t) = \int_0^\infty \exp\{2\pi i[\eta(t + T) - \eta(T)]\}\, dT \tag{3.55}$$

is the *correlation function.*

Equations (3.54) and (3.55), or their quantal analogues, are a common starting point for modern theories of spectral line shapes. This statement of the problem has the advantage that approximations based upon physical insight can be introduced directly into the correlation function, from which the line shape follows. As an example, we may

introduce an average phase shift $\bar{\eta}(t)$ defined by the relationship

$$\exp[2\pi i\bar{\eta}(t)] = \int_0^\infty \exp\{2\pi i[\eta(t+T) - \eta(T)]\}\, dT. \tag{3.56}$$

Whatever the exact meaning of $\bar{\eta}(t)$, it is reasonable to suppose that, averaged over many events, an average phase shift should increase linearly with time, or $\bar{\eta}(t) = (A + iB)t$ where A and B are constants. With this reasonable assumption,

$$C(t) \approx \exp[2\pi i(A + iB)t]. \tag{3.57}$$

Substituting (3.57) in (3.54),

$$f(\nu - \nu_0) = (B/\pi)/[(\nu - \nu_0 - A)^2 + B^2], \tag{3.58}$$

a Lorentz profile with a line width, B, shifted by a frequency, A. The simple assumption leading to (3.58) illustrates the reason for the wide applicability of the Lorentz profile.

The coefficients A and B define the rate of change of the average phase shift with time. Since phase shifts are caused by collisions, both coefficients must be proportional to the frequency of collisions and hence to the total pressure of the gas. Line shift and line width are both proportional to the pressure and, although it cannot be concluded from this analysis, they are often of similar magnitudes.

We now return to the complete theory of Lindholm for a van der Waals energy interaction. Interpreted in terms of frequency shifts as a function of approach distance, the van der Waals interaction is

$$\Delta\nu = -\beta r^{-6}, \tag{3.59}$$

where r is the distance between absorber and perturber and β is an interaction constant.

Lindholm obtained a correlation function for this van der Waals interaction. His result can be stated in terms of a Lorentz width, α_L, and a frequency, ν_P, that measures the displacement from the line center beyond which the Lorentz profile is no longer valid. The theory yields

$$\nu_P = 0.0361\beta^{-1/5}u^{6/5}, \tag{3.60}$$

$$\alpha_L = 1.34\beta^{2/5}u^{3/5}n, \tag{3.61}$$

where n is the number density of perturbers and u is the mean molecular velocity. Equation (3.61) reinforces an earlier conclusion that collisional widths are proportional to the pressure.

It is instructive to calculate ν_P by eliminating β between (3.60) and (3.61) and using observed line widths. This suggests a $\nu_P \sim 2\,\mathrm{cm}^{-1}$ for typical line widths. This result is of great practical importance. We have already reached the conclusion that vibration–rotation lines should have a shifted Lorentz line shape as a central core, the width and shift of which are proportional to the pressure; now we have an indication of how far out into the wings this profile will extend. ν_P is approximately 40 times greater than the line width at 1 bar. According to (3.60), it does not depend upon the pressure. This result is confirmed both by measurements and by more elaborate theories.

The Lindholm profile can be expressed in terms of its ratio to a Lorentz shape with line width α_L and a center shifted by $-0.726\alpha_L$. This ratio (F = Lindholm/Lorentz) is expressed in terms of the dimensionless parameter,

$$\mu = \frac{\nu - \nu_0 + 0.726\alpha_L}{\nu_P}.$$

The numerical results correspond quite closely to the following algebraic forms:

line center approximation $-2.5 \leq \mu \leq +1.5$,

$$F(\mu) = 1 - 0.286\mu; \tag{3.62}$$

low-frequency wing approximation $\mu < -2.5$,

$$F(\mu) = 0.932\,|\mu|^{1/2} + 0.319\,|\mu|^{-1/3}; \tag{3.63}$$

high-frequency wing approximation $\mu > 1.5$,

$$F(\mu) = 0.638\mu^{-1/3}. \tag{3.64}$$

There is a simple physical explanation for the shape of the low-frequency wing. For very large displacements from the line center, the first term on the right-hand side of (3.63) will dominate, giving a factor proportional to $(\nu - \nu_0)^{1/2}$. The Lorentz shape (3.48) contributes, in the far wing, a term proportional to $(\nu - \nu_0)^{-2}$ and the two, when multiplied together, give a frequency variation $(\nu - \nu_0)^{-3/2}$. This variation follows from a simple statistical theory. If there were a frozen configuration of molecules interacting with an inverse-m power law, statistical theory gives a low-frequency wing with the frequency variation, $(\nu - \nu_0)^{-(m+3)/m}$. For $m = 6$, as for (3.59), the statistical profile agrees with (3.63). It is a valid inference that, far from the center of a line, a statistical treatment may be relevant.

3.3.4. *The Anderson–Tsao–Curnutte theory*

Atmospheric absorbers and perturbers (principally nitrogen and oxygen) both possess vibrational and rotational energy. A joint transition during a collision may be close to resonance and, if permitted, should have a large transition probability. The Anderson–Tsao–Curnutte (ATC) theory is a quantum-mechanical treatment of collisions specialized to these conditions and is the standard method for calculating line widths and line shifts that have not been measured in the laboratory. The theory makes no attempt to calculate the line profile.

Near-resonance implies that transitions will take place before molecules approach too closely and it is a reasonable assumption that only the far-field interaction between the dominant electric multipole moments need be included in the perturbation Hamiltonian. These interactions are comparatively easy to handle analytically, which is why they are used, but they are the weak point of the theory. As an example, that we shall use in the subsequent discussion, the interaction between the dipole moment of a water molecule, μ, and the quadrupole moment of a nitrogen molecule, q, gives rise to an angle-dependent interaction proportional to

$$V(r) = \mu q r^{-4}, \tag{3.65}$$

where r is the distance between the molecules.

If the energy interactions are small, both absorber and perturber will follow straight-line paths with a distance of closest approach, or *impact parameter*, r_m. This is the *classical-path* approximation and, without it, the calculation would be very difficult. Using it, the perturbation Hamiltonian can be calculated along a path and the probability of transition evaluated as a function of the impact parameter.

The water molecule (our example) can make radiative transitions between an initial (i) and a final (f) state with the absorption or emission of an unperturbed photon of frequency

$$\nu_{0,if} = (E_f - E_i)/\mathbf{h}. \tag{3.66}$$

Since we are considering one particular spectral line, i and f are given for the calculation.

During a collision, both the water vapor and nitrogen molecules can undertake a variety of nonradiating transitions, any one of which can interrupt the $i \rightarrow f$ transition. For water vapor i,f can go to i',f', while for nitrogen the rotational levels can change $J \rightarrow J'$. The change in energy in the transition is either

$$\Delta E = (E_i - E_{i'}) + (E_J - E_{J'})$$

or (3.67)

$$\Delta E = (E_f - E_{f'}) + (E_J - E_{J'}).$$

Given all the possible choices of i', f', J, J' some ΔE's are likely to be very small indeed. The closeness to resonance is measured by the parameter

$$k = |(2\pi r \Delta E)/\mathbf{h}u|, \tag{3.68}$$

where u is the approach speed.

Given the classical-path assumption and the interaction potential (3.65) the perturbation Hamiltonian can be calculated and the probability of transition integrated over the whole path. The independent variables are the distance of closest approach, r_m, and the various quantum numbers. Of the latter, i,f are given and a sum can be taken over all of the primed variables. This permits us to express the transition probability $P_J(r_m)$ as a function of J and r_m only. From the transition probability a collision cross section can be defined,

$$\pi\sigma_J^2 = \int_0^\infty 2\pi r_m P_J(r_m)\, dr_m. \tag{3.69}$$

The collision diameter is approximately equal to the approach distance for which the transition probability is unity. The ATC theory is an impact theory and it leads to a Lorentz profile with the real part of (3.69) as the optical collision diameter. Equation (3.69) also has an imaginary part that the discussion of § 3.3.3 allows us to identify with a line shift parameter. The ATC theory gives both line shift and line width to the same degree of approximation.

At this point in the theory some serious numerical difficulties arise. The integral in (3.69) runs from 0 to ∞ and yet the interaction potential is certainly incorrect for small values of the impact parameter. Further, the theory is not applicable to transition probabilities greater than unity but there is nothing in the calculation to prevent such values from occurring. So, for $r_m < r_0$, the transition probability is assumed to be constant with r_0 taken to be the larger of either the kinetic collision diameter or the impact parameter that gives unit transition probability.

For a dipole–quadrupole interaction the collision diameter is

$$\sigma_{ifJ}^2 = r_0^2\left\{1 + (8\pi/45)(\mu q/\mathbf{h}u)^2\left[\sum_{i',J'} D(i,i')Q(J,J')F(k_0) + \sum_{f',J'} D(f,f')Q(J,J')F(k_0)\right]\right\}, \tag{3.70}$$

where D and Q are the known dipole and quadrupole transition probabilities for water and nitrogen, respectively, and $F(k_0)$ is a tabulated function of $k(r = r_0)$.

The Lorentz width follows from (3.51)

$$\alpha_{if} = \sum_J n(J)\sigma_{ifJ}^2[(2\mathbf{k}\theta/\pi)(1/m_{H_2O} + 1/m_{N_2})]^{1/2}, \tag{3.71}$$

where $n(J)$ is the population of the perturber J state. The line width depends upon the quantum numbers of the upper and lower states through the dipole transition probabilities in (3.70) and line widths therefore vary from line to line.

Figure 3.19 shows a comparison of measured and calculated widths for some wider lines of water vapor in three different bands. The agreement is fair, but the theory may be better than appears because many of the errors could be in the measurements rather than in the theory. There are significant discrepancies, however, particularly among the narrow lines that occur for high J values. It is for these lines that the choice of the impact parameter r_0 is particularly important. Early calculations with the ATC theory made use of the kinetic theory impact parameter (3.2 Å), which leads to a line width of 0.03 cm^{-1} at s.t.p.

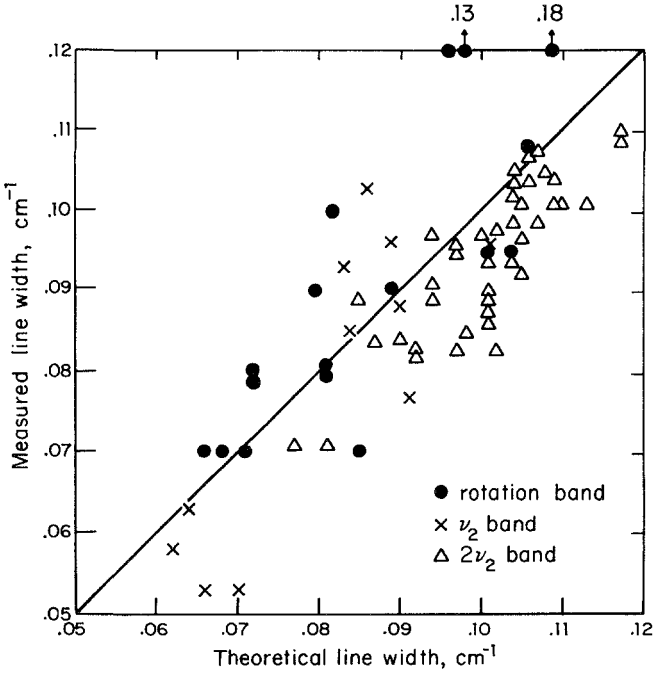


FIG. 3.19. Measured and calculated widths of water vapor lines. The lines shown are relatively wide and are from low-J transitions. Agreement between measurement and observation is not as good for high-J lines. The solid line is drawn at 45° and is not intended to be a best fit. These data are taken from Davies and Oli (1978).

According to the theory, this should be the least line width in a band but line widths as narrow as $0.007\,cm^{-1}$ at s.t.p. have been measured. Better agreement has been obtained using a smaller value of r_0 but the assumptions of the theory break down for collisions closer than kinetic theory allows.

Relatively few measurements of line shifts are available but, from those that do exist, it is clear that the agreement between theory and observation is very poor. Measured shifts at s.t.p. have published in the range -0.0066 to $+0.0044\,cm^{-1}$ with comparable theoretical values in the range -0.0062 to $+0.012\,cm^{-1}$ but with little detailed correspondence between the two. Many line shifts appear to be about half the corresponding line widths.

Measurements and calculations are also available for the case of water vapor self-broadening. The water molecule has a large dipole moment and collisions between like molecules are resonant. For both reasons, the line width for self-broadening of water vapor can be expected to be very large. The average ratio of the line widths for self-broadening to those for nitrogen broadening is about 5.

3.3.5. *The far wings of pressure-broadened lines*

The atmospheric absorption process is strongly nonlinear. In § 2.4.2 we obtained different expressions for the heating under transparent and opaque conditions. For very large absorption coefficients the heating rate becomes very small, (2.127), as it also does for very small coefficients, (2.120). The information that we require about molecular absorption bands is not, therefore, restricted to the centers of strong lines. There are circumstances that occur in the lower atmosphere for which the profile of water vapor lines is important as far as 10^3 line widths from the line centers.

A theoretical approach to far line wings involves large interaction energies and very close collisions. Since we have little exact knowledge of such circumstances, quantitative theoretical results cannot be expected. Fortunately, a semiempirical approach is feasible.

Line-wing absorption has some simple features. The absorption coefficient varies slowly with frequency compared to line centers, and it may be defined by measurements made at a few frequencies. If theory can offer some insight into the dependence of the absorption on the physical state, laboratory data may be extrapolated to atmospheric conditions. The most important state variable is the pressure; information about the influence of pressure on absorption can be obtained from order-of-magnitude considerations.

A number of recent studies have sought approximations to the correlation function that can take account of close collisions. Equation (3.57) is an approximation that is essentially equivalent to the impact

approximation. In order to study line wings, it must be modified to include information about the time, $\tau_c \sim (u/r_m)^{-1}$, that was spent during the collision.

If we modify (3.57) to include times less than τ_c we change the Fourier transform (3.54) for frequencies $(\nu - \nu_0) > 2\pi/\tau_c$. This introduces a new frequency scale

$$\alpha_c = 2\pi/\tau_c = 2\pi r_m/u. \tag{3.72}$$

For $(\nu - \nu_0) < \alpha_c$ we anticipate a shifted Lorentz profile.

For $(\nu - \nu_0) > \alpha_c$ the frequency perturbations are greater than the inverse transit time. Several oscillations can now take place in the duration of the collision; in this sense, the interaction is slow and may be approximately treated as static. For frequency displacements greater than α_c, we may anticipate that the results of statistical theory should apply. Our qualitative picture of a spectral line now includes a shifted Lorentz core (parameters calculated from ATC theory, for example) changing to statistical wings at a frequency displacement equal to α_c. Unlike α_L, α_c should not depend on the number density of perturbers, because the transit time spent in the collision does not do so. It will depend upon the temperature, however, through the molecular velocity, if not for other reasons.

The Lorentz line width is inversely proportional to the mean time between collisions so that

$$\frac{\alpha_c}{\alpha_L} \sim \frac{u}{r_m} \Big/ \frac{u}{l} \sim \frac{l}{\sigma}. \tag{3.73}$$

where l is the molecular mean free path and σ is the optical collision diameter. For atmospheric gases at s.t.p., $l/\sigma \sim 1\text{–}2 \times 10^{-2}$ and α_c should be in the range $5\text{–}10\,\mathrm{cm}^{-1}$.

An important conclusion to be drawn from this discussion is that, in line wings, whether the frequency displacement is greater or less than α_c, the absorption coefficient is proportional to the pressure. For the Lorentz part of the wings

$$k(\nu - \nu_0) \propto \frac{\alpha_L}{(\nu - \nu_0)^2}, \tag{3.74}$$

where α_L is proportional to the collision frequency. For the statistical section the absorption coefficient is proporitional to the probability of occurrence of the spatial configuration of perturbers leading to that particular frequency perturbation. This probability is proportional to the number of perturbers at a given distance from an absorbing molecule and, therefore, also proportional to the molecular number density.

The two important results that we have achieved are an estimate of the magnitude of α_c together with the information that it does not depend upon the pressure, and the prediction that line wing absorption will be proportional to the pressure, whether the frequency displacement is greater than or less than α_c. These two results are of practical importance for atmospheric calculations.

This qualitative discussion is supported by all aspects of the behavior of the Lindholm model (§ 3.3.2). α_c and ν_P are closely related, although there is a discrepancy in magnitude that is within the uncertainty of the discussion. Our present conclusions are not restricted to adiabatic collisions, however. There is a quantal analogue to the correlation function for which the foregoing discussion is equally valid.

To go further we turn to laboratory measurements. It is difficult to make measurements a few cm^{-1} from a line because few are that far away from other lines. Figure 3.20 shows one careful set of measurements involving two isotopic lines of hydrochloric acid. Measurements were made at a number of pressures and combined by plotting the

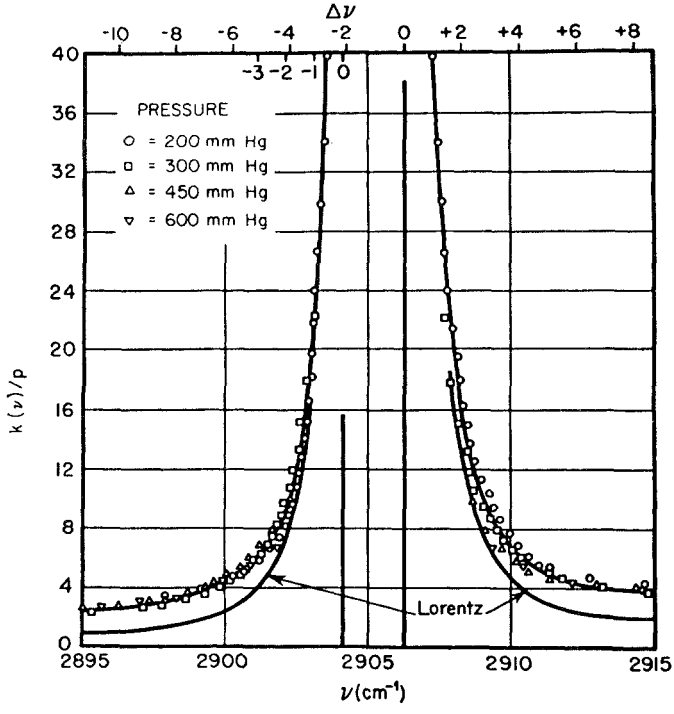


FIG. 3.20. The wings of the R(0) lines of the fundamental band of $H^{35}Cl$ and $H^{37}Cl$. The two vertical lines indicate the centers and intensities of the two isotopic lines. The Lorentz profile is indicated. The quantity plotted, $k(\nu)/p$, should not be a function of pressure. After Benedict et al. (1965a,b).

absorption coefficient divided by the pressure. This method of plotting demonstrates the anticipated dependence of absorption on pressure. Departures from the Lorentz profile occur for frequency displacements greater than about 2.5 cm^{-1}; this transition frequency, as predicted, does not depend upon pressure.

The far wings of the lines in Fig. 3.20 fall off more slowly than expected from the Lorentz profile. An empirical fit is

$$k(\nu - \nu_0) \propto |\nu - \nu_0|^{-1.73}. \tag{3.75}$$

The brief theoretical discussion of statistical profiles in § 3.3.3 suggests an energy interaction varying as r^{-4}.

Other laboratory studies have given different results for the far-wing profile, both larger than and smaller than Lorentz. A well studied case is the low-frequency wing of the ν_3 fundamental of carbon dioxide. There is a convergence of lines in the wing so that, past a certain frequency, the absorption coefficient is solely from line wings. One empirical fit gives the ratio to a Lorentz profile,

$$\begin{aligned} F(\nu - \nu_0) &= \exp[-0.46(|\nu - \nu_0| - 5)^{0.46}] \qquad \text{for } |\nu - \nu_0| > 5, \\ &= 1 \qquad \text{for } |\nu - \nu_0| < 5, \end{aligned} \tag{3.76}$$

where ν is in cm^{-1}.

An added complication is that, when line-wing profiles are important, a large number of wings inevitably overlap. This implies energy perturbations much larger than the energy gaps between rotational energy levels. The rotational wave functions are no longer independent and it is not permissible to add contributions from many lines calculated on the assumption of independence. To take an example based upon the Lorentz profile, if we treat the lines as independent the absorption coefficient far from an array of lines is

$$k(\nu - \nu_0) = \sum_i \frac{S_i \alpha_{L,i}}{\pi(\nu - \nu_{0,i})^2}, \tag{3.77}$$

where S_i, $\alpha_{L,i}$, and $\nu_{0,i}$ are the parameters for the ith line.

A more acceptable approximation has been shown to be

$$k(\nu - \nu_0) = \sum_{j,k} \frac{S_{jk} \alpha_{L,jk}}{\pi(\nu - \nu_{0,j})(\nu - \nu_{0,k})}. \tag{3.78}$$

The diagonal terms in (3.78) correspond to (3.77) but the off-diagonal terms are also important for inelastic collisions.

This mixing of wave functions provides one more reason why we are

unlikely, in the near future, to obtain absolute theories of line-wing absorption. Equation (3.78) still indicates proportionality between the absorption coefficient and the Lorentz widths (proportional to the pressure), and this is the only important result that we require.

The temperature dependence of line intensity differs from line to line. The two sums in (3.77) and (3.78) should give different results for the temperature variation of the wing absorption. Again we may conclude, for a different reason, that this aspect of the problem must be treated empirically.

3.3.6. *Doppler effects*

In the absence of collisional effects, spectral lines have a finite width because of the Doppler shifts from random molecular motions. The probability that there is a relative velocity (u) between absorber and observer is, from Maxwell's law,

$$p(u) = \left(\frac{m}{2\pi \mathbf{k}\theta}\right)^{1/2} \exp\left(-\frac{mu^2}{2\mathbf{k}\theta}\right). \tag{3.79}$$

If $u/\mathbf{c} \ll 1$, the Doppler shift is

$$\nu - \nu_0 = \frac{\nu_0 u}{\mathbf{c}}. \tag{3.80}$$

From (3.79) and (3.80), we obtain the *Doppler profile*,

$$f_D(\nu - \nu_0) = \left(\frac{1}{\alpha_D \pi^{1/2}}\right) \exp\left\{-\left[\frac{(\nu - \nu_0)}{\alpha_D}\right]^2\right\}, \tag{3.81}$$

where

$$\alpha_D = (\nu_0/\mathbf{c})(2\mathbf{k}\theta/m)^{1/2} \tag{3.82}$$

is the Doppler line width. It is defined in terms of the half frequency width to e^{-1} of the maximum of the profile, rather than to 2^{-1} as for the Lorentz profile.

Typical Doppler widths are

1. the 5577 Å forbidden line of atomic oxygen, at 300 K, $\alpha_D = 3.3 \times 10^{-2}\,\text{cm}^{-1}$;
2. a water vapor rotation line near $200\,\text{cm}^{-1}$, also at 300 K, $\alpha_D = 3.5 \times 10^{-4}\,\text{cm}^{-1}$.

The Doppler width does not depend upon the pressure and, at low enough pressures, it must exceed the Lorentz width. For a Lorentz width

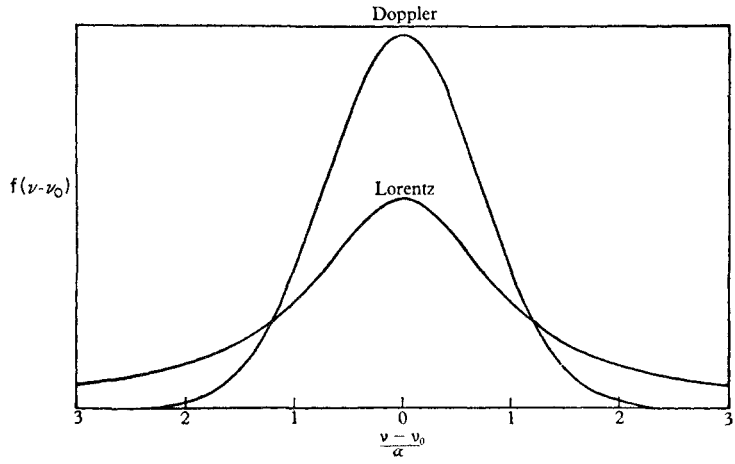


FIG. 3.21. Doppler and Lorentz line shapes for the same intensities and widths.

of $0.05\,\mathrm{cm}^{-1}$ at s.t.p. equal Doppler and Lorentz widths occur at a pressure of 0.66 bar (at 3–4 km altitude) for case (1) and 7 mb (at 34 km) for case (2).

The shapes of the two profiles are compared in Fig. 3.21. The greatest difference is in the wings where the Doppler profile falls off rapidly, as $\exp\{-[(\nu-\nu_0)/\alpha_D]^2\}$, while the Lorentz profile falls off much more slowly, as $(\nu-\nu_0)^2$.

If it were possible to treat collisional and Doppler broadening as completely independent (we shall show that it is not strictly correct to do so), the collision broadened line shape should be shifted by the Doppler shift (3.80) and averaged over the Maxwell distribution (3.79). If this procedure is applied to the Lorentz profile (3.48) we obtain the *Voigt profile,*

$$f_\nu(\nu-\nu_0)=\int_{-\infty}^{+\infty} du(\alpha_L/\pi)[(\nu-\nu_0-u\nu_0/\mathbf{c})^2+\alpha_L^2]^{-1} \times (m/2\pi\mathbf{k}\theta)^{1/2}\exp(-mu^2/2\mathbf{k}\theta). \tag{3.83}$$

The Voigt profile is extensively tabulated in the literature in terms of the parameter $d=2\alpha_L/\alpha_D$ and fast numerical algorithms are available for its computation (see Bibliography).

Figure 3.21 suggests the general nature of the Voigt profile. At high pressures ($d \gg 1$), the Doppler profile is narrow compared to the Lorentz profile and behaves in the convolution (3.83) as if it approximated a δ-function; under these conditions the Voigt and Lorentz profiles are identical.

At low pressures ($d \ll 1$) the situation is more complicated. Near the line center, the core of the Lorentz line will approximate a δ-function in the convolution (3.83) but in the wings the Lorentz profile will eventually vary more slowly with the frequency displacement; then it will be the Doppler profile that approximates to a δ-function. We are led to anticipate a hybrid line with a Doppler center but with Lorentz wings.

The half intensity width for the Voigt profile, normalized to the Doppler width, is shown by the uppermost curve in Fig. 3.22. The independent variable, $d/2 = \alpha_L/\alpha_D$, is proportional to the pressure. This representation obscures the differences that exist between the Doppler, Lorentz, and Voigt profiles and deals with a single parameter only, the half width to half intensity, but it is convenient for a discussion of departures from the Voigt profile. The uppermost curve represents the Voigt profile with the Doppler width when $d \ll 1$ and the Lorentz width, proportional to the pressure, when $d \gg 1$. The curves below the uppermost in Fig. 3.22 show the normalized line width for the *Dicke profile*, which takes account of the interaction between Doppler and collisional broadening. The curves are distinguished by the size of the parameter

$$z = \frac{8}{3\pi}\left(\frac{\sigma_o}{\sigma_k}\right)^2, \tag{3.84}$$

where σ_o is the optical collision diameter and σ_k is the kinetic collision diameter. For $z \ll 1$, and for d neither very large nor very small, the line

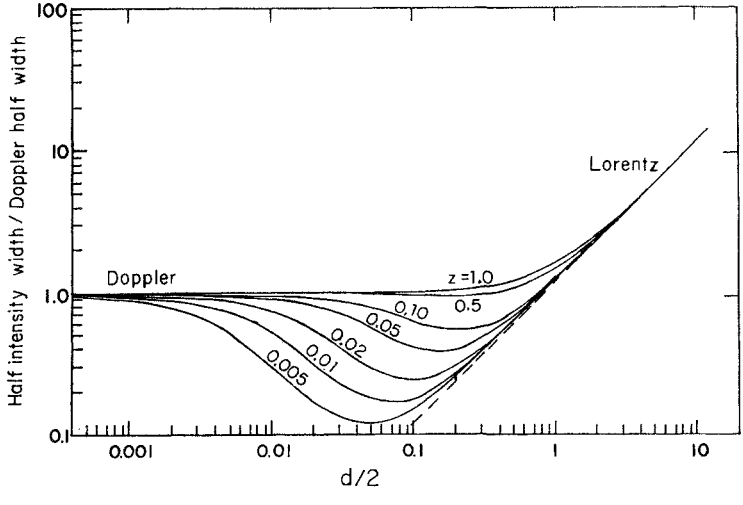


FIG. 3.22. The half width to half intensity for the Dicke profile. The parameter z is defined in the text. After Fink and Belton (1969).

is narrower than predicted by the Voigt profile; when it is narrower, the profile is Lorentz, even if $d < 1$, when a Doppler core might be expected.

This curious behavior of the line width was unsuspected until line widths less than the Doppler width were found in the microwave spectrum. Dicke supplied the appropriate quantal explanation; but we present here a simple qualitative discussion in terms of the correlation function, (3.55).

The solution to the line broadening problem, as stated in (3.54) and (3.55), requires an estimate of the average phase change after a given time, see the definition of $\bar{\eta}(t)$ in (3.56). If we make use of the average phase shift caused by Doppler effects, $\bar{\eta}_D(t)$, this procedure will yield the Doppler profile, (3.81). Collisions reorient velocities randomly and the average Doppler phase can, therefore, evolve only over the time between kinetic collisions, τ_k. If the mean molecular speed is $\bar{u}$, we have, making use of (3.82),

$$\bar{\eta}_D \sim \bar{u}\nu_o\tau_k/\mathbf{c} \sim \alpha_D\tau_k. \tag{3.85}$$

We may similarly approximate the collisional phase, in the manner of (3.57). If we accept that A and B are of similar magnitude and both approximately equal to α_L, we have

$$\bar{\eta}_c \sim \alpha_L\tau_o, \tag{3.86}$$

where τ_o is the time between optical collisions. From (3.85) and (3.86),

$$\begin{aligned}\bar{\eta}_c/\bar{\eta}_D &\sim (\alpha_L/\alpha_D)(\tau_o/\tau_k),\\ &= (\alpha_L/\alpha_D)(\sigma_k/\sigma_o)^2,\\ &\sim (\alpha_L/\alpha_D)z^{-1}.\end{aligned} \tag{3.87}$$

When the line shape is determined from the phase-shift approximation, (3.54), the dominant mechanism will be that giving the largest phase shift. Our treatment here is capable of giving information only about the line center. For $z \sim 1$ we conclude from (3.87) that the line center is dominated by collisions if $\alpha_L/\alpha_D > 1$ and by Doppler effects is $\alpha_L/\alpha_D < 1$. This is in full agreement with our previous discussion of the Voigt profile, although there is almost no common ground between the two treatments.

Dicke's line shape becomes important when $z \ll 1$, i.e., for optical collision diameters much smaller than kinetic collision diameters or, alternately stated, when many kinetic collisions occur during one optical collision. Then, according to (3.87), collisions may dominate the line center even when $\alpha_L/\alpha_D \ll 1$. The width of the resulting Lorentz line will

then be less than the Doppler width appropriate to the gas temperature. This is the phenomenon of *line narrowing*.

Our discussion explains some of the stranger features of Fig. 3.22. It has a limited importance for atmospheric calculations. In § 3.3.4, we noted that certain water vapor lines have widths less than predicted from kinetic theory by a factor of three. For these lines $z = 1/3$ and the Voigt profile may be slightly in error over a limited range of pressures. In practice, the Dicke line shape has not been used for terrestrial applications, but there are cases in the laboratory and in planetary spectra of the outer planets when it is important to use the correct formulation of the interaction between Doppler and collisional phase shifts.

3.4. Collision-induced and polymer spectra

In § 3.3 we treated interactions between molecules in terms of perturbations to permitted transitions of an otherwise isolated molecule. There are, however, many other aspects to collisional interactions that fall outside this simple prescription. For example, a homonuclear diatomic molecule, such as oxygen or nitrogen, that has no dipole moment in an isolated state, may have a dipole moment induced in it by a colliding molecule, in which case dipole transitions may occur during the short period occupied by the collision. Or, what amounts to the same thing, a transition, forbidden by a selection rule because of a high degree of symmetry, may be weakly permitted if that symmetry is destroyed during a collision. Alternatively, depending upon the interaction potential, colliding molecules may form a dimer or larger complex that can survive a few collisions. Such a complex is, from an optical point of view, a new species with its own vibration–rotation characteristics. Both circumstances, collision-induced spectra and dimers or polymers, have some relevance to atmospheric calculations.

First we consider collision-induced, forbidden transitions. The symmetry of either the vibrational or rotational wave functions may be altered in a collision. In the former category are such questions as the forbidden Q-branches of parallel bands of linear molecules (§ 3.2.4) that may appear as weak lines in spectra measured at high pressures. More important are pressure-induced vibrational transitions in the abundant oxygen and nitrogen molecules. These may need to be taken into account in precise calculations of atmospheric heating. On the outer planets, where hydrogen and helium are the dominant species and where pressures can be very high, pressure-induced vibration–rotation absorptions control the thermal state of the atmosphere.

The oxygen and nitrogen (forbidden) fundamental vibration bands lie at 6.42 and 4.29 μm, respectively. Laboratory spectra of the oxygen band are shown in Fig. 3.23. Pressures are high, ~10 bar for the spectra

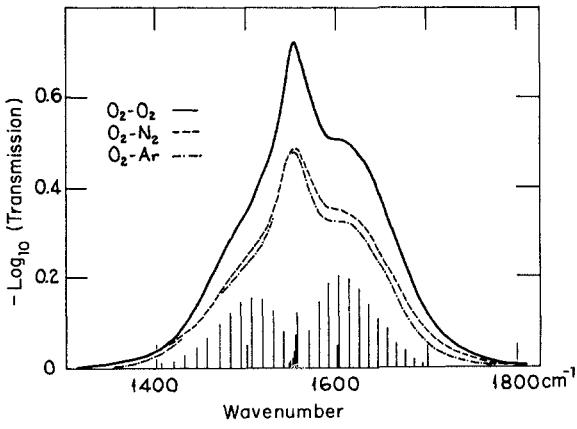


FIG. 3.23. The fundamental vibration–rotation band of oxygen. The path length is 40 m. The densities are as follows: for pure oxygen, 9.59 Amagat; for the oxygen–nitrogen mixture, oxygen 1.09 Amagat, nitrogen 56 Amagat; for the oxygen–argon mixture, oxygen 1.12 Amagat, argon 56.9 Amagat (see Appendix 3 for the definition of the Amagat). The vertical lines represent calculated rotation line intensities and positions. After Shapiro and Gush (1966).

of pure oxygen and ~60 bar for the gas mixtures. A dipole moment is induced in the oxygen molecule by quadrupole–quadrupole interaction or by overlap induction. In the former case, quadrupole selection rules apply, $\Delta J = -2,\ 0,\ +2$, giving rise to an O-branch, a Q-branch, or an S-branch, respectively. Transitions can take place in either molecule (assuming both to have induced dipoles) and two simultaneous absorptions can occur, e.g., a pure rotation band in addition to a vibration–rotation band. The overlap induction contributes mainly to the Q-branch.

In Fig. 3.23, the theoretical spectrum shown is for a single, quadrupole–quadrupole induced transition. No trace of the predicted rotational structure can be seen in the observed spectrum for the reason that the rotational lines have widths of about 30 cm^{-1} and are merged into one another. The cause of these broad lines is the very short time for which the dipole exists, giving rise to a line width approximately equal to the line parameter, α_c, discussed in § 3.3.4 and defined in (3.72). As was the case for α_c, this broadening is not a function of pressure and the lines cannot be resolved under any conditions.

The band intensity of a pressure-induced band is, unlike a normal band, proportional to the perturber pressure. The absorption intensity depends upon the amount of time for which the dipole exists; that is to the product of the time spent in a collision with the rate of collisions; the latter factor brings in the perturber pressure. The *volume* absorption coefficient is now proportional both to the number density of absorbers, n_A, and to the number density of perturbers, n_P, (or to the square of the

number density of absorbers in the case of collisions between like molecules). This is the same as the behavior in the far wings of a pressure-broadened line, and there is some analogy between the two situations. Because of this fact, it is usual to define *binary absorption coefficients* for pressure-induced bands

$$k_n^{A,P} = \frac{k_v}{n_A n_P} \text{ cm}^5 \text{ for unlike molecules,}$$
$$k_n^{A,A} = \frac{k_v}{n_A n_A} \text{ cm}^5 \text{ for like molecules.} \tag{3.88}$$

Given information about the multipole interactions between molecules, the induced dipole moment can be calculated. The following is the first-order term in the absorption coefficient,

$$k_n^{A,P} = \frac{\pi}{3mc^2}\left[\left(\alpha\frac{\partial Q_A}{\partial r_A}\right)^2 + \left(Q_A\frac{\partial \alpha}{\partial r_A}\right)^2\right]\int_0^\infty r^{-6}\exp\left[\frac{-V(r)}{k\theta}\right]dr, \tag{3.89}$$

where Q_A is the quadrupole moment of the absorber, r_A is the nuclear separation, α is the electric polarizability, $V(r)$ is the interaction potential, r is the distance between the absorber and perturber, and m, $\mathbf{c}$, and $\mathbf{k}$ have their usual meanings. Note the Boltzmann term involving the interaction potential. Since the interaction potential is negative, the effect of this term is to cause the absorption coefficient to decrease as temperature increases; in physical terms, the higher the molecular velocities, the shorter is the time that the molecules are in close contact.

The necessary condition for dimer formation is the existence of attractive forces for very close collisions. The resulting complex will then survive until the next collision or possibly it may survive a few collisions. The important difference from pressure-induced bands is that dimer lifetimes are greater than or equal to the time between collisions and, therefore, lines are not broadened as they are in Fig. 3.23. A dimer possesses its own characteristic vibrational modes with little relationship to those of the constituent molecules. It can have a resolvable vibration–rotation structure and it was the presence of such a structure in laboratory bands that led to the distinction between dimer bands and pressure-induced bands in spectra of nitrogen and oxygen at high pressures. The intensities of dimer bands, like those of pressure-induced bands for like molecules, are proportional to the square of the concentration of the absorbing molecule. The only dimer absorption that has been claimed to exist in the atmospheric spectrum is that of water vapor, in the continuum absorption of the 10-μm window. The alternative explanation of this continuum is absorption by the wings of strong lines in the 6.3 μm

and rotation bands of water vapor, as discussed in § 3.3.5. Until this controversy is resolved the importance of dimer absorption is moot.

3.5. Overview

Figures 3.24a and b, 3.25, and 3.26 show comparisons between solar spectra measured by an interferometer installed on the Space Shuttle and theoretical calculations based on the AFGL data and a Voigt line profile. Figures 3.24 and 3.25 were recorded in orbit. The absorption paths were tangential and the atmosphere was traversed twice; the tangent height is indicated. Figure 3.26 was recorded on the ground.

The agreement between observed and synthetic spectra is remarkable, but may be misleading because the eye emphasizes the near perfect agreement between line positions. This agreement goes beyond the needs of thermal calculations but is required for the identification and measurement of trace gases.

Absorption intensities do not agree as well as do line positions. Occasional features are to be seen in one spectrum but not the other and the strong lines differ up to ~10% in absorption. No lines are strongly absorbed and, for Figs. 3.25 and 3.26, the lines are dominated by Doppler broadening, for which we may be confident about the profile. Our knowledge of line profiles is not severely tested by these data. The data on the AFGL tape were not assembled for the purpose of precise absorption calculations. In order to form a judgment as to the reliability of atmospheric calculations, we must rely heavily upon theoretical conclusions drawn from §§ 3.2–3.4, as summarized below.

(i) Isolated molecules. Line positions and relative line intensities for a vibration–rotation band of an isolated molecule can be calculated to any desired precision, although high-order perturbation theory can be very complex. Since atmospheric molecules spend at least 99% of the time in isolation, the theory of isolated molecules forms an excellent basis for atmospheric calculations. Vibrational transition probabilities cannot be calculated accurately, but they can be determined from only a few measurements in the laboratory. Line intensities have not been calculated to the same precision as line positions, but this could be done. The temperature variation of line intensities is known precisely once the lower state has been assigned.

(ii) Collisional perturbations. Only the central core of a line profile can be strictly interpreted as a perturbation to the theory of isolated molecules. With some reservations, the profile of the core of a vibration–rotation line can be treated in terms of a shifted Lorentz line convolved with a Doppler profile. The Voigt profile is correct if the line is unshifted. The theory of Lorentz widths is understood to first order, but the measured and theoretical data available may be in significant error.

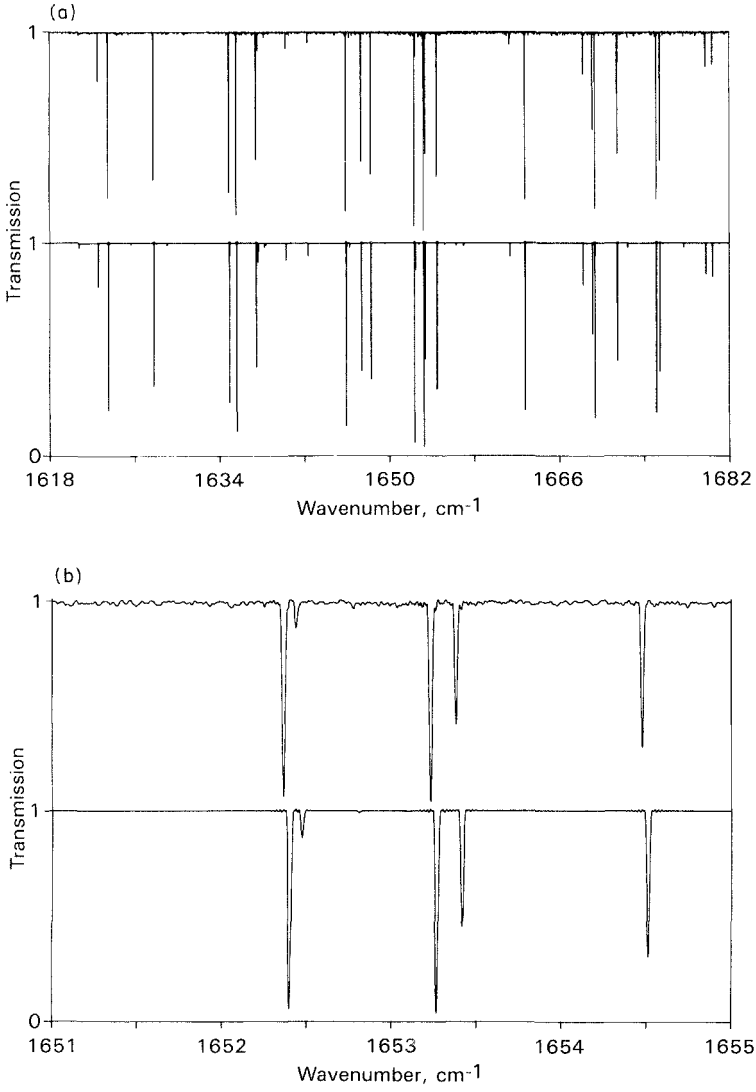


FIG. 3.24. Observed and calculated lines of the ν_2 band of water vapor in the solar spectrum. (a) Spectral range, 1618 to 1682 cm^{-1}. (b) Spectral range, 1651 to 1655 cm^{-1}. Tangent height, 71.5 km. The upper panels are observed and the lower panels are calculated from the AFGL tape and an atmospheric model. In this figure and in Figs. 3.25 and 3.26, the scale for the observed transmission (upper spectrum) is indicated by the 0 and the upper 1. The synthetic spectrum is shifted bodily downward by an amount equal to the difference between the upper and the lower 1's. A small frequency shift between the two spectra is attributable to second-order uncorrected Doppler shifts from the spacecraft motion. By permission of Dr. Crofton Farmer, Jet Propulsion Laboratory.

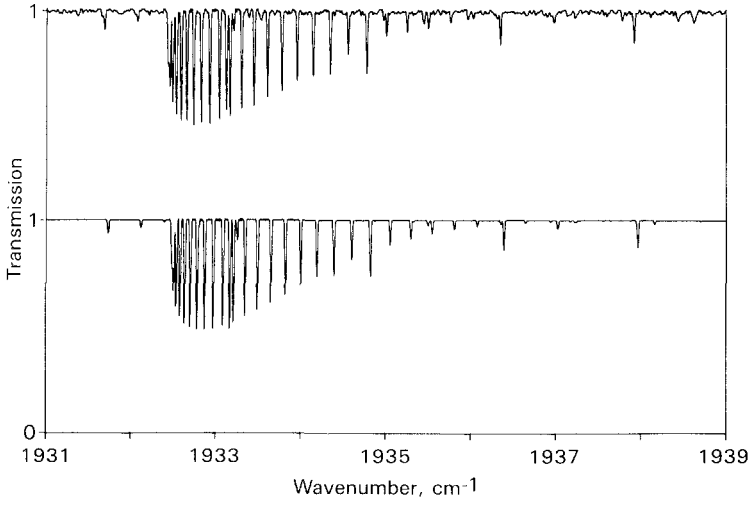


FIG. 3.25. Observed and calculated lines in the Q-branch of the $3\nu_2$ band of carbon dioxide in the solar spectrum. Spectral range, 1931 to 1939 cm^{-1}. Tangent height, 45.8 km. The upper panel is observed while the lower panel is calculated using the AFGL line parameters and an atmospheric model. The Q-branch lines are separated by a strong Fermi resonance. Even J values appear, from Q(2) at 1932.48 cm^{-1} to Q(60) at 1937.55 cm^{-1}. By permision of Dr. Crofton Farmer, Jet Propulsion Laboratory.

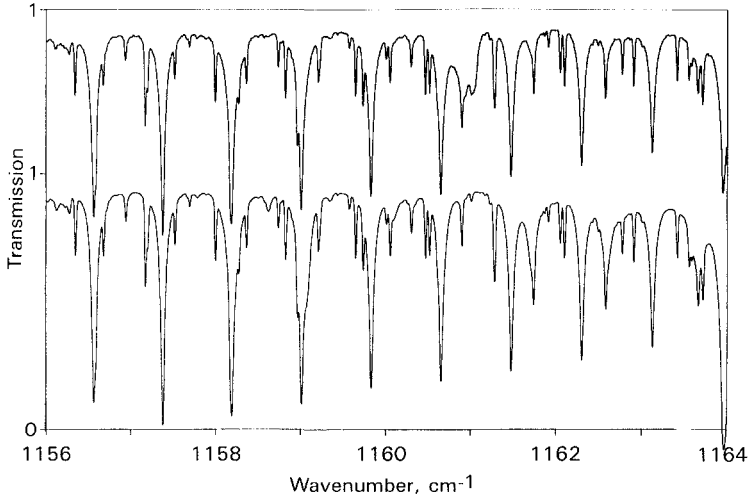


FIG. 3.26. Observed and calculated lines of the ν_1 band of ozone and the $2\nu_2$ band of nitrous oxide in the solar spectrum. Spectral range, 1156 to 1164 cm^{-1}. Level of observation, 0 km. Zenith angle of observation, 60°. The upper panel is the observed spectrum while the lower panel is calculated from the AFGL line parameters and an atmospheric model. The 10 strong, regularly spaced lines are P(5) to P(14) of the $2\nu_2$ band of nitrous oxide. The many ozone lines are weaker and unevenly spaced. By permission of Dr. Crofton Farmer, Jet Propulsion Laboratory.

Pressure shifts are poorly understood. Doppler widths can be calculated precisely.

(iii) Close collisions. For close collisions, the theory of isolated molecules is no more than an approximation. The most important question for atmospheric calculations is the shape of the far wings of absorption lines. All indications are that the absorption is proportional to the perturber pressure and, on this basis, a semiempirical approach is possible. Pressure-induced, forbidden transitions have been detected in atmospheric spectra and may be handled approximately with a combination of perturbation theory and empirical data. The possibility of contributions from dimers has not yet been eliminated.

BIBLIOGRAPHY

3.1. Introduction
Figure 3.1 is from

Valley, S. L. (Ed.), 1965, *Handbook of geophysics and space environment.* Washington D.C.: U.S. Air Force Geophysics Directorate.

3.2. Vibration–rotation spectra
The classical texts are

Herzberg, G., 1944, *Atomic spectra and atomic structure.* New York: Dover Publications.

———, 1950, *Molecular spectra and molecular structure. I. Spectra of diatomic molecules.* New York: Van Nostrand.

———, 1945, *Molecular spectra and molecular structure. II. Infrared and Raman spectra of polyatomic molecules.* New York: Van Nostrand.

The following book is principally concerned with rotational spectra but provides an excellent summary of the entire subject:

Townes, C. H., and Schawlow, A. L., 1975, *Microwave spectroscopy.* New York: Dover Publications.

Recent discussions of perturbation theory are by

Papoušek, D., and Aliev, M. R., 1982, *Molecular vibrational–rotational spectra.* Amsterdam: Elsevier.

Aliev, M. R., and Watson, J. K. G., 1985, "Higher-order effects in the vibration–rotation spectra of semirigid molecules," *Molecular spectroscopy: Modern research* (K. N. Rao, Ed.). New York: Academic Press, p. 1.

3.3. The shape of a spectral line
A classical and quantal literature of great complexity is reviewed by

Breene, R. G., 1961, *The shift and shape of spectral lines.* London: Pergamon.

The theory of strong encounters for a single collision time was initiated by

Michelson, A. A., 1895, "On the broadening of spectral lines," *Astrophys. J.* **2,** 251.

Nearly thirty years passed before the theory was put into final form by

Lenz, W., 1924, "Einige Korrepondenzmässig Betrachtungen," *Z. Phys.* **25,** 299.

Meanwhile a complete theory was developed in a different way by

Lorentz, H. A., 1906, "The absorption and emission of lines of gaseous bodies," *Proc. R. Acad. Sci. (Amsterdam)* **8,** 591.

The line profile (3.52) was obtained by

van Vleck, J. H., and Weisskopf, V. F., 1945, "On the shape of collision-broadened lines," *Rev. Mod. Phys.* **17,** 227.

The original ATC theory was by

Anderson, P. W., 1949, "Pressure broadening in the microwave and infra-red regions," *Phys. Rev.* **76,** 647,

and it subsequently appeared in a more extended form by

Tsao, C. J., and Curnutte, B., 1962, "Line-widths of pressure-broadened spectral lines," *J. Quant. Spectrosc. Radiat. Transfer* **2,** 41.

The first application to atmospheric molecules was by

Benedict, W. S., and Kaplan, L. D., 1959, "Calculation of line widths in H_2O–N_2 collisions," *J. Chem. Phys.* **30,** 388.

The data presented in Fig. 3.19 are based on a different approach that is equivalent to ATC, see

Davies, R. W., and Oli, B. A., 1978, "Theoretical calculations of H_2O line widths and pressure shifts: Comparison of the Anderson theory with quantum many-body theory for N_2- and air-broadened lines," *J. Quant. Spectrosc. Radiat. Transfer* **20,** 95.

The data in Fig. 3.20 are from

Benedict, W. S., Herman, R., Moore, G. E., and Silverman, S., 1956a, "The strengths, widths and shapes of infrared lines. I. General considerations," *Can. J. Phys.* **34,** 830.

———, 1956b, "The strengths, widths and shapes of infrared lines. II. The HCl fundamental," *Can. J. Phys.* **34,** 850.

Measurements in the far wings of the 4.3 μm carbon dioxide band were made by

Benedict, W. S., and Silverman, S., 1954, "Line shapes in the infrared," *Phys. Rev.* **94,** 752,

and the results were more fully reported by

Kaplan, L. D., 1954, "A quasi-statistical approach to the calculation of atmospheric transmission," *Proceedings of the Toronto Meteorological Conference 1953.* London: Royal Meteorological Society, p. 49.

A thorough investigation of line shapes for the atmosphere of Venus is by

Varanasi, P., Sarangi, S. K., and Tejwani, G. D. T., 1972, "Line shape parameters for HCl and HF in a CO_2 atmosphere," *J. Quant. Spectrosc. Radiat. Transfer* **12,** 857.

Examples of theoretical discussions of line-wing profiles are

Fomin, V. V., and Tvorogov, S. D., 1973, "Formation of the far wing contour of spectral lines broadened by a foreign gas; analysis of exponential decrease of continuous absorption beyond the band head of the 4.3 μm band of CO_2," *Appl. Optics* **12,** 584.

Boulet, C., Robert, D., and Galatry, L., 1978, "On the calculation of the molecular line shape in the wings," *J. Quant. Spectrosc. Radiat. Transfer* **20,** 371.

Birnbaum, G., 1979, "The shape of collision broadened lines from resonance to the far wings," *J. Quant. Spectrosc. Radiat. Transfer* **21,** 597.

Clough, S. A., Kneizys, F. X., Davies, R., Gamache, R., and Tipping, R., 1980, "Theoretical line shape for H_2O vapor: Application to the continuum," *Atmospheric water vapor* (A. Deepak et al., Eds.). New York: Academic Press, p. 25.

The question of line overlap is discussed by

Baranger, M., 1958, "Problem of overlapping lines in the theory of pressure broadening," *Phys. Rev.* **111,** 494.

Rapid and accurate computations of the Voigt profile are the subject of

Hui, A. K., Armstrong, B. H., and Wray, A. A., 1978, "Rapid computation of the Voigt and complex error functions," *J. Quant. Spectrosc. Radiat. Transfer* **19,** 509.

Humlíček, J., 1982, "Optimized computation of the Voigt and complex probability functions," *J. Quant. Spectrosc. Radiat. Transfer* **27,** 437.

The first paper on line narrowing in Doppler lines was

Dicke, R. H., 1953, "The effects of collisions on the Doppler width of spectral lines," *Phys. Rev.* **89,** 472.

Further contributions to the theory of the collision narrowing of Doppler lines are by

James, T. C., 1969, "Calculations of collision narrowing of the quadrupole lines in molecular hydrogen," *J. Opt. Soc. Am.* **59,** 1602.

Gersten, J. I., and Foley, H. M., 1968, "Combined doppler and collision broadening," *J. Opt. Soc. Am.* **58,** 933.

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.

3.4. Collision-induced and polymer spectra

The first quantitative analysis of pressure-induced vibrational transitions was

Van Kranendonk, J., 1958, "Induced infrared absorption in gases: Calculation of the binary absorption coefficients of symmetrical diatomic molecules," *Physica* **24,** 347.

More recent work is by

Borysow, A., Moraldi, M., and Frommhold, L., 1984, "Modelling of collision-induced absorption spectra," *J. Quant. Spectrosc. Radiat. Transfer* **31,** 235.

Timofeyev, Yu. M., and Tonkov, M. V., 1978, "Effect of the induced oxygen band on the transformation of radiation in the 6 μm region in the Earth's atmosphere," *Isv. Atmos. Ocean. Phys.* **14,** 437.

Birnbaum, G., and Cohen, E. R. 1976, "Theory of line shape in pressure-induced absorption," *Can. J. Phys.* **54,** 593.

The spectrum in Fig. 3.23 is from

Shapiro, M. M., and Gush, H. P., 1966, "The collision-induced fundamental and first overtone bands of oxygen and nitrogen," *Can. J. Phys.* **44,** 949.

There are relatively fewer advocates for polymers than there are for line wings as the explanation of the 10 μm atmospheric continuum; the case for dimers and polymers is made by

Gebbie, H. A., 1980, "Observations of anomalous absorption in the atmosphere," *Atmospheric water vapor* (A. Deepak et al., Eds.). New York: Academic Press, p. 133.

Dimer absorption by oxygen in the presence of pressure-induced absorption has been recorded in the laboratory by

Long, C. A., and Ewing, G. E., 1973, "Spectroscopic investigation of van der Waals molecules. I. The infrared spectra of $(O_2)_2^*$," *J. Chem. Phys.* **58,** 4824.

4

BAND MODELS

4.1. Introduction

Radiative heating calculations in the atmosphere involve four distinguishable scales of frequency. First, there is the comparatively slow variation with frequency of the Planck function and its derivative with respect to temperature (see Fig. 2.3). About one-half of the radiation from a black body at terrestrial temperatures lies in a wave number range of $500\,\mathrm{cm}^{-1}$.

The second scale is that of the unresolved contour of a band. For atmospheric molecules other than water vapor, the Planck function is effectively constant over a single band; water vapor bands must be divided into sections of the order of $50\,\mathrm{cm}^{-1}$ wide before this is so.

For a rotating molecule, the next relevant scale of frequency is that of the spacing between rotation lines, approximately $1\text{–}5\,\mathrm{cm}^{-1}$. Finally, there is the monochromatic scale on which the absorption coefficient may be treated as a constant, and for which Lambert's absorption law is obeyed: of the order of one-fifth of a line width $\simeq 2 \times 10^{-2}\,\mathrm{cm}^{-1}$ for a gas at atmospheric pressure, down to $2 \times 10^{-4}\,\mathrm{cm}^{-1}$ for a Doppler line in the middle atmosphere. This step takes us to a division of the frequency scale that, when taken together with other features of the calculation, presents a formidable computation task.

Calculations can, of course, be made and are made at this limiting spectral resolution (*line-by-line calculations*) but, despite the fact that they are technically feasible with modern computers, such calculations are rare and are usually performed to provide a few reference cases. The great majority of investigations make use of averages over many lines, embracing spectral ranges that are small compared to a band contour (*narrow-band models*), or over complete bands (*wide-band models*), or over the entire thermal spectrum (*emissivity models.*)

There are a number of reasons for working with spectral averages. Practical considerations are that important classes of laboratory measurements, and most atmospheric observations (e.g., satellite radiometry) are made with some spectral averaging, often comparable to that of narrow-band models. On the numerical side, even if the computing power exists to perform line-by-line calculations, there are usually other aspects of the overall calculation upon which available resources may be

better spent, e.g., by decreasing the grid size of time step for a general circulation model. In fact, even though available computing power increases rapidly with time there has been a simultaneous trend toward the simplest class of radiation calculation—the emissivity methods (see § 6.4.2).

In this chapter we shall examine the theory of narrow-band models in some detail; wide-band and emissivity models derive directly from this treatment. We shall also discuss the *k distribution* technique, which is closely related to band models, but which can be simply extended to conditions for which band models cannot be used.

Narrow-band models deal with averages over spectral regions for which the band contour and the Planck function are both approximately constant, but that, nevertheless, contain a large number of rotation lines. This situation is realized by the abstraction of an infinite array of absorption lines of uniform statistical properties. An interval of this infinite array is assumed, with a suitable choice of parameters, to simulate the properties of the spectral region under consideration. In the model, each interval is flanked by statistically similar intervals, but this will not be so in a real band. This is a major source of discrepancy between abstraction and actuality.

We shall limit our discussion to average transmissions, or *transmission functions,*

$$\bar{T}_i = \frac{1}{\Delta\nu_i} \int_i T_\nu \, d\nu, \tag{4.1}$$

where

$$\Delta\nu_i = \int_i d\nu$$

is the width of the ith frequency interval and

$$T_\nu = \exp(-\tilde{\tau}_\nu)$$

is the monochromatic transmission. All calculations involving thermal source functions can be stated in terms of mean transmissions. If, for example, we require the intensity integrated over all frequencies we may write, from (2.87),

$$\begin{aligned} I = \int_0^\infty I_\nu \, d\nu &= \int_0^\infty d\nu \int_0^\infty B_\nu e^{-\tilde{\tau}_\nu} \, d\tilde{\tau}_\nu \\ &= \int_0^\infty d\nu \int_0^1 B_\nu \, dT_\nu \\ &\approx \sum_i \Delta\nu_i \int_0^1 B_i \, d\bar{T}_i, \end{aligned} \tag{4.2}$$

where B_i is the (almost constant) value of the source function in the ith interval. All other properties of the radiation field can be obtained from (4.2).

In this chapter, we shall limit discussion to homogeneous absorption paths for which we may write $\tilde{\tau}_\nu = k_\nu m$, where m is the amount of absorbing material in units that must be consistent with those of the absorption coefficient (see the discussion in § 2.1.2). For non-homogeneous paths, the absorption coefficient varies along the path and does so differently for each frequency. We shall defer consideration of this problem until § 6.2; the usefulness of spectral averaging methods for atmospheric applications depends upon the existence of satisfactory solutions.

Before treating narrow-band models in more detail, we first consider an important observed property of spectrally averaged transmissions, *the multiplication property*. Figure 4.1 shows low-resolution absorption spectra for the ν_1 and ν_3 fundamentals of water vapor and for the resonating combination bands of carbon dioxide near 2.7 μm, for the two gases separately, and also when they are mixed together in a single absorption tube. The mean transmission of the mixture, averaged over the spectrometer slit width, is shown to be the product of the transmissions of the two components separately. It is not obvious that this should be so. If we suppose that the two gases do not interact, then, for monochromatic radiation, the transmission of the mixture will be the product of the transmissions of the two components. If we distinguish the two components by the numbers (1) and (2), we have

$$T_\nu(1,2) = T_\nu(1) \times T_\nu(2).$$

The multiplication property now implies that, for the ith spectral interval,

$$\bar{T}_i(1,2) = \bar{T}_i(1) \times \bar{T}_i(2). \tag{4.3}$$

For (4.3) to be correct implies, as a matter of definition, that the two spectra are uncorrelated, a condition that cannot be exact for any finite frequency interval. However, if the width of the interval is great and if the lines of either array are randomly arranged (*random bands*), (4.3) may be correct to any required degree of accuracy. Many important bands have this random property.

Another circumstance for which the correlation coefficient is small is for two arrays with regularly spaced lines (*regular bands*) but with line spacings that are noncommensurate. This condition can be realized in practice with arrays that are almost regular, but with line spacings that vary slowly over a band.

It is difficult to conceive of a state of affairs in which one or other of these two conditions does not hold, and there is a great amount of

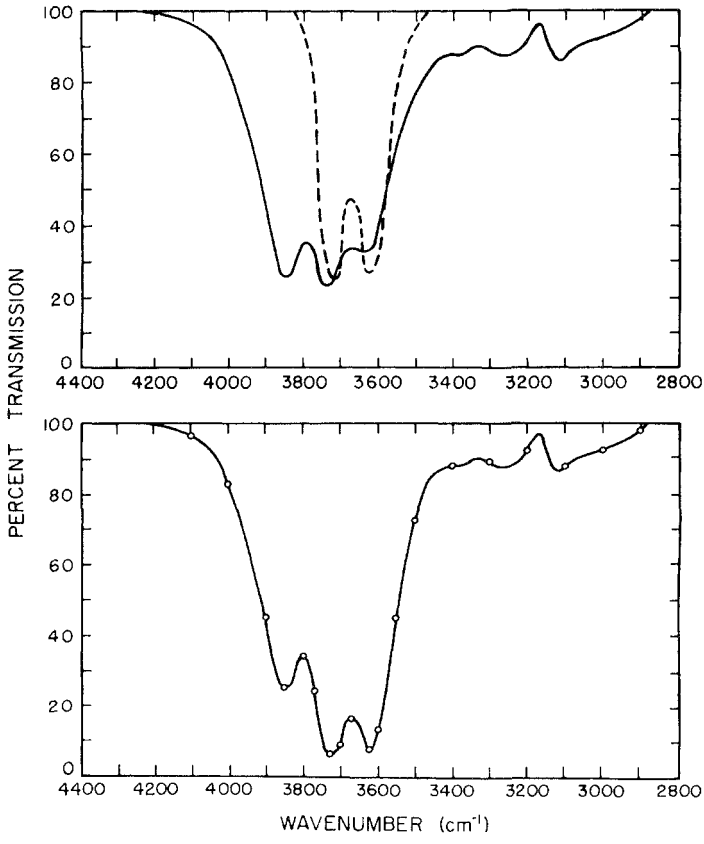


FIG. 4.1. The multiplication property of band transmission. The upper spectra are for CO_2 (broken line) and H_2O (solid line) individually. The lower spectrum is for the mixture; the solid line is observed and the points are obtained by multiplying together the two transmissions in the upper panel. The absorption path is 88 m, the H_2O partial pressure is 5 mm Hg, the CO_2 partial pressure is 4 mm Hg, and the total pressure is made up to 140 mm Hg with nitrogen. After Burch et al. (1956).

experimental evidence that bears this out. We shall, therefore, accept the multiplication property for two different arrays as a fundamental property of band transmission.

The two arrays mentioned above—regular and random—are the only two-parameter arrays that are physically distinct.

4.2. Isolated lines

4.2.1. *Single line of Lorentz shape*

Before considering regular and random models, we discuss a condition common to both, when lines are so far apart that they may be treated as

isolated. The conditions sufficient for this asymptotic case to be valid cannot be simply stated. Clearly there must be large gaps between lines and the average absorption must be small,

$$\bar{A}_i = 1 - \bar{T}_i \ll 1.$$

But the question of sufficient conditions can be discussed only in the context of models involving line overlap.

The monochromatic absorption for a homogeneous path is

$$A_\nu = 1 - T_\nu = 1 - \exp(-k_\nu m). \tag{4.4}$$

If k_ν is contributed by a single line, the integral[1] of (4.4) over all frequencies (the *absorption area*) is finite. For historical reasons, connected with astronomical spectroscopy, it is alternatively known as the *equivalent width*[2] of the line,

$$W(m) = \int_{-\infty}^{+\infty} A_\nu \, d\nu = \int_{-\infty}^{+\infty} [1 - \exp(-k_\nu m)] \, d\nu. \tag{4.5}$$

The relationship between $W(m)$ and m is called the *curve of growth.* These curious terms are now so deeply embedded in the literature that they cannot be avoided.

W has the dimension of frequency. In the following treatment it is convenient to work with the dimensionless average absorption,

$$\bar{A} = \frac{W}{\delta} = \frac{1}{\delta} \int_{-\infty}^{+\infty} [1 - \exp(-k_\nu m)] \, d\nu, \tag{4.6}$$

where δ is the average spacing between neighboring lines. The equivalent width can be recovered from this expression for the average absorption by setting $\delta = 1$ in dimensionless coefficients.

For the Lorentz shape (3.48),

$$k_\nu m = \frac{Sm\alpha_L}{\pi(\nu^2 + \alpha_L^2)}. \tag{4.7}$$

[1] Except where specifically stated, the zero frequency is placed at, ν_0, the center of a line or band. The limits for the integral in (4.5) are, therefore, $-\nu_0$ and $+\infty$, but for all cases of band models in the infrared spectrum, there is negligible contribution to the integral from the lower limit, which may be replaced by $-\infty$; for microwave lines this might be incorrect.

[2] The name refers to the width of a rectangular line, whose center is completely absorbed, having the same absorption area.

Introducing dimensionless variables,

$$\begin{aligned} x &= v/\delta, \\ y &= \alpha_L/\delta, \\ u &= Sm/2\pi\alpha_L, \end{aligned} \tag{4.8}$$

we may write

$$\bar{A} = \int_{-\infty}^{+\infty} dx \left[1 - \exp\left(\frac{-2uy^2}{x^2+y^2}\right)\right]. \tag{4.9}$$

This integral may conveniently be expressed in terms of Bessel functions of the first kind with imaginary arguments,

$$\begin{aligned} \bar{A} &= 2\pi y u e^{-u}[I_0(u) + I_1(u)] \\ &= 2\pi y L(u), \end{aligned} \tag{4.10}$$

or

$$W = 2\pi\alpha_L L(u).$$

Some values of the Ladenburg and Reiche function, $L(u)$, are given in Appendix 7.

$L(u)$ is a comparatively simple function of its argument. For small values of u it is linear, while for large values of u it varies as $u^{1/2}$. The following series expansion is valid for small u:

$$L(u) = u\left[1 - \sum_{n=1}^{\infty} (-1)^{n+1}(2n-1)(2n-3)\cdots 5\cdot 3\cdot 1\cdot u^n/n!\,(n+1)!\right]. \tag{4.11}$$

For large u, there is an asymptotic expansion,

$$L(u) = (2u/\pi)^{1/2}\left[1 - \sum_{n=1}^{\infty} (2n-1)^2(2n-3)^2(2n-5)^2\cdots 3^2\cdot 1^2/n!\,(8u)^n\right]. \tag{4.12}$$

A number of approximations have been proposed for the rapid computation of $L(u)$. The following has maximum errors of about 1% near $u = 1$:

$$L(u) = u[1 + (\pi u/2)^{5/4}]^{-2/5}. \tag{4.13}$$

Equation (4.12) may be approximated by its limit $(2u/\pi)^{1/2}$ for $u > 3$. This is the *square-root law*, while $u \ll 1$ defines the *linear law*. The changeover from one to the other is illustrated in Fig. 4.2.

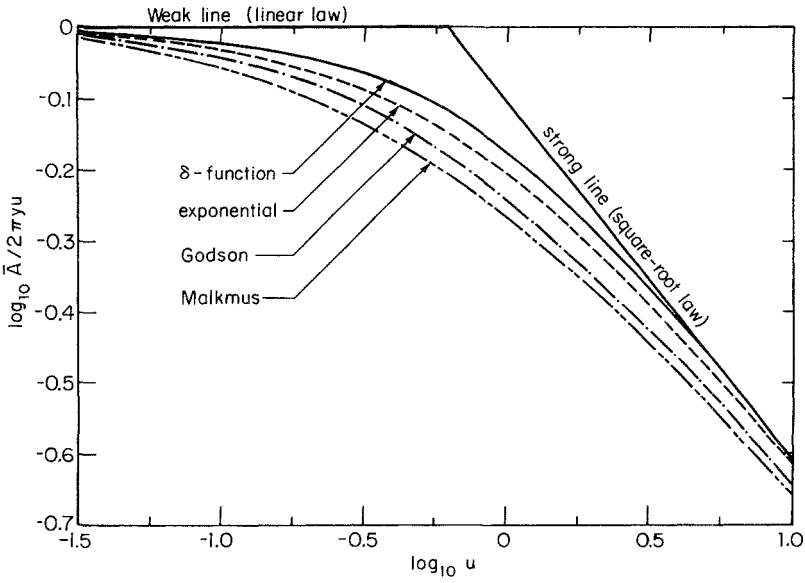


FIG. 4.2. Mean absorption by isolated Lorentz lines. Equation (4.10) corresponds to the curve marked δ-function. The other three curves are for averages over arrays of isolated lines (see § 4.3.2).

The physical significance of the two asymptotic laws is clearer if we note from (4.7) and (4.8) that u is half the optical path at the line center. Thus, for $u \ll 1$, the optical path is small at all frequencies and the exponential in (4.6) can be replaced by the first two terms in its expansion,

$$\bar{A} \approx \frac{1}{\delta} \int_{-\infty}^{+\infty} k_\nu m \, d\nu = Sm/\delta = 2\pi\gamma u. \tag{4.14}$$

Equation (4.14) has been derived without reference to the line shape and is valid for all line shapes. It is also valid for overlapping lines provided that $k_\nu m \ll 1$ for all ν, and that some means is available to identify the contribution from a single line. Thus (4.14) should be valid for all band models; only the conditions for validity should differ from one model to the other.

If $u \gg 1$, the center of the line is strongly absorbed and the integral in (4.9) becomes insensitive to changes in the integrand near $x = 0$, provided that these changes increase the exponent. If we neglect y^2 in the denominator of the exponent in (4.9), the absorption is increased for all x, but negligibly for $x \gg y$. If $u \gg 1$, when the line is strongly absorbed further out than the line width (Fig. 4.3), we may write approximately

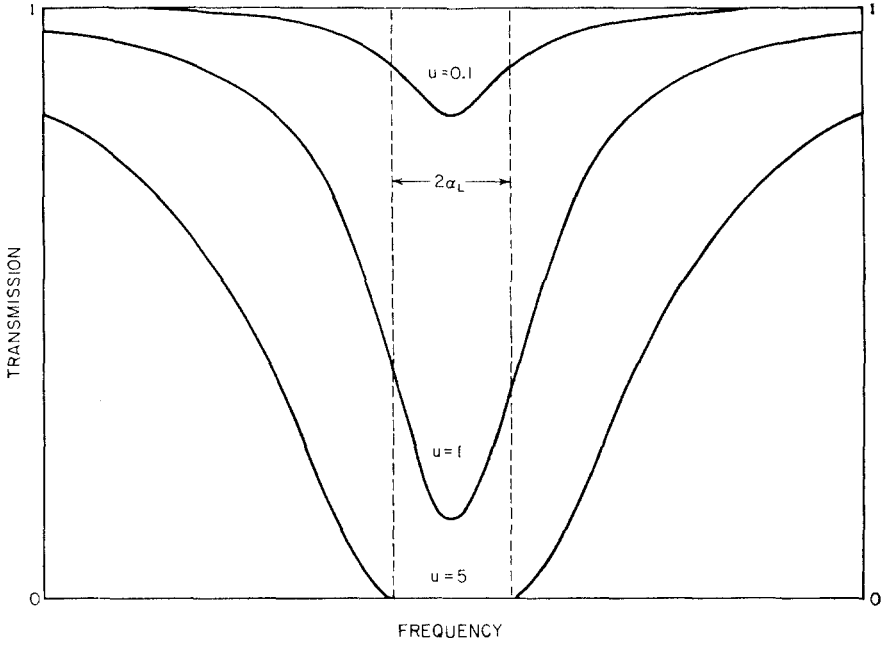


FIG. 4.3. The physical significance of the linear and square-root laws.

(with $\xi = 2uy^2/x^2$)

$$\bar{A}/y \approx (2u)^{1/2} \int_0^\infty (1 - e^{-\xi})\, d\xi/\xi^{3/2}$$
$$= 2(2\pi u)^{1/2}, \tag{4.15}$$

in agreement with the limit of (4.12).

This discussion suggests a more significant nomenclature for the two asymptotic limits, namely the *weak line* and the *strong line* limits.

The nature of the absorption law for strong lines explains two phenomena that perplexed the early investigators. It was observed that band absorption tended to vary as a low power of the amount of absorbing material rather than exponentially, as suggested by Lambert's law, and that the amount of a neutral dilutant gas appeared to be just as important as the amount of absorbing gas. From (4.15) we find

$$\bar{A} = (4Sm\alpha_L)^{1/2}/\delta. \tag{4.16}$$

α_L is, in dilute mixtures, proportional to the pressure of the dilutant gas [see (3.52)], and (4.16) predicts that, for strong lines, the partial pressure

of the absorber and the total pressure of the dilutant stand on the same footing.

Agreement between (4.10) and laboratory data is one of the best ways to test the Lorentz profile. Figure 4.4 shows a test for variable amounts of carbon monoxide diluted to constant pressure. Agreement is within the measurement errors. Comparisons of this nature are appropriate because they test a quantity of direct importance to atmospheric calculations. From the two asymptotes in Fig. 4.4, (4.14) and (4.16) can be used to determine the line intensity and the line width.

A specific test for the Lorentz profile can be made by enclosing a pure gas in an absorption tube of fixed length and varying the pressure. The amount of gas and the pressure are now proportional to each other and the dimensionless parameter, u, is constant, from (4.9). According to (4.10), therefore, the ratio of equivalent width to pressure should be constant. This prediction can be put to a careful test and a number of experimenters have done so. Within the experimental error it appears to be correct, provided that the pressure is high enough to avoid Doppler effects and low enough to avoid line overlap.

According to the discussion in § 3.3.4, the far wings of a pressure-broadened line will normally differ from the Lorentz profile. As can be inferred from the appearance of Fig. 3.20, these departures should begin to be noticed when the equivalent width is comparable to the frequency displacement α_c or ν_P, beyond which the statistical wings begin to

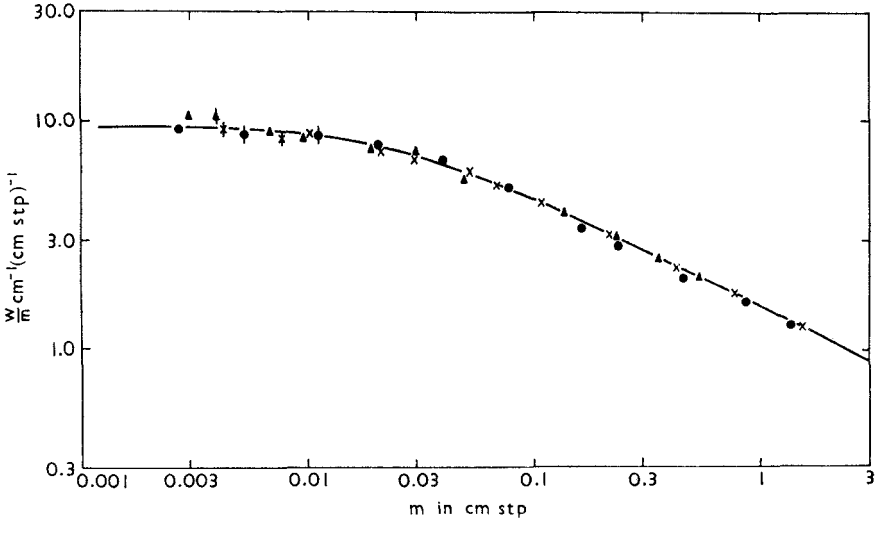


FIG. 4.4. Absorption by the P(6) line of CO. The total pressure was held constant at 700 mm Hg by adding dilutant gas. The solid line is from (4.10). After Shaw and France (1956).

develop. We may make a rough estimate of departures from the Lorentz expression, (4.10), in the following way.

Suppose that a line has the Lorentz shape out to a frequency ν_P from the line center, at which frequency a discontinuous change takes place to a wing shape,

$$k_\nu \propto (\nu - \nu_0)^{-n}.$$

From (3.75), a possible value for n is 1.73. With this hybrid profile, the equivalent width can be calculated to order α_L/ν_P with the result

$$W - W_L = \frac{1}{2\pi}\frac{2-n}{n-1}\frac{W_L - 2\alpha_L}{\nu_P}, \tag{4.17}$$

where W_L is the Lorentz equivalent width (4.10). We shall return to (4.17) when we consider line overlap.

4.2.2. *Single line with a Voigt profile*

If we substitute the Doppler profile, (3.81), in (4.6) we find

$$\bar{A} = \frac{\alpha_D}{\delta}\int_{-\infty}^{+\infty} \{1 - \exp[-w\exp(-x^2)]\}\, dx, \tag{4.18}$$

where $x = \nu/\alpha_D$ and $w = Sm/\alpha_D\pi^{1/2}$ is the optical path at the line center. Equation (4.18) can be expanded in ascending powers of w and integrated term by term, but the expression converges slowly for large w,

$$\bar{A} = \frac{\pi^{1/2} w \alpha_D}{\delta}\left[1 + \sum_{n=1}^{\infty}\frac{(-1)^n w^n}{n!\, n^{1/2}}\right]. \tag{4.19}$$

If $w \to 0$,

$$\bar{A} \to Sm/\delta,$$

which agrees, as it should, with (4.14).

For large w, there is an asymptotic expansion,

$$\bar{A} = \frac{2\alpha_D}{\delta}[(\ln w)^{1/2} + 0.2886(\ln w)^{-1/2} - 0.2473(\ln w)^{-3/2} + 0.3403(\ln w)^{-5/2} + \cdots]. \tag{4.20}$$

Equation (4.20) gives a very slow increase of absorption with amount. The sharp cut off in the wings of the Doppler profile requires a

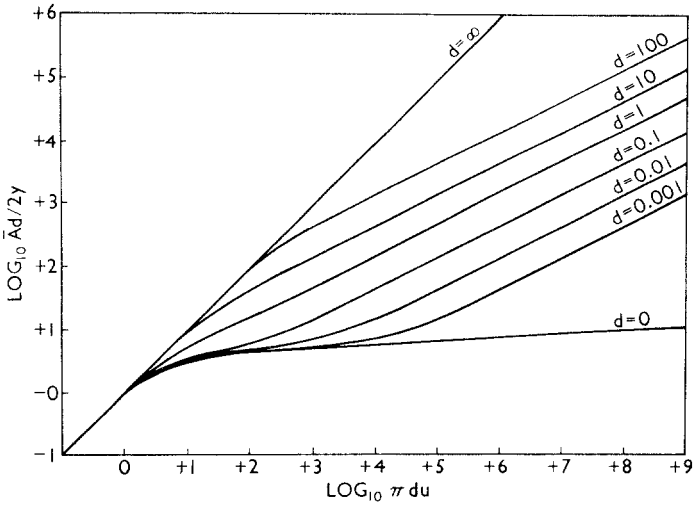


FIG. 4.5. Absorption for a Voigt profile. The ordinate and abscissa are chosen to give an instructive diagram but have no physical implications. After van der Held (1931).

very large increase of amount to bring about a significant increase in the equivalent width.

The far wings of a Voigt profile are determined by molecular collisions. If the absorption path is long enough, the absorption must ultimately be dominated by the line-wing profile, and will be given by (4.10).

Figure 4.5 is based upon numerical integration of (4.6) using the Voigt profile (3.83). The parameter d was introduced in § 3.2.5 and is equal to $2\alpha_L/\alpha_D$. The ordinate and abscissa in Fig. 4.5 are chosen so that weak lines are represented by a single line of unit slope. The inclusion of the factor d in both axes means that the strong-line limit for Lorentz lines, (4.15),is represented by a series of parallel lines of slope 1/2. The curve for Doppler lines is labeled $d = 0$.

Let us follow the curve marked $d = 0.001$ in Fig. 4.5. For small u, the absorption varies linearly with amount. When $\pi ud = \pi^{1/2}w \approx 1$, the Doppler core begins to be strongly absorbed and, from (4.20), the absorption varies very slowly with amount. When $\pi ud \approx 10^3$ the entire Doppler core is absorbed and the Lorentz wings start to be important. For $\pi ud > 10^5$ the curve of growth behaves as if there were no Doppler core at all. For $d > 1$ the influence of the Doppler core on absorption is negligible, under all conditions.

Equivalent widths for the Voigt profile (W_V) are tabulated in the literature (see Appendix 7). For purposes of computational economy, a number of approximations have been proposed to provide an interpolation between the equivalent widths for Lorentz and Doppler broadening

(W_L and W_D). One of the most satisfactory is

$$W_{V'} = [W_L^2 + W_D^2 - (W_L W_D / Sm)^2]^{1/2}. \tag{4.21}$$

Equation (4.21) is compared with exact calculations in Fig. 4.6. The maximum error in the equivalent width for the entire range of line parameters is less than 8%.

An alternative approach is to modify the line profile so that the equivalent width is integrable in terms of known functions. The following approximation replaces the Doppler core by a rectangular profile, with

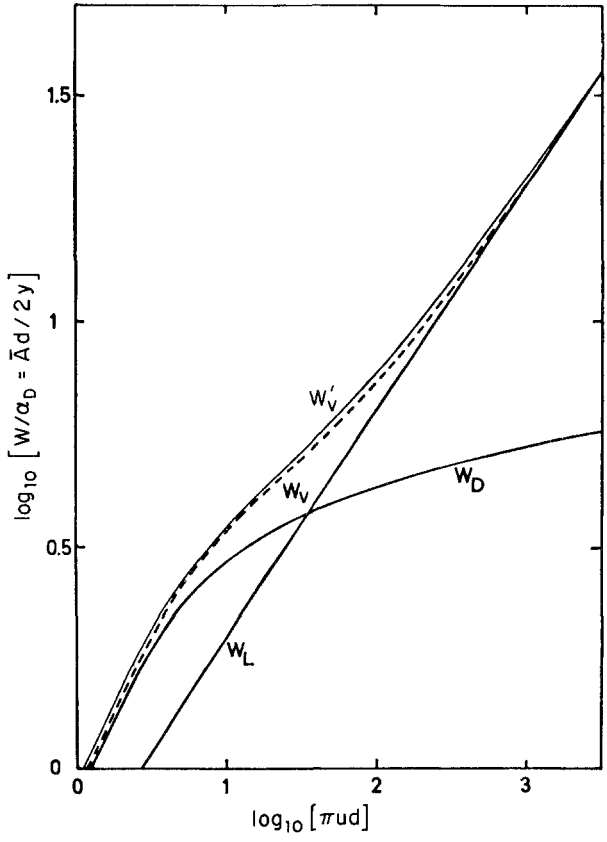


FIG. 4.6. Comparison of (4.21) with exact calculations for the Voigt profile ($d = 0.2$). W_V is the exact equivalent width. W'_V is calculated from (4.21). W_L and W_D are Lorentz and Doppler equivalent widths, respectively. After Rodgers and Williams (1974).

appropriate normalization,

$$f_\nu(\nu) = \frac{1}{2\nu_0} - \frac{\alpha_L}{\pi\nu_0^2} \qquad |\nu| \leqslant \nu_0$$
$$= \frac{\alpha_L}{\pi\nu^2} \qquad |\nu| > \nu_0, \tag{4.22}$$

where

$$\nu_0 = \frac{4\alpha_L}{\pi} + 1.25\alpha_D.$$

The two numbers in the definition of ν_0 were chosen for a good fit to exact calculations under practical circumstances.

With the line profile (4.22), the equivalent width can be integrated to give

$$W_{\nu'}(m) = 2\nu_0\left\{1 - \exp\left[-Sm\left(\frac{1}{2\nu_0} - \frac{\alpha_L}{\nu_0^2}\right)\right]\right\}$$
$$+ 2\alpha_L(2u)^{1/2}\left\{\pi^{1/2}\,\mathrm{erf}\left[\frac{(2u)^{1/2}\alpha_L}{\nu_0}\right]\right.$$
$$\left. - \left[\frac{\nu_0}{\alpha_L(2u)^{1/2}}\right]\left[1 - \exp\left(-\frac{\alpha_L^2 2u}{\nu_0^2}\right)\right]\right\}, \tag{4.23}$$

where erf[ ] is the error function. Excellent agreement has been demonstrated between (4.23) and the exact expression for the Voigt profile.

4.3. Distributed line intensities

In this section we shall discuss averages of the results obtained in § 4.2 over three distributions of line intensities. In this context, the results of § 4.2 can be looked upon as the case of a δ-function distribution function. The reason for generalizing this result will become clear in § 4.6.

We shall consider variations of line intensities from line to line, but we shall not include variations of line width, even though line widths can also vary significantly. There are two reasons, besides the desire to avoid analytical difficulty, why this is common practice. First, as we shall show later in this section, variations in line width can be incorporated exactly into band parameters in the strong-line and weak-line limits. Second, the range of variation of line widths is very small compared to the range for line intensities. The important physical problem is caused by the coexistence of lines with intensities differing by orders of magnitude. A single study by Godson (1955) showed that line-width variations were of trivial importance in comparison.

4.3.1 *Distribution functions*

Consider a range of frequencies, of width $N\delta$, containing N lines with mean spacing δ. We assume that the lines do not overlap each other. If the equivalent width of the ith line is W_i, the total absorption of all the lines is $\sum_{i=1}^{N} W_i$ and the mean absorption is

$$\begin{aligned}\bar{A} &= (1/N\delta)\sum_{i=1}^{N} W_i \\ &= \bar{W}/\delta,\end{aligned}$$

where $\bar{W}$ is the arithmetic mean of the W_i. Our aim is to determine $\bar{W}$ for certain distributions of line intensities in terms of their statistics.

Let $p(S)\,dS$ be the fraction of lines having intensities between S and $S + dS$. If we write

$$k_\nu = f(\nu)S,$$

the mean equivalent width can be expressed in terms of the distribution function

$$\bar{W} = \int_0^\infty p(S) \int_{-\infty}^{+\infty} \{1 - \exp[-Smf(\nu)]\}\, d\nu\, dS. \tag{4.24}$$

Three distribution functions have been studied:
exponential

$$p(S) = (1/\sigma)\exp(-S/\sigma), \tag{4.25}$$

Godson

$$\begin{aligned} p(S) &= \frac{\sigma}{SS_m} \qquad S < S_m \\ &= 0 \qquad S > S_m, \end{aligned} \tag{4.26}$$

Malkmus

$$p(S) = (1/S)\exp(-S/\sigma). \tag{4.27}$$

To these we may add, formally,
δ-function

$$p(S) = \delta(S - \sigma), \tag{4.28}$$

where δ is the Dirac delta-function,

$$\int_0^\infty \delta(S-\sigma)\,dS = 1.$$

Each of the above distributions has the same *mean line intensity*,

$$\sigma = \int_0^\infty Sp(S)\,dS. \tag{4.29}$$

With this normalization, (4.24) gives, in the limit of small m,

$$\bar{A} = m\sigma/\delta,$$

in agreement with expectation from the weak-line limit, (4.14).

The distributions (4.25), (4.26), and (4.27) each involve a range of line intensities. The very strong lines will have most effect for short path lengths while the weak lines will be important for long paths. Thus, the behavior of gaseous transmission over all path lengths can be correctly reproduced by a model only if the distribution function is correctly chosen. The exponential distribution, (4.25), yields the simplest results but it does not provide a good representation of the line intensities for known atmospheric bands. The Godson distribution, (4.26), has been shown to be appropriate for a number of bands but, despite its apparent simplicity, it is difficult to handle. The Malkmus distribution, (4.27), combines the simplicity of the exponential distribution with the realism of the Godson distribution and is the most widely used for analytical studies of band absorption.

Figure 4.7 illustrates the value of the Godson distribution for selected regions of the water vapor spectrum. The function plotted is the cumulative distribution,

$$\int_S^{S_m} p(S)\,dS = \frac{\sigma}{S_m}\ln\frac{S_m}{S}. \tag{4.30}$$

The points in Fig. 4.7 were obtained from lists of spectral line intensities, while the straight lines follow the Godson distribution. The exponential distribution fails to account for the large number of weak lines that occur in this spectrum.

If we exchange the order of integration in (4.24) and substitute the expressions (4.25) to (4.28), we find

δ-function

$$\bar{A} = (1/\delta)\int_{-\infty}^{+\infty} \{1-\exp[-m\sigma f(\nu)]\}\,d\nu, \tag{4.31}$$

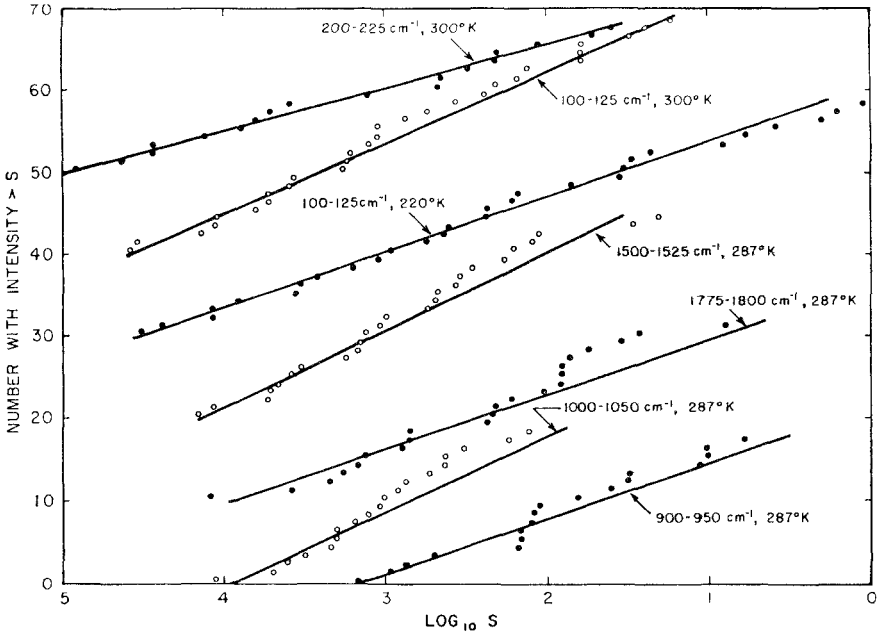


FIG. 4.7. Cumulative distributions of water vapor line intensities. Note the comparison of two temperatures for the spectral range 100–125 cm^{-1}. After Godson (1954).

exponential

$$\bar{A} = (1/\delta) \int_{-\infty}^{+\infty} m\sigma f(\nu)/[1 + m\sigma f(\nu)]\, d\nu, \tag{4.32}$$

Godson

$$\bar{A} = (1/\delta) \int_{-\infty}^{+\infty} (\sigma/S_m) G[mS_m f(\nu)]\, d\nu, \tag{4.33}$$

$$G(x) = \int_0^x dy[1 - \exp(-y)]/y,$$
$$= x - x^2/2 \cdot 2! + x^3/3 \cdot 3! - x^4/4 \cdot 4! + \cdots,$$

Malkmus

$$\bar{A} = (1/\delta) \int_{-\infty}^{+\infty} \ln[1 + m\sigma f(\nu)]\, d\nu. \tag{4.34}$$

4.3.2. *Applications to the Lorentz profile*

All of the expressions (4.31) to (4.34) are integrable for the Lorentz

shape,

$$f(\nu) = \frac{\alpha_L}{\pi(\nu^2 + \alpha_L^2)}.$$

The results are

δ-function

$$\bar{A}_L = 2\pi y L(u_\sigma), \tag{4.35}$$

exponential

$$\bar{A}_L = 2\pi y u_\sigma (1 + 2u_\sigma)^{-1/2}, \tag{4.36}$$

Godson

$$\bar{A}_L = 2\pi y(\sigma/S_m)\{I_0(u_m)\exp(-u_m) + 2u_m \exp(-u_m)[I_0(u_m) + I_1(u_m)] - 1\}, \tag{4.37}$$

Malkmus

$$\bar{A}_L = (\pi y/2)[(1 + 8u_\sigma)^{1/2} - 1]. \tag{4.38}$$

In the above expressions y has its usual meaning, (α_L/δ),

$$u_\sigma = \frac{\sigma m}{2\pi\alpha_L}, \tag{4.39}$$

$$u_m = \frac{S_m m}{2\pi\alpha_L} = \frac{S_m u_\sigma}{\sigma}, \tag{4.40}$$

and I_0 and I_1 are Bessel functions of the first kind with imaginary arguments.

In order to use these results we require numerical values for the parameters u_σ, y, and S_m/δ (for the Godson distribution only). Since the assumed distributions will not provide an exact fit to the observed line intensities, a method for making a best fit is required. One way is to require exact agreement in the asymptotic strong-line and weak-line limits. When this is done, it is possible to make simultaneous allowance for variations of line width.

We return to the sum form of (4.24) and assume that all lines are simultaneously either strong or weak. There is a possible difficulty here with the singularity at $S = 0$ for the Godson and Malkmus distributions, (4.26) and (4.27). This turns out to be unimportant for a Lorentz profile but it does give rise to a difficulty with the Doppler profile (§ 4.3.3). For

the Lorentz profile, the strong-line and weak-line limits can be written from (4.14) and (4.15),

$$\bar{A} = (1/N\delta) \sum_{i=1}^{N} 2(S_i \alpha_{\mathrm{L},i} m)^{1/2} \qquad \text{for } u \gg 1, \tag{4.41}$$

$$\bar{A} = (1/N\delta) \sum_{i=1}^{N} mS_i \qquad \text{for } u \ll 1. \tag{4.42}$$

If we take appropriate limits to (4.36), (4.37), and (4.38), and equate the results to (4.41) and (4.42), we find

$$u_\sigma = \pi u_0/4, \qquad y = 4y_0/\pi, \qquad u_m = 4u_0 \qquad (\text{or } S_m/\sigma = 16/\pi), \tag{4.43}$$

where

$$u_0 = \frac{m}{2\pi} \left[\frac{\sum_{i=1}^{N} S_i}{\sum_{i=1}^{N} (S_i \alpha_{\mathrm{L},i})^{1/2}} \right]^2, \tag{4.44}$$

$$y_0 = \frac{1}{N\delta} \frac{[\sum_{i=1}^{N} (S_i \alpha_{\mathrm{L},i})^{1/2}]^2}{\sum_{i=1}^{N} S_i}. \tag{4.45}$$

u_0 and y_0 are the appropriate values of u_σ and y to use with the δ-function distribution.

The mean absorptions for all four distributions can now be written in the form

$$\frac{\bar{A}_{\mathrm{L}}}{2\pi y_0 u_0} = f(u_0). \tag{4.46}$$

The four forms of $f(u_0)$ were plotted in Fig. 4.2.

4.3.3. *Application to the Doppler and Voigt profiles*

Series expansions for short absorption paths and asymptotic expansions for long paths are available for the Doppler profile and the distributions discussed in § 4.3.1. We may write

δ-function

$$\bar{A}_{\mathrm{D}} = (2\pi^{1/2} y/d) D(w_\sigma), \tag{4.47}$$

exponential

$$\bar{A}_{\mathrm{D}} = (2\pi^{1/2} y/d) E(w_\sigma), \tag{4.48}$$

Godson

$$\bar{A}_{\mathrm{D}} = (2\pi^{1/2} y\sigma/dS_m) G(w_m), \tag{4.49}$$

Malkmus

$$\bar{A}_D = (2\pi^{1/2}y/d)H(w_\sigma). \tag{4.50}$$

where

$$y/d = \alpha_D/\delta,\ w_\sigma = \sigma m/\alpha_D\pi^{1/2},\ \text{and}\ w_m = S_m m/\alpha_D\pi^{1/2}.$$

The expansions for $w \ll 1$ are

$$D(x) = x\left[1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n!\, n^{1/2}}\right], \tag{4.51}$$

$$E(x) = x\left[1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{(n+1)^{1/2}}\right], \tag{4.52}$$

$$G(x) = x\left[1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{(n+1)!\,(n+1)^{3/2}}\right], \tag{4.53}$$

$$H(x) = x\left[1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{(n+1)^{3/2}}\right]. \tag{4.54}$$

The asymptotic expansions for large w are

$$D(x) = 2\pi^{-1/2}(\ln x)^{1/2}[1 + 0.2886(\ln x)^{-1} - 0.2473(\ln x)^{-2} \cdots], \tag{4.55}$$

$$E(x) = 2\pi^{-1/2}(\ln x)^{1/2}[1 - 0.4112(\ln x)^{-2} - 3.5514(\ln x)^{-4} \cdots], \tag{4.56}$$

$$G(x) = (4\pi^{-1/2}/3)(\ln x)^{3/2}[1 + 0.8659(\ln x)^{-1} + 0.7417(\ln x)^{-2} \cdots], \tag{4.57}$$

$$H(x) = (4\pi^{-1/2}/3)(\ln x)^{3/2}[1 + 1.2337(\ln x)^{-2} + 1.0654(\ln x)^{-4} \cdots]. \tag{4.58}$$

The functions $D(x)$, $E(x)$, $G(x)$, and $H(x)$ are plotted in Fig. 4.8, for $10^{-1} < x < 10^4$. All have the same form for $x \to 0$ but, unlike the comparable expressions for the Lorentz shape, they behave differently as $\ln x \to \infty$. The Godson and Malkmus distributions give $G(x)$, $H(x) \to (4\pi^{1/2}/3)(\ln x)^{3/2}$, while the δ-function and exponential distributions give $D(x)$, $E(x) \to 2\pi^{1/2}(\ln x)^{1/2}$.

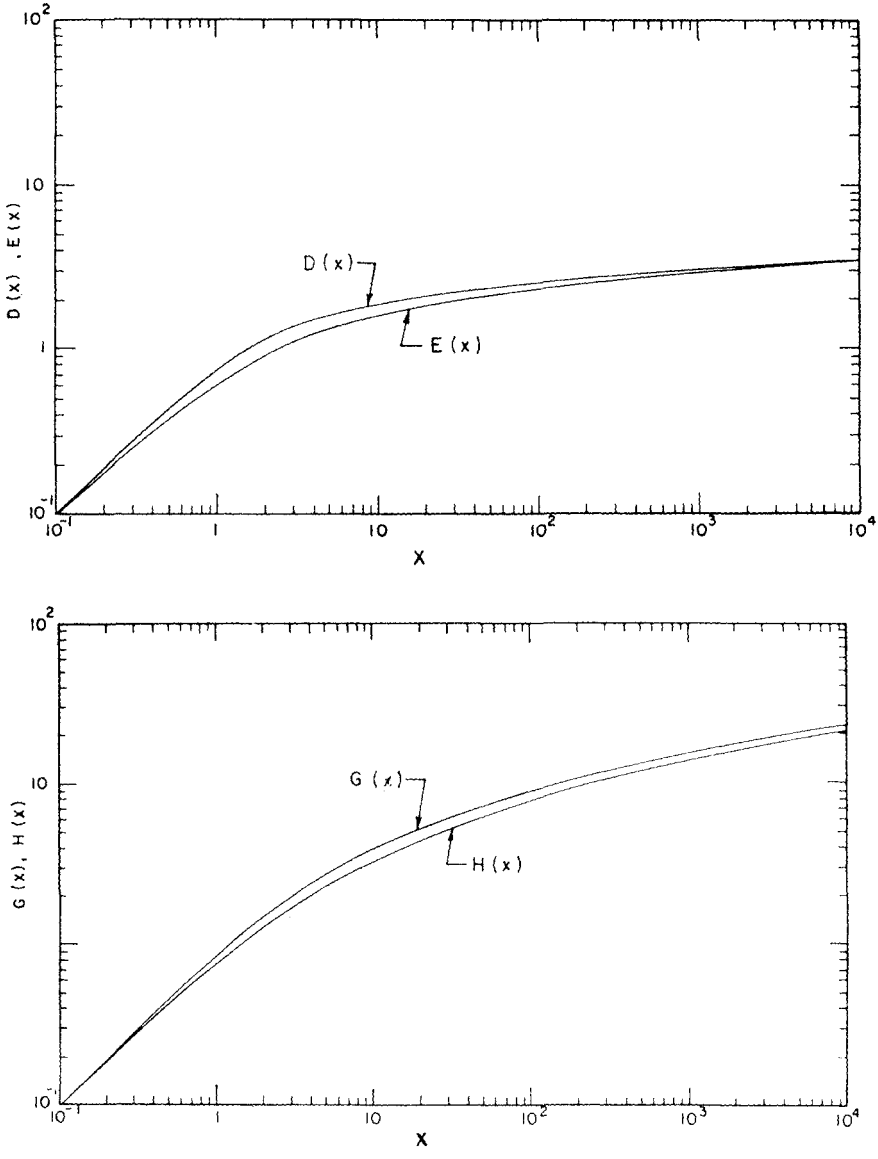


FIG. 4.8. The functions $D(x)$, $E(x)$, $G(x)$, and $H(x)$. After Malkmus (1968).

The reason for this difference can be traced to the behavior of the distribution functions as $S \rightarrow 0$. If the absorption varies sufficiently rapidly with the line intensity, the effect of many weak lines is unimportant. For the Lorentz profile, strong-line absorption varies as $S^{1/2}$ but, for the Doppler profile, it varies as $(\ln S)^{1/2}$. This is sufficient difference to alter the asymptotic behavior for long paths, when weak lines are important.

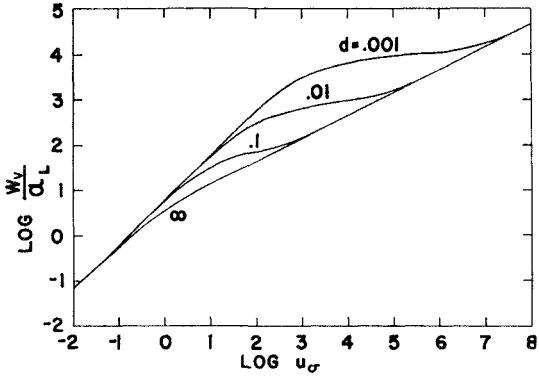


FIG. 4.9. The average equivalent width (W_V) for an exponential distribution of line intensities and a Voigt profile. After Gille and Ellingson (1968).

The Voigt profile was applied to the δ-function distribution in § 4.2.2. Numerical integration of (4.32), (4.33), and (4.34) using the Voigt profile presents no difficulties, but results are available only for the exponential distribution. These are shown in Fig. 4.9 in the form of a plot of W_V/α_L as a function of the variables u_σ and d. The curve for $d = \infty$ in Fig. 4.9 is the result for a Lorentz profile.

The approximate Voigt profile with a rectangular core, (4.22), can also be applied to distributed line intensities. For the Malkmus distribution a rather complex expression is available, involving only elementary functions (see Bibliography).

4.4. The effect of overlap

Attempts to modify the theory of isolated lines to include some effect of line overlap have not, on the whole, proved to be fruitful. They have been superseded in practice by models that treat an array of lines as a statistical entity, rather than as a group of interacting individuals, but two techniques are sufficiently interesting to warrant a brief account.

4.4.1. *Schnaidt's model*

Schnaidt assumed that the effect of line overlap was simply to terminate each line at frequency displacements $\pm\delta/2$ from its center. From (4.6)

$$\bar{A} = (1/\delta)\int_{-\delta/2}^{+\delta/2} [1 - \exp(-k_\nu m)]\, \mathrm{d}\nu. \tag{4.59}$$

This expression takes no account of contributions to the absorption from lines outside the range $\pm\delta/2$. One effect of this is that the weak-line limit, (4.13), is no longer obtained in the limit $m \to 0$.

For a symmetrical line, (4.59) can be written in the form

$$\bar{A} = (1/\delta)\int_{-\infty}^{+\infty}[1-\exp(-k_\nu m)]\,d\nu - (2/\delta)\int_{\delta/2}^{\infty}[1-\exp(-k_\nu m)]\,d\nu. \tag{4.60}$$

For the Lorentz shape, we have, from (4.10),

$$\bar{A} = 2\pi y L(u) - 2\int_{1/2}^{\infty}\left\{1-\exp\left[\frac{2uy^2}{(x^2+y^2)}\right]\right\}dx. \tag{4.61}$$

For most atmospheric problems the line spacing greatly exceeds the line width ($y \ll 1$). Since $x \geqslant 1/2$ in the integral in (4.61), we may neglect y^2 in the denominator of the exponent. After some rearrangement, there results

$$\bar{A} = 2\pi y L(u) + [1-\exp(-8uy^2)] - (8\pi u y^2)^{1/2}\,\mathrm{erf}[2y(2u)^{1/2}]. \tag{4.62}$$

In Fig. 4.10 the absorption according to (4.62) is compared to the absorption from regular and random models with uniform line intensities (see below, all for $y = 0.05$).

4.4.2. *The method of Matossi, Meyer, and Rauscher*

This is an attempt to increase indefinitely the number of overlapping Lorentz lines. The equations are ingeniously reduced to a series of approximate, simultaneous, partial differential equations having the independent-line solution, (4.10), as a boundary condition.

The analysis is complicated and only the result will be stated. The mean absorption over a frequency interval of width $\Delta\nu$, containing N overlapping, Lorentz lines is

$$\bar{A} = \frac{\pi y}{N}\left[\sum_{i=1}^{\infty} L\left(u_i - \sum_{j=i+1}^{\infty}\frac{2}{\pi}u_j X_{ij}^{+}\right) + \sum_{i=1}^{N-1} L\left(\sum_{j=i+1}^{N}\frac{2}{\pi}u_j X_{ij}^{+}\right) + \sum_{i=1}^{N} L\left(u_i - \sum_{j=i+1}^{N}\frac{2}{\pi}u_j X_{ij}^{-}\right) - \sum_{i=1}^{N-1} L\left(\sum_{j=i+1}^{N}\frac{2}{\pi}u_j X_{ij}^{-}\right)\right], \tag{4.63}$$

where L is the function defined in (4.10), i and j are indices denoting individual lines, and

$$X_{ij}^{\pm} = \tan^{-1}\left\{\frac{\Delta\nu}{2\alpha_L}\left[1 \pm \frac{2(\nu_j-\nu_i)\,\Delta\nu}{(\nu_j-\nu_i)^2+\alpha_L^2}\right]\right\}. \tag{4.64}$$

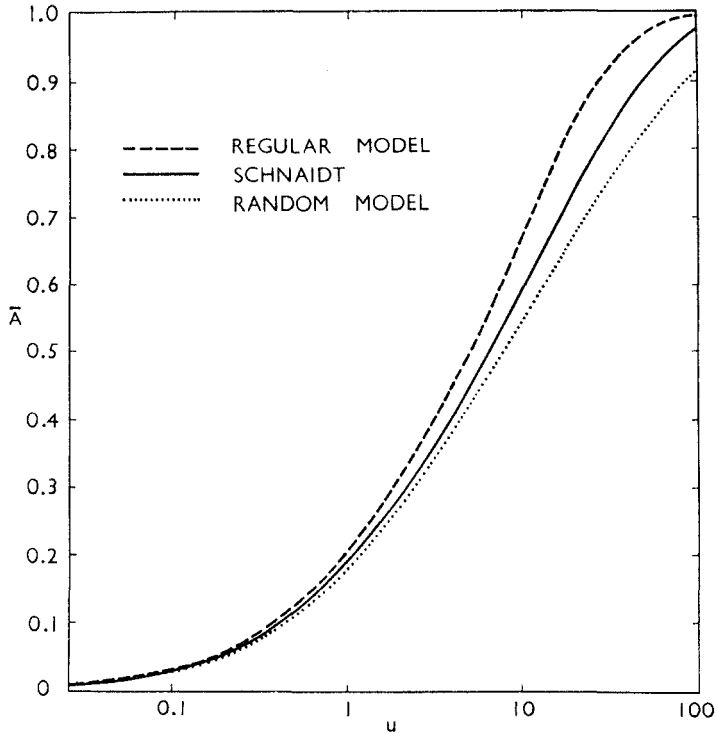


FIG. 4.10. Comparison of Schnaidt's model with regular and random models. All three models utilize the Lorentz shape. The regular and random models have uniform line intensities. $y = 0.05$ for all three models.

One restriction on this derivation is that y must be small. Other restrictions also exist, but they are not stated explicitly in the original paper. The effect of overlap occurs in the X factors. If all are zero, (4.63) reduces, as expected, to

$$\bar{A} = (1/N) \sum_{i=1}^{N} 2\pi y L(u_i). \tag{4.65}$$

According to the authors (4.63) can be used to compute the mean absorption for water vapor. However, the double sums are tedious and the method is cumbersome; it has not, in fact, been exploited, although, in principle, it is more general than the models discussed in the following sections.

4.5. Regular models

4.5.1. *The Elsasser model for Lorentz lines*

This model (Fig. 4.11) consists of an infinite array of Lorentz lines of equal intensity, spaced at equal intervals. This condition is most closely met for P- and R-branches of linear molecules; see, for example, the spectrum of the 7.78 μm band of nitrous oxide in Fig. 3.3. Even for linear molecules, however, the situation is usually complicated by the presence of upper state, combination, or difference bands that superimpose on the fundamental band. Figure 4.12 shows a part of the ν_2 band of carbon dioxide with the transitions differentiated. A band of this nature may reasonably be treated as six independent Elsasser bands, to be combined by multiplication (see § 4.1).

For the Lorentz profile, the absorption coefficient at a frequency displaced by ν from the center of one line in the array is

$$k_\nu = \sum_{n=-\infty}^{+\infty} \frac{S}{\pi} \frac{\alpha_L}{(\nu - n\delta)^2 + \alpha_L^2}, \tag{4.66}$$

where δ is the line spacing; with the definitions (4.8),

$$mk_x = \sum_{n=-\infty}^{+\infty} \frac{2uy^2}{(x-n)^2 + y^2}. \tag{4.67}$$

Equation (4.67) can be summed with the help of the Poisson sum rule. We may write

$$\frac{mk_x}{2uy^2} = \sum_{n=-\infty}^{+\infty} f_L(n). \tag{4.68}$$

where

$$f_L(n) = \frac{1}{(x-n)^2 + y^2}. \tag{4.69}$$

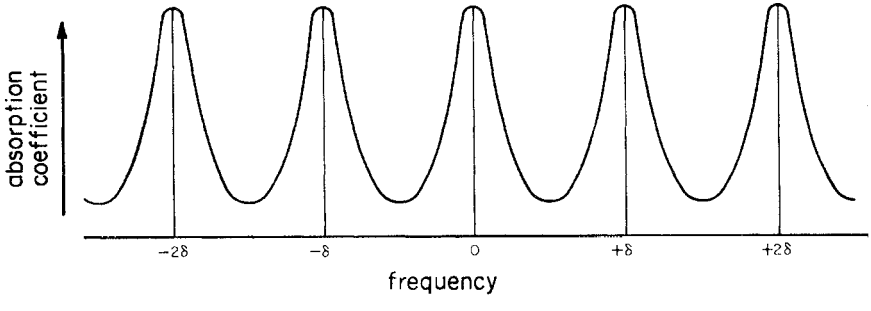


FIG. 4.11. The Elsasser model.

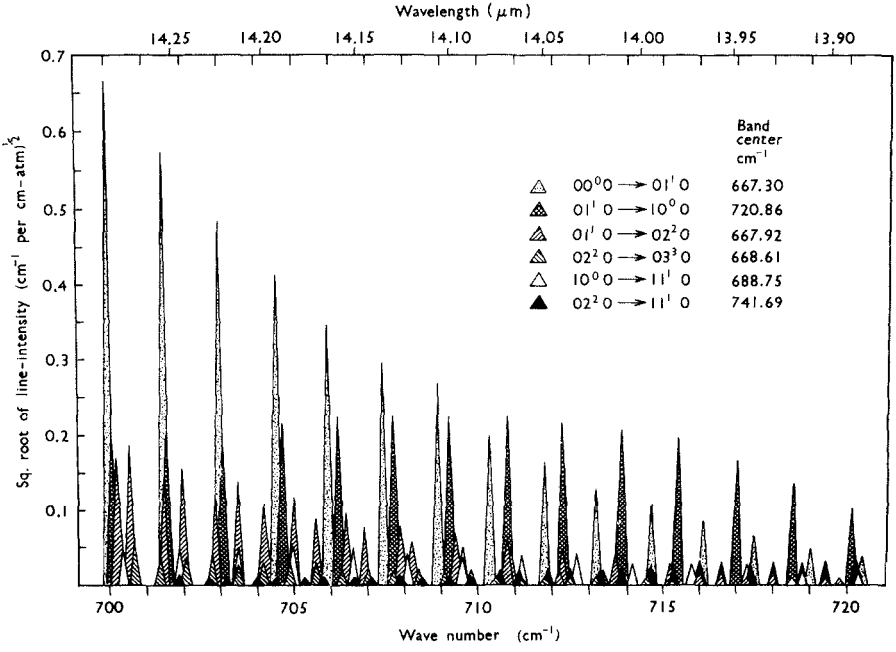


FIG. 4.12. A section of the ν_2 band of carbon dioxide.

The sum in (4.68) may be evaluated from the Fourier transform of f_L, with n treated as a continuous variable ($n = \xi$),

$$F_L(r) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} \exp(ir\xi) f_L(\xi)\, d\xi$$
$$= (1/y)(\pi/2)^{1/2} \exp(-y\,|r| + irx). \tag{4.70}$$

The Poisson sum rule permits us to write,

$$mk_x/2uy^2 = (2\pi)^{1/2} \sum_{n=-\infty}^{+\infty} F_L(2n\pi), \tag{4.71}$$

from which, after some manipulation, we find

$$mk_x = \frac{2\pi uy \sinh 2\pi y}{\cosh 2\pi y - \cos 2\pi x}. \tag{4.72}$$

The average absorption can now be obtained by integrating $(1 - T_x)$ with respect to x from $-1/2$ to $+1/2$,

$$\bar{A} = 1 - E(y, u),$$

where

$$E(y, u) = \int_{-1/2}^{+1/2} \exp\left(\frac{-2\pi u y \sinh 2\pi y}{\cosh 2\pi y - \cos 2\pi x}\right) dx, \tag{4.73}$$

is the *Elsasser function.* It must be evaluated numerically; some data are given in Appendix 8. These may be extended with one of a number of asymptotic forms.

In the limit $y \to \infty$, $\sinh 2\pi y \to \cosh 2\pi y \to \infty$, and

$$E(y, u) \to \exp(-2\pi y u) = \exp(-Sm/\delta). \tag{4.74}$$

In this limit, lines strongly overlap and there is no line structure; further increase of y (i.e., increase of pressure) has no effect upon this continuum. The transmission is now exponential in amount and independent of pressure (this is a special case of Lambert's law known as *Beer's law* of absorption). This behavior is found for all band models when the line width is greater than the line spacing (whatever the line shape may be); it contains the weak-line limit, (4.14), as a special case, for small u.

An approximation for the case of moderately large y (incompletely overlapping lines) can be obtained by expanding (4.72) to give

$$mk_x \approx 2\pi y u(1 + \alpha \cos 2\pi x), \tag{4.75}$$

where

$$\alpha = 2 \exp(-2\pi y).$$

The integral in (4.73) is now

$$E(y, u) = I_0(2\pi y u \alpha) \exp(-2\pi y u). \tag{4.76}$$

Equation (4.76) tends to (4.74) as $y\alpha \to 0$.

This result is not true for all line shapes since explicit use was made of the Lorentz profile. There are, however, analogous results for other line profiles.

A second important asymptotic limit is that of isolated lines. As discussed in § 4.2.1, $\bar{A} \ll 1$ is a necessary condition for (4.10) to provide an approximation to the behavior of a band. Another necessary condition, to avoid overlap, is $y \ll 1$. The sufficient condition for an Elsasser band to tend to the isolated line limit has, however, not been given in a simple form.

The third asymptotic limit is the strong-line limit; it can be obtained, as it was when we discussed isolated lines, by neglecting y^2 in the denominator of the Lorentz profile, (4.68). This is equivalent to allowing

y to tend to zero in the expression (4.73) for the mean absorption. If we write

$$\sinh 2\pi y \approx 2\pi y,$$

$$\cosh 2\pi y \approx 1,$$

the integral in the Elsasser function can be transformed (with a little effort) into

$$E(y, u) = 1 - 2/\pi^{1/2} \int_0^{\pi y\sqrt{2u}} \exp(-\xi^2)\, d\xi, \tag{4.77}$$

or

$$\bar{A} = \operatorname{erf}[\pi y(2u)^{1/2}]. \tag{4.78}$$

These three asymptotic limits provide envelopes to experimental measurements of mean absorption as a function of amount or of pressure, if the axes are correctly chosen (Fig. 4.13). If $\bar{A}$ is plotted against mp ($\propto uy^2$), the strong-line limit, (4.78), is an envelope. If $\bar{A}$ is plotted against m, Beer's law, (4.74), is an envelope. If $\bar{A}p$ or $\bar{A}/m$ is plotted against m/p ($\propto u$) the independent-line limit, (4.10), is an envelope.

For most atmospheric conditions, spectral line widths are less than line spacings and the limit $y \ll 1$ has important applications. When he first introduced the idea of a regular band, Elsasser was under the impression that (4.78) was the limit for $y \ll 1$ but, in reality, it is the strong-line limit ($u \gg 1$). Because of its importance, there have been a number of attempts to identify the limit for $y \ll 1$. One proposal starts with the inverse of (4.78)

$$(1/uy) \operatorname{erf}^{-1} \bar{A} = \pi(2/u)^{1/2}. \tag{4.79}$$

If the argument of the error function is small,

$$\operatorname{erf}^{-1}(x) = (\pi)^{1/2} x/2.$$

If we apply this operator to the weak-line limit (4.14), we find

$$(1/uy) \operatorname{erf}^{-1} \bar{A} = \pi^{3/2}. \tag{4.80}$$

Equations (4.79) and (4.80) can both be written in the form

$$(1/uy) \operatorname{erf}^{-1} \bar{A} = F(u), \tag{4.81}$$

which suggests that this general form may perhaps be valid over a wide parametric range. If the absorption from the Elsasser model is plotted in

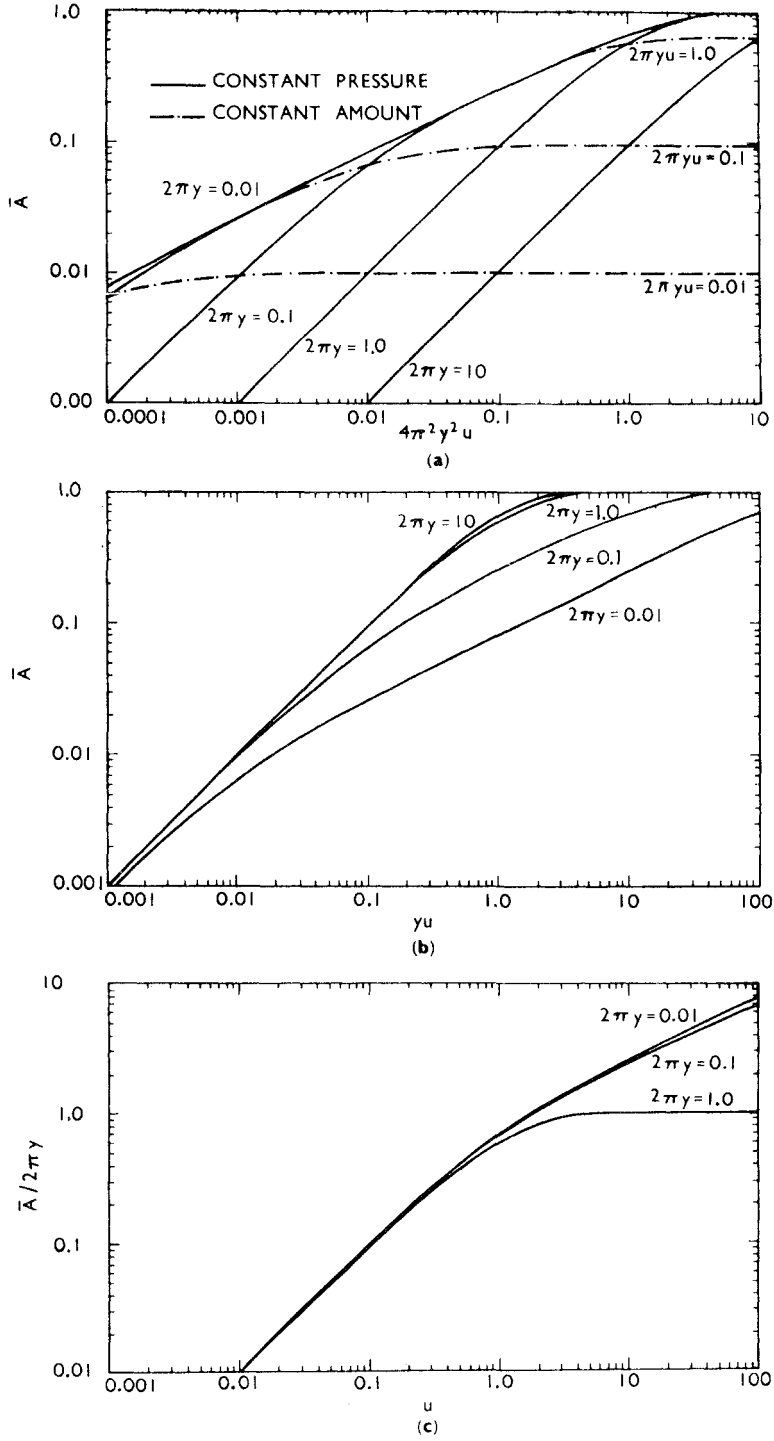


FIG. 4.13. Three methods of plotting the Elsasser function. (a) The left-hand curve is the strong-line limit, (4.78). (b) The left-hand curve is Beer's law, (4.74). (c) The left-hand curve is the isolated line limit, (4.10). After Plass (1960).

the form $(1/yu)\,\mathrm{erf}^{-1}\bar{A}$ as a function of u, it immediately becomes clear that curves coalesce when $y \ll 1$.

This identifies the appropriate asymptotic form for small y and it now remains to find the function $F(u)$. If we hold u constant and allow y to tend to zero (this requires appropriate adjustments to both the pressure and the amount), it is clear that lines will eventually cease to overlap and that the isolated line limit, (4.10), will be reached. At the same time $\bar{A} \to 0$ and

$$\mathrm{erf}^{-1}\bar{A} \to \pi^{1/2}\bar{A}/2 = \pi^{3/2} y L(u), \tag{4.82}$$

where L is the function defined in (4.10). Although (4.82) need only be true for y less than some function of u, the fact that it has the same form as (4.81) ensures that it is valid for all u if $y \ll 1$. We conclude therefore that the expression

$$\bar{A} = \mathrm{erf}[\pi^{3/2} y L(u)] \tag{4.83}$$

is the asymptotic limit to the regular model for $y \ll 1$.

The approximation (4.83) is compared to the exact expression in Table 4.1. Agreement is very good for $y < 1$. If, in addition, $u \ll 1$, we may also expect agreement with the isolated line limit, and this is shown to be so in Table 4.1.

Table 4.1. Exact and approximate forms for the absorption of a regular band when $y \leqslant 1$[a]

	Mean absorption					
	$u = 0.1$		$u = 1$		$u = 10$	
$\log_{10} y$	Approx.	Exact.	Approx.	Exact	Approx.	Exact
0	0.5464	0.4665	1.0000	0.9981	—	—
−0.2	0.3636	0.3273	0.9992	0.9809	—	—
−0.4	0.2344	0.2210	0.9652	0.9146	—	—
−0.6	0.1492	0.1451	0.8172	0.7718	—	—
−0.8	0.0946	0.0934	0.5992	0.5788	0.9984	0.9972
−1.0	0.0598	0.0595	0.4040	0.3975	0.9512	0.9464
−1.2	0.0378	0.0377	0.2622	0.2603	0.7838	0.7811
−1.4	0.0238	0.0238	0.1672	0.1668	0.5648	0.5639
−1.6	0.0150	0.0150	0.1060	0.1059	0.3776	0.3774
−1.8	0.0095	0.0095	0.0670	0.0670	0.2440	0.2440
−2.0	0.0060	0.0060	0.0423	0.0423	0.1555	0.1555
−2.2	0.0038	0.0038	0.0267	0.0267	0.0984	0.0984
−2.4	0.0024	0.0024	0.0169	0.0169	0.0623	0.0623

[a] The approximate expression, (4.83) is compared with the exact form (4.73). Below the double line the isolated-line limit, (4.10), is also valid to better than 0.01.

An alternative proposal for the limit $y \ll 1$, also based on a nonrigorous argument, is

$$\bar{A} = (\pi/2u)^{1/2} L(u) \operatorname{erf}[\pi y (2u)^{1/2}]. \tag{4.84}$$

The differences between (4.83) and (4.84) are very small when $y \ll 1$.

Tests of Elsasser's theory against laboratory data are rare. One reason for this is that few bands are free from overlapping upper-state or other weak bands. The carbon monoxide fundamental is one case for which overlapping bands are unimportant at room temperature, and good agreement between theory and measurement can be expected. Results of a comparison at the R-branch maximum are shown in Fig. 4.14. The method of plotting the results is such as to give a single curve if $y \ll 1$; the theoretical curve follows from (4.83). Pressures employed in Fig. 4.14 range from 5 to 3000 mm Hg. For the highest pressure, the value of y is 0.08. Not only does Fig. 4.14 show excellent qualitative agreement but the line widths and line intensities derived from the intercepts are in good agreement with high-resolution spectrographic data.

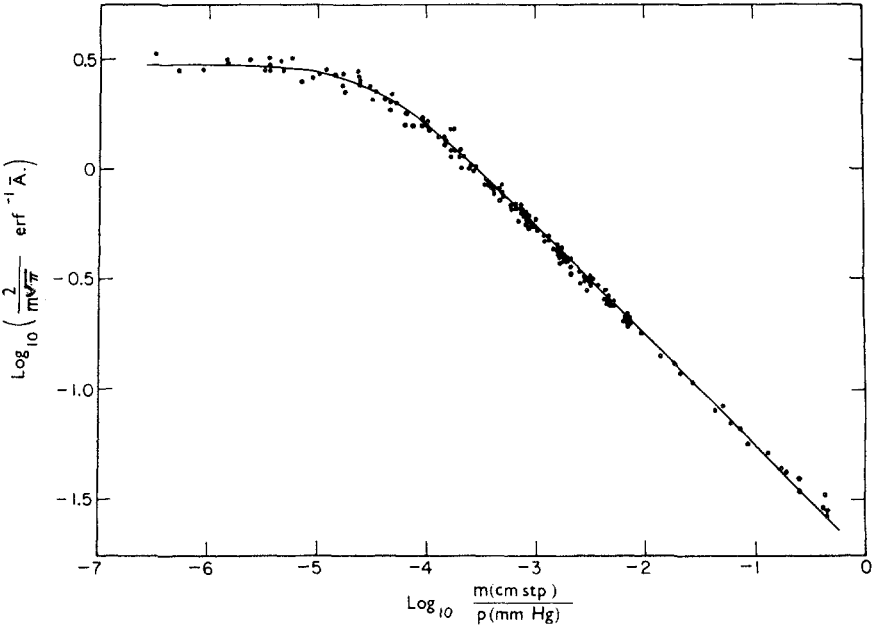


FIG. 4.14. A comparison between theory and measurement for the absorption at the maximum of the R-branch of the fundamental band of CO. The amount of CO ranged from 0.00096 to 45.6 cm s.t.p., the total pressure of the CO–N_2 mixture from 5 to 3000 mm Hg, and the path length from 1.55 to 400 cm. The full line is given by (4.83).

4.5.2. *The Curtis model*

This model does not represent any known bands, but it has the characteristic line spacings of a regular model and is conveniently expressed in terms of tabulated functions.

Consider an array of Lorentz lines with the same widths, but arbitrary intensities. Equation (4.67) can be modified to

$$k_\nu m = \sum_{n=-\infty}^{+\infty} \frac{mS(n)}{\pi} \frac{\alpha_L}{(\nu - n\delta)^2 + \alpha_L^2}. \tag{4.85}$$

The absorption at frequency ν is

$$A_\nu = 1 - \exp\left[- \sum_{n=-\infty}^{+\infty} \frac{mS(n)}{\pi} \frac{\alpha_L}{(\nu - n\delta)^2 + \alpha_L^2} \right], \tag{4.86}$$

and its average value is

$$\begin{aligned} A_\nu^* = &\int_0^\infty dS(1)p[S(1)] \int_0^\infty dS(2)p[S(2)] \cdots \int_0^\infty \\ &\times dS(n)p[S(n)] \ldots \times A_\nu[S(1), S(2), \ldots, S(n), \ldots] \\ &= 1 - \prod_{n=-\infty}^{+\infty} \int_0^\infty p[S(n)] \exp \frac{-mS(n)\alpha_L}{\pi[(\nu - n\delta)^2 + \alpha_L^2]} dS(n), \end{aligned} \tag{4.87}$$

where $p(S)$ is an intensity distribution function. With the exponential distribution, (4.25), for the line intensities and with the definition of u_σ, (4.39), we have

$$A_x^* = 1 - \prod_{n=-\infty}^{+\infty} \frac{(x-n)^2 + y^2}{(x-n)^2 + y^2(1 + 2u_\sigma)}. \tag{4.88}$$

The Poisson sum rule can again be used to transform the infinite product into an expression involving periodic and hyperbolic functions,

$$A_x^* = \frac{\cosh 2\pi y(1 + 2u_\sigma)^{1/2} - \cosh 2\pi y}{\cosh 2\pi y(1 + 2u_\sigma)^{1/2} - \cos 2\pi x}. \tag{4.89}$$

The average absorption is now obtained by integrating this expression with respect to x from $-1/2$ to $+1/2$, with the result

$$\bar{A} = \frac{\cosh 2\pi y(1 + 2u_\sigma)^{1/2} - \cosh 2\pi y}{\sinh 2\pi y(1 + 2u_\sigma)^{1/2}}. \tag{4.90}$$

It is of interest to compare asymptotic forms of (4.90) with those for the Elsasser model. If y becomes large,

$$\bar{A} \to 1 - \exp 2\pi y[1 - (1 + 2u_\sigma)^{1/2}]. \tag{4.91}$$

In the laboratory, this limit is achieved by increasing the pressure. If, at the same time, the path length is not increased indefinitely u_σ must go to zero and

$$\bar{A} \to 1 - \exp(2\pi y u_\sigma), \tag{4.92}$$

in agreement with (4.74).

Now consider the limit to (4.90) as $2\pi y(1 + 2u_\sigma)^{1/2} \to 0$. It follows, necessarily, that $2\pi y \to 0$ and

$$\bar{A} \to \frac{2\pi y u}{(1 + 2u_\sigma)^{1/2}}. \tag{4.93}$$

This is the isolated-line limit, (4.36), for an exponential distribution of line intensities.

Finally, if $u_\sigma \to \infty$, we have the strong-line case, equivalent to the error function limit, (4.78). If the transmission is not to be vanishingly small, the limit must be approached with $2\pi y(1 + 2u_\sigma)^{1/2}$ finite. Thus we must let $y \to 0$. Then,

$$\bar{A} \to \frac{\cosh 2\pi y(2u_\sigma)^{1/2} - 1}{\sinh 2\pi y(2u_\sigma)^{1/2}}. \tag{4.94}$$

This expression differs from (4.78) when the absorption is close to unity, giving a smaller absorption.

4.5.3. *The Elsasser model for the Voigt profile*

The Poisson sum rule can be used to express the absorption for an Elsasser array of Voigt lines in terms of periodic and hyperbolic functions in the same way as for the Lorentz profile. With the Voigt profile, (3.83), the expressions (4.69) and (4.70) become

$$mk_x/2uy^2 = \beta\pi^{-1/2} \sum_{n=-\infty}^{+\infty} f_V(n), \tag{4.95}$$

and

$$f_V(n) = \int_{-\infty}^{+\infty} \frac{\exp[-(\beta\xi)^2]}{(x - n - \xi)^2 + y^2}\, d\xi, \tag{4.96}$$

where

$$\beta = \frac{\delta}{\alpha_D} = \frac{d}{2y}.$$

f_V is a convolution of $\exp[-(\beta\xi)^2]$ with the Lorentz profile f_L, (4.70), and its Fourier transform is 2π times the product of the transforms of f_L and $\exp[-(\beta\xi)^2]$,

$$F_V(k) = (2\pi^{3/2}/d)\exp(-r^2/4\beta^2 - 2y|r| + irx). \tag{4.97}$$

Using the Poisson sum rule, after some manipulation,

$$mk_x = 2\pi yu\left[1 + 2\sum_{n=1}^{+\infty} \exp\left(\frac{-4\pi^2 n^2 y^2}{d^2} - 2\pi yn\right)\cos 2\pi xn\right]. \tag{4.98}$$

As $y \to \infty$, only the first term in the sum in (4.98) need be considered. This yields the same expression for nearly overlapping lines

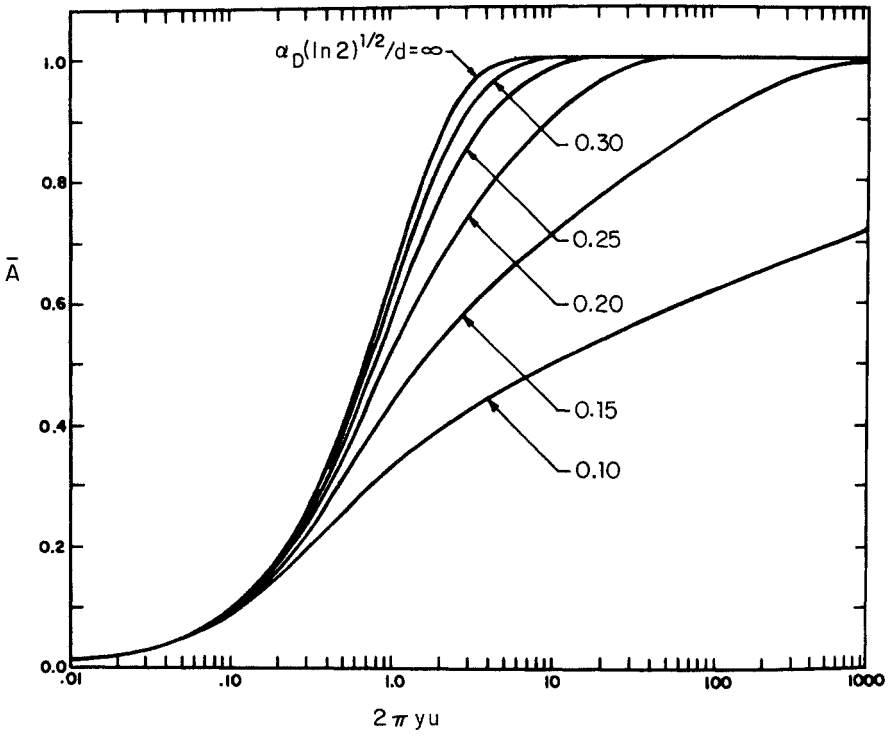


FIG. 4.15. Absorption for an Elsasser band with Doppler lines. The abscissa is equal to Sm/δ and is proportional to the amount of the gas. After Golden (1968).

that was found for the Lorentz profile, (4.75), except that

$$\alpha = 2\exp\left(-\frac{4\pi^2 y^2}{d^2} - 2\pi y\right). \tag{4.99}$$

The Lorentz result (4.76) is valid, as it should be, in the limit $d \to \infty$.

In the Doppler limit ($d \to 0$) the mean absorption,

$$\bar{A} = \int_{-1/2}^{+1/2} [1 - \exp(-mk_x)]\, dx, \tag{4.100}$$

has been evaluated numerically. Some results are shown in Fig. 4.15 and further tabulations are available in the literature.

4.6. Random models

4.6.1. *Introduction*

In 1950, Cowling computed the average absorption in the rotation band of water vapor by what would now be called a line-by-line method, using theoretical line positions and intensities with the Lorentz profile. His computations were for six independent frequency ranges, each covering 25 cm^{-1}, and after he had performed them, he was led to the conclusion that "in atmospheric work complication is avoided, and remarkably little error is involved, if a single absorption curve is used at all wavelengths." Inspection of the spectra in Figs. 3.6 and 3.24 suggests that the only feature common to 25 cm^{-1} ranges is the apparently random line positions, and, therefore, that we should enquire into the absorption of a band with random line positions.

Two approaches have been made to the problem. The first considers the average absorption at the center of a finite array of N lines and examines the limit as N goes to infinity. The second takes as a starting point any infinite array, such as the Elsasser model, and combines a number of arrays by multiplication; in the limit of a large number of arrays, the same result as the first method is achieved.

The main difference between the two methods is that with the second, the intermediary stages, when a finite number of bands is combined, are real physical situations whereas with the first approach, a finite number of lines is a mathematical fiction. The essential feature common to both is the existence of all possible phase relations between lines, implying that lines or arrays are placed at random with respect to frequency. The probability of a line lying between ν and $\nu + d\nu$ is then proportional to $d\nu$, and the inverse of the proportionality factor is, by definition, the mean line spacing δ.

This implies a Poisson distribution for the spacings between neigh-

boring lines. Consider a band of frequencies between $\nu = -N\delta/2$ and $\nu = +N\delta/2$ containing N lines. Consider a subrange of frequencies of width Δ. If one line is placed randomly in the larger range, the probability that it will fall in the subrange is $\Delta/N\delta$. The probability that it will not fall in the subrange is $1 - \Delta/N\delta$, and the probability that N lines will not do so is

$$\pi(\Delta) = (1 - \Delta/N\delta)^N. \tag{4.101}$$

But,

$$\lim_{N\to\infty} (1 + x/N)^N = \exp(x), \tag{4.102}$$

and, therefore, for a large number of lines in the interval, the probability that there is a gap or spacing of width Δ is

$$\pi(\Delta) = \exp(-\Delta/\delta). \tag{4.103}$$

When a limited number of lines is considered, (4.103) may not hold, and even for 200 water vapor lines significant deviations from a Poisson distribution have been found. In one investigation there was a significant lack of very large gaps and, for very small mean transmissions, such gaps can be of great importance. Since line positions are determined by quantum-mechanical formulas, some order must be present in the line spacings and the hypothesis of randomness can be judged only by the results achieved.

4.6.2. *Constant line intensity*

Consider an array of identical lines whose shapes are described by the absorption coefficient k_ν. Let N lines be distributed randomly between $-N\delta/2$ and $+N\delta/2$. The absorption coefficient at the center of the array, caused by lines at frequency displacements ν_i, is

$$\sum_{i=1}^{N} k_i.$$

The resultant transmission is

$$\begin{aligned} T &= \exp\left(-m \sum_{i=1}^{N} k_i\right) \\ &= \prod_{i=1}^{N} \exp(-mk_i). \end{aligned} \tag{4.104}$$

The probability that a line lies in the interval $d\nu_i$ is $d\nu_i/\delta$ and the

joint probability that there are lines between ν_1 and $\nu_1 + d\nu_1$, ν_2 and $\nu_2 + d\nu_2, \ldots,$ and ν_N and $\nu_N + d\nu_N$ is

$$\prod_{i=1}^{N} \frac{d\nu_i}{\delta}.$$

If we consider all possible arrangements of lines we must allow each line to lie anywhere in the range $-N\delta/2$ to $+N\delta/2$, and the appropriate average of (4.104) is

$$\bar{T} = \frac{\Pi_{i=1}^{N} \int_{-N\delta/2}^{+N\delta/2} (d\nu_i/\delta) \exp(-mk_i)}{\Pi_{i=1}^{N} \int_{-N\delta/2}^{+N\delta/2} (d\nu_i/\delta)}. \tag{4.105}$$

The N integrations in both numerator and denominator of (4.105) are identical and

$$\begin{aligned} \bar{T} &= \left[(1/N) \int_{-N\delta/2}^{+N\delta/2} \exp(-mk_\nu)\, d\nu/\delta \right]^N \\ &= \left\{ 1 - (1/N) \int_{-N\delta/2}^{+N\delta/2} [1 - \exp(-mk_\nu)]\, d\nu/\delta \right\}^N. \end{aligned} \tag{4.106}$$

As $N \to \infty$,

$$\begin{aligned} \bar{T} &\to \exp\left\{ (-1/\delta) \int_{-\infty}^{+\infty} [1 - \exp(-mk_\nu)]\, d\nu \right\} \\ &= \exp(-W/\delta), \end{aligned} \tag{4.107}$$

or

$$\bar{A} = 1 - \exp[-\bar{A}(\text{isolated line})]. \tag{4.108}$$

It is tempting not to proceed to the limit of large N, as we have done here, but to adopt (4.106) as a generalized transmission function, that includes (4.107) as a special case. The integral in the exponent of (4.106) is available for the Lorentz profile and was discussed in § 4.4.1. However, the derivation of (4.105) assumed all possible line positions and this can be so only for an infinite number of lines. A finite number of lines is a mathematical fiction.

No specific line profile was assumed in the derivation of (4.107) which is, therefore, valid for all profiles. Thus, all of the theory of isolated-line models can now be utilized in band models with overlapping lines.

We may now compare the properties of random and regular models

having a single line intensity. Both have independent-line and Beer's law asymptotic limits, and a little consideration shows that the two models must be identical in these limits, provided that the same line intensities and widths are employed. The most important differences exist for strong lines. If we substitute (4.15) into (4.108) we find, for the strong-line limit,

$$\bar{A}(\text{random}) = 1 - \exp[-2y(2\pi u)^{1/2}]. \tag{4.109}$$

For the Elsasser model, the comparable expression is (4.78). If the argument of the error function is large, we may use an asymptotic limit, and (4.78) becomes

$$\bar{A}(\text{regular}) = 1 - \frac{\exp(-2u\pi^2 y^2)}{\pi^{3/2} y (2u)^{1/2}}. \tag{4.110}$$

Equations (4.109) and (4.110) may be compared when $\bar{A}$ approaches unity. There is little resemblance between them. In particular, the regular model tends to unity much more rapidly than does the random model as $yu^{1/2} \to \infty$. This difference arises from the occasional large gaps in a random array that require very large amounts of absorbing material before they cease to transmit.

There are no known cases of bands with random line spacings but with only one line intensity. In observed spectra, intensities are distributed over a wide range, and this was the reason for the discussion of distributed line intensities in § 4.3. In the next section, we show that (4.107) can be extended to any distribution of line intensities.

4.6.3. *The general random model*

A frequency range contains N lines, each a single member of a different infinite array of random lines of equal intensity and line spacing $N\delta$. The mean transmission of one of the N arrays is, from (4.107),

$$\bar{T}_i = \exp(-W_i/N\delta), \tag{4.111}$$

where W_i is the equivalent width of one line in the array under consideration.

The conditions for the multiplication property (§ 4.1) are met for random arrays and the N arrays can be combined by multiplication,

$$\begin{aligned} \bar{T} = \prod_{i=1}^{N} \bar{T}_i &= \exp\left[-(N\delta)^{-1} \sum_{i=1}^{N} W_i\right] \\ &= \exp(-\bar{W}/\delta), \end{aligned} \tag{4.112}$$

where $\bar{W}$ is the average equivalent width for the N lines in the selected frequency range.

Equation (4.112) was derived without reference to a line profile and it is not even restricted to a single profile for all of the lines. The only limitation is that the frequency interval should be large enough for the multiplication property to be valid.

Evaluation of the argument of the exponential in (4.112) was the purpose of § 4.3. Solutions are available for the Voigt profile and for four different distributions of line intensities. All data discussed in that section may be taken over and used in the appropriate random model. An alternative approach is not to use the formal distributions discussed in § 4.3 but to calculate the equivalent widths of each line in an array, using the isolated line result, (4.10). The average of these equivalent widths may then be employed in (4.112). Although this may appear to be a time-consuming procedure, it is very much less so than the line-by-line methods in which the transmission is calculated at each frequency and then averaged.

4.6.4. *Verification of the theory*

The hypothesis of random line positions can be tested by comparing model transmission with line-by-line calculations in specific cases. Figure 4.16 shows Cowling's original calculations for the water vapor rotation band. Best agreement is shown with the universal curve (a) which represents an average over all the computed ranges. The agreement is good, even bearing in mind that two free parameters can be adjusted to give a best fit. The range 100–125 cm^{-1} was computed for four different pressures and the agreement shown in (b) indicates how well the effect of pressure broadening is taken into account by the model. The range 175–250 cm^{-1} is of interest because the three individual ranges of 25 cm^{-1} that were combined each failed to fit the model well, but, when averaged, the fit is fairly good. The range 300–350 cm^{-1} contains only 14 lines and agreement with the random model is not good as the transmission approaches zero.

Other comparisons with line-by-line transmission calculations also show that, under appropriate conditions, the random model can provide a good approximation. In Fig. 4.17 the random model is derived from line strengths and widths for the 9.6 μm band of ozone using the best-fit parameters from (4.43), (4.44), and (4.45). The line-by-line calculations use the same line parameters and are averaged over the same frequency intervals (2.5 cm^{-1}) as those used for the random model. The agreement is excellent and similarly good agreement was found in the same study for the 2.7 μm band of water vapor. Comparative calculations were also made for both the exponential distribution (used in Fig. 4.17) and the Malkmus distribution, with the appropriate best-fit parameters; very little

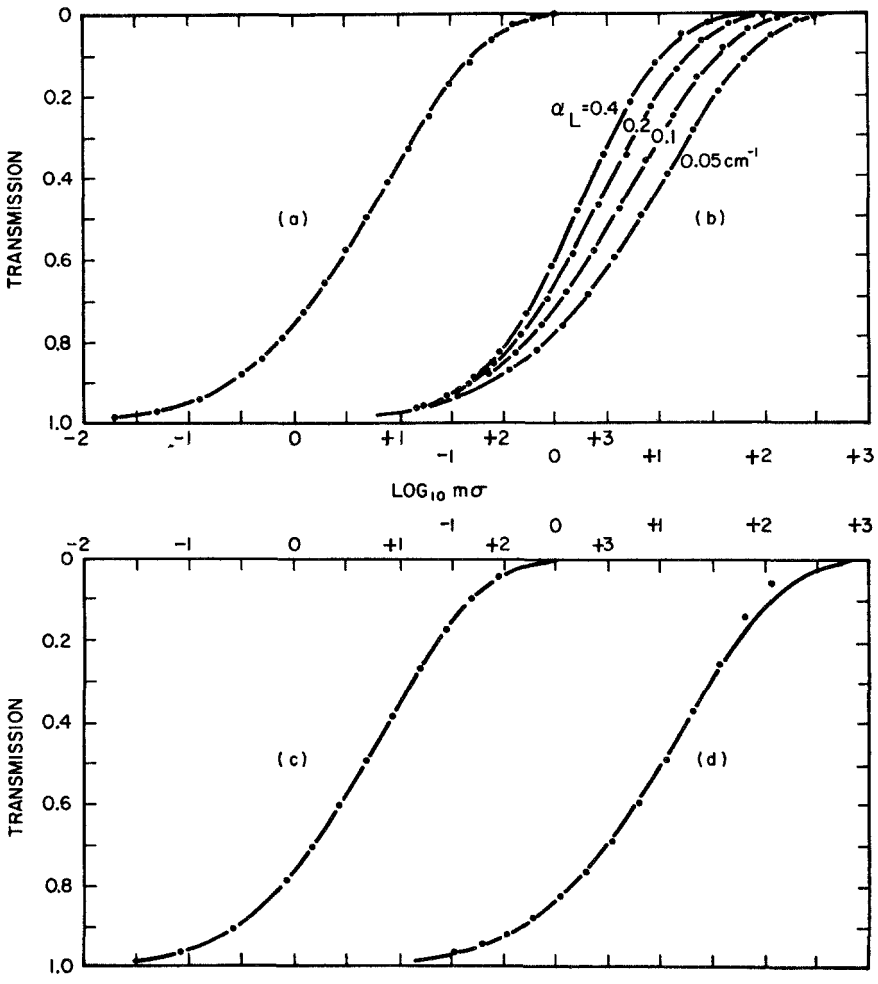


FIG. 4.16. Comparison of a random model with line-by-line calculations for the water vapor rotation band. The points are line-by-line calculations for the water vapor rotation band by Cowling (1950), and the curves are for the random model (4.107) with an exponential distribution of line intensities (4.36), with y and u chosen to give best fit. (a) Cowling's *universal curve*, his estimate of the average behavior of all of the ranges calculated by him; (b) 100 to 125 cm^{-1}, calculated for four line widths. The theoretical curves use the same value of σ and δ in each case; (c) 175 to 250 cm^{-1}; (d) 300 to 350 cm^{-1}.

difference was found between the two. In the case of the 2.7 μm band of water vapor, laboratory measurements were also available. Line-by-line and model calculations, while in agreement with each other, differed considerably from the laboratory data. This makes the important point that errors inherent in the adoption of a band model may not be very important for atmospheric calculations, but that much remains to be done to find the best band parameters from laboratory data.

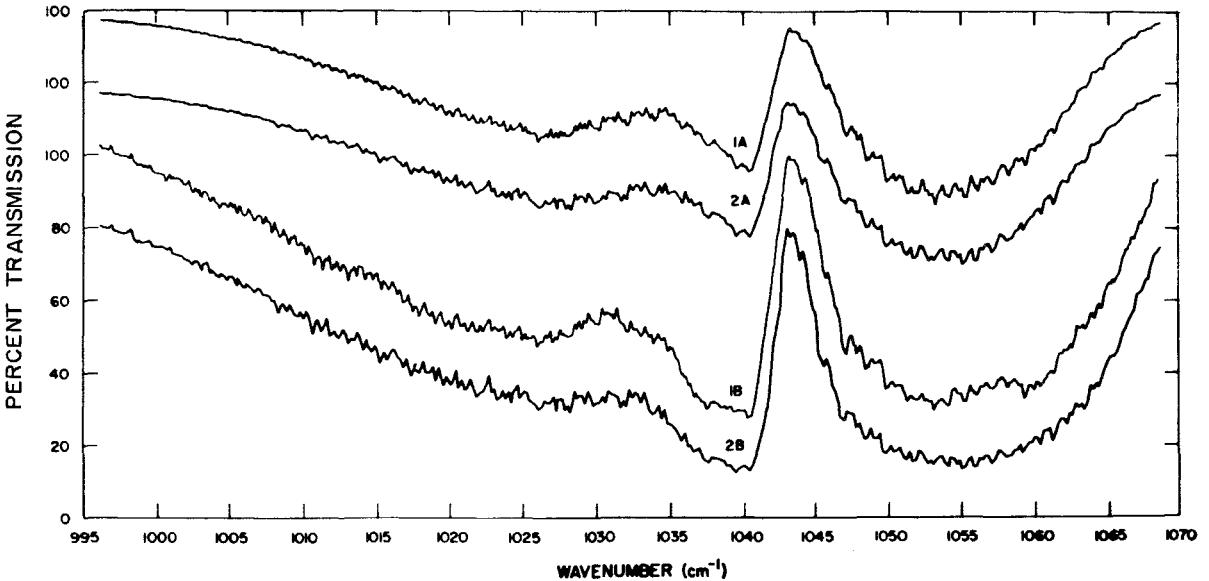


FIG. 4.17. Comparison of a random model with line-by-line calculations for the 9.6 μm band of ozone. 1A and 1B are line-by-line calculations, while 2A and 2B use the same line parameters in a random model with an exponential distribution of line intensities. Calculations are for overlapping intervals, each 25 cm^{-1} wide; the line width is 0.08 cm^{-1} at s.t.p. and the temperature is 233 K. The upper curves are for a pressure of 0.197 bar and a path length of 0.2447 cm; the lower curves are for 0.0197 bar and 97.88 cm. The model calculations are displaced downward by 20%. After Goldman and Kyle (1968).

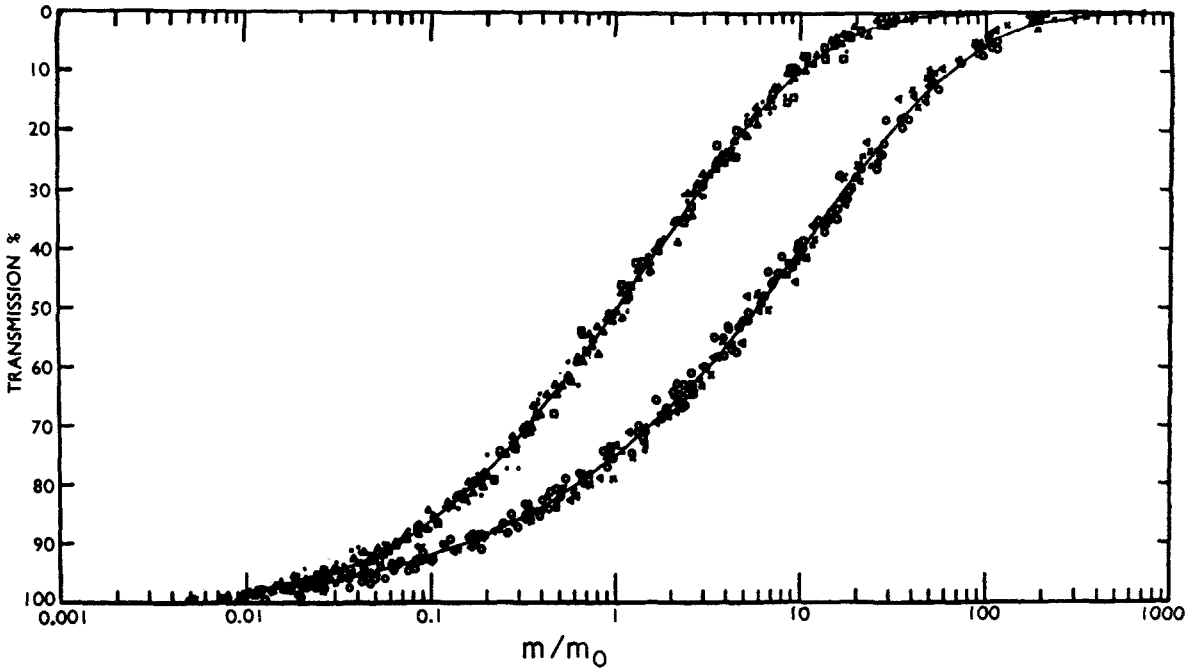


FIG. 4.18. Comparison of the random model with laboratory measurements in water vapor bands. The left-hand curve is for a pressure of 740 mm Hg of dilutant nitrogen and the right-hand curve is for 125 mm Hg. m_0 is the amount of water required to give a transmission of 0.5 at 740 mm Hg. 6.3 μm band: 740 mm Hg (□), 125 mm Hg (×); 2.6 and 3.2 μm bands: 740 mm Hg (△), 125 mm Hg (▽); 1.87, 1.38, and 1.1 μm bands: 740 mm Hg (●), 125 mm Hg (○). After Howard et al. (1956).

Figure 4.18 shows a comparison between laboratory data on water vapor bands and a random model using an exponential distribution of line intensities. The abscissa is m/m_0, where m_0 is the amount of water vapor required to give an absorption of one-half at 740 mm Hg pressure. Since amounts and line intensities always occur as a product, this method of plotting eliminates differences from band to band, or between parts of a band, that arise from differences in the mean line intensity. If α_L/δ is the same for all spectral ranges, all experimental points, from all bands, should lie on the same curve. This is so within the experimental error, although the introduction of m_0 as an adjustable parameter makes this test less discriminating than it may appear to be.

Comparison of theory and experiment for pressure-broadened spectra is facilitated by a method of plotting that does not depend on the distribution of line intensities. From (4.112) and (4.10),

$$\ln \bar{T} = -\bar{W}/\delta = -2\pi\overline{y_i L(u_i)}.$$

If, in accordance with the Lorentz theory, we write $\alpha_i = \alpha_i^0 p/p_0$, we have

$$\begin{aligned}(\ln \bar{T})/m &= -2\pi\overline{(\alpha_i^0/\delta p_0)(p/m)L(S_i p_0 m/2\pi\alpha_i^0 p)} \\ &= -F(m/p),\end{aligned} \tag{4.113}$$

regardless of the nature of the average represented by the overbar. If we plot experimental results for $\log(-\ln \bar{T}/m)$ as a function of $\log(m/p)$ the points should lie on a single curve.

The inherent plausibility of the random model, its simplicity, and the many comparisons that have been made with line-by-line calculations and with laboratory data have led to its wide adoption for atmospheric computations. The most popular distribution of line intensities is that of Malkmus, since it usually gives a better account of the weak lines in the atmospheric spectrum than does the exponential distribution. Either distribution can be used as an effective substitute for line-by-line calculations; whichever method is used, the limiting factor in the accuracy of a thermal calculation is likely to be the line parameters employed rather than the precision of the model. If these parameters have not been obtained by fitting to good low-resolution, laboratory transmission data, there should be cause for concern.

4.7. Generalized transmission functions

4.7.1. *Superimposed regular and random bands*

In this and the next two sections, we discuss means for generalizing band models. The simplest approach is to superimpose random and regular bands, using the multiplication property in cases where analysis of the

fine structure of a band shows this to be appropriate. For example, in the case of the band illustrated in Fig. 4.12, the value of analyzing into Elsasser subbands is self-evident.

This idea can be generalized to give a simple three-parameter model that includes both random and regular models as asymptotic cases. Consider the superimposition of n bands, each with a line spacing $\delta' = n\delta$. The transmission of one array is

$$\bar{T}(m, \delta') = 1 - \bar{A}(m, n\delta),$$

and the transmission for n superimposed arrays is

$$\bar{T}(n, m, \delta) = \bar{T}(m, \delta')^n = [1 - \bar{A}(m, n\delta)]^n. \tag{4.114}$$

Regardless of the form of $\bar{A}(m, n\delta)$, we know from previous discussion that a random model results in the limit $n \to \infty$. Thus, we may create a series of models lying between any model we choose and a random model. The transition from an Elsasser model ($n = 1$) to a random model with equal line intensities ($n = \infty$) is illustrated in Table 4.2.

Although this connection between the two fundamental models is interesting, its usefulness is restricted. There are an unlimited number of three-parameter models that include random and regular models as asymptotic limits. Equation (4.114) is perhaps the simplest, but its use must be justified on either empirical or theoretical grounds.

Table 4.2. Transition from an Elsasser to a random model[a]

		$\log_{10} y$			
$\log_{10} u$	n	0	−0.4	−0.8	−1.2
	1	0.273	0.109	0.043	0.017
−1.0	10	0.267	0.105	0.042	0.016
	∞	0.260	0.103	0.041	0.016
	1	2.729	1.068	0.376	0.131
0.0	10	2.200	0.797	0.301	0.118
	∞	1.836	0.731	0.291	0.116
	1	—	—	2.550	0.660
+1.0	10	—	—	1.215	0.450
	∞	—	—	1.077	0.429

[a] The quantity tabulated is $\log_{10} \bar{T}(n, u, y)$ from (4.114), using the Elsasser model, (4.73).

4.7.2. *Deviations from the Voigt profile*

The theory of band absorption can be developed for any line profile although, in practice, it is worthwhile to do so only for the Voigt profile. Interpretation of data would be greatly complicated should these results not have wide validity. However, we have seen that the Voigt profile is only an approximation, and it is important to understand the effect of deviations from it.

Consider the equivalent width of a Lorentz line with modified wings, (4.17), in the context of a random band with equal line intensities, and compare the result to that obtained with the Lorentz profile. From (4.112) we have

$$\frac{\Delta \bar{T}}{\bar{T}} = \frac{\Delta W}{\delta}, \tag{4.115}$$

provided that the changes indicated by Δ are small. Using the first term in (4.17), $n = 1.73$ [see (3.75)], and $W \gg \alpha_L$ (which applies to all but trivial cases), we have

$$\Delta W = \frac{0.27 W^2}{2\pi 0.73 \nu_P},$$

and

$$\Delta \bar{T} = \frac{\bar{T}(\ln \bar{T})^2 0.27 \delta}{2\pi 0.73 \nu_P}.$$

But,

$$\bar{T}(\ln \bar{T})^2 \leq 0.54$$

and

$$\Delta \bar{T} \leq 3.2 \times 10^{-2} \delta / \nu_P. \tag{4.116}$$

Estimates of δ/ν_P are 0.82 for H_2O, 0.050 for O_3, and 0.36 for the 7.6 μm CH_4 band, for which the transmission errors will be less than 2.6, 0.16, and 1.1%, respectively. Detailed computations using the Lindholm line shape and an exponential distribution of line intensities give almost identical results.

Errors of this magnitude are probably not important for most atmospheric calculations, particularly if band parameters are selected, or at least adjusted, by fitting the model to precise low-resolution laboratory data. When this is done, small errors in the model may be compensated in the selection of model parameters.

This discussion should not, however, be taken to mean that deviations from the Lorentz profile are always unimportant. If we are working in a spectral region in which most of the absorption is caused by line wings, the wing shape is evidently important. Such is the case for continua between bands, such as the 10 μm water vapor window. If we

had to depend upon a priori calculations for such windows we would need to know much more about the physics of close collisions than we do at the present time. However, a semiempirical approach is possible and will be described in the next section.

4.7.3. *Background continuum*

When we attempt to represent a frequency interval in an actual band by an infinite model an important source of error arises from the fact that, for the real band, the statistics of the line intensities outside the interval will differ from those inside the interval. The absorption coefficients contributed by lines outside the interval are from line wings and may be approximated by a structureless continuum. We may regard this particular source of error as equivalent to making an error in this continuum absorption. We treat this situation empirically by adding or subtracting a continuum coefficient from the absorption coefficient for the infinite model.

The effect of a small continuum (positive or negative) may be very large if the transmission is small (see Fig. 4.19). This figure has been plotted in such a way [see (4.113)] that all data should lie on a single curve for any random model. In this theoretical example, the random model curve is the same as that shown for small absorber amounts ($m \ll 1$). A continuum has been added with a coefficient proportional to $m \times p$, as we anticipate for pressure broadened line wings. With this method of plotting, the deviations from the random model are obvious and the data can be used to calculate the required continuum correction. This appears to have been done only for a single band, the 9.6 μm band of ozone. It is important that it be done systematically for all relevant atmospheric bands.

Note that this semiempirical approach deals equally well with deviations from the Lorentz profile in line wings. It is irrelevant whether the continuum is in error because of varying line statistics or because of an incorrect line shape.

4.8. *k* distributions

4.8.1. *Band models and spectral representations*

As far as band transmission is concerned, it is of no importance where in a frequency interval, $\Delta\nu$, a particular absorption coefficient occurs; it is sufficient to know what fraction of the frequency domain, $f(k)\,dk$, is occupied by absorption coefficients between k and $k + dk$. It is equally acceptable to write the mean transmission for a homogeneous path in either of the two forms

$$\bar{T}(m) = (1/\Delta\nu) \int_{\Delta\nu} \exp(-k_\nu m)\, d\nu, \tag{4.117}$$

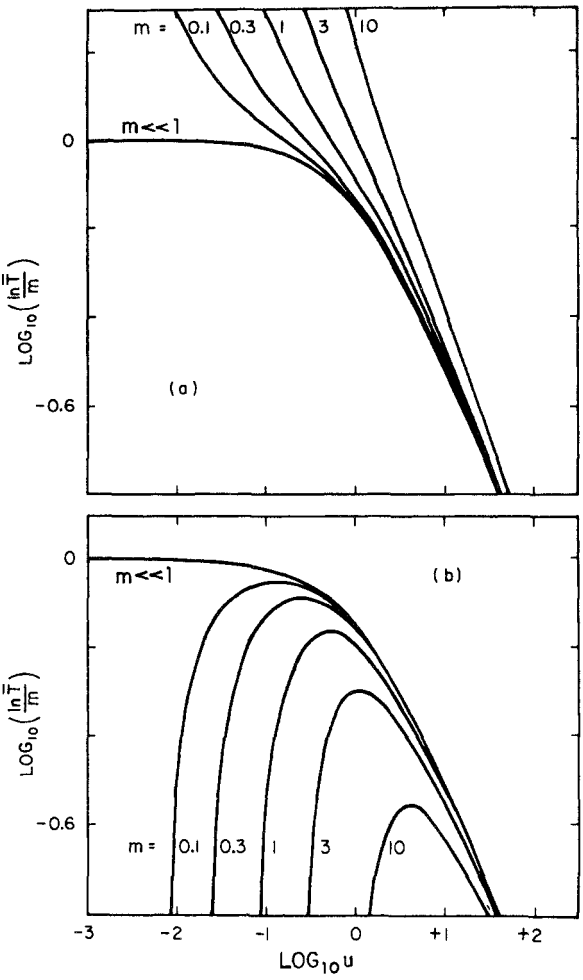


FIG. 4.19. Random model for Lorentz lines with an added continuum. (a) $\tau_c = +0.1(2\pi y^2 u)$. (b) $\tau_c = -0.07(2\pi y^2 u)$. τ_c is the optical path of the continuum. The lines are for constant values of $2\pi yu$, i.e., constant amount. After Walshaw (1954).

or

$$\bar{T}(m) = \int_0^\infty f(k) \exp(-km)\, dk. \tag{4.118}$$

Equation (4.117) is the form appropriate to band model theory provided that $\Delta\nu$ contains many lines. Equation (4.118) defines the *spectral* or *k distribution* approach.

The *k* distribution has a long history in both the astrophysical and meteorological literatures (see Bibliography). It was an obvious extension

of the earliest work with grey models. A few terms in a series,

$$\bar{T}(m) = \sum_{i=1}^{N} \alpha_i \exp(-k_i m), \tag{4.119}$$

must be an improvement on a grey model, and, with $\alpha_i = f(k_i)\delta k_i$, (4.119) tends to the correct expression, (4.118), as $N \to \infty$.

k distribution theory has been used relatively less than the theory of band models for two main reasons. First, Elsasser's early work created an impetus in favor of band models that was reinforced by the discovery of the hierarchy of random models. Second, means were discovered to extend band-model theory to the important case of nonhomogeneous atmospheric paths. This question will be discussed in Chapter 6. At this point, we need only comment that no comparable techniques were known for the k distribution, until recently. With this difficulty removed, the spectral approach has some important advantages over band models.

In order to discuss both advantages and disadvantages we must look more closely at the relationship between the expressions (4.117) and (4.118). This relationship can be looked upon in two different ways. The first makes no reference to the details of the absorption spectrum. The transmission function (4.117) is regarded as an established function of amount, either from band model theory or as the result of low spectral resolution measurements in the laboratory. Equation (4.118) then defines $f(k)$ as a spectral function that has $\bar{T}(m)$ as its Laplace transform,

$$\bar{T}(m) = \mathscr{L}[f(k)]. \tag{4.120}$$

Since $\bar{T}(m)$ may be taken to be continuous and analytic, we may invert (4.120) to find $f(k)$,

$$f(k) = \mathscr{L}^{-1}[\bar{T}(m)]. \tag{4.121}$$

If (4.121) can be solved, it provides a convenient way to find the spectral function.

The second way to look upon the relationship between (4.117) and (4.118) is in terms of the details of the absorption spectrum. The absorption coefficient has a large but finite number of maxima and minima. We define the ith subrange as lying between the ith maximum and the ith minimum, at which the absorption coefficients are $k_i(\max)$ and $k_i(\min)$. In this subrange we may make the change of variable

$$\frac{d\nu}{\Delta\nu} \to \frac{dk}{\Delta\nu}\left|\frac{d\nu}{dk_\nu}\right|_i. \tag{4.122}$$

But this change of variable need not be restricted to evaluation of the

transmission. The expression

$$\bar{G}(m) = (1/\Delta\nu) \int_{\Delta\nu} G[k_\nu]\, d\nu, \tag{4.123}$$

where G is any analytic function, may be transformed to

$$\bar{G}(m) = \int_0^\infty f(k)G(k)\, dk, \tag{4.124}$$

with

$$f(k) = \sum_{i=1}^{N} \frac{1}{\Delta\nu} \left| \frac{d\nu}{dk_\nu} \right|_i \times \{h[k - k_i(\min)] - h[k_i(\max) - k]\}, \tag{4.125}$$

where h is the step-function.

Equation (4.125) can be used in two ways. If we have an analytic form for the frequency spectrum it can be differentiated directly. Two examples for which this has been done are the Elsasser and Schnaidt models, for both of which there is only one independent subrange and

$$f(k) = \frac{2}{\delta} \left| \frac{d\nu}{dk_\nu} \right| \qquad \text{for } 0 \leq \nu \leq \delta/2. \tag{4.126}$$

This equation may be regarded as an ingenious way to evaluate the inverse Laplace transform in certain special cases.

The second way to use (4.125) is more general, as a binning algorithm for use with a spectral compilation, such as the AFGL tape. The evaluation of each absorption coefficient requires a sum to be taken over many neighboring and even distant lines and it is a major computational task; but, if the distribution function $f(k)$ can be reused for many purposes, it may be worthwhile.

If we had stated the problem of this section only in terms of the inverse Laplace transform (4.121), we would have missed the real importance of the k distribution approach. As applied to the mean transmission, the two techniques are alternatives, either of which may be used, depending upon mathematical convenience. However, we see from (4.124) that the same spectral function can be applied to any property of the radiation field that is a function of the absorption coefficient. The significance of (4.121) is that it allows us to exploit the theory of band models to derive k distributions that may be applied to other situations.

We are now in a position to discuss some of the advantages and disadvantages of band models and k distribution theory.

1. Scattering problems can be treated by k distribution theory but not by band models. To remind ourselves of the problem with band models we return to the derivation of (4.2) in which we replaced the mean value of a product $\overline{B_\nu T_\nu}$ by the product of mean values $B_i \bar{T}_i$. This is acceptable for a thermal source function since it may be treated as a constant over the ith spectral interval. A scattering source function, (2.32), on the other hand, involves the local radiation intensity, which is a function of the absorption coefficient. J_ν and T_ν are now partially correlated, and (4.2) is incorrect. But if $G(k)$ is any function of the radiation field, including the scattered intensity, that can be calculated for a single value of k, (4.124) can be used to give the average over a spectral interval.
2. k distribution theory can be used for wide spectral regions that may not be statistically homogeneous. For example, the 15 μm carbon dioxide band has wide P- and R-branches with a very different Q-branch. Provided that the Planck function does not vary significantly over the band, there is no reason not to apply a numerical binning algorithm to the whole band. (In § 4.9 we shall discuss entire bands in terms of band model theory.)
3. In practice, we always use approximations to the atmospheric transmission. In the case of band models, approximations are made to the behavior of the absorption coefficient as a function of frequency. Since we do not need to know where, in a chosen spectral interval, a particular absorption coefficient occurs, frequency is usually not a very significant physical parameter, and the importance of an approximation in frequency space can be assessed only by trial and error. On the other hand, the k distribution is approximated by replacing the integral, (4.118), by a discrete sum. Here we are better able to assess the effect of errors because the absorption coefficient is directly related to the required radiation field. As an example, it is possible to find simple analytic approximations for large and for small absorption coefficients using the transparent and opaque limits (§ 2.4.2). This involves an external length scale specific to a particular problem. Given such knowledge, remarkable accuracy is possible for the k distribution using as few as 10 well-chosen terms in the discrete sum.
4. The other side of this is that band models are able to give an explicit treatment of the important effects of pressure on absorption, just because they do treat the relationships between lines in frequency space. With the frequency scrambling that accompanies the calculation of a k distribution these relationships are destroyed and, in the general case, the k distribution must be recalculated for each value of the pressure (and the temperature).

In those cases for which an inverse Laplace transformation of an analytic transmission function can be made, explicit pressure dependence can, of course, also be given for the k distribution.

5. The discussion in (4) also has relevance to the problem of nonhomogeneous paths. We will consider the advantages and disadvantages of band models and k distribution methods for nonhomogeneous paths in Chapter 6.

4.8.2. *Calculations of k distributions*

The k distribution has been obtained for the Malkmus random model, the Elsasser model with Lorentz lines, and the Schnaidt model, the first by means of an inverse Laplace transformation and the other two from the differential of the absorption coefficient, (4.126). The results are

Malkmus

$$kf(k) = \frac{1}{2}\left(\frac{\bar{k}y}{k}\right)^{1/2} \exp\left[\frac{\pi y}{4}\left(2 - \frac{k}{\bar{k}} - \frac{\bar{k}}{k}\right)\right], \tag{4.127}$$

Elsasser

$$kf(k) = \frac{1}{\pi}\left[2\frac{k}{\bar{k}}\coth(2\pi y) - 1 - \frac{k^2}{\bar{k}^2}\right]^{-1/2}, \tag{4.128}$$

for

$$\frac{\sinh 2\pi y}{\cosh 2\pi y + 1} \leqslant \frac{k}{\bar{k}} \leqslant \frac{\sinh 2\pi y}{\cosh 2\pi y - 1},$$

Schnaidt

$$kf(k) = \frac{1}{\pi}\left(\frac{1}{\pi y}\frac{k}{\bar{k}} - \frac{k^2}{\bar{k}^2}\right)^{-1/2}, \tag{4.129}$$

for

$$\frac{1}{\pi(y + 1/4y)} \leqslant \frac{k}{\bar{k}} \leqslant \frac{1}{\pi y}.$$

In these expressions $\bar{k} = S/\delta$. For the Malkmus and Elsasser models $\bar{k}$ is the mean absorption coefficient, but not for the Schnaidt model because it is incorrectly normalized.

k distributions for the Malkmus and Elsasser models, (4.127) and (4.128), are compared in Fig. 4.20. The differences between the two sets of curves are striking, one being convex upward and the other convex

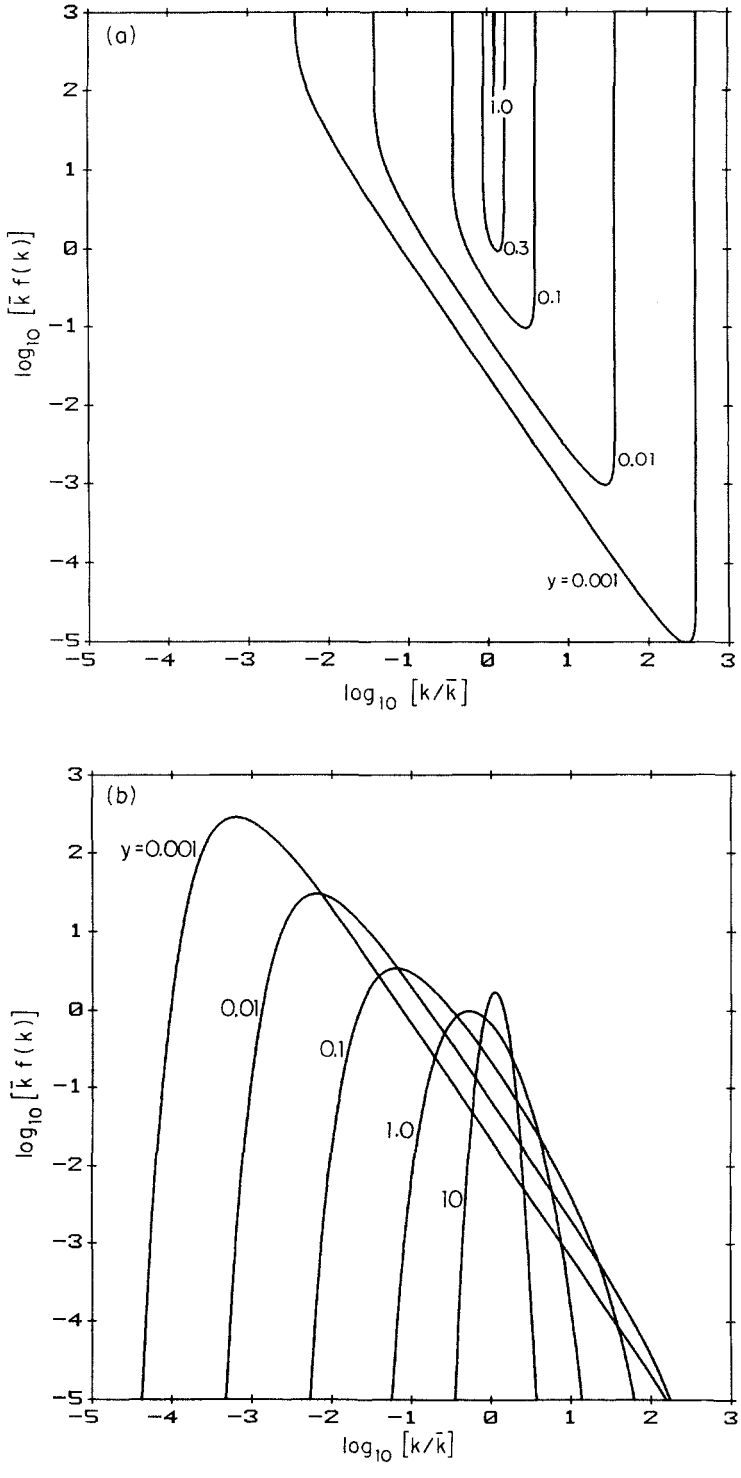


FIG. 4.20. *k* distributions. (a) Elsasser model. (b) Malkmus model.

downward. There are similarities, however. For $y \gg 1$ (overlapping lines), both distributions tend to δ-functions, corresponding to a featureless continuum. The manner of doing so differs greatly. For $y = 1$ the Elsasser model is essentially a continuum while the Malkmus model is not for $y = 10$. This difference of behavior was discussed in § 4.6.2.

For well-separated lines ($y \ll 1$), both distributions show a region in which

$$f(k) \propto k^{-3/2}.$$

This distribution corresponds to the wings of Lorentz lines, for which

$$k \propto \nu^{-2}.$$

For both models, this region is bounded at both high and low absorption coefficients. The behavior of the Elsasser model is obvious. For the Malkmus model, the high-k cutoff is caused by the exponential term in the distribution of line intensities (4.27). The low-k cutoff is a consequence of the rarity of large gaps in the spectrum allowed by the Poisson distribution (4.103) for a random band. The larger a gap the lower can be the saddle in the absorption coefficient in between two flanking lines, and the further can the k distribution extend to low values of k. As a result, the Malkmus distribution gives some lower k values than does the Elsasser model. But the Poisson distribution ensures that as k tends to zero low saddles between lines will be so rare that $f(k)$ will tend to zero.

The data shown in Fig. 4.21 for the 15 μm band of carbon dioxide fit the Malkmus model much better than the Elsasser model. However, the Malkmus model appears to be deficient at low-k, suggesting that the number of wide gaps in the observed spectrum exceeds that allowed by the Poisson distribution. It is also possible that the Malkmus distribution of line intensities is deficient in weak lines. There is some evidence that this is so for the 15 μm band, but it probably cannot account for the sharp divergence at low-k that is shown in Fig. 4.21. Since absorption coefficients less than $10^{-2}\,\bar{k}$ appear to have little effect on atmospheric calculations, the Malkmus model is usually regarded as a very good fit for the 15 μm band.

4.8.3. *Overlapping bands*

For thermal source functions, the transmission of overlapping bands can be dealt with, as for band models, by calculating separate band transmissions and subsequently multiplying them together. Attempts have been made to treat overlapping bands in k space; if this could be done, the results would apply to any source function and to some

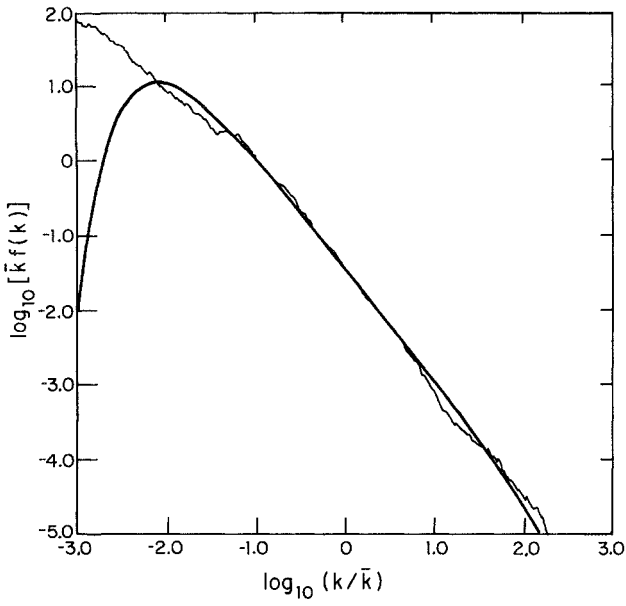


FIG. 4.21. Comparison between k distributions for the 15 μm CO_2 band and the Malkmus model. $f(k)$ (thin line) is computed for the frequency range 550 to 800 cm^{-1}, embracing most of the band. The pressure is 0.1 bar and the temperature 296 K. The Malkmus model (cf. Fig. 4.20b, $y = 0.01$) was fitted to line-by-line calculations of the transmission and the curve shown is the inverse Laplace transform of that transmission function. After Lacis and Oinas (1986).

complete bands. There is, however, no evidence that this is possible. For any single gas mixture, a combined k distribution can be calculated, but it cannot be related to the k distributions of the individual gases and must, as matters now stand, be recalculated for each different mixture.

4.9. Models of complete bands

4.9.1. *Band absorption areas*

If the source function varies only slightly over a vibration–rotation band contour, it is economical to treat the band as a whole, using an average value of the source function. The relevant absorption is then the band absorption area,

$$\mathcal{A}_{\text{band}} = \int_{\text{band}} A_\nu \, d\nu. \tag{4.130}$$

In § 4.8 we mentioned the application of the k distribution to an entire band; here we examine a large body of work from the point of view of band-model theory, from the engineering and climate literatures.

Three approaches are available, based, respectively, on approximate theory, a combination of theory and laboratory measurements, and on empirical measurements alone. All three have been used in numerical climate models with somewhat similar results. The choice among them is a matter for the preference of the individual investigator.

Absorption band areas are sometimes represented as a mean absorption,

$$\bar{A}_{\text{band}} = \frac{\mathscr{A}_{\text{band}}}{\mathscr{A}_0}, \tag{4.131}$$

where $\mathscr{A}_0$ is a measure of the width of the entire band. Since the wings of bands can spread indefinitely as the amount of absorbing matter increases, there is no rational choice for $\mathscr{A}_0$ that allows $\bar{A}_{\text{band}}$ to have unity as an upper limit. In practice, $\mathscr{A}_0$ is simply an additional and arbitrary parameter. $\bar{A}_{\text{band}}$, as defined by (4.131), is called a *wide-band model* (see § 4.1).

4.9.2. *Empirical models*

A large body of laboratory data for atmospheric gases has been fitted to the empirical expressions

$$\mathscr{A} = cm^d p^k \qquad \mathscr{A} \leqslant \mathscr{A}_0, \tag{4.132}$$

$$\mathscr{A} = C + D \ln(m) + K \ln(p) \qquad \mathscr{A} > \mathscr{A}_0. \tag{4.133}$$

$\mathscr{A}_0$, c, d, k, C, D, and K are all empirical, fitting parameters. Equations (4.132) and (4.133) are awkward to use numerically because of the discontinuity at $\mathscr{A} = \mathscr{A}_0$, and the following theoretical and semiempirical models are designed to avoid this difficulty.

It is evident that (4.132) and (4.133) cannot correctly represent all aspects of band absorption even though they may provide a good fit for a limited range of parameters employed in the laboratory. If lines are very wide, so that they overlap strongly, the absorption will cease to depend upon the pressure and Beer's law is obeyed for limited regions in the band. Both k and K are zero for this condition and yet there are other circumstances under which the pressure will be an important parameter and the data cannot be represented with k and K equal to zero.

4.9.3. *Exponential band contour*

An important feature of (4.133) is the logarithmic dependence of band area on the amount of absorbing material. This is a very common approximate behavior for band areas when absorption is intense; it is a

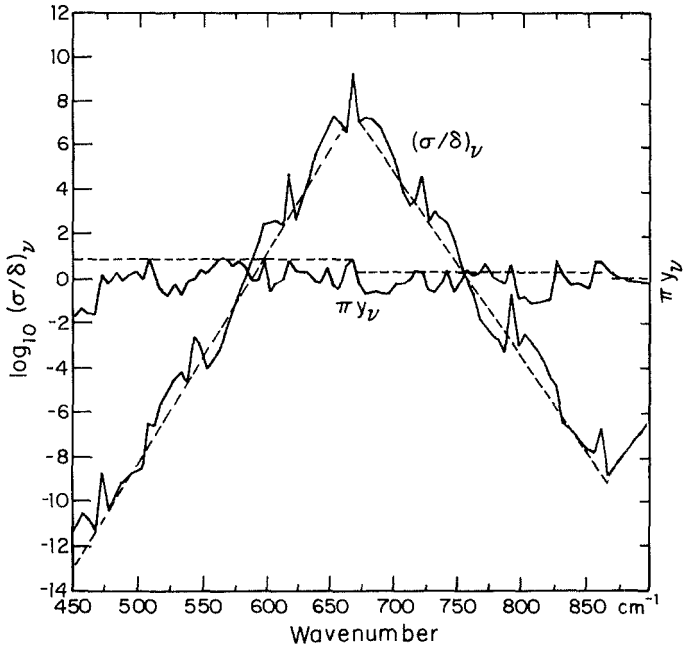


FIG. 4.22. Narrow-band parameters for the 15 μm band of CO_2. Band parameters have been calculated from the AFGL tape using the relationships (4.43), (4.44), and (4.45). Line widths are calculated at 1 bar and the line intensities for a temperature of 250 K. The broken lines represent approximations used in the exponential contour model. The slight rise in $(\sigma/\delta)_\nu$ between 860 and 900 cm^{-1} is caused by an upper-state band at 10.4 μm that is not included in the discussion. After Crisp et al. (1986).

consequence of the shape of the P-, Q-, and R-branch contours of typical vibration–rotation bands.

In Chapter 6, we shall apply band-model theory to spectral regions that are narrow compared to a band and yet contain a large number of lines. Figure 4.22 shows band parameters for the 15 μm band of carbon dioxide, estimated from (4.43), (4.44), and (4.45), and applied to overlapping spectral intervals, 5 cm^{-1} wide. These data suggest that each half of the band may be taken to have a constant value of y and an exponential dependence of (σ/δ) on frequency displacement from the band center (subscript 0),

$$\left(\frac{\sigma}{\delta}\right)_\nu = \left(\frac{\sigma}{\delta}\right)_0 \exp\frac{-|\nu - \nu_0|}{\mathscr{A}_0}. \tag{4.134}$$

The subscript ν is used here to represent the slow change of the 5 cm^{-1} averages over the band contour, and not monochromatic conditions.

If $\bar{A}[y, (\sigma/\delta)_\nu]$ is the mean absorption for the Malkmus model (for example), we may write for each half of the band

$$\mathscr{A}_{1/2\,\text{band}} = \int_{1/2\,\text{band}} \bar{A}[y, (\sigma/\delta)_\nu]\, d\nu. \tag{4.135}$$

Since $(\sigma/\delta)_\nu$ is a monotonic function of ν, (4.135) can also be written in the form

$$\begin{aligned}\mathscr{A}_{1/2\,\text{band}} &= \int_0^{(\sigma/\delta)_0} \bar{A}(y, \sigma/\delta)\,|\,\partial(\sigma/\delta)_\nu/\partial\nu|^{-1} d(\sigma/\delta) \\ &= \mathscr{A}_0 \int_0^{(\sigma/\delta)_0} A(y, \sigma/\delta) d\ln(\sigma/\delta).\end{aligned} \tag{4.136}$$

With the mean absorption for a Malkmus model, (4.38) and (4.112), the integration can be performed, with the result

$$\begin{aligned}\mathscr{A}_{1/2\,\text{band}} = \mathscr{A}_0 \Bigg\{ &\exp(\pi y)\left\{E_1\left\{\frac{\pi y}{2}[1 + (1 + 8u_0)^{1/2}]\right\} - E_1(\pi y)\right\} \\ &+ E_1\left\{\frac{\pi y}{2}[(1 + 8u_0)^{1/2} - 1]\right\} \\ &+ \gamma + \ln(2\pi y u_0)\Bigg\},\end{aligned} \tag{4.137}$$

where $u_0 = m(\sigma/\delta)_0/2\pi y$, E_1 is the first exponential integral, and γ is Euler's constant.

It is difficult to follow parametric dependence for an expression as complicated as (4.137). The strong-line behavior ($u_0 \gg 1$) is shown in Table 4.3.

Both dimensional and dimensionless forms are given in Table 4.3. The empirical expression, (4.133), is accounted for in a general way in rows (a) and (b), while (4.132) corresponds to row (d). Logarithmic dependence upon the amount of absorber is shown by rows (a) and (b), but the table also shows that there can be cases in which the absorber amount is large, without such dependence. This is at variance with statements in the engineering literature, based upon numerical studies, to the effect that logarithmic dependence is the invariable limit for large amounts of absorber.

We can obtain strong-line and weak-line limits to the exponential contour model by writing $A(y, \sigma/\delta)$ in (4.136) in the appropriate limit, before performing the integral. Since the integral extends to $\sigma/\delta = 0$, this procedure is consistent for the weak-line limit but not for strong lines. In the latter case there is a remainder term.

Table 4.3. Asymptotic forms for the exponential contour model when $u_0 \gg 1$

	$\pi y(2u_0)^{1/2}$ or $[\pi(\sigma/\delta)_0(\alpha_{L,0}/\delta)(p/p_0)m]^{1/2}$	y or $(\alpha_{L,0}/\delta)(p/p_0)$	Asymptotic form
(a)	$\gg 1$	$\gg 1$	$\gamma + \ln(2\pi y u_0)$ $= \ln(m) + \text{const.}$
(b)	$\gg 1$	$\ll 1$	$2\gamma + \ln(2\pi^2 y^2 u_0)$ $= \ln(m) + \ln(p) + \text{const.}$
(c)	$\ll 1$	$\ll 1$	Impossible
(d)	$\ll 1$	$\ll 1$	$2\pi y(2u_0)^{1/2}$ $= \text{const}\,(pm)^{1/2}$

The integrals can be performed in both limits with the help of the expression

$$\begin{aligned}\phi(x) &= \int_0^1 \zeta^{-1}[1 - \exp(-x\zeta)]\, d\zeta \\ &= \gamma + \ln(x) + E_1(x),\end{aligned} \tag{4.138}$$

with the results

$$\mathscr{A}_{1/2\,\text{band}} = \mathscr{A}_0 \phi(2\pi y u_0) \qquad u_0 \ll 1, \tag{4.139}$$

$$\mathscr{A}_{1/2\,\text{band}} = 2\mathscr{A}_0 \phi[\pi y(2u_0)^{1/2}] \qquad u_0 \gg 1. \tag{4.140}$$

The following expansions are valid (Appendix 6):

$$\phi(x) \to x \qquad \text{as } x \to 0,$$

$$\phi(x) \to \gamma + \ln(x) \qquad \text{as } x \to \infty.$$

The weak-line expression, (4.139), contains both a linear law and a logarithmic law, as the argument of ϕ goes from small values to large. The strong-line expression, (4.140), exhibits the two limits (b) and (d) in Table 4.3.

4.9.4. *Semiempirical treatment*

It has been proposed, based upon an examination of the results of numerical computations, that the band area can be expressed approximately as a function of the sum of the equivalent widths of the lines

forming the band,

$$\mathscr{A}_{\text{band}} = F\left(\sum_i W_i\right). \tag{4.141}$$

This proposal draws some support from the general form for random models, (4.112), and from the narrow-line limit to the Elsasser model, (4.83).

In the limit of weak lines, the independent-line limit must be achieved, so that for weak absorption

$$\mathscr{A}_{\text{band}} = \sum_i W_i. \tag{4.142}$$

If, in addition, we require a logarithmic law for strong absorption, the following form suggests itself:

$$\mathscr{A}_{\text{band}} = C \ln\left(1 + \sum_i W_i/C\right), \tag{4.143}$$

where C is an adjustable constant.

We cannot expect any equation as simple as (4.143) to satisfy all asymptotic forms of an expression such as that for the exponential band contour, (4.137). We are free to choose C and the distribution of line parameters by fitting to observed and theoretical data. Very good results have been achieved by doing so, but the result can be accepted only for the parametric range of the observations.

Figure 4.23 shows a comparison of (4.143), using the exponential distribution of line intensities, (4.36), with a line-by-line calculation; in this case, only C was adjusted to give a good fit. The agreement is excellent; over this range of parameters no purpose is served in using the full line-by-line calculation.

As we have already anticipated, the expression used in Fig. 4.23 cannot satisfy all asymptotic limits. In order to satisfy both (4.139) and (4.140) in their logarithmic limits, we must identify C with $\mathscr{A}_0$ in (4.139) and with $2\mathscr{A}_0$ in (4.140). This discrepancy arises only if the parameter $N\alpha_L/C$ (in effect, this is the familiar parameter, y) can be both large and small. For the data shown in Fig. 4.24, and for all pressures less than 1 bar for the 15 μm band of carbon dioxide, this parameter is small and no problem arises.

Proposals have been made for modifying $\sum_i W_i$ in such a way that the limits for both large and small $N\alpha_L/C$ are consistent. Such modifications are ad hoc and can be justified only by fitting to an extended range of empirical data. All expressions as simple as (4.143) will break down for certain values of the parameters.

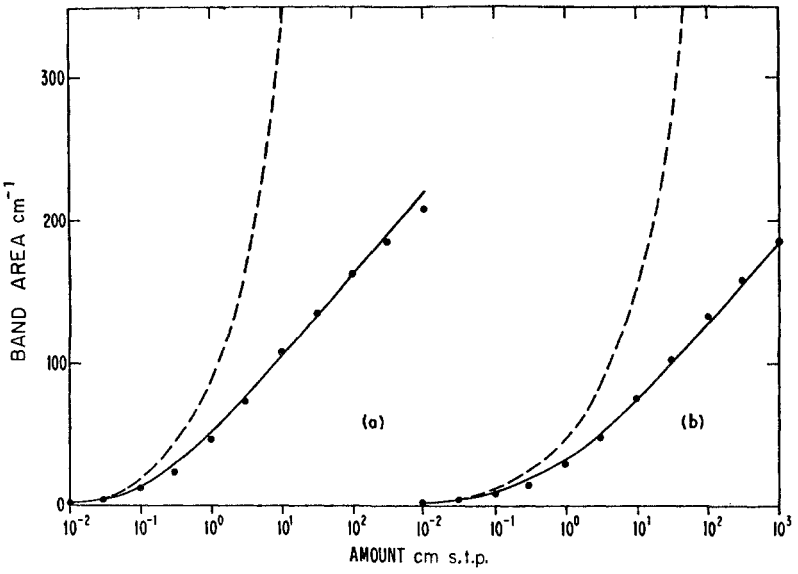


FIG. 4.23. Comparison of a semiempirical model with line-by-line calculations for the 15 μm carbon dioxide band. The broken lines represent the independent-line expression (4.142) and the solid lines are given by the approximation (4.143) with C chosen to be 55 cm^{-1}. The points are line-by-line calculations made by Yamamoto and Sasamori (1957). (a) $\alpha_L = 0.064\ cm^{-1}$ (1 bar), (b) $\alpha_L = 0.0128\ cm^{-1}$ (0.2 bar). After Goody and Belton (1967).

BIBLIOGRAPHY

4.1. Introduction

Experimental evidence in favor of the multiplication property of band transmission (Fig. 4.1) is given by

Burch, D. E., Howard, J. N., and Williams, D., 1956, "Infrared transmissions of synthetic atmospheres. V. Absorption laws for overlapping bands," *J. Opt. Soc. Am.* **46,** 452.

Its correctness is also evidenced by the proper functioning of all commercial double-beam spectrometers. As a formal problem it appears first to have been mentioned by Kaplan (1954), § 3.3.

4.2. Isolated lines

Definitions of the Bessel functions I_0 and I_1 are to be found in

Whittaker, E. T., and Watson, G. N., 1915, *A course of modern analysis.* London: Cambridge University Press.

Abramowitz, M., and Stegun, T. A., 1964, *Handbook of mathematical functions, with formulas, graphs and mathematical tables.* Washington D.C.: U.S. Government Printing Office.

The expression (4.10) was first derived by

Ladenburg, R., and Reiche, F., 1913, "Über selektive Absorption," *Ann. Phys.* **42,** 181.

Numerical values for $L(u)$ and references for both the Lorentz and Voigt profiles are given in Appendix 7.

The data shown in Fig. 4.4 are from

Shaw, J. H., and France, W. L., 1956, "Intensities and widths of single lines of the 4.7 micron CO fundamental band," Unpublished report AFCRC-TN-56-466, The Ohio State University Research Foundation.

The expansions (4.19) and (4.20) were first given by

Struve, O., and Elvey, C. T., 1934, "The intensities of stellar absorption lines," *Astrophys. J.* **79,** 409.

but (4.20) has been corrected to agree with (4.55).

One of the earliest treatments of the equivalent width of a Voigt line (from which Fig. 4.5 was obtained) was

van der Held, E. F. M., 1931, "Intensität und natürliche Breite von Spektrallinien," *Z. Phys.* **70,** 508.

For further references to more recent work, see Appendix 7. The modified Voigt profile, (4.22), was proposed by

Fels, S. B., 1979, "Simple strategies for the inclusion of Voigt effects in infrared cooling rate calculations," *Appl. Optics* **18,** 2634.

The expression (4.23) is given in

Crisp, D., Fels, S. B., and Schwarzkopf, M. D., 1986, "Approximate methods for finding CO_2 band transmission functions in planetary atmospheres," *J. Geophys. Res.* **91,** 11,851.

Fig. 4.6 is from

Rodgers, C. D., and Williams, A. P., 1974, "Integrated absorption of a spectral line with the Voigt profile," *J. Quant. Spectrosc. Radiat. Transfer* **14,** 319.

4.3. Distributed line intensities

The distribution functions (4.25), (4.26), and (4.27) were proposed by

Goody, R. M., 1952, "A statistical model for water-vapour absorption," *Quart. J. R. Meteorol. Soc.* **78,** 165.

Godson, W. L., 1954, "Spectral models and properties of transmission functions," *Proc. Toronto Meteorol. Conf. 1953.* London: Royal Meteorological Society, p. 35.

———, 1955a,b "The computation of infrared transmission by atmospheric water vapour. I and II," *J. Meteorol.* **12,** 272 and 533.

Malkmus, W., 1967, "Random Lorentz model with exponential-tailed S^{-1} line intensity distribution function," *J. Opt. Soc. Am.* **57,** 323.

The Doppler expansions in § 4.3.3 are all given by

Malkmus, W., 1968, "Random band models with lines of pure Doppler shape," *J. Opt. Soc. Am.* **58,** 1214.

Fig. 4.9 is from

Gille, J. C., and Ellingson, R. G., 1968, "Correction of random exponential band transmissions for Doppler effects," *Appl. Optics* **7,** 471.

4.4. The effect of overlap

Schnaidt's model was given by

Schnaidt, F., 1939, "Über die Absorption von Wasserdampf und Kohlensäure mit besonderer Berucksichtigung der Druck- und Temperatur-abhängigkeit," *Beitr. Geophys.* **54,** 203.

A modified version was given by

Plass, G. N., 1958, "Models for spectral band absorption," *J. Opt. Soc. Am.* **48,** 690.

The general model of § 4.4.2 was developed by

Matossi, F., Mayer, R., and Rauscher, E., 1949, "On total absorption in spectra with overlapping lines," *Phys. Rev.* **76,** 760.

Matossi, F., and Rauscher, E., 1949, "Zur Drückabhängigkeit der Gesamtabsorption in ultraroten Bandspectren," *Z. Phys.* **125,** 418.

4.5. Regular models

The Elsasser model was the first physically satisfying model of the average absorption by a vibration–rotation band, although it was, at first, used only in the strong–line limit, see

Elsasser, W. M., 1943, *Heat transfer by infrared radiation in the atmosphere.* Harvard Meteorological Studies No. 6. Cambridge, Mass.: Harvard University Press.

A table of the Elsasser function is given in Appendix 8, together with further references.

Details of the Poisson sum rule can be found in

Carrier, G. F., Krook, M., and Pearson, C. E., 1966, *Functions of a complex variable: Theory and technique.* New York: McGraw-Hill, p. 310.

Plass (1958), see § 4.4, has discussed at length the three asymptotic limits for random and regular bands. The data in Fig. 4.14 are from

Plass, G. N., 1960, "Useful representations for measurements of spectral band absorption," *J. Opt. Soc. Am.* **50,** 868.

The nearly overlapping line result, (4.75), has been rediscovered by a number of authors; the first appears to have been

Malkmus, W., 1967, "Criteria for accuracy of measurements of line intensities and mean absorption coefficients," *Appl. Optics* **6,** 349.

The limit, (4.83), for $y \ll 1$, was given by

Goody, R. M., 1961, "A note on the regular model of an absorption band," *Quart. J. R. Meteorol. Soc.* **87**, 428.

The alternate limit, (4.84), is from

Craig, R. A., 1963, "A note on the transmissivity of an Elsasser band," *J. Atmos. Sci.* **20**, 66.

We are indebted to A. R. Curtis for the unpublished material in § 4.5.2.
The Voigt profile is discussed by

Golden, S. A., 1967, "The Doppler analog of the Elsasser band model," *J. Quant. Spectrosc. Radiat. Transfer* **7,** 483.

Kyle, T. G., 1967, "Absorption of radiation by uniformly spaced Doppler lines," *Astrophys. J.* **148,** 845.

Golden, S. A., 1969, "The Voigt analog of an Elsasser band," *J. Quant. Spectrosc. Radiat. Transfer* **9,** 1067.

Absorption by a Doppler–Elsasser band is tabulated by

Golden, S. A., 1968, "The Doppler analog of the Elsasser band model. II. The integrated emissivity," *J. Quant. Spectrosc. Radiat. Transfer* **8,** 877.

4.6. Random models

The principle of the random model was first suggested by Goody (1952) (§ 4.3). This particular treatment was concerned with the limit of a large number of lines. Several authors have demonstrated how the random model can result from the superposition of many subbands. The superposition of Elsasser bands is discussed by Kaplan (1954), (§ 4.1), and Plass (1958), (§ 4.4), while the superposition of random bands was discussed by Godson (1955), (§ 4.3).

The Poisson distribution of line spacings was first discussed by

Yamamoto, G., and Sasamori, T., 1957, "Numerical study of water vapour transmission," *Sci. Rept. Tohoku Univ., Ser. 5, Geophys.* **8**(2), 146.

The first line-by-line calculations for water vapor were made by

Cowling, T. G., 1950, "Atmospheric absorption of heat radiation by water vapour," *Phil. Mag.* **41,** 109.

The calculations for the 9.6 μm ozone band, shown in Fig. 4.17, are by

Goldman, A., and Kyle, T. G., 1968, "A comparison between statistical model and line-by-line calculation with application to the 9.6-μ ozone and the 2.7-μ water vapor bands," *Appl. Optics* **7,** 1167.

The experimental data in Fig. 4.18 are from

Howard, J. N., Burch, D. E., and Williams, D., 1956, "Infrared transmission of synthetic atmospheres. IV. Application of theoretical band models," *J. Opt. Soc. Am.* **46,** 334.

4.7. Generalized transmission functions

Figure 4.19 is from

Walshaw, C. D., 1954, *An empirical investigation of the* 9.6 μ *band of ozone,* Ph.D. Thesis, Cambridge University.

4.8. *k* distributions

k distributions have a long history in the astrophysical and meteorological literatures. The first generally recognized use was by

Ambarzumian, V. A., 1936, "The effect of absorption lines on the radiative equilibrium of the outer layers of the stars," *Pub. Observ. Astronom. Univ. Leningrad* **6,** 7.

Soviet authors have frequently used a finite sum of exponential terms to describe the absorption of water vapor. An account of some of this work is to be found in

Kondratiev, K. Ya., 1965, *Radiative heat exchange in the atmosphere.* Oxford: Pergamon Press, p. 147.

The first attempt to study systematically the relationship between band models and *k* distributions was by

Arking, A., and Grossman, K., 1972, "The influence of line shape and band structure on temperatures in planetary atmospheres," *J. Atmos. Sci.* **29,** 937.

Some aspects of the fundamental theory, and the Laplace transform for the Malkmus model were given by

Domoto, G. A., 1974, "Frequency integration for radiative transfer problems involving homogeneous non-gray gases: The inverse transmission function," *J. Quant. Spectrosc. Radiat. Transfer* **14,** 935.

Figure 4.21 is from

Lacis, A. A., and Oinas, V., 1986, "Application of correlated-*k* methodology to gaseous absorption, thermal emission and multiple scattering in vertically inhomogeneous atmospheres," *J. Atmos. Sci.*, submitted.

4.9. Models of complete bands

The following is a good statement of the engineering viewpoint:

Cess, R. D., and Tiwari, S. N., 1972, "Infrared radiative energy transfer in gases," *Adv. Heat Transfer* **8,** 229.

The following review has an exhaustive bibliography:

Tiwari, S. N., 1978, "Models for infrared atmospheric radiation," *Adv. Geophys.* **20,** 1.

The empirical formulas, (4.132) and (4.133), are from

Howard, J. N., Burch, D. E., and Williams, D., 1956, "Infrared transmission of synthetic atmospheres," *J. Opt. Soc. Am.* **46,** 186, 237, 334, 452.

A careful numerical study of the absorption by an entire band is

Edwards, D. K., and Maynard, W. A., 1964, "Comparison of models for correlation of total band absorption," *Appl. Optics* **3,** 621.

The analytical result, (4.137), is from

Wang, W.-C., 1983, "An analytical expression for the total band absorptance of infrared-radiating gases," *J. Quant. Spectrosc. Radiat. Transfer* **29,** 279.

Two reviews of the climate literature containing many references to narrow- and wide-band models are Crisp et al. (1986) (§ 4.2) and

Kiehl, J. T., and Ramanathan, V., 1983, "CO_2 radiative parameterization used in climate models: Comparison with narrow band models and with laboratory data," *J. Geophys. Res.* **88,** 5191.

The discussion leading to (4.141) is to be found in Yamamoto and Sasamori (1957), (§ 4.6). The comparison in Fig. 4.23 is by

Goody, R. M., and Belton, M. J. S., 1967, "Radiative relaxation times for Mars," *Planet. Space Sci.* **15,** 247.

Another, similar treatment, based upon semiempirical band models is

Aoki, T., 1980, "An accurate representation of the transmission of the H_2O and CO_2 infrared bands," *J. Quant. Spectrosc. Radiat. Transfer.* **24,** 191.

5

ABSORPTION BY ATMOSPHERIC GASES

5.1. Introduction

Digital archives offer the investigator an up-to-date analysis of an extensive and specialized literature. Periodic revisions are reported in the open literature and it seems unlikely that future investigators will attempt to use any other source where archives can provide the required data.

For this reason, we shall confine our comments on permitted vibration–rotation transitions to describing the AFGL tape contents, but we shall add two areas not contained in it: first, electronic bands, and second, the related topics of forbidden transitions, collision-induced transitions, and polymer spectra.

The AFGL tape lists data on one important set of electronic transitions, those giving rise to the near-infrared *atmospheric bands* of molecular oxygen. These bands behave in the same way as vibration–rotation bands, except for the frequency displacement caused by the change in electronic energy and the symmetry conditions imposed by the electronic wave functions. Other electronic transitions usually involve larger differences between energy levels and cannot be understood as completely as the lower energy, vibrational and rotational transitions. Fortunately, visible and ultraviolet bands of importance for atmospheric problems are less complicated than vibration–rotation bands and they are usually less affected by state parameters. Atmospheric absorption calculations in the visible and ultraviolet spectrum are commonly made on the basis of empirical data without requiring the level of understanding developed in Chapters 3 and 4 for vibration–rotation bands.

The altitude of unit optical depth for ultraviolet atmospheric bands is illustrated in Fig. 5.1. The intensity of solar radiation falls off rapidly with decreasing wavelength in the spectral region shown (the irradiance at 2000 Å compared to that at 3000 Å is 10^{-2} whereas at 1000 Å it is 10^{-5}, see Appendix 9). For heating rate calculations at altitudes less than 100 km, only O_2 and O_3 are important, except under special conditions when the atmosphere contains large amounts of volcanic aerosols, or polar stratospheric clouds at high latitudes. All of the absorptions shown in Fig. 5.1 are important for other reasons that do not directly concern us here.

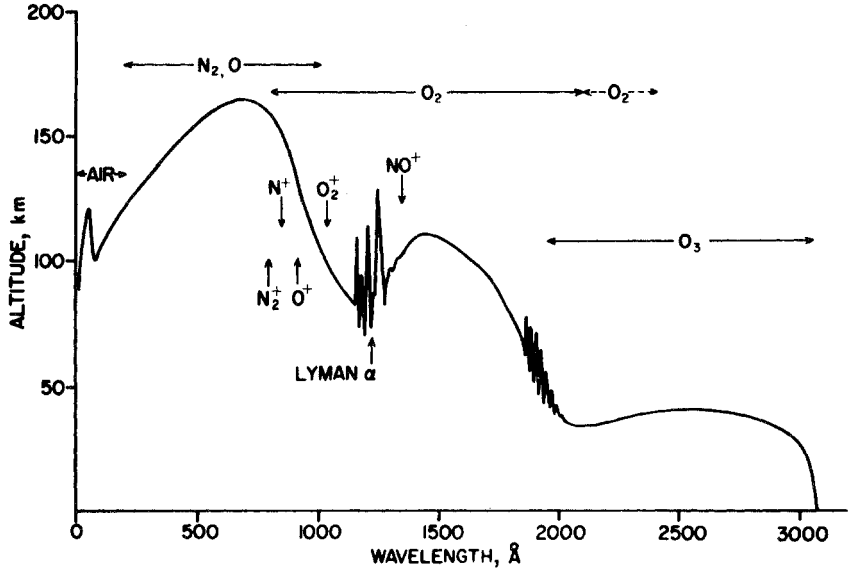


FIG. 5.1. Depth of penetration of solar radiation in the ultraviolet spectrum as a function of wavelength. The line shows the altitude of unit optical depth. The vertical arrows indicate ionization limits. The broken line represents predissociation for molecular oxygen. After Herzberg (1965).

Figure 5.1 does not show the near-ultraviolet Huggins bands or the visible Chappuis bands of ozone. Both of these electronic bands are of some importance, although they are weak, with optical depths at ground level of the order of unity or less.

The second area to which we shall give some attention is that of forbidden transitions, collision-induced transitions, and polymers. Forbidden transitions, in the sense of magnetic dipole and electric quadrupole transitions, are as well understood as electric dipole transitions, except that they are much weaker. Where relevant, we shall mention such transitions. More important are transitions that are allowed during collisions, but not otherwise. This is the topic of far wing line shapes and pressure-induced transitions; the two subjects are related and it is convenient to consider the more speculative question of polymers at the same time.

We may briefly repeat the discussion in the overview to Chapter 3 (§ 3.5) because it is important. Results presented in that chapter (see especially Figs. 3.24, 3.25, and 3.26) showed remarkable agreement between observations and the AFGL data, under appropriate circumstances. As long as we are concerned with line center information for isolated molecules, the theory of molecular spectra is impressively accurate. However, if we are concerned with long paths and elevated

pressures the underlying continuum can become important. This continuum is caused by far wings of lines, pressure-induced transitions, and, possibly, polymers. The AFGL tapes have no information on these matters.

The current AFGL listing is in two parts. The *main listing* is titled "AFGL atmospheric absorption line parameters" and includes data for seven major atmospheric absorbers: oxygen, water vapor, carbon dioxide, ozone, nitrous oxide, carbon monoxide, and methane. The 1982 revision includes 181,000 vibration–rotation lines. The principal parameters tabulated for each line are frequency (cm^{-1} at zero pressure), molecular line intensity (cm), Lorentz width (cm^{-1} at 1 bar), and energy of the lower state (cm^{-1}). In addition, the line is identified as to molecule, isotope, and quantum numbers (including l value and Fermi resonance parameters). Missing are pressure-shift coefficients and any information on the line shape, except for the Lorentz core.

The second part of the AFGL listing ("AFGL trace gas compilation") concerns the pollution problem. Data are listed (1982 edition) on 21 gases and their isotopes, from the millimeter region of the spectrum to 1 μm. The gases are (numbers of lines in parentheses): NO (7385), SO_2 (18,169), NO_2 (9468), NH_3 (5556), HNO_3 (12,777), OH (8490), HF (62), HCl (199), HBr (256), HI (145), ClO (6020), OCS (737), H_2CO (2701), HOCl (7723), N_2 (117), HCN (772), CH_3Cl (6687), H_2O_2 (2389), C_2H_2 (306), C_2H_6 (4328), and PH_3 (2886).

All of these gases can have minor effects on the atmospheric heat balance and are commonly included in climate calculations, but we shall discuss only the data on nitrogen.

5.2. Nitrogen

The symmetry of the $^{14}N^{14}N$ molecule forbids vibrational–rotational electric dipole transitions. Because of the large amount of nitrogen in the atmosphere, however, both forbidden and pressure-induced transitions must be taken into account, although they are only strong enough to rate an AFGL listing as a "trace gas"!

The fundamental vibration frequency of $^{14}N^{14}N$ lies at 2329.9 cm^{-1}; the bond length is 109.76 pm and the equilibrium rotational constant is 2.01 cm^{-1}.

Quadrupole transitions in the fundamental vibration band have been detected in atmospheric spectra and are included in the AFGL listing with a molecular band intensity of 6.4×10^{-27} cm (see Appendix 2 for units). Even allowing for the large amounts of nitrogen in the atmosphere, this is a very small band intensity. The selection rule for a quadrupole transition in a homonuclear molecule is

$$\Delta J = 0, \pm 2, \qquad J = 0 \text{ to } J = 0 \text{ excepted.} \tag{5.1}$$

This gives rise to three branches, O, Q, and S; the O- and S-branches have twice the usual frequency spacing of P- and R-branches.

More important are the collision-induced bands of nitrogen. The fundamental band overlaps the high-frequency wing of the ν_3 band of carbon dioxide and affects the use of this band for the remote sounding of atmospheric temperature (Chapter 6). In the stratosphere, where water vapor concentrations are very low, the collision-induced rotation band of nitrogen may be of some importance for thermal calculations (at lower levels the water vapor absorption is overwhelming).

Figure 5.2 shows the rotation band for nitrogen–nitrogen collisions. The rotation lines are not resolved because they are wider than the line spacing. The absorption coefficient generally reflects the rotation line intensities at the frequency concerned and the variation with temperature follows the expected behavior of the underlying rotation lines. The lines may be as wide as $50\,\mathrm{cm}^{-1}$, however, and, for $\nu > 200\,\mathrm{cm}^{-1}$, the contribution from the wings of distant lines may exceed the contribution from local lines. The rotational band intensity at 300 K is

$$S_n^{N_2N_2} = \int k_n^{N_2N_2}\, d\nu = 6.5 \times 10^{-43}\,\mathrm{cm}^4.$$

The atmospheric absorption at the maximum of the rotation band has been estimated for a stratospheric path from 12 km to space for a

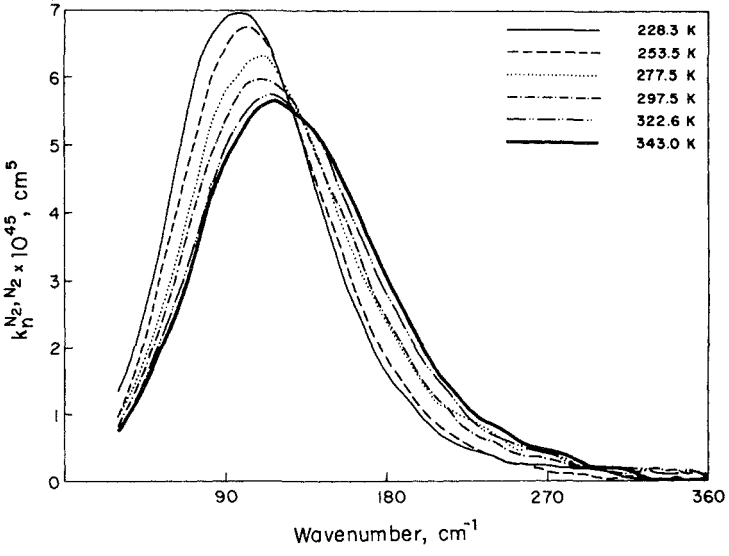


FIG. 5.2. The collision-induced rotation band of pure nitrogen (nitrogen–nitrogen collisions). The vertical axis is the binary absorption coefficient, as defined in (3.88). After Stone et al. (1984).

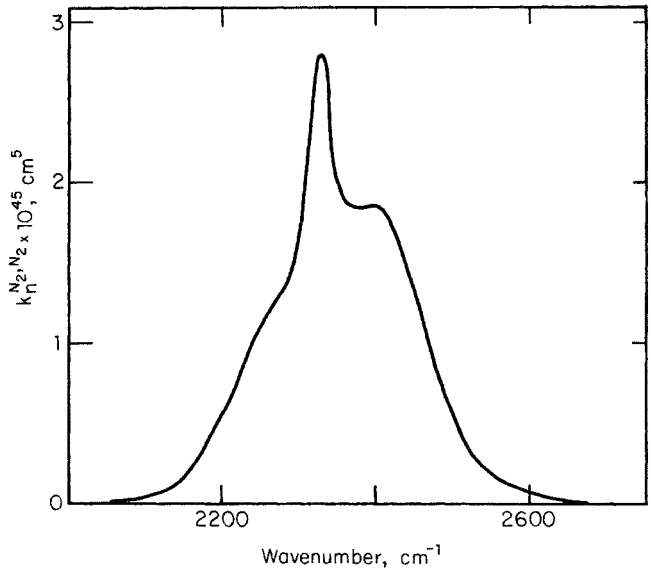


FIG. 5.3. The collision-induced fundamental band of pure nitrogen (nitrogen–nitrogen collisions). The temperature is 290 K. After Shapiro and Gush (1966).

temperature of 220 K. The calculation gave 8% absorption for a zenith angle of 25° and 38% absorption for a zenith angle of 80°, comparable to absorption by the water vapor rotation band.

The pressure-induced fundamental band and first overtone band of nitrogen have been measured in the laboratory; data for the fundamental band are shown in Fig. 5.3. The first overtone band is of comparatively small importance. No rotational structure appears for the same reasons as for the rotation band. The band intensities are

$$S_n^{N_2N_2} = 4.8 \times 10^{-43} \text{ cm}^4 \text{ for the fundamental}$$
$$= 2.6 \times 10^{-45} \text{ cm}^4 \text{ for the overtone band.}$$

For oxygen–nitrogen collisions, the binary coefficient is about 0.85 times that for nitrogen–nitrogen collisions.

5.3. Oxygen

Nitrogen and oxygen are both stable, homonuclear molecules and there are many similarities between their spectra; nevertheless, the differences that do exist make the spectrum of oxygen much more important for atmospheric calculations than that of nitrogen.

5.3.1. *Ultraviolet, molecular absorptions*

The ultraviolet absorption spectrum of $^{16}O^{16}O$ commences with the weak Herzberg band at 2600 Å. Below 2420 Å the transition becomes dissociative, with end-products $^{16}O(^3P) + {}^{16}O(^3P)$, and the weak Herzberg continuum sets in. The molecular absorption coefficient is very small, between 10^{-23} and 10^{-24} cm^2 at the threshold, and of little importance for energy absorption. It is, however, important for the formation of atmospheric ozone.

The Schumann–Runge bands occupy the spectral region 1950 to 1750 Å (Fig. 5.4). At 1750 Å the bands merge into a stronger dissociation continuum with the end-products $^{16}O(^3P) + {}^{16}O(^1D)$, which extends to 1300 Å and is the most important single feature of the absorption spectrum of molecular oxygen (the Schumann–Runge continuum). The three features at 1295, 1332, and 1352 Å may indicate dissociation products with more energetic end-products.

The bands between 1060 and 1280 Å have not yet been identified. Particular attention has been paid to the absorption coefficient at the solar Lyman-α line (1215.7 Å), which happens to lie in a deep absorption minimum. The absorption coefficient at low pressures is 1.00×10^{-20} cm^2 with a self-broadening coefficient of 1.47×10^{-23} cm^2 mb^{-1}. The mechanism involved in this pressure effect is unclear but is unimportant for our purposes.

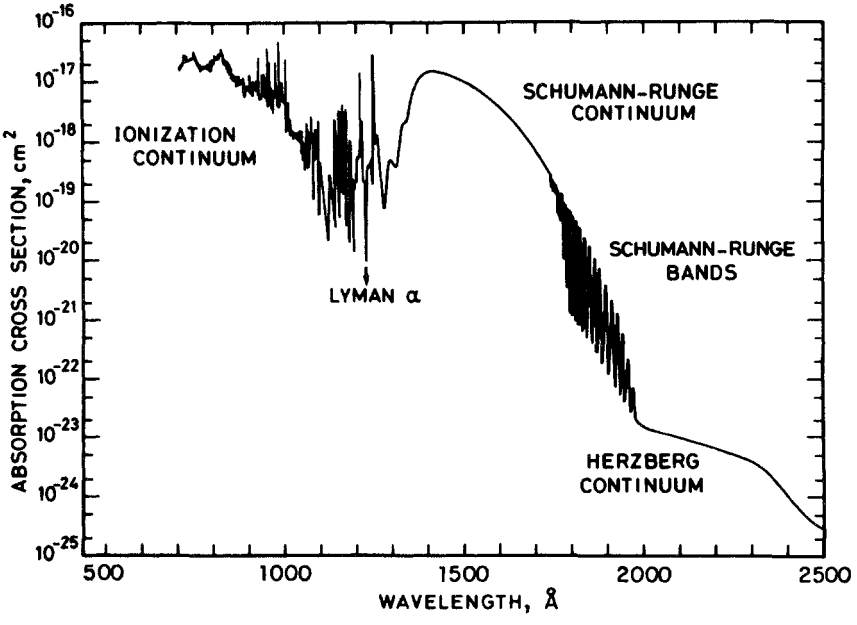


FIG. 5.4. Absorption cross section of $^{16}O^{16}O$ in the ultraviolet spectrum. After Brasseur and Solomon (1984).

Between 850 and 1100 Å are a series of distinct Rydberg bands, known as the Hopfield bands, with peak cross sections near 950 Å as great as 5×10^{-17} cm^2. Below 1026.5 Å (12.08 eV), the absorption is partly caused by bound-free ionizing transitions.

Below 850 Å, ionization dominates the absorption. Below 300 Å, the absorption is probably the same as that of two atoms of oxygen.

5.3.2. *Forbidden bands in the vibration–rotation spectrum*

Oxygen is a paramagnetic gas with an unusually large magnetic dipole moment. Forbidden magnetic dipole transitions in the rotation band are familiar to microwave spectroscopists who have made extensive studies of some of the low-J lines. The selection rules are the same as for electric dipole transitions. Intensities can be calculated from the known magnetic dipole moment. The band intensity for the rotation band is 7.23×10^{-24} cm and 161 lines are given in the 1982 AFGL listing. The equilibrium rotational constant in the ground electronic state is 1.4457 cm^{-1}, corresponding to an O–O bond length of 120.74 pm.

An electric quadrupole transition in the fundamental vibration band is barely observable in the atmospheric spectrum. The fundamental frequency is 1556.379 cm^{-1} and the band intensity is 6.15×10^{-27} cm; 146 lines are listed in the 1982 AFGL compilation.

5.3.3. *The "atmospheric" bands*

The ground electronic state of molecular oxygen is a triplet with a ground state designated X, and two excited states, a and b. The $X \rightarrow a$ and $X \rightarrow b$ transitions involve energy changes of 7882 and 13,120 cm^{-1}, respectively. These electronic transitions are accompanied by vibrational–rotational transitions and give rise to two band systems, the *infrared bands* and the *red bands*, respectively, some of which absorb strongly in the atmospheric spectrum at ground level (particularly the so-called A, B, and γ bands). Details of eight bands that appear in the AFGL listing are given in Table 5.1 and a synthetic spectrum of one is shown in Fig. 5.5.

5.3.4. *The collision-induced spectrum*

Six collision-induced or dimer bands have been reported in the literature. Three are in the visible spectrum and are responsible for the blue color of liquid oxygen. In the infrared spectrum, the fundamental, the first overtone, and the rotation band have been observed. Only the rotation band and one of the visible bands have been observed in atmospheric spectra; the remainder are laboratory identifications.

A question that has been debated without clear conclusions is whether we are dealing with dimers of finite lifetime or complexes that

Table 5.1. The strongest "atmospheric" bands of molecular oxygen

Band origin (cm^{-1})	Electronic transition	Vibrational transition	Band intensity (cm)	Comment
		Infrared bands		
6326.033	$a \leftarrow X$	$0 \leftarrow 1$	1.13×10^{-28}	
7882.425	$a \leftarrow X$	$0 \leftarrow 0$	1.82×10^{-24}	See Fig. 5.5
9365.877	$a \leftarrow X$	$1 \leftarrow 0$	8.63×10^{-27}	
		Red bands		
11,564.516	$b \leftarrow X$	$0 \leftarrow 1$	7.80×10^{-27}	
12,969.269	$b \leftarrow X$	$1 \leftarrow 1$	9.42×10^{-26}	A band
13,120.909	$b \leftarrow X$	$0 \leftarrow 0$	1.95×10^{-22}	A band
14,525.661	$b \leftarrow X$	$1 \leftarrow 0$	1.22×10^{-23}	B band
15,902.418	$b \leftarrow X$	$2 \leftarrow 0$	3.78×10^{-25}	γ band

exist only during collisions. From the empirical point of view, both will have absorption coefficients proportional to the ambient pressure. Dimers should exhibit new vibrational modes, but all observed oxygen bands can be accounted for in terms of modes of the monomer. Dimers could also exhibit rotational structure. Some weak fine structure has been reported in the visible bands. The question is unresolved, but unimportant for our purposes.

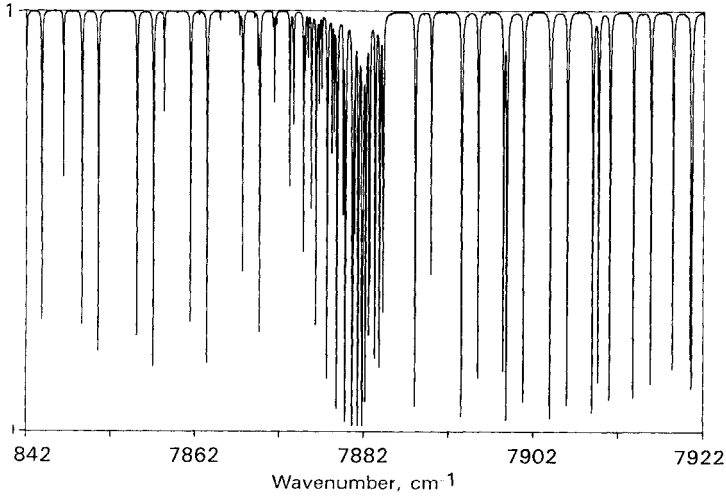


FIG. 5.5. The 7882 cm^{-1} infrared "atmospheric" band of molecular oxygen. This is a synthetic spectrum prepared from the AFGL tape. Level of observation, 0 km. Zenith angle of observation, 30°.

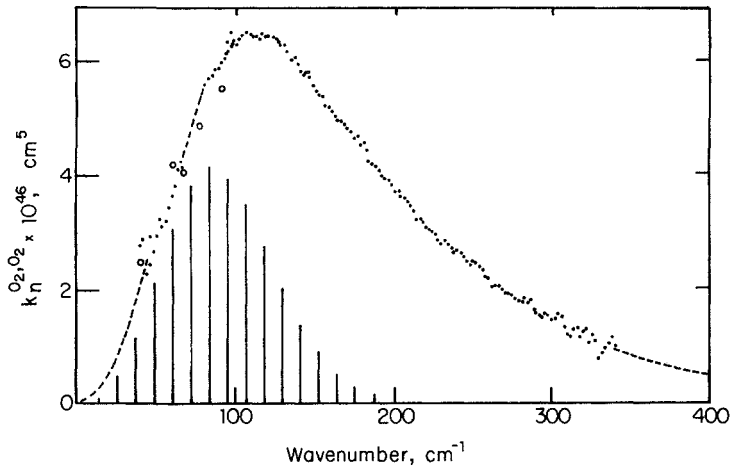


FIG. 5.6. The collision-induced rotation spectrum of oxygen. The temperature is 300 K. The vertical lines show the relative strengths of the underlying rotation lines. After Bosomsworth and Gush (1965).

Of the three visible bands, two are very weak and correspond to (2, 0) and (3, 0) vibrational transitions of the red "atmospheric" system. The third, at 21,000 cm^{-1}, shows an absorption of a few percent in the spectrum of the zenith sun. It has been tentatively identified as two transitions taking place simultaneously in the ground-state fine structures of the two oxygen atoms, in contrast to a normal molecular transition. The binary band intensity at 300 K is $S_n^{O_2O_2} = 2.3 \times 10^{-43}$ cm^4.

Careful laboratory measurements are available on the fundamental, the first overtone, and the rotation band of oxygen. Data on the fundamental were shown in Chapter 3 (Fig. 3.23). The band contour, which shows no fine structure, resembles that of the predicted intensities of the underlying rotation lines, showing that, while the line widths are large compared to the line spacing, they must be small compared to the band contour. Measurements with nitrogen as the broadening gas show almost identical binary absorption coefficients. The binary band intensities for the fundamental and the first overtone are 9.5×10^{-43} and 9.7×10^{-45} cm^4, respectively.

The collision-induced rotation band for pure oxygen is shown in Fig. 5.6. The correspondence between the band contour and the underlying rotational lines is less satisfactory than for the fundamental band (Fig. 3.23), suggesting that the lines are very broad indeed, broader than the band contour. The band intensity at 300 K is 1.20×10^{-43} cm^4.

5.3.5. *Atomic oxygen*

Oxygen atoms absorb in the far-ultraviolet spectrum but they also possess an interesting absorption (or emission) in the infrared caused by

transitions between fine-structure levels in the electronic ground state of the atom.

The ground state of atomic oxygen is a triplet, with components $^3P_0$, $^3P_1$, and $^3P_2$. Magnetic dipole transitions can take place between $^3P_0$ and $^3P_2$, and between $^3P_1$ and $^3P_2$; the Einstein coefficient and energy gap are $1.7 \times 10^{-5}\,s^{-1}$ and $226\,cm^{-1}$ for the former and $8.8 \times 10^{-5}\,s^{-1}$ and $161\,cm^{-1}$ for the latter. These are very small Einstein coefficients, but it is the small sizes of the energy gaps that make these transitions important. At a temperature of 300 K, the product $\mathbf{k}\theta$ is $208\,cm^{-1}$ and almost every collision can excite or deexcite the fine-structure levels, giving rise to a relaxation time approximately equal to the time between collisions (100 s at 400 km).

If we use these numbers in the discussion of § 2.2.2, we conclude that the fine-structure levels of atomic oxygen are in thermodynamic equilibrium to far higher levels than any other thermal radiator that has been discussed and that the 62 μm ($161\,cm^{-1}$) line of atomic oxygen will be the most important thermal radiator over a large part of the upper atmosphere.

5.4. Water vapor

5.4.1. *The vibration–rotation spectrum*

The water molecule is an asymmetric top with the oxygen atom in the middle; the bond length is 95.8 pm and the bond angle is 104.45°. The water molecule has a large electric dipole moment, 6.16×10^{-30} Coulomb-meter (C-m) in its equilibrium configuration, and has strong rotation bands. The three moments of inertia differ greatly from each other and they are all small, giving rise to a widespread and apparently disorderly array of rotation lines. These properties combine with the relatively large concentrations of water vapor in the lower atmosphere to account for the ubiquitous water vapor lines in every region of the solar spectrum and for the unique importance of this gas in atmospheric radiation problems.

Four isotopic forms have identifiable lines in the solar spectrum: $H^{16}OH$, $H^{18}OH$, $H^{17}OH$, and $H^{16}OD$ are, according to Table 1.2, present as 99.73, 0.2039, 0.0373, and 0.0298%, respectively. Each isotope has a different vapor pressure and the abundances depend slightly on the evaporation–condensation cycle.

The $H^{18}OH$ and the $H^{17}OH$ molecules have vibrational and rotational constants differing very little from those of $H^{16}OH$. Relative to the normal molecule, $H^{18}OH$ lines are shifted by -11 to $1\,cm^{-1}$, while shifts for $H^{17}OH$ are half as great. $H^{16}OD$ stands out alone because the vibrational frequencies differ markedly from those of the other isotopic forms.

Table 5.2. Observed frequencies of water vapor fundamentals

Band	Transition	Band center (cm^{-1})	
		$H^{16}OH$	$H^{16}OD$
ν_1	100←000	3657.05	2723.68
ν_2	010←000	1594.75	1403.49
ν_3	001←000	3755.93	3707.47

Table 5.3. Rotational constants for water vapor in the ground vibrational state (cm^{-1})

Axis	$H^{16}OH$	$H^{16}OD$
A	27.79	23.38
B	14.51	9.06
C	9.29	6.38

Water vapor has three fundamentals (see Fig. 3.9a for vibrations of the nonlinear XY_2 molecule), the ν_2 bending mode having the lowest frequency and ν_1 and ν_3 both having approximately twice this frequency. The close coincidence between ν_1, ν_3, and $2\nu_2$ leads to complex interactions between vibrational states. Table 5.2 shows the observed fundamental frequencies of $H^{16}OH$ and $H^{16}OD$. The HOH bands are type B if v_3 changes by an even number; otherwise they are type A. The lower symmetry of HOD allows hybrid bands of mixed type A and B. Rotational constants are given in Table 5.3.

5.4.2. *Listed data*

The AFGL data are based upon a Hamiltonian having 25 adjustable constants. The changes of the dipole moment with respect to the normal coordinates are obtained from observed band intensities of the fundamental bands. Estimates of the precision of vibrational–rotational line intensities vary from ±15% for most lines to a factor of two for high-J lines.

Lines widths are obtained from a few direct measurements combined with calculations based on a modified ATC theory (see Fig. 3.16). Difficulties with narrow, high-J lines were discussed in § 3.3.3. Even for this most carefully studied of all atmsopheric molecules, the line widths are not known with great accuracy.

The important bands of water vapor fall into a number of distinct

classes: the rotation band from 0 to 1000 cm^{-1}; the ν_2 (6.3 μm) bands from 900 to 2400 cm^{-1}; the 2.7 μm group, consisting of ν_1, ν_3, and $2\nu_2$, from 2800 to 4400 cm^{-1}; the near-infrared bands, with six distinguishable groups of lines between 4500 and 11,000 cm^{-1}; and the weak, visible bands from 11,000 to 18,000 cm^{-1}. The 6.3 μm and the rotation bands dominate the thermal region of the spectrum; the near-infrared bands, although much weaker, absorb a great deal of solar radiation in the lower atmosphere.

Table 5.4 shows the 2.7 μm, the 6.3 μm, and the rotation bands of water vapor. The 1982 AFGL compilation includes eight rotation bands, eleven 6.3 μm bands, and twenty-three 2.7 μm bands. In Table 5.4, an arbitrary cut-off for the band intensity between 10^{-20} and 10^{-21} cm has been imposed. No upper-state bands are included. These are generally rather weak, but that depends upon the temperature, to which they are very sensitive.

Table 5.5 shows some of the near-infrared and visible bands of water vapor. The AFGL list contains 42 near-infrared bands (more have been reported) and 27 visible bands. This list has been shortened by including only bands with S_n greater than 10^{-20} cm or the strongest band in each group. As for Table 5.4, no upper-state bands are included.

Table 5.4. The strongest infrared bands of water vapor[a]

Region	Band origin (cm^{-1})	Isotope	Upper state ($v_1 v_2 v_3$)	S_n cm $\times 10^{21}$ at 296 K	Number of lines listed
Rotation	0.00	$H^{16}OH$	000	52,700.0	1728
	0.00	$H^{17}OH$	000	19.4	622
	0.00	$H^{18}OH$	000	107.0	766
6.3 μm	1588.28	$H^{18}OH$	010	21.0	852
	1591.33	$H^{17}OH$	010	3.82	668
	1594.75	$H^{16}OH$	010	10,400.0	1807
2.7 μm	3151.63	$H^{16}OH$	020	75.4	1146
	3657.05	$H^{16}OH$	100	486.0	1381
	3707.47	$H^{16}OD$	001	1.42	1651
	3741.57	$H^{18}OH$	001	13.9	711
	3748.32	$H^{17}OH$	001	2.52	529
	3755.93	$H^{16}OH$	001	6930.0	1750

[a] The lower states are all 000. S_n for isotopes is calculated on the basis of the total number of molecules of all isotopic species. Only the 12 most important bands are listed.

Table 5.5. Overtone and combination bands of water vapor[a]

Region	Band origin (cm^{-1})	Isotope	Upper state ($v_1v_2v_3$)	S_n cm $\times 10^{21}$ at 296 K	Number of lines listed
Ω	5234.98	$H^{16}OH$	110	37.2	991
	5331.27	$H^{16}OH$	011	804.0	1306
ψ	6871.51	$H^{16}OH$	021	56.4	—[b]
	7201.48	$H^{16}OH$	200	52.9	—
	7249.93	$H^{16}OH$	101	747.0	—
ϕ	8807	$H^{16}OH$	111	49.8	—
τ[c]	10,239	$H^{16}OH$	121	2.0	—
σ[c]	10,613	$H^{16}OH$	201	10.0	—
ρ[c]	11,032	$H^{16}OH$	003	2.0	—
Visible	13,653	$H^{16}OH$	221	—	216
	13,828	$H^{16}OH$	202	—	169
	13,831 to 17,458	$H^{16}OH$	301	—	330
	17,496	$H^{16}OH$	500	—	108
		$H^{16}OH$	203	—	182

[a] S_n for isotopes is calculated on the basis of the total number of molecules of all species.
[b] —indicates that the information is not available although some must exist on magnetic tape.
[c] Not in the AFGL compilation.

5.4.3. *Continuum absorption*

The window between the water vapor 6.3 μm band and the rotation band occurs close to the peak of the Planck function at atmospheric temperatures and the transmission of thermal radiation through this window is of crucial importance for some atmospheric problems. The window region contains weak high-J lines of both bands but these are superimposed on a stronger continuum that absorbs about 10% of the incident radiation for a vertical path through an atmosphere containing 1 g cm^{-2} of water vapor.

There is a debate about the reason for this continuum; as for the collision-induced bands of oxygen (§ 5.3.4), the debate centers around the alternatives of dimers of finite lifetime or events taking place only

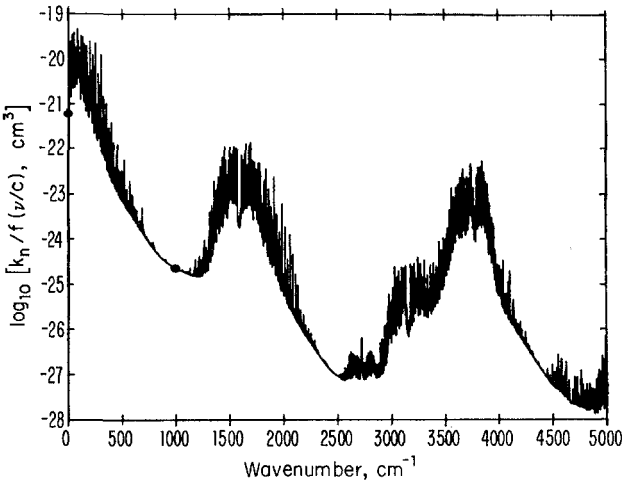


FIG. 5.7. Theoretical absorption coefficients for pure water vapor at 1 bar and 296 K. The vertical axis is the molecular absorption coefficient divided by a "radiation term,"

$$f(\nu/\mathbf{c}) = (\nu/\mathbf{c}) \tanh(\mathbf{h}\nu/2\mathbf{k}\theta),$$

which is approximately equal to $\nu/\mathbf{c}$ (the frequency in wave numbers) for $\nu/\mathbf{c} > 500\,\mathrm{cm}^{-1}$. After Clough et al. (1980).

during collisions. Both yield binary absorption coefficients, although results are sometimes given in terms of normal coefficients with linear pressure dependence, which amounts to the same thing. To attribute the continuum to events occurring during collisions is equivalent to saying that it is caused by the far wings of strong lines positioned close to band centers. This is the most popular explanation and there are reasonable line shapes that give a good account of the observed continuum absorption.

Figure 5.7 shows a calculation based upon a theoretical line shape. The envelope beneath the curve is the continuum; the fine structure by itself would normally be identified as rotation lines; but, in fact, both continuum and lines result from the same data on line positions and line shapes. The figure makes the point that continua occur at all frequencies. Window regions are those in which continuum absorption is stronger than line absorption (ca. 1000, 2400, and $4300\,\mathrm{cm}^{-1}$), but only the $1000\,\mathrm{cm}^{-1}$ region is important for atmospheric calculations.

Laboratory measurements of binary absorption coefficients for the $1000\,\mathrm{cm}^{-1}$ window are shown in Fig. 5.8. These are for water–water collisions. The same investigators found the coefficients for water–nitrogen collisions to be so small that, despite the much larger concentation of nitrogen molecules, the contribution of water–nitrogen collisions is probably unimportant in the atmosphere. This view corresponds to that

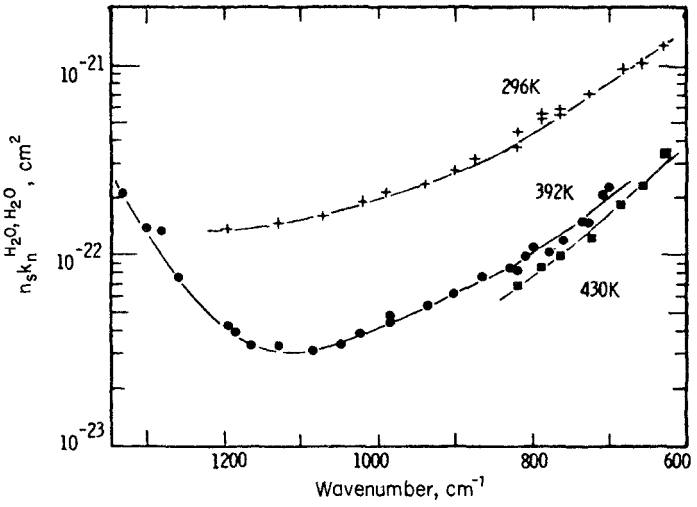


FIG. 5.8. Binary absorption coefficients for water–water collisions in the 1000 cm^{-1} window. n_s is Loschmidt's number. After Burch and Gryvnak (1980).

of some field investigators, based on studies of the correlation between atmospheric absorption and water vapor pressure. The term *e-type absorption* (from the meteorological symbol for water vapor pressure) is often used to describe the self-broadening nature of this phenomenon.

The increase of absorption with decreasing temperature shown in Fig. 5.8 has been cited as a reason to favor dimer theories of the continuum, but this temperature variation is also consistent with far wing line shapes, given appropriate interaction potentials.

Figure 5.9 shows binary absorption coefficients in the 2600 cm^{-1}

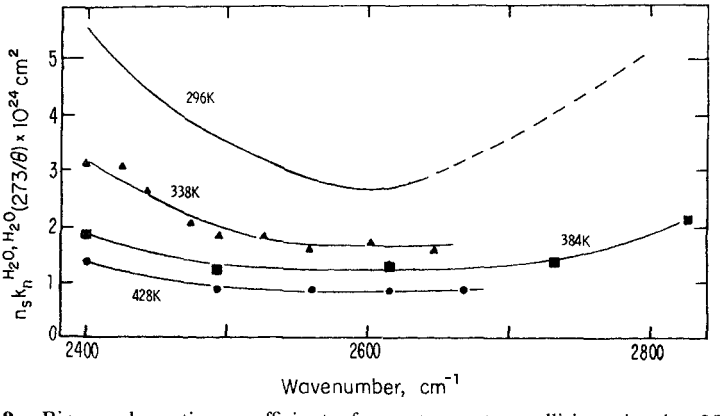


FIG. 5.9. Binary absorption coefficients for water–water collisions in the 2500 cm^{-1} window. n_s is Loschmidt's number. The data for 296 K are extrapolated. After Burch and Gryvnak (1980).

region of the spectrum. This region is not as free from rotation lines as is the $1000\,cm^{-1}$ window and the investigators have used numerical methods to correct for the absorption by nearby lines, so that only continuum absorption is represented.

The self-broadening coefficients shown in Fig. 5.9 are unimportant for atmospheric calculations. However, water–nitrogen collisions are more important than for the $1000\,cm^{-1}$ window, and it may be desirable to take account of them under some circumstances. The ratio of nitrogen–water to water–water coefficients is about 0.08, which is offset by the large ratio of nitrogen to water molecular densities.

5.5. Carbon dioxide

5.5.1. *The vibration–rotation spectrum*

The carbon dioxide molecule is linear and symmetric (OCO), with a bond length of 115.98 pm in the ground vibrational state, and a corresponding rotational constant of $0.3906\,cm^{-1}$. Because of the symmetry, the molecule has no permanent dipole moment and no permitted rotation band.

The fundamental modes of vibration of a linear molecule were discussed in § 3.2.2. Owing to the symmetry of the molecule, the ν_1 vibration involves no change in the dipole moment (Fig. 3.9) and is inactive. It has a frequency close to twice that of the ν_2 vibration, with the result that Fermi resonance occurs between groups of levels such as $(02^00, 10^00)$, $(03^10, 11^10)$, $(04^01, 12^01, 20^01)$, etc.

The ν_2 bending frequency is degenerate and the selection rules involve the quantized vibrational angular momentum. The ν_2 fundamental involves an $l=0$ to $l=1$ transition and is perpendicular. The ν_3 fundamental, on the other hand, involves an $l=0$ to $l=0$ transition, which is parallel, and the band lacks a Q-branch.

The oxygen atom has zero nuclear spin and the statistical weights of odd-J levels are zero, so that alternate rotation lines are missing from the spectrum. The same applies to the isotope $^{16}O^{13}C^{16}O$, forming 1.108% of the total carbon dioxide. $^{16}O^{12}C^{17}O$ and $^{16}O^{12}C^{18}O$ are present in concentrations of 0.0646 and 0.4078% and, being of lower symmetry, have a rotational structure that is different from that of $^{16}O^{12}C^{16}O$. Isotopic shifts can be estimated from the data in Table 5.6.

5.5.2. *Listed data*

The thermal spectrum of carbon dioxide is dominated by the very strong 15 μm (ν_2) and 4.3 μm (ν_3) bands (Table 5.6). Both bands are complicated by Fermi resonances and by the fact that the first ν_2 level is significantly populated at atmospheric temperatures. Bands with lower

Table 5.6. The strongest vibration–rotation bands of carbon dioxide[a]

Region	Band origin (cm^{-1})	Isotope	Upper state ($v_1 v_2^l v_3$)	Lower state ($v_1 v_2^l v_3$)	S_n cm × 10^{20} at 296 K
15 μm	618.03	$^{16}O^{12}C^{16}O$	10^00	01^10	14.4
	647.06	$^{16}O^{12}C^{16}O$	11^10	10^00	2.22
	648.48	$^{16}O^{13}C^{16}O$	01^10	00^00	8.60
	662.37	$^{16}O^{12}C^{18}O$	01^10	00^00	3.30
	667.38	$^{16}O^{12}C^{16}O$	01^10	00^00	826.0
	667.75	$^{16}O^{12}C^{16}O$	02^20	01^10	64.9
	668.11	$^{16}O^{12}C^{16}O$	03^30	02^20	3.82
	688.68	$^{16}O^{12}C^{16}O$	11^10	10^00	1.49
	720.81	$^{16}O^{12}C^{16}O$	10^00	01^10	18.5
4.3 μm	2271.76	$^{16}O^{13}C^{16}O$	01^11	01^10	8.18
	2283.49	$^{16}O^{13}C^{16}O$	00^01	00^00	96.0
	2311.68	$^{16}O^{12}C^{16}O$	03^31	03^30	1.23
	2319.74	$^{16}O^{12}C^{18}O$	01^11	01^10	2.58
	2324.15	$^{16}O^{12}C^{16}O$	02^21	02^20	30.8
	2326.59	$^{16}O^{12}C^{16}O$	10^01	10^00	11.8
	2327.43	$^{16}O^{12}C^{16}O$	10^01	10^00	19.3
	2332.11	$^{16}O^{12}C^{18}O$	00^01	00^00	33.3
	2336.64	$^{16}O^{12}C^{16}O$	01^11	01^10	766.0
	2349.15	$^{16}O^{12}C^{16}O$	00^01	00^00	9600.0
2.7 μm	3580.33	$^{16}O^{12}C^{16}O$	11^11	01^10	8.04
	3612.84	$^{16}O^{12}C^{16}O$	10^01	00^00	104.0
	3632.92	$^{16}O^{13}C^{16}O$	10^01	00^00	1.60
	3714.78	$^{16}O^{12}C^{16}O$	10^01	00^00	150.0
	3723.25	$^{16}O^{12}C^{16}O$	11^11	01^10	11.4
2.0 μm	4977.83	$^{16}O^{12}C^{16}O$	20^01	00^00	3.50
	5099.66	$^{16}O^{12}C^{16}O$	20^01	00^00	1.12

[a] The cut-off for S_n is 10^{-20} cm. Where bands have identical upper and lower states, they are resonating members of a Fermi triplet. Band intensities for upper state and isotopic bands are given in terms of the total number of molecules of all species and all levels.

states $v_2 = 1$, 2, or 3 and $v_1 = 1$ occur in Table 5.6. The intensities of these bands are, of course, strongly dependent upon temperature.

Also listed among the strong bands are two groups near 2.7 and 2.0 μm. The former consists of a number of combination bands that together have an intensity comparable to that of the fundamentals.

Table 5.7 shows data on some weaker bands of carbon dioxide that are important because they appear in gaps between other strong bands in the terrestrial spectrum. This is particularly true for the 10 μm bands ($\nu_3 - \nu_1$) because they appear in the middle of the water vapor window,

Table 5.7. Some weaker vibration–rotation bands of carbon dioxide[a]

Region	Band origin (cm^{-1})	Isotope	Upper state ($v_1 v_2^l v_3$)	Lower state ($v_1 v_2^l v_3$)	S_n cm × 10^{22} at 296 K
10 μm	960.96	$^{16}O^{12}C^{16}O$	00^01	10^00	4.9
	1063.73	$^{16}O^{12}C^{16}O$	00^01	10^00	6.3
5 μm	1932.47	$^{16}O^{12}C^{16}O$	11^10	00^00	4.1
	2076.87	$^{16}O^{12}C^{16}O$	11^10	00^00	22.0
	2093.36	$^{16}O^{12}C^{16}O$	12^20	01^10	4.0
	2129.78	$^{16}O^{12}C^{16}O$	20^00	01^10	1.3
1.6 μm	6227.92	$^{16}O^{12}C^{16}O$	30^01	00^00	4.3
	6347.85	$^{16}O^{12}C^{16}O$	30^01	00^00	4.3
1.4 μm	6935.15	$^{16}O^{12}C^{16}O$	01^13	01^10	1.1
	6972.58	$^{16}O^{12}C^{16}O$	00^03	00^00	15.0

[a] The cut-off for S_n is 10^{-22} cm. Where bands have identical upper and lower states they are members of a resonating Fermi triplet. Band intensities for isotopic and upper-state bands are calculated on the basis of the number of molecules of all species and in all levels.

close to the maximum of the Planck function at atmospheric temperatures. Their temperature dependence leads to an important feedback involving the atmospheric temperature.

The AFGL compilation contains 530 bands of carbon dioxide. The listing terminates at 9611 cm^{-1} although weak bands have been reported in the solar spectrum out to 12,774.4 cm^{-1} (10^05–00^00).

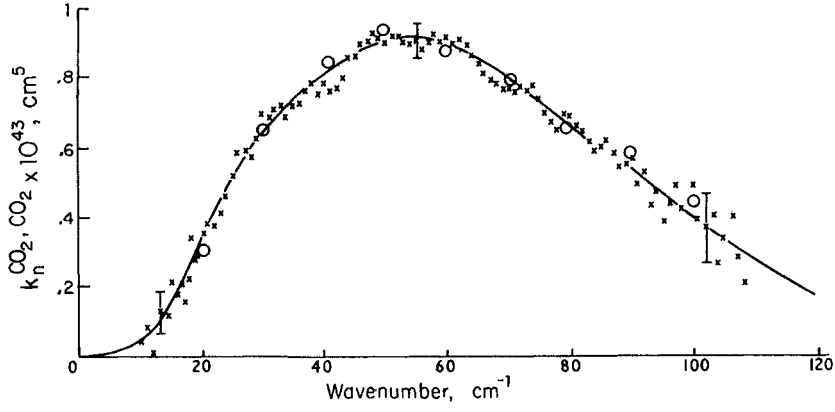


FIG. 5.10. The collision-induced rotation band of pure carbon dioxide at 293 K. The spectral resolution is 1.0 cm^{-1}. The circles and crosses are two independent sets of measurements. The vertical bars are error estimates. After Herries (1970).

5.5.3. *The collision-induced rotation spectrum*

Figure 5.10 shows the collision-induced rotation band for carbon dioxide–carbon dioxide collisions. No measurements are available for nitrogen collisions so that their impact on atmospheric calculations is difficult to gauge, but it is probably small. The effect of temperature has been measured over the range 200 K to 373 K. The binary coefficients vary in a manner consistent with the underlying rotational structure. The binary band intensity $S_n^{CO_2CO_2}$ is independent of temperature and equal to 6.5×10^{-42} cm^4.

5.6. Ozone

5.6.1. *Electronic bands*

The electronic spectrum of ozone is dominated by the Hartley bands, centered at 2553 Å, with a peak molecular absorption coefficient (cross section) of 1.15×10^{-17} cm^2. A typical solar beam reaching the ground traverses 1.4×10^{19} molecules of ozone cm^{-2} so that the transmission in the band center is about 10^{-70}.

The Hartley bands (Fig. 5.11) consist of some weak structure on a very strong continuum. A slight temperature dependence of absorption has been reported; compared to 291 K, the absorption coefficient at 2500 Å is 0.98 at 243 K and 0.97 from 227 K to 201 K. Larger temperature variations may occur in the wings.

On the short-wave wing of the Hartley bands, the absorption coefficient falls to a minimum of 3×10^{-19} cm^2 at 2000 Å. Thereafter it

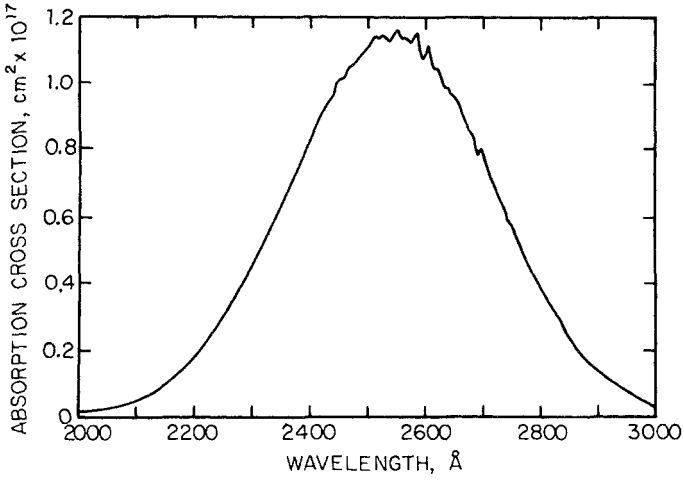


FIG. 5.11. The Hartley bands of ozone. The temperature is 303 K. After Griggs (1968).

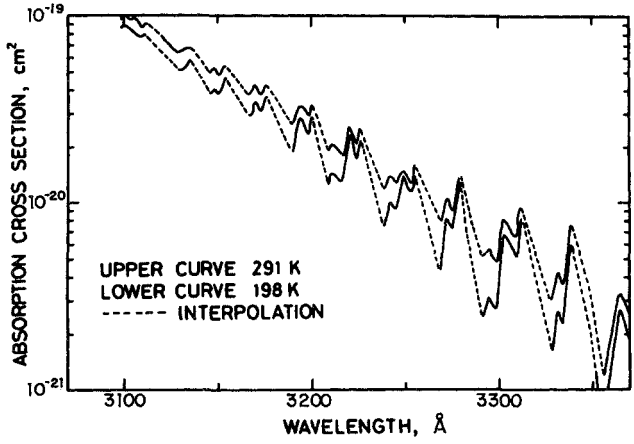


FIG. 5.12. The Huggins bands of ozone. After Vigroux (1953).

increases to a series of maxima, the highest being 2×10^{-17} cm^2 at 1220 Å.

On the long-wave wing of the Hartley bands lie a series of weak bands (the Huggins bands) that appear in the spectrum of the low sun and were responsible for the first positive identification of ozone in the atmosphere. A large temperature variation occurs in the minima of the Huggins bands (Fig. 5.12). There is probably no comparably important pressure effect.

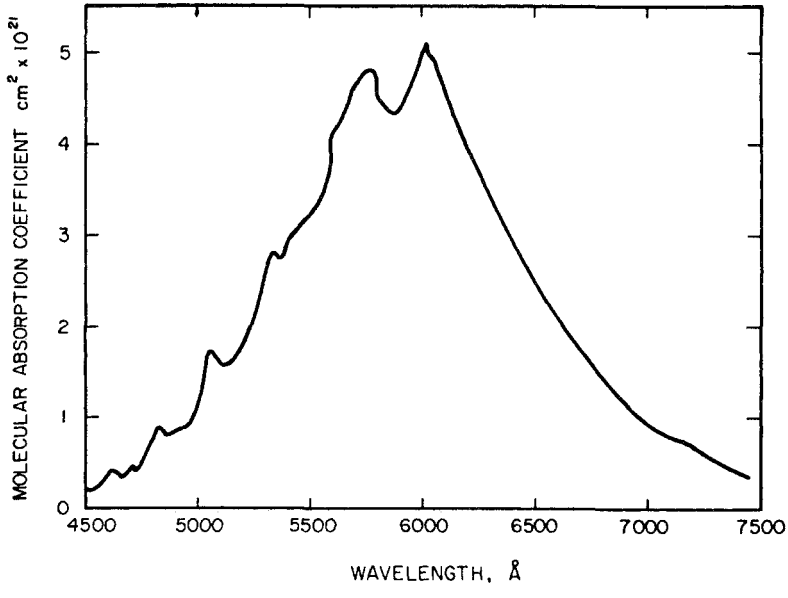


FIG. 5.13. The Chappuis bands of ozone at 291 K. After Vigroux (1953).

Between 3400 and 4500 Å lies a relatively transparent spectral region. Absorption begins again at 4500 Å with the Chappuis bands and extends to 7500 Å (Fig. 5.13). The maximum absorption coefficient in the Chappuis bands is 5×10^{-21} cm^2 and the peak absorption for solar radiation traversing two air masses is about 7%. This absorption can be significant under some circumstances because it occurs in the spectral region of maximum solar emission. Temperature coefficients for these bands appear to be negligible.

5.6.2. *The vibration–rotation spectrum*

Ozone is a nonlinear molecule with a bond length of 127.8 pm, a bond angle of 127°, and a permanent dipole moment of 1.77×10^{-30} C-m. The principal isotopic species $^{16}O^{16}O^{16}O$, $^{16}O^{18}O^{16}O$, and $^{16}O^{16}O^{18}O$ are present in proportions of 99.28, 0.203, and 0.406%, respectively, but only $^{16}O^{16}O^{16}O$ has bands of importance for atmospheric calculations.

The 1982 AFGL listing contains 19 bands of ozone. Of these, 9 have band intensities larger than 10^{-19} cm and are listed in Table 5.8.

An unusual feature of the ozone molecule is that the ν_1 and ν_2 fundamentals, at 1103.14 and 700.93 cm^{-1}, respectively, are very weak compared to ν_3 at 1042.06 cm^{-1}, weaker in fact than the combination band $\nu_1 + \nu_3$ at 2110.79 cm^{-1}. In addition, ν_1 is very close to ν_3 and strong resonances make it difficult to assign line positions and to calculate good line intensities. The 1982 AFGL listing claims that most line

Table 5.8. The strongest vibration–rotation bands of $^{16}O_3$[a]

Region	Band origin (cm^{-1})	Upper state ($v_1v_2v_3$)	Lower state ($v_1v_2v_3$)	S_n cm × 10^{20} at 296 K
Rotation	0.00	000	000	41.3
14 μm	700.93	010	000	62.8
9.6 μm	1015.81	002	001	17.4
	1025.60	011	010	45.0
	1042.08	001	000	1394.0
	1103.14	100	000	67.1
Overtone and combination	2057.89	002	000	11.1
	2110.79	101	000	113.4
	3041.20	003	000	11.0

[a] Only bands with intensities greater than 10^{-19} cm are listed. Band intensities for upper-state bands are calculated on the basis of the number of molecules in all levels.

intensities for ν_1 and ν_3 are known to better than 10%, but this may be too optimistic.

5.7. Nitrous oxide, carbon monoxide, and methane

5.7.1. *Nitrous oxide*

The nitrous oxide molecule is linear, asymmetric (NNO), with N–N and N–O bond lengths of 112.6 and 118.6 pm, respectively, and a rotational constant in the ground state of $0.4190\,\text{cm}^{-1}$. The permanent dipole moment is 0.557×10^{-30} C-m. The three fundamentals are ν_1 ($1284.907\,\text{cm}^{-1}$), ν_2 ($558.767\,\text{cm}^{-1}$), and ν_3 ($2223.756\,\text{cm}^{-1}$). Since ν_1 is approximately equal to $2\nu_2$ there is a strong Fermi resonance between the levels. The isotopic abundances are $^{14}N^{14}N^{16}O$ (99.043%), $^{15}N^{14}N^{16}O$ and $^{14}N^{15}N^{16}O$ (0.358%), $^{14}N^{14}N^{18}O$ (0.199%), and $^{14}N^{14}N^{17}O$ (0.040%), but few lines from isotopic species are of significance.

The spectrum of nitrous oxide has been thoroughly investigated. The AFGL listing includes 106 bands (not including the rotation bands) arranged in 27 band systems. The 9 strongest are given in Table 5.9.

Table 5.9. The strongest band systems of nitrous oxide[a]

Region	Band origin (cm^{-1})	Upper state ($v_1 v_2^l v_3$)	S_n cm $\times 10^{20}$ at 296 K
Rotation	0.00	00^00	Not listed
17 μm	588.77	01^10	118
7.8 μm	1168.13	02^00	39
	1284.91	10^00	996
4.5 μm	2223.76	00^01	5710
Combination bands	2462.00	12^00	33
	2563.34	20^00	135
	3363.97	02^01	11
	3480.82	10^01	197

[a] The lower states are all 00^00. Only the 4.5 μm band contains significant isotopic lines. The cut-off for S_n is 10^{-19} cm.

5.7.2. *Carbon monoxide*

The carbon monoxide molecule has a C–O bond length of 123 pm and an equilibrium rotational constant of $1.9313\,\text{cm}^{-1}$. The permanent dipole moment is 0.34×10^{-30} C-m, giving rise to a weak rotation band.

The AFGL listing contains lines from a number of isotopes from the rotation band, the fundamental and the first and second overtones. Of these, only the fundamental of the principal isotope is important. Its origin is at 2143.27 cm^{-1} and its band intensity is 9.81×10^{-18} cm.

5.7.3. *Methane*

The methane molecule is a spherical top with a C–H bond length of 109.3 pm. Of the nine fundamentals, only four are independent (ν_1, ν_2, ν_3, ν_4), and only two are active in the infrared spectrum (ν_3 and ν_4); these are both triply degenerate. Interactions resolve all of these degeneracies and give rise to a line structure of exceptional complexity (see, for instance, Fig. 3.2, which, in the absence of interactions, should look more like the carbon monoxide band in Fig. 3.4). The strongest bands of methane are given in Table 5.10.

Table 5.10. The strongest bands of methane[a]

Bands	Band origin (cm^{-1})	Isotope	Upper state ($\upsilon_1\upsilon_2\upsilon_3\upsilon_4$)	S_n cm $\times 10^{20}$ at 296 K
Fundamentals	1302.77	$^{13}CH_4$	0001	5.7
	1310.76	$^{12}CH_4$	0001	504.1
	1533.37	$^{12}CH_4$	0100	5.5
	3009.53	$^{13}CH_4$	0010	29.3
	3018.92	$^{12}CH_4$	0010	1022.0
Overtone and combination	2612	$^{12}CH_4$	0002	5.4
	2822	$^{13}CH_4$	0101	4.3
	2830	$^{12}CH_4$	0101	38.0
	3062	$^{12}CH_4$	0201	16.4
	4223	$^{12}CH_4$	1001	24.0
	4340	$^{12}CH_4$	0011	40.8
	4540	$^{12}CH_4$	0110	6.2

[a] The ground states are all 0000. The intensities of isotope bands are based on the total number of molecules of all species. ^{13}C atoms form 1.108% of the total carbon.

BIBLIOGRAPHY

5.1. Introduction

Figure 5.1 is taken from

Herzberg, L., 1965, "Solar optical radiation and its role in upper atmospheric processes," in *Physics of the Earth's upper atmosphere* (C. O. Hines, I. Paglis, T. R. Hartz, and J. A. Fejer, Eds.). Englewood Cliffs, N.J.: Prentice Hall, p. 31.

The AFGL (previously called AFCRL) *main listing* is documented in the

following references:

McClatchey, R. A., Fenn, W., Selby, J. E. A., Volz, F. E., and Garing, J. S., 1973, "AFCRL atmospheric absorption line parameters compilation," unpublished report to the US Air Force AFCRL-TR-73-0096, obtainable from the National Technical Information Service, Washington, D.C.

Rothman, L. S., and McClatchey, R. A., 1976, "Updating of the AFCRL atmospheric absorption line parameter compilation," *Appl. Optics* **15,** 2616.

Rothman, L. S., 1978, "Update of the AFGL atmospheric absorption line parameters compilation," *Appl. Optics* **17,** 3517.

———, 1981, "AFGL atmospheric absorption line parameters compilation: 1980 version," *Appl. Optics* **20,** 791.

———, et al., 1983, "AFGL atmospheric absorption line parameters compilation: 1982 edition," *Appl. Optics* **22,** 2247.

The AFGL trace gas compilation is documented in

———, et al., 1978, "AFGL trace gas compilation," *Appl. Optics* **17,** 507.

———, et al., 1981, "AFGL trace gas compilation: 1980 version," *Appl. Optics* **20,** 1323.

———, et al., 1983, "AFGL trace gas compilation: 1982 version," *Appl. Optics* **22,** 1616.

The tapes can be obtained from the National Climate Center of NOAA, Environmental Data Service, Federal Building, Asheville, North Carolina 28801.

A compilation of references to the literature on band intensities and line widths for gases, including those discussed in this Chapter, is

Smith, M. A. H., Rinsland, C. P., Fridovich, B., and Rao, K. N., 1985, "Intensities and collision broadening parameters for infrared spectra," in *Molecular spectroscopy: Modern research,* Vol. III (K. N. Rao, Ed.). New York: Academic Press, p. 111.

A recent review of electronic absorptions by atmospheric gases also containing many references is

Brasseur, G., and Solomon, S., 1984, *The aeronomy of the middle atmosphere: Chemistry and physics of the stratosphere and mesosphere.* Dordrecht: Reidel.

An older, but still valuable review of ultraviolet data is

Watanabe, K., 1958, "Ultraviolet absorption processes in the upper atmosphere," *Adv. Geophy.* **5,** 153.

5.2. Nitrogen

Electric quadrupole lines in the fundamental band have been observed in atmospheric spectra

Goldman, A., Reid, J., and Rothman, L. S., 1981, "Identification of electric quadrupole O_2 and N_2 lines in the infrared atmospheric absorption spectrum due to the vibration–rotation fundamentals," *Geophys. Res. Lett.* **8,** 77.

Figure 5.2 is from

Stone, N. W. B., Read, L. A. A., Anderson, A., Dagg, I. R., and Smith, W., 1984, "Temperature dependent collision-induced absorption in nitrogen," *Can. J. Phys.* **62,** 338.

Figure 5.3 is taken from Shapiro and Gush (1966), see § 3.4.
Additional references to the collision-induced rotation band are

Herries, J. E., 1979, "The temperature dependence of collision-induced absorption in gaseous N_2," *J. Opt. Soc. Am.* **69,** 386.

Bosomsworth, D. R., and Gush, H. P., 1965, "Collision-induced absorption of compressed gases in the far infrared, Parts I and II," *Can. J. Phys.* **43,** 729, 751.

The fundamental band is treated by

Reddy, S. P., and Cho, C. W., 1965, "Induced infrared absorption of nitrogen and nitrogen-foreign gas mixtures", *Can. J. Phys.* **43,** 2331.

The following papers are concerned with atmospheric absorption by the fundamental band:

Farmer, C. B., and Houghton, J. T., 1966, "Collision-induced absorption in the earth's atmosphere," *Nature (London)* **209,** 1341.

Bernstein, L. S., Robertson, D. C., Conant, J. A., and Sandford, B. P., 1979, "Measured and predicted atmospheric transmission in the 4.0–5.3-μm region, and the contribution of continuum absorption by CO_2 and N_2," *Appl. Optics* **18,** 2454.

Susskind, J., and Searl, J. E., 1977, "Atmospheric absorption near 2400 cm^{-1}," *J. Quant. Spectrosc. Radiat. Transfer* **18,** 581.

5.3. Oxygen

The data in Fig. 5.4 are from Brasseur and Solomon (1984), see § 5.1.

Forbidden quadrupole transitions in the fundamental band are discussed by Goldman et al. (1981), see § 5.2.

A useful collection of data on the A, B, and γ "atmospheric bands" of oxygen is by

Nakazawa, T., Yamanouchi, T., and Tanaka, M., (1982), "Line parameters of the oxygen A-band and calculation of the atmospheric transmission functions in the A, B and γ bands," *J. Quant. Spectrosc. Radiat. Transfer* **27,** 615.

Magnetic dipole transitions in the rotational band are discussed by Townes and Schawlow (1975), see § 3.2.

The collision-induced visible bands of oxygen have been the subject of many

investigations:

Blickensderfer, R. P. and Ewing, G. E., 1967, "Collision-induced absorption spectrum of gaseous oxygen at low temperatures and pressures. II. The simultaneous transitions $^1\Delta_g + {}^1\Delta_g \leftarrow {}^3\Sigma_g^- + {}^3\Sigma_g^-$ and $^1\Delta_g + {}^1\Sigma_g^+ \leftarrow {}^3\Sigma_g^- + {}^3\Sigma_g^-$," *J. Chem. Phys.* **51,** 5284.

For a discussion of the evidence in favor of true dimer absorption see Long and Ewing (1973), § 3.4.

The collision-induced fundamental has been observed in the stratosphere by

Rinsland, C. P., Smith, M. A. H., Seals, R. K., Goldman, A., Murcray, F. J., Murcray, D. G., Larsen, J. C., and Rarig, P. L., 1982, "Stratospheric measurements of collision-induced absorption by molecular oxygen," *J. Geophys. Res.* **87,** 3119.

A discussion of the importance of the collision-induced fundamental for atmospheric calculations has been given by Timofeyev and Tonkov (1978), see § 3.4.

The collision-induced rotation band data in Fig. 5.6 are from Bosomsworth and Gush (1965), see § 5.3.

The importance of transitions in the ground state of atomic oxygen was first pointed out by

Bates, D. R., 1951, "The temperature of the upper atmosphere," *Proc. Phys. Soc. London Sect. B* **64,** 805.

5.4. Water vapor

Elsasser (1941), see § 1.1, was the first to draw attention to the importance of the 10 μm water vapor continuum for atmospheric processes. More recent contributions are by

Clough, S. A., et al., 1980, "Theoretical line shape for H_2O vapor: Application to the continuum," in *Atmospheric water vapor* (A. Deepak, T. D. Wilkerson, and L. H. Ruhnke, Eds.). New York: Academic Press, p. 25.

Burch, D. E., and Gryvnak, D. A., 1980, "Continuum absorption by H_2O vapor in the infrared and millimeter regions," in *Atmospheric water vapor* (A. Deepak, T. D. Wilkerson, and L. H. Ruhnke, Ed.). New York: Academic Press, p. 47.

Thomas, M. E., and Nordstrom, R. J., 1985, "Line shape model for describing infrared absorption by water vapor," *Appl. Optics* **24,** 3526.

A case for the importance of the 2500 cm^{-1} continuum is made by

Barton, I. J., 1981, "Water vapor absorption in the 3.5 to 4.2 μm atmospheric window," *Quart. J. R. Meteorol. Soc.* **107,** 967.

5.5. Carbon dioxide

The collision-induced rotation band has been studied by

Herries, J. E., 1970, "The temperature variation of the far infrared absorption in compressed CO_2," *J. Phys. Ser. B: Atom. Molec. Phys.* **3,** 704.

5.6. Ozone

The Hartley band has most recently been studied by

Griggs, M., 1968, "Absorption coefficients of ozone in the ultraviolet and visible regions," *J. Chem. Phys.* **49,** 857.

Some old measurements on the Huggins and Chappuis bands have stood the test of time,

Vigroux, E., 1953, "Contribution à l'étude expérimentale de l'absorption de l'ozone," *Ann. Phys.* **8,** 709.

6

RADIATION CALCULATIONS IN A CLEAR ATMOSPHERE

6.1. Introduction

6.1.1. *Line-by-line calculations*

This chapter is concerned with the requirements of numerical weather prediction and general circulation models. These numerical models always assume a stratified atmosphere and utilize a limited number of grid points in the vertical direction. Computations are repeated at many horizontal grid points and at frequent time intervals; a premium is placed on computational economy. The nested integrals involved in radiative flux and heating calculations, particularly the frequency integration, can create an unacceptable computational burden unless approximated.

In this chapter we limit attention to *clear-sky conditions,* i.e., to absorbing constituents and a thermal source function (§ 2.2). For a Planck function, the formal solution, (2.86), is a definite integral involving measurable quantities, temperatures, and gaseous densities. Scattering problems, on the other hand, involve the intensity in the source function and cannot be solved by a single application of this integral. Scattering calculations will be discussed further in Chapter 8; it will be shown that scattering can be neglected if the volume scattering coefficient is not very much larger than the volume absorption coefficient. This is usually the case for aerosols in the thermal region of the spectrum.

As regards boundary conditions, it is usual for clear-sky calculations to assume that the earth's surface and the upper and lower surfaces of clouds can be treated as black surfaces in the thermal spectrum. Equations (2.86) and (2.87) are stated in terms of general boundary conditions. In the flux and heating integrals, (2.106) and (2.110), these conditions are specialized to a black surface at ground level, but they can be generalized without difficulty to include a black surface at any level or partial reflection from these surfaces, if appropriate.

The equations for which efficient algorithms are required are the flux equations, (2.107) and (2.108), the heating equations, (2.110) or (2.111),

and the solar flux equations, (2.115). The nested integrals are

1. the vertical integral, (2.92), for the optical depth;
2. the integral, (2.86), along the optical path;
3. the angular integral, (2.102);
4. an integral over all frequencies.

We may introduce the issues by considering a restricted example, that of the intensity recorded outside the atmosphere by a downward pointing satellite spectrometer. The spectrometer is approximately collimated, eliminating the angular integration. The calculation is made at only one atmospheric level, reducing the work involved in (1) and (2) together by the square of the number of vertical data points, and the calculation need be performed only once rather than at frequent time intervals. With these restrictions it is feasible to use a line-by-line calculation without further approximation.

The appropriate equations are (2.92) and (2.98) evaluated at $\tau_\nu = 0$,

$$\tau_\nu(z, \infty) = \int_z^\infty k_{\nu,\nu}(z')\, dz', \tag{6.1}$$

and

$$I_\nu(0, \xi) = B_\nu(\theta^*) \exp(-\tau_{1,\nu}/\xi) + \int_0^{\tau_{1,\nu}} B_\nu(\tau_\nu) \exp(-\tau_\nu/\xi)\, d\tau_\nu/\xi. \tag{6.2}$$

The volume absorption coefficient can be expressed in terms of the contributions by individual lines (suffix, j) of a number of chemical species (index, l),

$$k_{\nu,\nu} = \sum_{j,l} n^l S^l_{n,j} f^l_j(\nu - \nu^l_{0,j}) + k^c_\nu. \tag{6.3}$$

S_n is the line intensity, f is the normalized line profile, ν_0 is the line frequency, and n^l is the number density of molecules of species, l.

k^c_ν is a continuum coefficient that is handled differently by different investigators. Given a complete knowledge of line profiles it may be calculated as a part of the first term on the right-hand side of (6.3). As discussed in § 3.3.4, however, there is a difference between our confidence in our knowledge of the line profile close to the line center and our knowledge of the far wings. It is common practice to include only the line centers in the sum in (6.3) and to treat the remainder semiempirically as an added continuum. However this situation may be handled, we shall regard (6.3) as soluble, given the data on the AFGL or similar tape, and some knowledge of line shapes. The line-by-line sum in (6.3) adds another to the four nested operations already described.

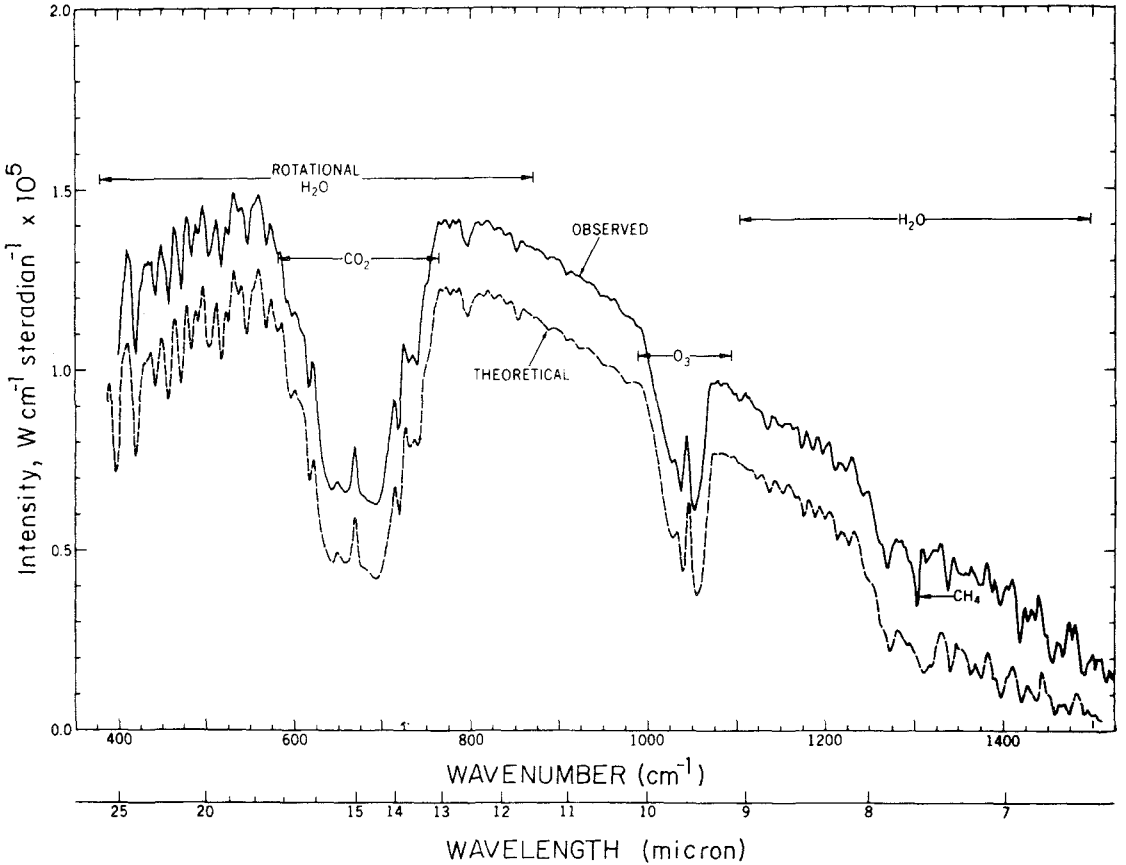


FIG. 6.1. Observed and theoretical spectra for clear skies over the Gulf of Mexico, April 23, 1969. The observed spectrum is displaced upward by 0.2×10^{-5} W cm^{-2} steradian^{-1} wave number^{-1}. After Conrath et al. (1970).

Given the temperature and densities of absorbing molecules at each atmospheric level, (6.2) can be solved numerically for a single frequency. Figures 6.1 and 6.2 show comparisons between observed and computed spectra for measurements made on the meteorological satellite, Nimbus 3, at a time and place when good atmospheric data were available. The spectra are averaged over the band-pass of the instrument ($\sim 5\,\mathrm{cm}^{-1}$), sufficient to smooth out most of the fine spectral detail.

The agreement between theory and observation in Figs. 6.1 and 6.2 is generally within about 10%. It is surprising, at first sight, that it is not better. Uncertainties in the spectroscopic data are partially responsible, but it is difficult to assign all the errors to this source. Local variations in temperature and departures from a strictly stratified atmosphere must also contribute errors. The radiosonde data used may not correctly apply

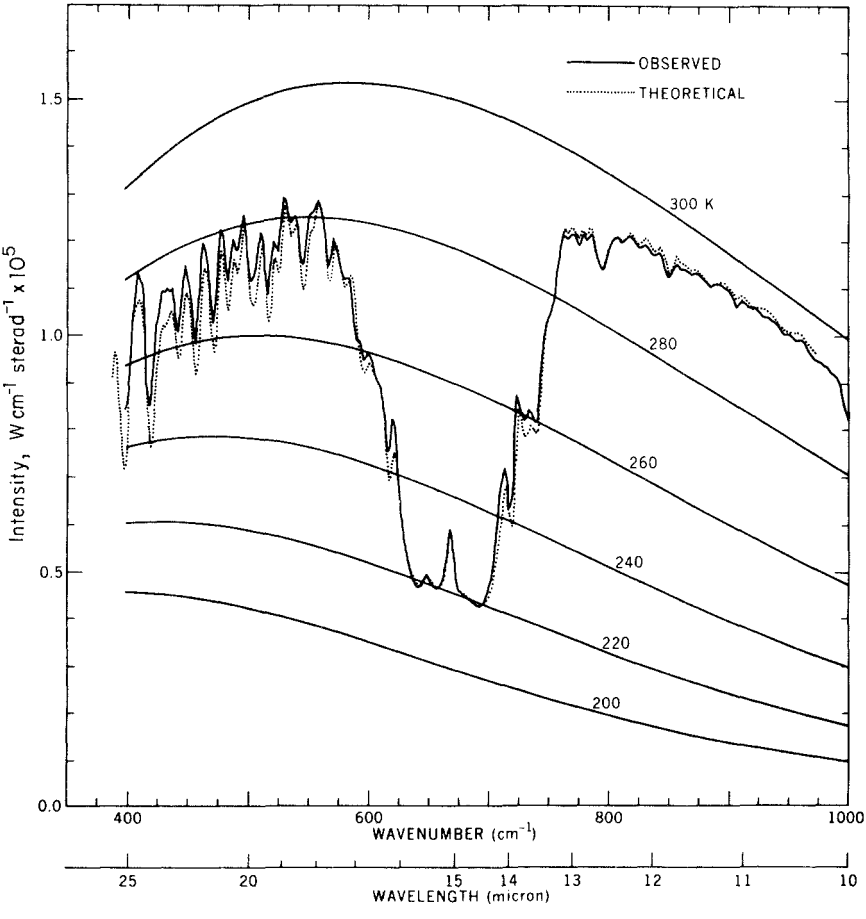


FIG. 6.2. Detail from Fig. 6.1, with black-body curves. After Conrath et al. (1970).

to the path of the radiation. The atmospheric temperatures could be adjusted slightly to give better agreement, a step in the direction of temperature inversions from observed intensities, a technique to be discussed in § 6.5.

It is difficult to reach any firm conclusion about errors such as are exhibited in Figs. 6.1 and 6.2. Line-by-line calculations are often adopted as a standard against which to test certain approximations. Their value in a relative context is indisputable, but that should not be taken to mean that line-by-line calculations are necessarily of high absolute accuracy.

This comment is relevant to an implicit assumption in much of the current literature: that more and more detailed physics encoded onto larger and larger computers will eventually yield accurate weather and climate predictions. This is more an article of faith than a demonstrable proposition. It is also possible to argue that numerical complexity hides or introduces its own sources of error, in addition to making it impossible to penetrate the algorithms of another investigator. The utility of radiation algorithms probably cannot be judged in general terms and is better discussed in the context of specific atmospheric processes (Chapters 9 and 10).

6.1.2. *The angular integration*

Given the intensity as a function of the directional cosine, ξ, the flux may be calculated from (2.3) and (2.101). We distinguish between upward and downward components in the manner of (2.107) and (2.108),

$$\begin{aligned} F_\nu^+(z)/\pi &= 2\int_0^1 I_\nu(z, \xi)\xi \, d\xi, \\ F_\nu^-(z)/\pi &= 2\int_0^{-1} I_\nu(z, \xi)\xi \, d\xi. \end{aligned} \tag{6.4}$$

The computing time for a flux calculation is proportional to the number of quadrature points used to calculate the integral in (6.4); experience suggests that sufficient accuracy can often be obtained with a single quadrature point. This class of approximation (a *diffusivity approximation*) is employed in many radiation algorithms.

First-order Gaussian quadrature, (2.148), leads to

$$\begin{aligned} F_\nu^+(z)/\pi &\simeq \frac{2}{3^{1/2}} I_\nu(z, 3^{-1/2}), \\ F_\nu^-(z)/\pi &\simeq \frac{2}{3^{1/2}} I_\nu(z, -3^{-1/2}). \end{aligned} \tag{6.5}$$

For historical reasons, (6.5) is not employed in this form. Instead, discussion has centered on the best value for the *diffusivity factor, r,* in the approximations

$$\begin{aligned} F_\nu^+(z)/\pi &\simeq I_\nu(z, \xi = r^{-1}), \\ F_\nu^-(z)/\pi &\simeq I_\nu(z, \xi = -r^{-1}). \end{aligned} \tag{6.6}$$

r can be chosen to make (6.6) exact for any particular values of ξ and z. The question is whether a value exists that makes (6.6) approximately correct for averages over spectral intervals and for all values of z? The consensus of many investigations is that errors in calculated heating rates should not exceed 2% if we use $r = 1.66$, a value first proposed by Elsasser in 1942 for a regular band.

The implication of (6.6) is that flux calculations can be reduced to intensity calculations made at a specific zenith angle. Since ξ occurs only as a divisor in the calculation of optical depth, (2.197), the intensity calculation may be made for a vertical path through an atmosphere in which absorber densities have been increased by a factor r.

A formal relationship between intensity and flux can also be demonstrated from the exact equations for a stratified atmosphere. We introduce a *flux transmission,*

$$T_\nu^{\mathrm{f}}(z', z) = 2E_3[\tau_\nu(z', z)], \tag{6.7}$$

to be compared to the normal transmission,

$$T_\nu(z', z; \xi) = \exp[-\tau_\nu(z', z)/\xi]. \tag{6.8}$$

If we substitute (6.8) into (2.98) and (2.99), and change the independent variables to z and z' (Fig. 2.5, but with $z'' = 0$),

$$\begin{aligned} I_\nu(z, \xi) &= B(\theta^*)T_\nu(0, z; \xi) \\ &\quad + \int_0^z J_\nu(z') \frac{\partial T_\nu(z', z; \xi)}{\partial z'} dz' \end{aligned} \tag{6.9}$$

for $1 \geqslant \xi > 0$, and

$$I_\nu(z, \xi) = -\int_z^\infty J_\nu(z') \frac{\partial T_\nu(z, z'; \xi)}{\partial z'} dz' \tag{6.10}$$

for $-1 \leqslant \xi < 0$.

If we now substitute (6.7) into (2.107) and (2.108) we find

$$\frac{F_\nu^+(z)}{\pi} = B_\nu(\theta^*)T_\nu^{\mathrm{f}}(0, z) + \int_0^z J_\nu(z')\frac{\partial T_\nu^{\mathrm{f}}(z', z)}{\partial z'}\,dz', \tag{6.11}$$

and

$$\frac{F_\nu^-(z)}{\pi} = -\int_z^\infty J_\nu(z')\frac{\partial T_\nu^{\mathrm{f}}(z, z')}{\partial z'}\,dz'. \tag{6.12}$$

The analogy between flux and intensity calculations is now explicit. If we wish to adopt the diffusivity approximation, we write

$$T^{\mathrm{f}}(x) \simeq T(rx). \tag{6.13}$$

In § 2.4.5 we stated that the diffusivity approximation leads to an approximate differential equation, (2.151), similar to the two-stream approximation. The two-stream approximation can be obtained from first-order Gaussian quadrature, (6.5), while the diffusivity approximation results from (6.6). The close relationship between these two equations accounts for the similarity between the two procedures.

6.1.3. *The frequency integration*

Thermodynamic calculations involve fluxes and heating rates integrated over the entire spectrum. In Chapter 4, we approximated the frequency integral by sums over discrete frequency ranges (suffix i, width $\Delta\nu_i$) that are narrow enough to neglect changes in the source function but wide enough to contain many lines. By analogy with the intensity integral, (4.2), the flux integrals (6.11) and (6.12) are, for a single spectral range,

$$\frac{F_i^+(z)}{\pi} = B_i(\theta^*)\bar{T}_i^{\mathrm{f}}(0, z) + \int_0^z J_i(z')\frac{\partial \bar{T}_i^{\mathrm{f}}(z', z)}{\partial z'}\,dz', \tag{6.14}$$

$$\frac{F_i^-(z)}{\pi} = -\int_z^\infty J_i(z')\frac{\partial \bar{T}_i^{\mathrm{f}}(z, z')}{\partial z'}\,dz', \tag{6.15}$$

where

$$\bar{T}_i^{\mathrm{f}} = \frac{1}{\Delta\nu_i}\int_{\Delta\nu_i} T_\nu^{\mathrm{f}}\,d\nu. \tag{6.16}$$

It is plausible to employ band model theory (Chapter 4) to evaluate

approximately integrals such as (6.16). This procedure, or a simplified version of it, has been the most commonly used approach to the frequency integral for about 40 years. It obliges us to look at another problem.

Band model theory has been developed for homogeneous absorption paths, usually in terms of two nondimensional variables involving the pressure and the amount of absorbing matter in the path. Atmospheric problems, on the other hand, involve paths along which pressure and temperature vary continuously. There is no obvious connection between these two circumstances, but if we wish to use the available theory of band models a connection must be demonstrated. Approximate treatments are examined in the next section.

6.2. Transmission through a nonhomogeneous atmosphere

6.2.1. *Exact solutions for constant mixing ratio*

A number of atmospheric gases have approximately constant mixing ratios in the lower atmosphere. If, in addition, temperature effects upon line intensity and width may be neglected exact solutions exist for averaged transmissions. Although this situation is an idealization, the effect of pressure varying along the path is correctly treated and the results are useful as standards against which to test approximations.

For hydrostatic equilibrium,

$$dp = -\rho g\, dz, \tag{6.17}$$

where p and ρ are the air pressure and density and g is gravity. For an absorbing gas with constant mixing ratio, c, the absorber density is $c\rho$. If we adopt the Lorentz profile, (3.48), the optical path between two levels z and z' is

$$\begin{aligned}\tau_\nu(z, z') &= \int_z^{z'} \frac{S_m \alpha_L(z'')}{\pi[\nu^2 + \alpha_L(z'')^2]}\, c\rho(z'')\, dz'' \\ &= \int_{\alpha_L(z')}^{\alpha_L(z)} \frac{2\tilde{u}\alpha_L\, d\alpha_L}{\nu^2 + \alpha_L^2} \\ &= \tilde{u} \ln \frac{\nu^2 + \alpha_L(z)^2}{\nu^2 + \alpha_L(z')^2},\end{aligned} \tag{6.18}$$

where $\tilde{u} = (p/\alpha_L)S_m c/2\rho g$ is a constant because the Lorentz width is proportional to the pressure.

The equivalent width for a single line evaluated over this atmos-

pheric path is

$$W(z, z') = \int_{-\infty}^{+\infty} \left\{ 1 - \left[\frac{\nu^2 + \alpha_L^2(z)}{\nu^2 + \alpha_L^2(z')} \right]^{\tilde{u}} \right\} d\nu. \tag{6.19}$$

The integral in (6.19) can be solved in terms of the hypergeometric function. The special case for $z' = \infty$ (a path to space) can be given in terms of gamma functions:

$$W(z, z') = \frac{2\alpha_L(z)\pi^{1/2}\Gamma(\tilde{u} + 1/2)}{\Gamma(\tilde{u})}. \tag{6.20}$$

This result can be generalized to give the mean transmission for a random model by substituting in (4.107).

An expression has also been derived for the mean transmission of an Elsasser model (§ 4.5)

$$\bar{T}(z, z') = \int_{-1/2}^{+1/2} \left(\frac{\cosh 2\pi y - \cos 2\pi x}{\cosh 2\pi y' - \cos 2\pi x} \right)^{\tilde{u}} dx. \tag{6.21}$$

Equation (6.21) has been integrated numerically and can also be expressed as a sum of hypergeometric functions.

6.2.2. *Scaling approximations*

The question posed in § 6.1.3, was how well can the transmission of a variable atmospheric path be represented by the transmission of a path at constant temperature and pressure? Let $\tilde{m}$, $\tilde{p}$, and $\tilde{\theta}$ be the parameters for the optimum homogeneous path. It is usual to fix $\tilde{\theta}$ by setting it equal to the average temperature for the path of integration, for example, and to seek the best values for the remaining two parameters. Before examining such two-parameter approximations, however, we first consider the one-parameter approximation for which the pressure is also preassigned (e.g., $\tilde{p} = 1$ bar) and only $\tilde{m}$ remains to be chosen. This is referred to as a *scaling approximation.*

A scaling approximation is exact if variables can be factorized in the following way:

$$k_\nu(p, \theta) = \phi(p, \theta)\eta(\nu). \tag{6.22}$$

The optical path may then be written

$$\begin{aligned} \tau_\nu &= \int_{\text{path}} k_\nu(p, \theta)\, dm \\ &= k_\nu(\tilde{p}, \tilde{\theta}) \int_{\text{path}} \phi(p, \theta)/\phi(\tilde{p}, \tilde{\theta})\, dm \\ &= k_\nu(\tilde{p}, \tilde{\theta})\tilde{m}, \end{aligned} \tag{6.23}$$

where

$$\tilde{m} = \int_{\text{path}} \frac{\phi(p, \theta)}{\phi(\tilde{p}, \tilde{\theta})} dm. \tag{6.24}$$

Given the factorization (6.22), (6.23) and (6.24) permit the optical path to be calculated from that for a homogeneous atmosphere. If the optical path is correct, so will be the fluxes and heating rates derived from it, and a calculation for the most general atmosphere is reduced to one for a homogeneous atmosphere with fixed pressure, $\tilde{p}$, and temperature, $\tilde{\theta}$, with a change of variable from the amount, m, to the *scaled amount*, $\tilde{m}$. The scaled amount can be calculated simply and independently from the rest of the calculation.

We must now discuss the function $\phi(p, \theta)$. We first consider its form for isolated Doppler and Lorentz lines and show that although an appropriate form exists for Doppler lines, different forms are required for strong and weak Lorentz lines. If a result is correct for an isolated line, then it is also correct for a random array of identical lines [see (4.108)] and for an important asymptotic limit of the Elsasser model [see (4.83)]. For a random array of differing lines, $\phi(p, \theta)$ must be chosen to give the best estimate of the average equivalent width for the lines in a given spectral range; this will satisfy requirements for all random models [see (4.112)] and for the semiempirical band model (4.143).

First consider the Doppler profile, (3.81). The Doppler width depends not at all upon pressure and only weakly on the temperature. Line intensities, on the other hand, can vary rapidly with temperature. It is, therefore, appropriate to identify $\phi(p, \theta)$ with the line intensity. For a finite spectral interval, an average line intensity is required. For weak lines, we shall demonstrate that a linear average is correct and it is usually adopted for all Doppler lines, either strong or weak,

$$\phi(p, \theta) = \sum_{i=1}^{N} S_i(\theta), \tag{6.25}$$

where the subscript, i, represents one of the N lines in the spectral interval.

Now consider the important case of the Lorentz profile. Equation (6.22) is not even approximately correct although we can obtain two different expressions independently for strong and for weak lines. Equations (4.41) and (4.42) are general expressions for the average absorption for arrays of independent strong and weak lines, respectively. These two equations can also be written for an optical path in a nonhomogeneous atmosphere. If we equate the homogeneous and

nonhomogeneous expressions we find

$$\sum_{i=1}^{N} \{S_i(\tilde{\theta})\alpha_{L,i}^0(\tilde{\theta})[\tilde{p}/p_0]\tilde{m}\}^{1/2} = \sum_{i=1}^{N} \left\{ \int_{\text{path}} S_i(\theta)\alpha_{L,i}^0(\theta)[p/p_0]\, dm \right\}^{1/2} \tag{6.26}$$

for strong lines, and

$$\sum_{i=1}^{N} S_i(\tilde{\theta})\tilde{m} = \sum_{i=1}^{N} \int_{\text{path}} S_i(\theta)\, dm \tag{6.27}$$

for weak lines. $\alpha_{L,i}^0$ is the line width at pressure p_0. Equation (6.27) is identical to (6.24) and (6.25), and is the method of averaging already chosen for weak Doppler lines.

We now consider the strong-line limit. If all the lines in the spectral interval are identical, (6.26) becomes

$$S(\tilde{\theta})\alpha_L^0(\tilde{\theta})\tilde{p}\tilde{m} = \int_{\text{path}} S(\theta)\alpha_L^0(\theta)p\, dm. \tag{6.28}$$

For an array of unequal lines we look for an expression analogous to (6.28), but with $S(\theta)\alpha_L^0(\theta)$ replaced by an average over all the lines in a spectral interval. The following has been proposed:

$$\psi(\tilde{\theta})\tilde{p}\tilde{m} = \int_{\text{path}} \psi(\theta)p\, dm, \tag{6.29}$$

where

$$\psi(\theta) = \left\{ \sum_{i=1}^{N} [S_i(\theta)\alpha_{L,i}^0(\theta)]^{1/2} \right\}^2. \tag{6.30}$$

Equation (6.30) has the correct form for identical lines. For lines of differing intensities, it gives each line a weight proportional to its equivalent width [see (4.16)]. This is a reasonable hypothesis. It was first made by W. L. Godson in an unpublished paper, but there appears to be no more rigorous derivation than has been given here. This is not to imply that the use of (6.30) will lead to errors in heating or flux calculations. Errors may well be very small, but no theoretical or numerical demonstration of its applicability has yet been given.

In § 6.4.2, we discuss "emissivity" techniques for evaluating directly quantities integrated over the entire thermal spectrum. Although pressure dependence is the same for all lines in the thermal spectrum, the temperature dependance can be in either direction in different spectral intervals. If we wish to use a scaling approximation with an emissivity calculation it is necessary to omit the temperature dependence from (6.27) and (6.29), with the results

$$\tilde{m} = \int_{\text{path}} dm \tag{6.31}$$

for weak lines and

$$\tilde{m} = \int_{\text{path}} (p/\tilde{p})\, dm \tag{6.32}$$

for strong lines.

The scaling approximation, applied to Lorentz lines, leaves us with the choice of two alternative procedures, one for weak lines and one for strong lines. Since strong lines are usually more important than weak lines, recent work has invariably made use of (6.28) or (6.29) with (6.30). Some of the early literature sought a compromise between these two possibilities. We may write both (6.31) and (6.32) in the form

$$\tilde{m} = \int_{\text{path}} (p/\tilde{p})^{\beta}\, dm, \tag{6.33}$$

where $\beta = 0$ for weak lines and 1 for strong lines. A common compromise was to set $\beta = \frac{1}{2}$.

The accuracy of scaling approximations, as is also the case for the more elaborate approximations that follow, can be judged only by means of numerical comparisons between fluxes and heating rates calculated with and without the approximation. For some gases, numerical calculations based on (6.29) with a fixed $\tilde{p}$ (e.g., 1 bar) have proved to be remarkably accurate, because the important lines are strongly absorbed. This need not always be the case, however.

6.2.3. *The H–C–G approximation*

We do not need to choose between strong- and weak-line limits if we allow two parameters to vary. Equations (6.24) and (6.29) can be solved for both $\tilde{m}$ and $\tilde{p}$. This procedure was proposed independently by van de Hulst, Curtis, and Godson. It ensures that a correct result is obtained for strong Lorentz lines, for Doppler lines, and for all weak lines.

We may repeat the results of the previous section in the form usually quoted for the H–C–G approximation,

$$\tilde{m} = \int_{\text{path}} \frac{\phi(\theta)}{\phi(\bar{\theta})}\, dm, \tag{6.24}$$

where

$$\phi(\theta) = \sum_{i=1}^{N} S_i(\theta), \tag{6.25}$$

and

$$\tilde{p} = \int_{\text{path}} \frac{\psi(\theta)}{\psi(\bar{\theta})} \frac{dm}{\tilde{m}}, \tag{6.29}$$

where

$$\psi(\theta) = \left\{ \sum_{i=1}^{N} [S_i(\theta)\alpha^0_{L,i}(\theta)]^{1/2} \right\}^2. \tag{6.30}$$

If we neglect temperature dependence, the H–C–G approximation becomes

$$\tilde{m} = \int_{\text{path}} dm = m, \tag{6.31}$$

and

$$\tilde{p} = \int_{\text{path}} p\, dm/\tilde{m}. \tag{6.32}$$

In this approximation, the amount of absorbing matter is unchanged, while an effective pressure is obtained by weighting pressures along the path by amounts. The effective pressure for a constant mixing ratio follows from (6.17), from which $dp \propto \rho\, dz = dm$; hence,

$$\tilde{p} = \tfrac{1}{2}(p_1 + p_2),$$

where p_1 and p_2 are the pressures at the two ends of the path.

The physical situation for pressure-broadened lines may be described as follows. A ray passing through a nonhomogeneous atmosphere forms an absorption line from the superimposition of many Lorentz profiles, which is inconsistent with any single Lorentz profile. The H–C–G approximation tells us how to choose the absorber amount and the line width so that the profile of a single Lorentz line corresponds as well as possible with the superimposed profile.

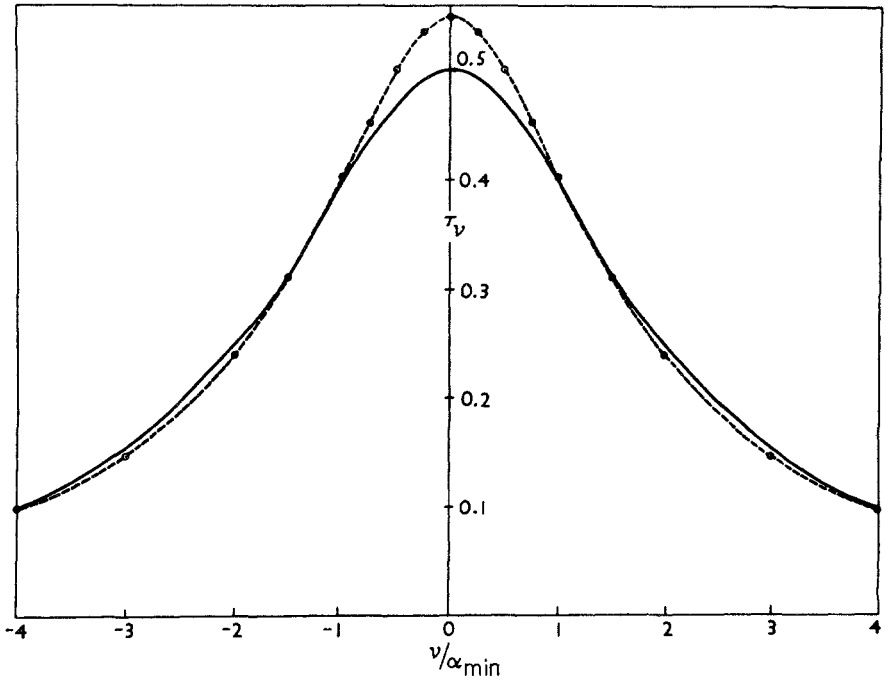


FIG. 6.3. Optical paths for an isothermal, mixed layer. The path is vertical, through an isothermal layer with a basal pressure three times that of the upper boundary. The absorbing gas is mixed evenly through the atmosphere. The broken line is an exact calculation from (6.18). The solid line is based on the H–C–G approximation.

Figure 6.3 illustrates the H–C–G approximation, applied to a mixed layer with a pressure ratio from bottom to top of 3:1. The exact result was obtained from (6.18). The approximation provides for the profiles to coincide in the wings (the strong-line limit) and assures that the areas are the same (the weak-line limit).

The H–C–G approximation is widely accepted as the best available technique for numerical models short of line-by-line calculations. Two studies of its effect upon heating rate calculations have reached the following conclusions.

For carbon dioxide, the H–C–G approximation is very accurate. Since carbon dioxide is the principal radiator in the mesosphere, this is an important result for middle-atmosphere calculations. Since the approximation is satisfactory for both Lorentz and Doppler lines, it also applies to the Voigt profile.

The approximation may, however, be unnecessarily complicated for mesopheric carbon dioxide. For very strong lines, the radiation fluxes at a particular level all originate from very close levels. For such paths, the H–C–G mean pressure is very close to the pressure of the level itself. For

very strong absorptions, we may avoid the effort of calculating the H–C–G mean pressure by substituting the pressure at the level of calculation.

For water vapor, heating rates calculated with the H–C–G approximation are satisfactory, but can be in error by about 10% for spectral regions in which the lines are neither strong nor weak. Such regions are more important for water vapor than for carbon dioxide.

Ozone is the least likely candidate for the use of the H–C–G approximation. The combination of large amounts of ozone at low stratospheric pressures with small amounts of ozone at high tropospheric pressures ensures that, for a ray traversing the whole atmosphere, the resulting line profile will differ greatly from that of a single Lorentz line. Errors of 10% in ozone heating rates are common.

There have been a number of attempts to improve the H–C–G approximation by introducing a third variable parameter; for example, the Doppler width in a Voigt profile may be treated as an independent variable for which a best fit is sought, or two Lorentz lines may be superimposed with a relationship expressed between the two intensities in order to reduce the number of variables from four to three. It is possible to obtain good results for ozone with any one of a number of three-parameter models. References to some treatments are given in the Bibliography. They all involve much more effort than the H–C–G approximation, and have, therefore, not proved to be attractive for atmospheric calculations.

Figure 6.4 shows calculations analogous to those of Fig. 6.1, except that band models and the H–C–G approximation were used in place of line-by-line calculations. The spectral intervals were 40 cm^{-1} wide. There are significant differences between observed and theoretical spectra, but these differences are not obviously larger than those shown in Fig. 6.1. It is possible that common errors arising from the meteorological and spectroscopic data exceed those arising from approximations in the calculation.

6.2.4. *Correlated k*

In this section we discuss extension of the k distribution technique to nonhomogeneous paths. Scaling approximations involve modifications to the amount of absorbing material without changing the frequency distribution of the absorption coefficient [see (6.22)]. The results can be as well applied to the k distribution as to band models. The k distribution is formed for a standard pressure and the amount of absorbing matter scaled by (6.27) or (6.29) for weak or strong lines, respectively; from the experience gained with band models, the strong-line scaling (6.29) is most useful.

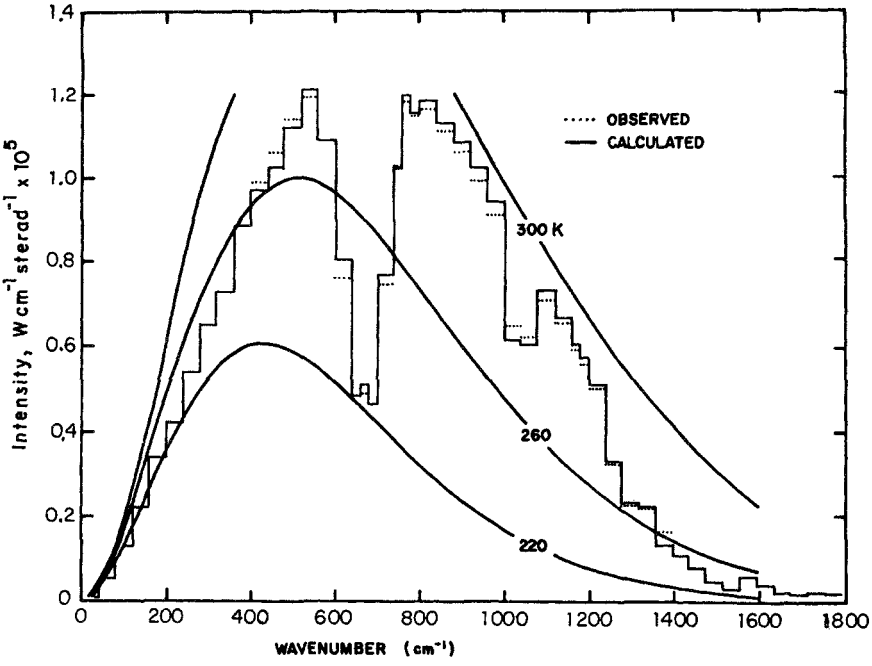


FIG. 6.4. Observed and theoretical spectra for clear skies over the Gulf of Mexico, June 1, 1969. The data were observed under similar conditions to those for Fig. 6.1. They have been averaged over 40 cm^{-1} spectral intervals for convenient comparison with calculations based upon band models and the H–C–G approximation. After Ellingson and Gille (1978).

There is, however, no direct analogy to the H–C–G approximation. The spectral information is scrambled in the k distribution method and there is no simple way to refer to the relationship between different parts of the same line without recalculating the k distribution.

The *correlated k* (c-k) method is a very effective alternative to the H–C–G approximation. It is also based upon exact results in certain asymptotic limits. Numerical calculations are required to establish the accuracy for intermediate cases.

The circumstances under which c-k gives exact results are

1. Whenever the similarity condition, (6.22), is valid, in particular, for Doppler lines and for strong Lorentz lines;
2. for weak absorption, regardless of the line shape;
3. for an isolated line, for the Schnaidt model, or for an Elsasser band, again, regardless of the line shape.

Conditions (1) and (2) are the same as those for the H–C–G approximation. The third condition additionally applies to intermediate absorptions that are neither strong nor weak and to any line shape. This

additional constraint upon errors appears to give the c-k method distinct advantages over the H–C–G approximation.

The c-k method employs the cumulative distribution function,

$$g(k) = \int_0^k f(k')\, dk' \tag{6.34}$$

[see § 4.8.1 for the definition of $f(k)$].

With this definition, the average transmission, (4.129), can be written

$$\bar{T}_i(m) = \int_0^1 \exp(-k_g m)\, dg. \tag{6.35}$$

g is a monotonic function of k, and the relationship between them can be inverted; this is the significance of the use of a g subscript for k_g. Typical $k \sim g$ relationships are shown in Fig. 6.5. The smooth curves are obtained by inverting the Malkmus model for the mean transmission, (4.138), for three widely spaced pressures. These curves are illustrative of the data that might be available for three atmospheric layers in a numerical calculation.

The division of the g axis in Fig. 6.5 into 10 intervals illustrates a computational advantage enjoyed by this technique. For these examples,

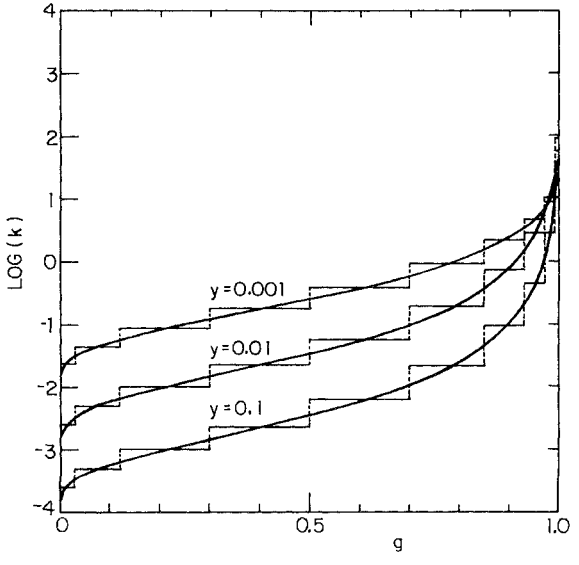


FIG. 6.5. Cumulative k distributions. The curves are obtained from the expression, (4.138), for the Malkmus model, with $\bar{k} = 1$. After Lacis and Oinas (1986).

high precision in computing the integral, (6.32), was obtained by numerical quadrature using these intervals of g only. Moreover, the same intervals may be used for all atmospheric layers. The use of g as an independent variable for integrations along nonhomogeneous paths is a useful numerical procedure, but its use must be justified on physical grounds, as we shall show to be possible.

The mean transmission over a spectral interval, $\Delta\nu_i$ (4.128), may be written in the form

$$\bar{T}_i(m) = \int_0^1 \exp(-k_\nu m)\, d(\nu/\Delta\nu_i). \tag{6.36}$$

The close analogy between (6.35) and (6.36) suggests that $\nu/\Delta\nu_i$ and g may be treated as interchangeable variables for homogeneous paths.

Turning now to nonhomogeneous paths, the physically appropriate independent variable is the radiation frequency, ν, since all of the processes that we consider are frequency coherent. Atmospheric problems involve the exchange of photons of a particular frequency between one layer and another. But, by working with frequency as an independent variable, we have been obliged to expend a lot of effort because the frequency spectrum of atmospheric gases is so complicated. If g can replace ν as an independent variable, an immense numerical simplification is possible, in which thousands of frequency intervals may be replaced by a few g intervals. The question is whether this is possible for a nonhomogeneous atmospheric path?

Write (6.36) in the form appropriate to a vertical atmospheric path in the discrete form of a sum over a number of atmospheric layers (subscript j)

$$\bar{T}_i = \int_0^1 \exp\left(-\sum_j k_{\nu,j}\,\Delta m_j\right) d(\nu/\Delta\nu_i). \tag{6.37}$$

The hypothesis underlying the c-k method is that, following the analogy between (6.35) and (6.37), we may write (6.37) in the form

$$\bar{T}_i = \int_0^1 \exp\left(-\sum_j k_{g,j}\,\Delta m_j\right) dg. \tag{6.38}$$

From our discussion of Fig. 6.5, it is clear that the c-k method will be very convenient, if justified. All that is then required for a numerical evaluation of (6.38) is a $g \sim k$ relationship for each atmospheric layer. It is unimportant how this relationship comes about; line shapes and temperature variations are important only to the degree that they influence this relationship. In practice, it may be convenient to use

analytic forms based upon the Lorentz profile, such as the Malkmus model, (4.138), but it is not necessary to do so.

We now examine the conditions under which (6.38) may be valid.

(i) A scaling approximation is obeyed. Equation (6.22) can be written in the discrete form, $k_{\nu,j} = k_{\nu,0} f_j$, where the subscript, 0, indicates one particular layer, and (6.37) becomes

$$\bar{T}_i = \int_0^1 \exp\left(-k_{\nu,0} \sum_j f_j \, \Delta m_j\right) d(\nu/\Delta\nu_i). \tag{6.39}$$

The equality between (6.35) and (6.36) holds for any value of m. Consequently, (6.39) can be written in the form

$$\bar{T}_i = \int_0^1 \exp\left(-k_{g,0} \sum_j f_j \, \Delta m_j\right) dg, \tag{6.40}$$

where $k_{\nu,0}$ and $k_{g,0}$ apply to the same atmospheric layer. It is straightforward to show that $k_{g,j} = k_{g,0} f_j$, so that (6.40) is identical to (6.38); the c-k method is correct whenever conditions for a scaling approximation occur, in particular, for strong Lorentz lines and for Doppler lines (see § 6.2.2).

(ii) In the weak-line limit. From (4.123), (4.124), and (6.34), the following relationship is true for any analytic function, G,

$$\int_0^1 G(k_\nu) \, d(\nu/\Delta\nu_i) = \int_0^1 G(k_g) \, dg. \tag{6.41}$$

Expand (6.37) and (6.38) term by term,

$$\bar{T}_i = \int_0^1 \left[1 - \sum_j k_{\nu,j} \, \Delta m_j + \frac{1}{2!}\left(\sum_j k_{\nu,j} \, \Delta m_j\right)^2 - \frac{1}{3!}\left(\sum_j k_{\nu,j} \, \Delta m_j\right)^3 + \cdots\right] d(\nu/\Delta\nu_i), \tag{6.42}$$

$$\bar{T}_i = \int_0^1 \left[1 - \sum_j k_{g,j} \, \Delta m_j + \frac{1}{2!}\left(\sum_j k_{g,j} \, \Delta m_j\right)^2 - \frac{1}{3!}\left(\sum_j k_{g,j} \, \Delta m_j\right)^3 + \cdots\right] dg. \tag{6.43}$$

The linear terms in (6.42) and (6.43) are the terms that occur in the weak-line limit. According to (6.41) they are identical. For any weak-line condition, the c-k method is exact.

Equation (6.41) also implies that some higher order terms in (6.42)

and (6.43) are equal since, for example,

$$\int_0^1 (k_{\nu,j}\,\Delta m_j)^n\, d(\nu/\Delta\nu_i) = \int_0^1 (k_{g,j}\,\Delta m_j)^n\, dg. \tag{6.44}$$

The validity of the c-k method extends over a wider range of conditions than does the weak-line approximation.

(iii) Regular and Schnaidt models. For an isolated line, or for a regular band, we shall show that

$$dg = d(\nu/\Delta\nu_i), \tag{6.45}$$

from which it follows that the suffices ν and g can be interchanged and that (6.37) and (6.38) are identical.

For the two situations shown in Fig. 6.6, k_ν is a monotonic function of ν between a line center and $\delta/2$ from the center; according to (4.126), we may write

$$f(k) = \frac{2}{\delta}\left|\frac{dk_\nu}{d\nu}\right|^{-1}. \tag{6.46}$$

From (6.34)

$$g(k) = \int_{k_{\min}}^{k} \frac{2}{\delta}\left|\frac{dk_\nu}{d\nu}\right|^{-1} dk_\nu, \tag{6.47}$$

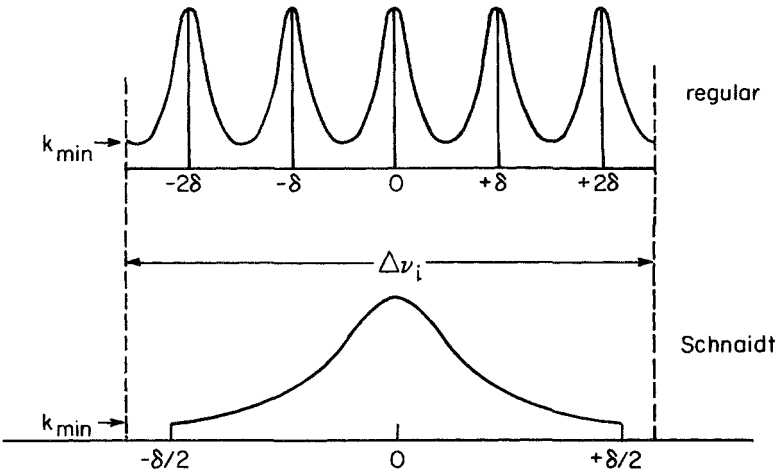


FIG. 6.6. Regular and Schnaidt models. If the Schnaidt line is wholly within the spectral interval $\Delta\nu_i$, or if many of the regular band lines occur in the same interval, an average over the band is, apart from a multiplying factor, equivalent to an average over the range 0 to $\delta/2$.

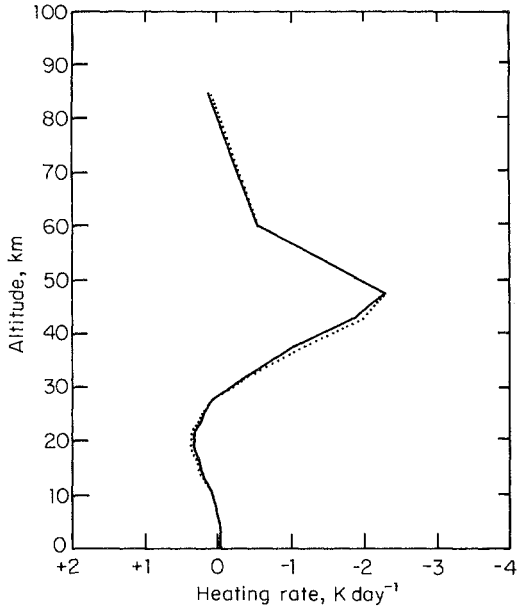


FIG. 6.7. Heating rates for ozone (975–1175 cm^{-1}). The solid line is the result of a line-by-line calculation. The dotted line uses the c-k method. After Lacis and Oinas (1986).

where k_{min} is the minimum value of k_ν. If the integrand in (6.47) has no singularities,

$$g(k) = |\nu(k) - \nu(k_{min})|\, 2/\delta, \tag{6.48}$$

and (6.45) is satisfied exactly (the sign is unimportant).

There is a question of singularities associated with zeros of $dk_\nu/d\nu$ (one for the Schnaidt model and two for the regular model). The integral may be carried as close to them as we please. Equation (6.48) applies over all but a negligible part of the frequency domain.

The comparisons with line-by-line calculations in Figs. 6.7 and 6.8 show that the c-k method is remarkably accurate, as well as being convenient for numerical methods. The agreement is particularly good for ozone, which gave difficulties for the H–C–G approximation. The region of disagreement above 60 km in Fig. 6.8 is concerned with a single atmospheric level.

6.3. Topics concerning heating rates

We have discussed the steps necessary to calculate fluxes and heating rates in an absorbing, stratified atmosphere. It is usual for such

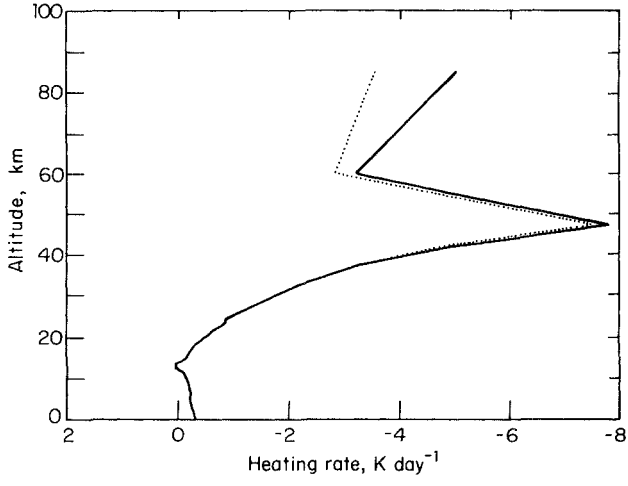


FIG. 6.8. Heating rates for carbon dioxide (500–1000 cm^{-1}). The solid line is the result of a line-by-line calculation. The dotted line uses the c-k method. Only one data point is used above 60 km. After Lacis and Oinas (1986).

calculations to be made numerically, but with different approximations depending upon the choice and resources of the investigator; comments on some investigations are made in the Bibliography.

In the remainder of this section we shall discuss some special topics; in § 6.4 we shall consider drastic approximations that change the nature of the investigation.

6.3.1. *The Chapman layer*

Here we consider the transmission of solar radiation through an atmosphere in which only absorption occurs. The thermal source function can be neglected for wavelengths less than 3 or 4 μm. There is, however, an observable amount of scattering in the atmosphere, even for a clear sky, and the scattering source function cannot be omitted without further discussion.

We shall show in Chapter 9 that, in an atmosphere that both absorbs and scatters, scattering may be neglected if the volume absorption coefficient is not small compared to the volume scattering coefficient. When the optical depth for absorption is larger than the optical depth for scattering, it is usually satisfactory to treat the absorption alone. For two important circumstances this condition is satisfied: absorption by ultraviolet bands in the upper atmosphere and absorption by the near-infrared bands of water vapor in the troposphere.

We consider the first component of the irradiance Stokes vector, (2.115). The energy flux (negative) is the product of the irradiance

(positive) and the direction cosine, ξ_0 (negative). In terms of the spectral intervals, $\Delta\nu_i$, introduced in § 6.1.3,

$$F = \sum_i F_i = -\sum_i |\xi_0| f_i(0) \int_{\Delta\nu_i} \exp(-\tau_\nu/|\xi_0|)\, d\nu. \tag{6.49}$$

We have introduced a change of sign by using the absolute value, $|\xi_0|$, because it is common practice to do so in the atmospheric literature (the reader should be warned, however, that the literature is full of sign inconsistencies). Equation (6.49) can be solved numerically but it is important also to be familiar with a simple analytic approximation introduced by Chapman in his early investigations of the ionosphere.

If we calculate the heating rate, (2.10) from (2.115),

$$h_\nu(z) = f_\nu(0) k_{\nu,\nu}(z) \exp\left[-\int_z^\infty k_{\nu,\nu}(z')\, dz'/|\xi_0|\right]. \tag{6.50}$$

Now assume that the molecular absorption coefficient is constant and that the number density of absorbers follows a barometric law with scale height, H,

$$k_{\nu,\nu} = n(z) k_{n,\nu} = n(z') k_{n,\nu} \exp[-(z - z')/H]. \tag{6.51}$$

The heating rate has a maximum value h_{max} at a level z_{max} where the number density is n_{max}. After some manipulation,

$$\frac{h(z)}{h_{max}} = \exp\left[1 - \frac{(z - z_{max})}{H}\right] \exp\left[-\exp\frac{-(z - z_{max})}{H}\right], \tag{6.52}$$

$$h_{max} = \frac{f(0)\,|\xi_0|}{eH}, \tag{6.53}$$

$$n_{max} = \frac{|\xi_0|}{k_n H}. \tag{6.54}$$

The ν suffices are omitted. The larger the absorption coefficient and the larger the zenith angle (the smaller $|\xi_0|$) the smaller is n_{max} and the higher is the Chapman layer.

Equation (6.52) is shown in graphical form in Fig. 6.9. The figure illustrates the important qualitative result that monochromatic solar radiation is deposited in a layer whose thickness is approximately two scale heights, the height of which varies with the absorbing molecule and the incidence angle of the solar beam. When a range of values of the absorption coefficient is involved in the absorption, the heating function will be a superimposition of a number of elementary Chapman layers and

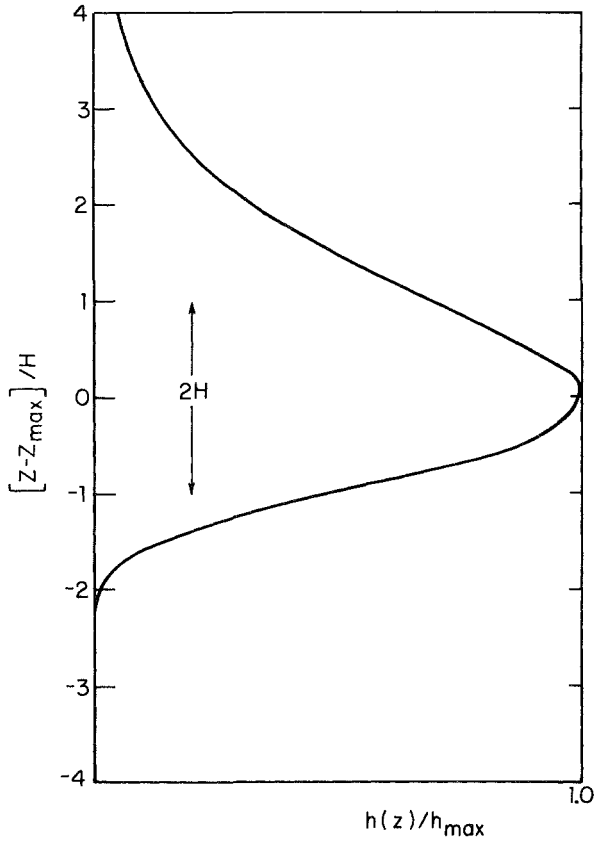


FIG. 6.9. A Chapman layer. In this example, z_{max} is far above the earth's surface.

will probably also exhibit a layered structure, but wider than that illustrated in Fig. 6.9.

6.3.2. *The Curtis matrix*

The flux equations, (6.11) and (6.12), are stated in terms of continuous variables but, for numerical applications, the integrals are replaced by discrete sums. The number of vertical grid points must be kept to a minimum and will usually be set by the requirements of some other feature of the calculation, e.g., by the dynamic equations.

The Curtis matrix offers the best approach to this numerical problem, paticularly when the mean free path of the radiation is small compared to the grid spacing. It is a useful technique under other circumstances but it is the only self-consistent treatment for opaque atmospheres.

The numerical difficulty with opaque atmospheres lies in the cancellation between upward and downward flux components [(2.107) and (2.108)] and the corresponding components of heating rates. Since the upward and downward components have different boundary conditions, they are often calculated independently and care must be taken to give consistent treatments to the canceling terms.

The importance of this question is illustrated by (2.127), for the opaque approximation. The heating rate is proportional to the curvature of the Planck function. Much larger terms, proportional to the Planck function (for fluxes) or its first derivative (for heating rates), must cancel between the upward and downward integrals. In order to calculate the second derivative to the Planck function, we must use data points both above and below the level of calculation. If the two halves of the integral are approximated independently, there is no guarantee that this information will be correctly incorporated.

Even though the Curtis matrix provides the optimum solution to this problem, it is not generally used in weather prediction or general circulation models. Terrestrial calculations can obtain sufficient accuracy from independent upward and downward flux calculations because the most important contributions to radiative heating come from translucent spectral regions. The reason will be further discussed in § 6.4.1. For atmospheres with no translucent regions, such as the atmospheres of Venus or the outer planets, the Curtis matrix is irreplaceable.

We start from (2.111), with the altitude included explicitly

$$H_i(z) = \left(\int_i k_{\nu,\nu}\, d\nu \right)^{-1} \int_i k_{\nu,\nu}(z) \left\{ \int_0^1 [J_\nu(z') - J_\nu(z)] \times dE_2[\tau_\nu(z') - \tau_\nu(z)] + \int_0^1 [J_\nu(z') - J_\nu(z)] \times dE_2[\tau_\nu(z) - \tau_\nu(z')] \right\} d\nu. \tag{6.55}$$

E_2 does not exist over the entire range of both integrals in (6.55), but does so if we define

$$J_\nu(z') = B_\nu(\theta^*)$$

for

$$E_2[\tau_\nu(z') - \tau_\nu(z)] \leq E_2[\tau_\nu(0) - \tau_\nu(z)], \tag{6.56}$$

and

$$J_\nu(z') = 0$$

for

$$E_2[\tau_\nu(z) - \tau_\nu(z')] < E_2[\tau_\nu(z)]. \tag{6.57}$$

This statement bears the same relationship to (2.111) that (2.87) does to (2.86).

We now introduce the assumption that J_ν is constant over the ith spectral interval, and we define a new function

$$K_i(z', z) = \left[\int_i k_{\nu,\nu}(z)\, d\nu\right]^{-1} \int_i k_{\nu,\nu}(z) E_2[\tau_\nu(z') - \tau_\nu(z)]\, d\nu. \tag{6.58}$$

$K_i(z', z)$ is proportional to the derivative of the flux transmission (§ 6.1.2). With this definition, (6.55) becomes

$$H_i(z) = \int_0^1 [J_i(z') - J_i(z)]\, dK_i(z', z) + \int_0^1 [J_i(z') - J_i(z)]\, dK_i(z, z'). \tag{6.59}$$

Equation (6.59) displays the relationship between heating function and source functions in the simplest possible form.

Any numerical approximation to an integral gives its value as a linear function of the values of the integrand at a chosen set of points. For the sake of discussion, we identify these points with the integers (r, s), equal to $-5 \log_{10}[p(z)/p(0)]$, where $p(0)$ is 1 bar. The two integrals in (6.55) can now be combined into a single sum

$$H_i^{(r)} = \sum_s A_i^{(r,s)} J_i^{(s)}. \tag{6.60}$$

Since pressure is the independent variable, the matrix elements $A_i^{(r,s)}$ are functions of absorber densities and temperature only. Their values depend on the chosen method of numerical integration, but they combine consistently data from above and from below the reference level, and any cancellation of terms is correctly taken into account when the $A_i^{(r,s)}$ are evaluated.

Table 6.1 illustrates the use of one column of a Curtis matrix. Any source functions may be entered into the third column; the heating function is then the sum of the fourth column. The entry marked "space" is the sum of the matrix elements and can be shown to equal $-K_i(z, \infty)$. Multiplied by J_i(2.4), it gives the contribution to the heating function arising from the exchange of radiation between the reference level and space. Since it has already been included in the calculation, it is assigned

Table 6.1. Curtis' scheme for CO_2 illustrated for one level[a]

s	$A_i^{(r,s)}$	$J_i^{(s)}$	$A_i^{(r,s)} \times J_i^{(s)}$
0	+0.0499	$J_i(0)$	+ ———
1	+0.0230	$J_i(-0.2)$	+ ———
2	+0.0364	$J_i(-0.4)$	+ ———
⋮	⋮	⋮	+ ⋮
10	+2.0142	$J_i(-2.0)$	+ ———
11	+5.8745	$J_i(-2.2)$	+ ———
12 ($=r$)	−23.3593	$J_i(-2.4)$	− ———
13	+3.7395	$J_i(-2.6)$	+ ———
14	+0.8185	$J_i(-2.8)$	+ ———
⋮	⋮	⋮	+ ⋮
19	+0.0113	$J_i(-3.8)$	+ ———
⋮	⋮	⋮	+ ⋮
Space	−7.9228	0	0

[a] The reference level is $r = 12$. The argument of the source function is $\log_{10}[p(z)/p(0)]$.

zero source function. This *radiation-to-space* term is discussed in more detail in § 6.4.1.

The Curtis matrix separates the source function from the transmission in a heating calculation. To the degree that the matrix elements do not change from one set of circumstances to another, it provides an economical and accurate means of calculation, and it is often used for carbon dioxide calculations in the middle atmosphere. Carbon dioxide is usually assumed to have a fixed mixing ratio and gaseous densities can be taken into account in the matrix coefficients, once and for all. Temperature influences the transmission, however, and were this effect large it would require recalculation of the matrix elements for each different distribution of temperature, in which case the value of the method would be reduced. Fortunately, as is illustrated by the computations shown in Fig. 6.10, the first-order effect of temperature is on the source function rather than the transmission. For carbon dioxide, accurate interpolation schemes have been developed based upon calculations for a few standard distributions of atmospheric temperature.

6.3.3. *Calculations for the middle atmosphere*

The calculations in Fig. 6.10b use the two-level source function (2.75). For carbon dioxide below 75 km, this source function is effectively the Planck function, but at higher levels it is given by (2.79), an incoherent scattering source function, exhibiting no net heating. To examine this situation further we consider (2.74), with the suffix i used to denote the

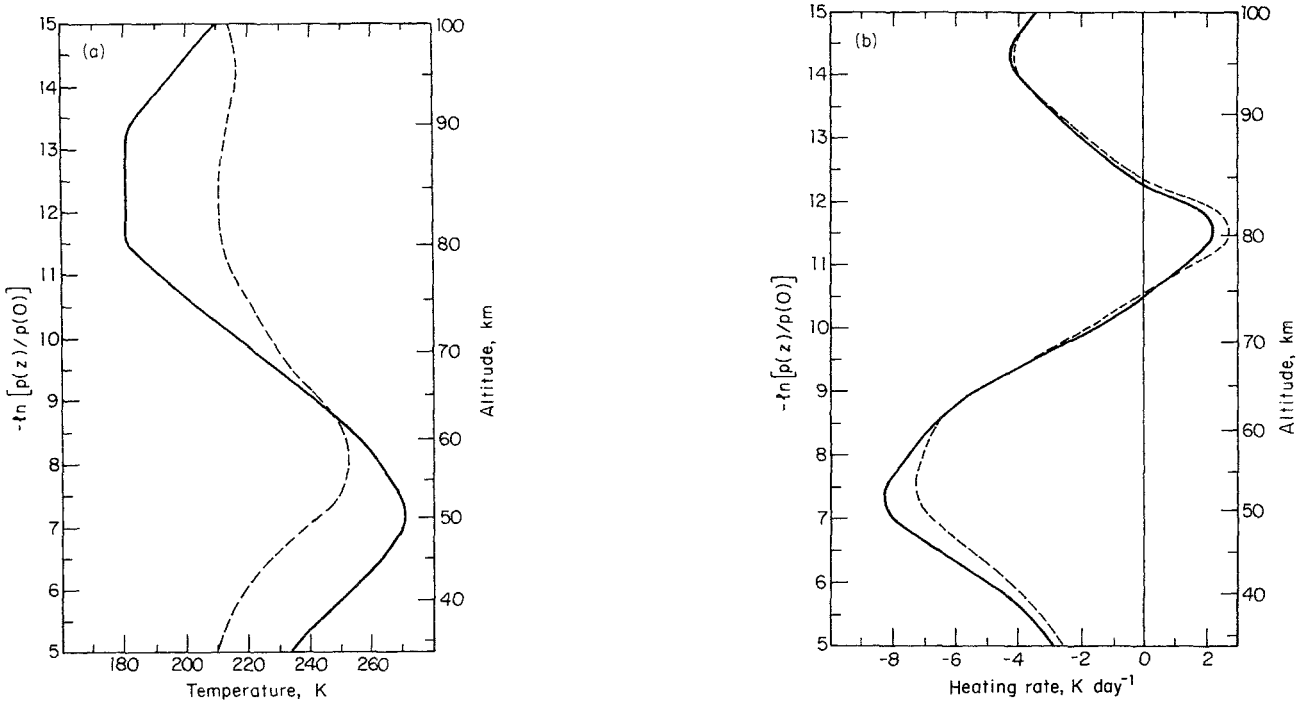


FIG. 6.10. Illustrating the effect of temperature on Curtis matrix elements. (a) *Assumed temperatures* (the height scale applies only to the U.S. Standard Atmosphere). Solid line: U.S. Standard Atmosphere (1962). Broken line: arctic night, 70°N, 1 January. (b) *Heating rates*. Solid line: source function, U.S. Standard Atmosphere; transmission, U.S. Standard Atmosphere. Broken line: source function, U.S. Standard Atmosphere; transmission, arctic night. The heating rates closely reflect the variation of source function with height but are little influenced by the effect of the large change of temperature on the transmission. After Williams (1971).

(u, l) band,

$$J_i(z) = B_i(z) + [\eta_i(z)/2\phi_i]H_i(z).$$

If we substitute (2.74) into (6.60) we find

$$H_i^{(r)} = H_{i,0}^{(r)} + \sum_s A_i^{(r,s)}[\eta_i^{(s)}/2\phi_i]H_i^{(s)}, \tag{6.61}$$

where $H_{i,0}^{(r)}$ is the heating function that would result if the source function were the Planck function at every level. $H_{i,0}^{(r)}$ can be calculated, given the temperature at each level, and the heating function can be obtained from (6.61) by means of a matrix inversion. To do so requires a good first approximation with the correct asymptotic properties at high levels. This is provided by the radiation-to-space term

$$A_i^{(r)}\,(\text{space}) = -K_i(r, \infty), \tag{6.62}$$

giving, as a first approximation,

$$H_{i,1}^{(r)} = H_{i,0}^{(r)}\{1 + [\eta_i^{(r)}/2\phi_i]K_i(r, \infty)\}^{-1}. \tag{6.63}$$

Table 6.2 shows an early calculation of the heating function for thermodynamic equilibrium, (H_0), the first approximation, (H_1), and the actual heating function, (H), for the 15 μm CO_2 band. The interesting comparisons are between the source function and the Planck function and between the three heating functions. There are only small differences, as expected, below 75 km. Above this level the two source functions, and the heating functions derived from them, diverge rapidly. The heating rate, H, tends to zero at high levels in marked contrast to H_0. The first approximation, H_1, provides a remarkably good approximation at all levels, although it is expected to do so only at the highest levels.

The two-level source function offers valuable insights into the effects of the breakdown of thermodynamic equilibrium among vibrational levels. Elaboration of the model to include other departures from thermodynamic and thermochemical equilibrium leads gradually into a different class of studies from those considered in this book, because lower atmosphere studies (which we emphasize) usually take equilibrium states for granted. The constraints imposed by equilibrium assumptions are very powerful; chemical species are conserved, and only macroscopic quantities such as the temperature, pressure, and the density need be included in the thermal and dynamic equations.

At the opposite extreme are aeronomical studies that are specifically concerned with ionized, excited, and disequilibrium atomic and molecu-

Table 6.2. Middle atmosphere calculations for the 15 μm CO_2 band[a]

$\log_{10} p$ (dyne cm^{-2})	z (km)	πB	$-\pi H_0$	$-\pi H_1$	$-\pi H$	πJ
3.0	49.2	329.1	5.4	5.4	5.4	329.1
2.8	52.9	317.3	5.0	5.0	5.0	317.3
2.6	56.5	291.5	4.2	4.2	4.2	291.5
2.4	60.1	252.0	3.2	3.2	3.2	252.0
2.2	63.4	208.4	2.2	2.2	2.2	208.2
2.0	66.6	169.4	1.4	1.4	1.4	169.2
1.8	69.8	139.0	0.7	0.7	0.7	138.9
1.6	72.7	121.6	0.6	0.6	0.8	121.2
1.4	75.6	111.6	1.3	1.3	2.0	110.2
1.2	78.5	105.2	2.3	2.2	3.8	101.0
1.0	81.3	101.5	4.2	3.7	5.9	91.2
0.8	84.2	99.9	6.8	5.2	7.3	79.6
0.6	87.1	101.5	12.2	7.0	7.7	67.6
0.4	90.1	106.4	21.6	7.9	7.2	56.4
0.2	93.1	113.1	35.0	7.3	6.0	47.1
0.0	96.3	121.9	54.2	6.1	4.7	39.7
−0.2	99.6	132.9	(73.8)	(4.5)	(3.7)	(31.4)
−0.4	103.0	144.5	(94.1)	(3.2)	(2.7)	(24.9)
−0.6	106.5	159.2	(116.9)	(2.2)	(2.0)	(19.0)
−0.8	110.1	185.7	(149.0)	(1.7)	(1.5)	(15.1)
−1.0	114.1	227.3	(194.5)	(1.3)	(1.2)	(11.8)
−1.2	118.4	299.0	(268.0)	(1.1)	(1.0)	(9.1)

[a] B is the Planck function at 667 cm^{-1}. The units of B, H, and J are erg cm^{-2} s^{-1} steradian^{-1} wavenumber^{-1}. In these units, $\pi H = 1$ corresponds to a temperature change of approximately 1.1 K day^{-1}. Values in parentheses depend upon doubtful values of the CO_2 concentration. The Planck functions are based upon temperatures that differ somewhat from those in Fig. 6.10a. After Curtis and Goody (1956).

lar species, the main cause of which is the absorption of high-energy photons from the sun. Solar photons, particularly those with wavelengths less than 2000 Å, can dissociate molecules, ionize atoms and molecules, and excite electronic, vibrational, and rotational levels to disequilibrium distributions. In an equilibrium context, absorbed solar energy is transformed entirely into translational energy, the state of which is defined by the temperature. An aeronomical perspective considers instead a collisionally interacting ensemble of electrons, neutral and ionized atoms, and molecules in all modes of electronic, vibrational, and rotational excitations, in a steady state with a flux of photons of many wavelengths. The species possess translational energies that, for reasons discussed in § 2.2.2, may be close to thermodynamic equilibrium at a

kinetic temperature. Given this situation, the microscopic steady state can be expressed in terms of the law of mass action applied to all possible interactions, with temperature-dependent rate coefficients. Given the rate coefficients, a specification of all possible reaction paths, and a large enough computer, the steady-state equations can usually be solved. There is, however, sufficient controversy about some of the data that investigators often differ significantly, and the complexity of the calculation may make the origin of discrepancies hard to trace.

If we are concerned with a specific transition, such as the (u, l) transition in Fig. 2.4, the aeronomical calculations are required to determine the populations of the "other levels" and hence the transitions to the upper and lower levels, other than those exclusively between the two. Given these transitions, we may then use a radiative transfer formulation by defining the source function from (2.61). This source function is defined in terms of the two-level populations; of these, the upper is usually the most variable and there is some equivalence between the source function and the upper state population.

A systematic discussion of all of these issues is not yet possible and would raise many questions outside the scope of this book. Instead, we briefly discuss a few results from a recent study of radiative transfer in the ν_2 fundamental of CO_2, including all upper-state transitions and isotopic bands.

The interacting levels chosen for this particular study are shown in Fig. 6.11. On the left are the ν_2 levels. The central region shows ν_3 levels, which may be excited by solar absorption. On the right are levels of N_2, O_2, and H_2O that are also involved only under daytime conditions.

Figure 6.12 shows the nighttime heating rate broken down into ground-state, upper-state, and isotopic contributions. Below 70 km, all transitions have the same source function (i.e., the same ratio of upper to lower state populations). Above 70 km, the source functions differ, but only transitions connected with the ground state are important. The same source function could be used for all levels without error.

Figure 6.12 illustrates a question of crucial importance for middle-atmosphere calculations, namely their sensitivity to the presence of atomic oxygen. When atomic oxygen is abundant it dominates collisional excitation and deexcitation of the ν_2 band, probably because of its chemical reactivity; in its absence the heating rate differs greatly (see the two curves in Fig. 6.12, calculated with and without atomic oxygen). The presence of atomic oxygen does not bring about thermal equilibrium, however, and the heating rate depends critically on poorly known rate coefficients.

The absorption of solar radiation can affect the ν_2 level populations by transitions in ν_3 levels and levels of N_2, O_2, and H_2O (the intensity of solar radiation at 15 μm is, in itself, negligible). The net effect upon the ν_2 transitions is, however, small. A more important question is whether

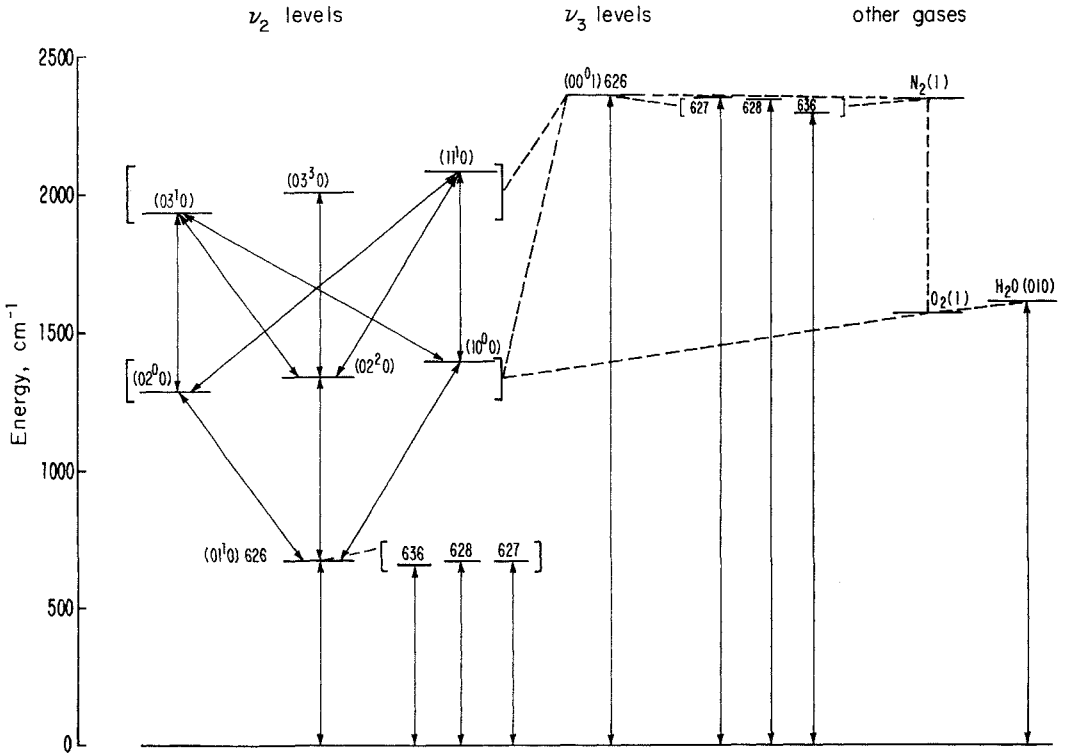


FIG. 6.11. Communicating states for CO_2, ν_2 transitions. The solid lines indicate radiating transitions (they can also be caused by collisions). The broken lines are nonradiating interactions with other excitations or species. The isotope code is $^{16}O^{12}C^{16}O$, (626); $^{16}O^{13}C^{16}O$, (636); $^{16}O^{12}C^{18}O$, (628). The left side of the diagram shows the ν_2 fundamental bands ($\Delta v_2 = \pm 1$). The ground-state transition (010) $\leftarrow$ (000) is unique, but the upper-state bands involve Fermi triplets. After Lopez-Puertas et al. (1986).

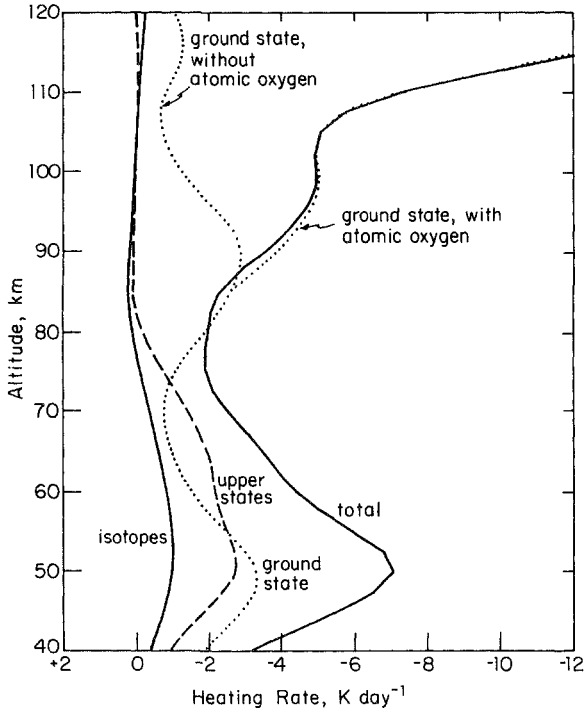


FIG. 6.12. Heating rates for CO_2, ν_2 transitions, nighttime conditions. The total heating rate is broken down into contributions from ground-state, upper-state, and isotopic transitions. A calculation of the ground-state contribution in the absence of atomic oxygen is also shown. The differences between these results and those in Table 6.2, above 100 km, are principally attributable to different temperatures used in the two calculations. After Lopez-Puertas et al. (1986).

all absorbed solar radiation ends up as translational energy, as has been tacitly assumed in previous chapters, or whether the energy is partially reemitted from other excited energy levels? The latter turns out to be very important. Figure 6.13 shows that, above 80 km, most of the radiation absorbed in the important 2.7 μm band ($\nu_1 + \nu_3$) is reemitted in ν_2 and ν_3 bands.

6.4. Approximate methods

6.4.1. *Exchange of radiation with the boundaries*

The heating rate expression, (6.59), exhibits the relationship between the heating at a level, z, and the conditions at another level z'. From the symmetry of this equation, the mutual effect of two atmospheric levels is equal but opposite; if $J_i(z) = J_i(z')$, their mutual effect is zero.

A contribution to each integral also comes from the exchange of

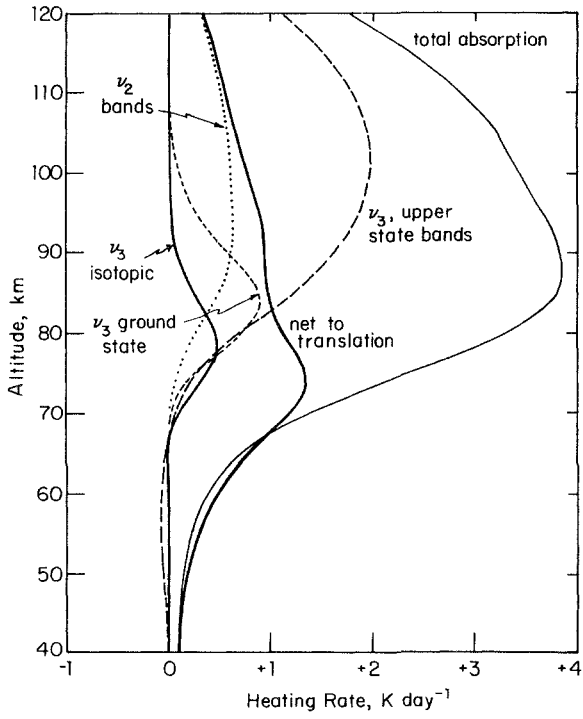


FIG. 6.13. Redistribution of solar energy absorbed by CO_2 at 2.7 μm. The 2.7 μm band is the $\nu_1 + \nu_2$ transition. The curves marked ν_2 and ν_3 show reemissions in these bands. The net energy that goes into translation is shown by the heavy curve. After Lopez-Puertas et al. (1986).

radiation between z and the boundaries. We may isolate these terms by setting $J_i(z') = J_i(z)$ at all levels in the atmosphere. From (6.56), (6.57), and (6.59), with a thermal source function,

$$H_i(z) \simeq [B_i(\theta^*) - B_i(z)]K_i(0, z) - B_i(z)K_i(z, \infty). \tag{6.64}$$

The first term on the right-hand side involves exchange of radiation with the lower boundary. Its sign depends on the relationship between the ground temperature and the temperature at z. If the atmosphere is colder than the surface it will be positive. The second term on the right-hand side involves exchange of radiation with space. It is always negative and usually larger in absolute value than the first term, partly because $B_i(z)$ and $B_i(\theta^*)$ may not differ greatly and partly because $K_i(z, \infty)$ exceeds $K_i(0, z)$, except near to the ground. The importance of the term involving $K_i(z, \infty)$ has already been mentioned in connection with the calculation in Table 6.2. Equation (6.64) is an important approximation

to the heating rate equations, which we may call the *boundary exchange approximation.*

If we use the definitions (2.21), (6.7), (6.58), and a property of exponential integrals (Appendix 6), the surface and space terms may be separately written

$$h_i(z)_{\text{surf.}} = -\pi[B_i(\theta^*) - B_i(z)]\frac{d\bar{T}_i^{\text{f}}(0, z)}{dz}, \tag{6.65}$$

$$h_i(z)_{\text{space}} = +\pi B_i(z)\frac{d\bar{T}_i^{\text{f}}(z, \infty)}{dz}. \tag{6.66}$$

These expressions can be extended to include the effect of an opaque cloud. If the lower surface of the cloud is at a level $z_c > z$, and if it emits with a Planck function $B_i(c)$, (6.66) becomes

$$h_i(z)_{\text{cloud}} = -\pi[B_i(c) - B_i(z)]\frac{d\bar{T}_i^{\text{f}}(z, z_c)}{dz}. \tag{6.67}$$

Equation (6.66), used by itself, is the *cooling-to-space approximation.* As a simplification to the equation of transfer for the earth's atmosphere, its importance cannot be overstated. Its accuracy in one context has been demonstrated by the agreement between H and H_1 in Table 6.2. Comparisons with exact calculations for carbon dioxide, ozone, and water vapor under a number of conditions are shown in Fig. 6.14. The agreement between exact and approximate calculations shown in Table 6.2 and in Fig. 6.14, is imperfect but remarkably good in view of the simplicity of the cooling-to-space approximation. The most serious disagreement in Fig. 6.14 is near to the ground under mid-winter, arctic conditions. This is hardly surprising, for there is a very large ground inversion in the arctic winter. The data employed show a ground temperature of −45.8°C and an 850 mb temperature of −19.5°C. In addition, the lower atmosphere is unusually transparent and the surface term (6.65) must be large; it was not, however, included in the calculation.

The importance of radiative exchange with the boundaries can be related to our discussion of transparent and opaque conditions in § 2.4.2. It is assumed in the derivation of the opaque approximation that the radiation field originates near the point of reference and that none comes from the boundaries. Such opaque conditions are encountered for all wavelengths in the lower atmosphere of Venus, for example, and the approximations (6.65) and (6.66) are simply irrelevant. In many spectral regions, the terrestrial atmosphere is also opaque. The conclusion to be drawn from the computations shown in Fig. 6.14 is that it is the relatively

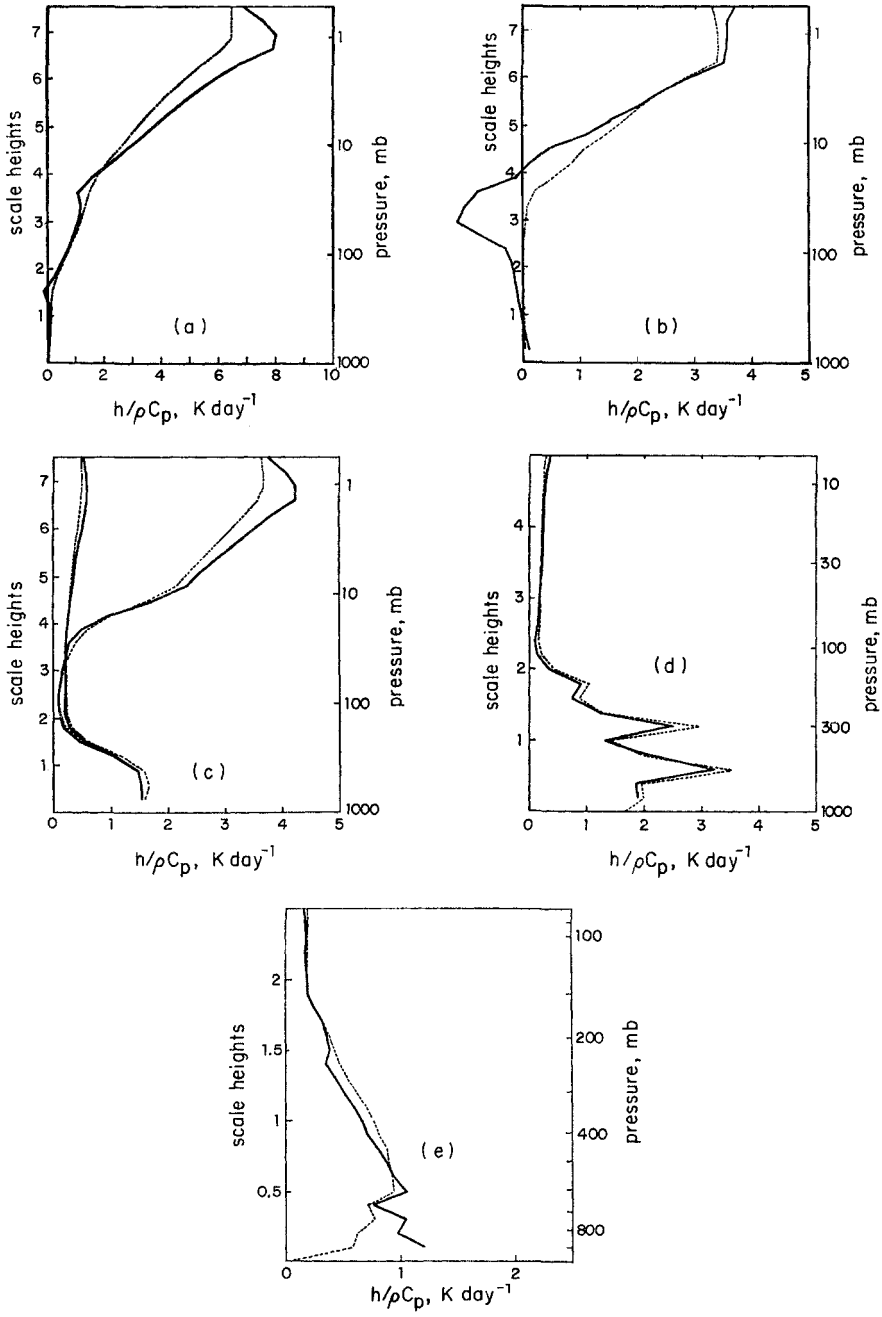


FIG. 6.14. Cooling to space compared to total cooling. The solid lines include all terms in the heating equation; the dotted line is calculated from the approximation (6.66). The vertical scale is $-\ln[p(z)/p(0)]$. (a) CO_2 15 μm band for a mid-latitude temperature profile. (b) O_3 9.6 μm band for a tropical temperature profile. (c) H_2O for a mid-latitude temperature profile with *wet* and *dry* stratospheres. (d) H_2O for a tropical temperature profile. (e) H_2O for a mid-winter, arctic temperature profile. For details of the temperature and water vapor distributions, see Rodgers and Walshaw (1966).

transparent regions of the spectrum that contribute most strongly to the heating rate.

Consider the expressions (2.120) and (2.127), for heating under the transparent and opaque approximations, respectively. From the form of their dependence on the absorption coefficient, it is anticipated that regions of intermediate opacity should contribute most strongly to the heating. For the present discussion it is the opacity to the boundaries that is important. Since atmospheric gases possess a wide range of absorption coefficients there are always some spectral ranges for which the optical path to the surface or to space is of the order of unity, and it is these regions that contribute strongly to heating in the earth's atmosphere. The particular wavelengths involved will change with the pressure, the temperature, and the reference level, but, since the transparent approximation always has the same form, the only variable will be a multiplying factor defining the extent of the nearly transparent region of the spectrum; that, in qualitative terms, is the significance of the derivative of the flux transmission in the approximations (6.65), (6.66), and (6.67).

A quantitative comparison between radiation-to-space and the transparent approximations is possible if we write (6.66) in monochromatic form and allow the optical depth to tend to zero. From the definition of the flux transmission, (6.7), and properties of the exponential integrals (Appendix 6),

$$h_\nu(z)_{\text{space}} = -2\pi B_\nu(z) k_{\nu,\nu}(z). \tag{6.68}$$

If we apply a small perturbation to the Planck function, (6.68) becomes one-half of (2.120); the other half comes from the surface contribution, (6.65).

The approximations (6.65) and (6.66) are of great importance for hydrodynamic theories because they make it possible to express a radiative heating perturbation linearly in terms of the local temperature perturbation (*Newtonian cooling,* see also § 2.4.2 and § 10.1). This relationship can be expressed in terms of a *radiative relaxation time,* t_{rad},

$$\delta h_i(z) = \frac{\rho c_p \delta\theta(z)}{t_{\text{rad}}}, \tag{6.69}$$

where

$$t_{\text{rad}}^{-1} = \frac{\pi dB_i/d\theta}{\rho c_p \, d\bar{T}_i^{\text{f}}(z, \infty)/dz}. \tag{6.70}$$

ρ and c_p are the density and specific heat at constant pressure of the air.

There is an interesting relationship between the Chapman layer, (6.50), and the radiation-to-space approximation, (6.66). If, in (6.50),

ξ_0^{-1} is replaced by the diffusivity factor, if (6.66) is written in the form for a single frequency, and if the source function is assumed to vary only slowly over a scale height, the two equations are the same except for a numerical factor. Radiation to space, from a single frequency, takes place from a Chapman layer (Fig. 6.9). Radiation to space from a band will be represented by the superimposition of a number of Chapman layers, that will still form a layer, albeit broader than a single Chapman layer. Equation (6.66) and its layered structure will appear again in § 6.5, when we consider sampling functions for the inversion of satellite data.

6.4.2. *Use of emissivities*

We have discussed the computation of fluxes by means of line-by-line calculations and the use of narrow-band transmission functions; we now consider the approximations that must be made in order to perform the entire frequency integration (or sum over spectral intervals) once and for all. Before the ready availability of fast computers, there was little option but to attempt such a simplification because it was impractical to calculate fluxes independently for many spectral ranges. We shall show that the spectrally integrated functions that must be employed are closely related to laboratory measurements of gas emissivities. To the extent that good laboratory data exist, and that they are in an appropriate form for the calculation of fluxes, this avoids the need to go into the details of gaseous absorption. With the help of empirical adjustments and compromise methods, there is evidence that emissivity methods can be almost as accurate as much more elaborate schemes.

We first define the *gas emissivity,* as measured in the laboratory, in terms of quantities used in earlier sections. Consider a slab of gas, amount m per unit area, at constant temperature (θ) and constant pressure (p), backed by a cold boundary. Equation (6.14) with $\theta^* = 0$ and $J_i(z') = B_i(\theta)$ gives

$$F_i/\pi = B_i(\theta)[1 - \bar{T}_i^{\mathrm{f}}(m, p, \theta)]. \tag{6.71}$$

The emissivity $\epsilon(m, p, \theta)$ is the total flux from the slab divided by the black-body emission

$$\epsilon(m, p, \theta) = \sum_i \frac{\pi B_i(\theta)[1 - \bar{T}_i^{\mathrm{f}}(m, p, \theta)]\,\Delta\nu_i}{\pi B(\theta)}. \tag{6.72}$$

Closely related to the emissivity is a second quantity,

$$\epsilon^*(m, p, \theta) = \sum_i \frac{\pi dB_i(\theta)}{d\theta} \frac{[1 - \bar{T}_i^{\mathrm{f}}(m, p, \theta)]\,\Delta\nu_i}{\pi\, dB(\theta)/d\theta}. \tag{6.73}$$

This quantity cannot be measured in the laboratory but is almost as well determined as is the gas emissivity. The two quantities differ only in the weighting functions $B_i(\theta)$ and $dB_i(\theta)/d\theta$, which are similar functions of frequency. If the flux transmissions are sufficiently well known to give results in agreement with laboratory emissivity measurements, they should allow us to determine ϵ^* with similar precision. It is, in fact, not a drastic approximation to assume that the two quantities are the same.

We now restrict attention to water vapor because it has wide bands extending over most of the thermal spectrum. Other gases have relatively narrow bands, over which the Planck function can usually be regarded as constant. The contribution of an isolated, narrow band to the gas emissivity is simply the band area, $\mathscr{A}$ (§ 4.9), multiplied by B_i/B.

One form of the flux equations is obtained by integrating (6.14) and (6.15) by parts

$$F_i^+(z)/\pi = B_i(z) + [B_i(\theta^*) - B_i(0)]\bar{T}_i^{\mathrm{f}}(0, z) - \int_{B(0)}^{B(z)} \bar{T}_i^{\mathrm{f}}(z', z)\, dB_i(z'), \tag{6.74}$$

$$F_i^-(z)/\pi = B_i(z) - B_i(H)\bar{T}_i^{\mathrm{f}}(z, H) + \int_{B(z)}^{B(H)} \bar{T}_i^{\mathrm{f}}(z, z')\, dB_i(z'). \tag{6.75}$$

These equations have been written in a form appropriate to a finite discontinuity of temperature at the ground $[\theta(0) \neq \theta^*]$ and at an altitude, $z = H$, above which there is no emission $[\theta(z') = 0]$. The second statement is required for any numerical calculation with a finite grid but there are alternatives to the first. The form of (6.74) and (6.75) has been selected because it is convenient for extracting the radiation-to-boundaries term.

In the spirit of the approximations discussed in § 6.2.2, we may write

$$\bar{T}_i^{\mathrm{f}}(z, z') \simeq \bar{T}_i^{\mathrm{f}}[\tilde{m}_i(z, z'), \tilde{p}_i(z, z'), \tilde{\theta}_i(z, z')]. \tag{6.76}$$

We did not discuss an optimum value for $\tilde{\theta}$ in § 6.2.2 but took it to have a presassigned value, e.g., $\tilde{\theta} = 273$ K. Better choices can be made. One is to assume $\tilde{\theta}(z, z') = \theta(z)$. Another is to use a weighted average temperature for the path (z, z'), for example, a mass-weighted temperature,

$$\theta_{\mathrm{P}}(z, z') = \frac{\int_z^{z'} \theta(z'')\rho(z'')\, dz''}{\int_z^{z'} \rho(z'')\, dz''}. \tag{6.77}$$

To go further, it is now necessary to assume that the evaluations of $\bar{m}_i$ and $\bar{p}_i$ are dominated by the densities and pressures in the sums (6.26) and (6.27) and, therefore, that they are the same for each spectral interval, so that we may omit the suffix; essentially, and this is the crux of the emissivity formulation, we must accept some degree of approximation to the influence of temperature on flux transmission. This approximation can be stated in the form of the requirement

$$\bar{T}_i^{\mathrm{f}}(z, z') \simeq \bar{T}_i^{\mathrm{f}}[\tilde{m}(z, z'), \tilde{p}(z, z'), \theta_{\mathrm{P}}(z, z')]. \tag{6.78}$$

Given (6.78), the sums of (6.74) and (6.75) over all spectral intervals become

$$\begin{aligned} F^{+}(z)/\pi = {} & B(z) + B(\theta^*)\{1 - \epsilon[\tilde{m}(0, z), \tilde{p}(0, z), \theta_{\mathrm{P}}(0, z), \theta^*]\} \\ & - B(0)\{1 - \epsilon[\tilde{m}(0, z), \tilde{p}(0, z), \theta_{\mathrm{P}}(0, z), \theta(0)]\} \\ & - \int_{B(0)}^{B(z)} \{1 - \epsilon^*[\tilde{m}(z', z), \tilde{p}(z', z), \theta_{\mathrm{P}}(z', z), \theta(z')]\}\, dB(z'), \end{aligned} \tag{6.79}$$

$$\begin{aligned} F^{-}(z)/\pi = {} & B(z) - B(H)\{1 - \epsilon[\tilde{m}(z, H), \tilde{p}(z, H), \theta_{\mathrm{P}}(z, H), \theta(H)]\} \\ & + \int_{B(z)}^{B(H)} \{1 - \epsilon^*[\tilde{m}(z', z), \tilde{p}(z', z), \theta_{\mathrm{P}}(z', z), \theta(z')]\}\, dB(z'). \end{aligned} \tag{6.80}$$

In (6.79) and (6.80), we have extended the definitions of ϵ and ϵ^* to include two different temperatures, one, θ_{P}, in the argument of the transmission, and the other, θ, in the argument of the Planck function. This takes us away from the use of measured emissivities, but the quantities involved can be calculated if the transmissions are known.

Many different approaches have been taken to (6.79) and (6.80). Among them are the following:

1. Tabulate ϵ and ϵ^* as functions of both θ_{P} and θ. Depending upon the method used to introduce a homogeneous path, this involves either three or four independent variables. Very good results have been reported from this approach.
2. The *isothermal emissivity approximation* assumes $\theta_{\mathrm{P}} = \theta$. This assumption has been widely used in the construction of radiation charts (§ 6.4.3).
3. Manual calculations are greatly simplified by assuming that ϵ and ϵ^* are equal to each other, and that both are independent of temperature (these assumptions are mutually consistent).

It is important to realize that the presence of two functions, ϵ and

ϵ^*, can be avoided if we introduce thermal boundary layers to make a steady transition between the temperatures of the atmosphere and its boundaries. Provided that we maintain consistency, we may then write $\theta(0) = \theta^*$ and $\theta(H) = 0$ so that terms involving ϵ vanish from (6.79) and (6.80). In this discussion, we allow a boundary discontinuity so that we may obtain the boundary-exchange limit to these equations by allowing $\theta(z') \rightarrow \theta(z)$, except at the boundaries. The integrals in (6.79) and (6.80) then vanish and we are left with terms in ϵ only.

The heating rate is

$$h(z) = \frac{-\partial[F^+(z) - F^-(z)]}{\partial z}, \tag{6.81}$$

from which the boundary contribution can be obtained

$$\begin{aligned} h(z)_{\text{bound}} = {} & [\pi B(\theta^*)\, \partial/\partial z]\epsilon[\tilde{m}(0, z), \tilde{p}(0, z), \theta(z), \theta^*] \\ & -[\pi B(z)\, \partial/\partial z]\epsilon[\tilde{m}(0, z), \tilde{p}(0, z), \theta(z), \theta(z)] \\ & -[\pi B(z)\, \partial/\partial z]\epsilon[\tilde{m}(z, H), \tilde{p}(z, H), \theta(z), \theta(z)]. \end{aligned} \tag{6.82}$$

Equation (6.82) corresponds to the sum over spectral intervals of (6.65) and (6.66) and should, therefore, account for most of the flux divergence. The difference between the total flux divergence calculated from (6.79) and (6.80) and the boundary term (6.82) represents the exchange with other atmospheric layers, which will normally be small. But this small term is the only one for which we must use the approximation, (6.78), inherent in the emissivity method. The correct boundary terms (6.65) and (6.66) can be calculated without approximation. The exchange terms can be calculated from (6.81) and (6.82) using consistent emissivity approximations; the quantity [(6.81) − (6.82) + (6.65) + (6.66)] is then an expression for the heating rate that makes optimum use of both emissivity and boundary exchange approximations.

Figure 6.15 shows errors for two emissivity calculations, assessed against an "exact" calculation (a line-by-line calculation). The emissivity calculations make consistent use of the same data: the first is the isothermal emissivity approach; the second is the same, but modified with an exact cooling-to-space term. The errors involved in the second calculation are small.

6.4.3. *Radiation Charts*

Prior to the ready availability of large computers, radiation charts offered the best way to calculate fluxes or heating rates. Although they are no longer commonly used it is desirable that they be understood because they provide a visualization of the process of flux calculation.

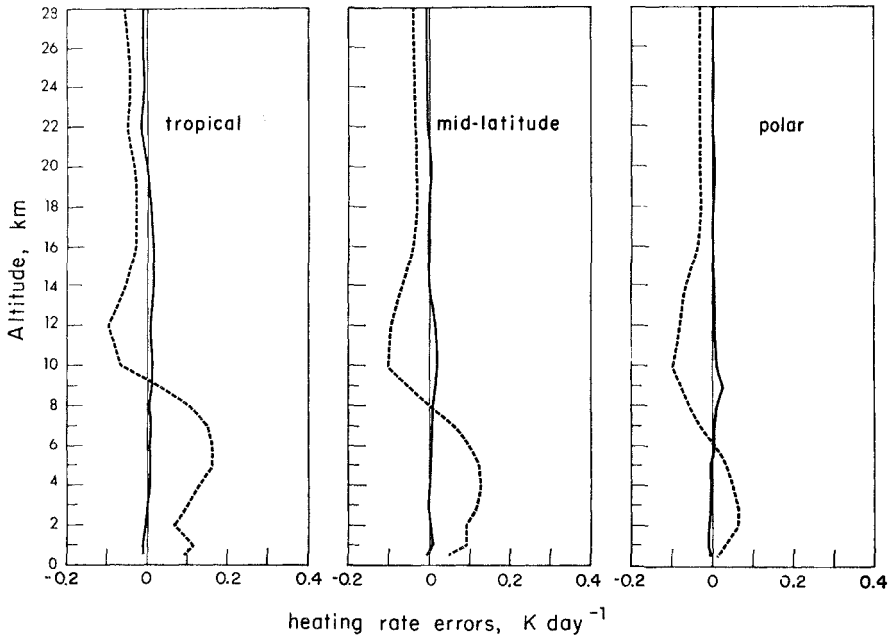


FIG. 6.15. Errors in emissivity calculations with and without corrections for cooling-to-space. The two lines give errors with respect to "exact" (line-by-line) calculations. The broken line uses an isothermal emissivity procedure; the solid line corrects the isothermal emissivity procedure with an exact radiation-to-space term. After Fels and Schwartzkopf (1975).

With a graphical method, we may use continuous vertical variables and it is possible to apply the continuous boundary conditions that eliminate terms in ϵ from (6.79) and (6.80). In addition, we make use of isothermal emissivities and the scaling approximation (§ 6.2.2) for the nonhomogeneous path. With these approximations (6.79) and (6.80) become

$$F^{+}(z) = \pi B(z) + \int_{B(z)}^{B(0)} \{1 - \epsilon^{*}[\tilde{m}(z, z'), \theta(z')]\}\, d\pi B(z'), \quad (6.83)$$

$$F^{-}(z) = \pi B(z) + \int_{B(z)}^{B(H)} \{1 - \epsilon^{*}[\tilde{m}(z, z'), \theta(z')]\}\, d\pi B(z'). \quad (6.84)$$

The limits of the integral in (6.83) have been reversed to emphasize the similarity to (6.84). Both expressions can evaluated from a graphical device with orthogonal axes πB and $1 - \epsilon^{*}$. The particular procedure represented by (6.83) and (6.84) is employed in the Yamamoto chart, shown in Fig. 6.16.

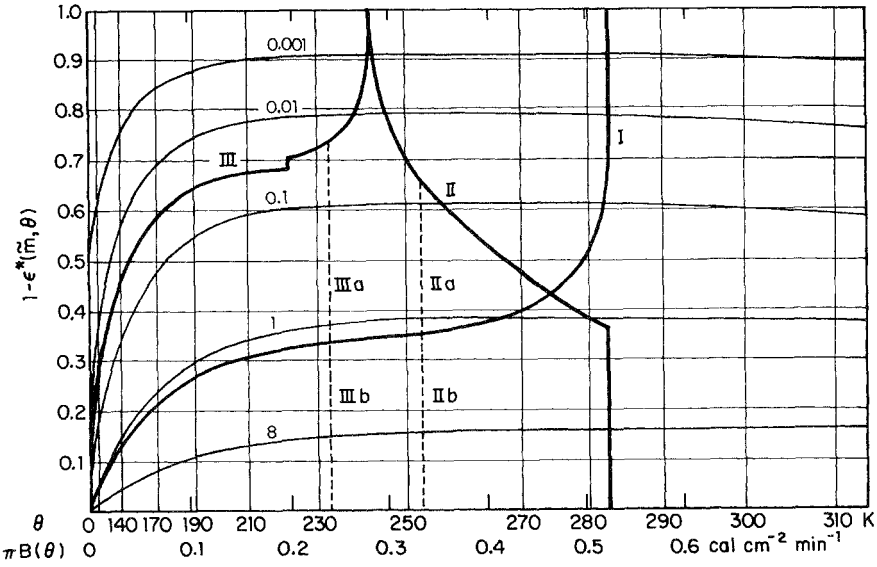


FIG. 6.16. Yamamoto's flux chart. The light vertical lines are isotherms. The light, curved lines are isopleths for water vapor with amounts in g cm^{-2}. The original chart has been modified by omitting constructions used to evaluate the radiation flux from carbon dioxide; as the chart stands it yields the water vapor flux, as if carbon dioxide were absent. (I) Ascent curve for evaluating the downward flux at the surface. (II) Ascent curve for the upward flux component at 6.7 km. (III) Ascent curve for the downward flux component at 6.7 km. Ascent curves IIa, IIb, IIIa, and IIIb are modified to include the effect of cloud (see text). After Yamamoto (1952).

Many other radiation charts have been proposed; all are transforms of, or approximations to, (6.83) and (6.84) and they differ only in convenience for specific problems. Flux divergence charts can also be constructed on similar principles.

In order to use Yamamoto's chart it is desirable to add curvilinear axes for $\theta(z')$ and $\tilde{m}(z, z')$. The isotherms are vertical lines with nonlinear spacing. The isopleths ($\tilde{m}$ = constant) are almost logarithmically spaced and with strong curvature only at low temperatures.

Relationships between θ and $\tilde{m}$ (*ascent curves*) must be evaluated separately for the upward and downward radiation fluxes. For both, $\tilde{m}$ is zero for $z = z'$ and increases monotonically as z' moves away from the reference level, z. At the same time, $\theta(z')$ will change; in the troposphere it will generally increase as m increases for the upward flux component (from the atmosphere below z) and decrease as m increases for the downward flux component (from the atmosphere above z). In the example shown in Fig. 6.18, the two limits are $\theta^* = 283$ K (the surface temperature) for the upward flux and $\theta(H) = 223$ K (the stratosphere temperature) for the downward flux.

Upward and downward branches of an ascent curve are illustrated in Fig. 6.16. The reference level is at 6.7 km, where the temperature is 243 K. Curve II is the ascent curve for the upward flux component. The surface is reached at $\tilde{m} = 1.34\ \mathrm{g\ cm^{-2}}$. Continuation of the ascent curve to $\tilde{m} = \infty$ along the isotherm $\theta = 283$ K adds nothing to the integral, but allows us to complete the contour. $F^+(z)$ is the area to the left of curve II.

The ascent curve for F^- terminates in the stratosphere at $\theta(H) = 223$ K, $\tilde{m}(z, H) = 1.38\ \mathrm{g\ cm^{-2}}$ (the kink is the tropopause). The ascent curve must now be terminated with a boundary layer containing a negligible amount of water vapor, bridging the gap between 223 K and 0 K. The ascent curve must follow the $\tilde{m} = 1.38\ \mathrm{g\ cm^{-2}}$ isopleth to the 0 K isotherm, as shown in the radiation chart. $F^-(z)$ is now the area to the left of curve III. The net flux, $F^+ - F^-$, is the area between curves II and III.

I is the ascent curve for the downward flux at the surface.

The effects of clouds (treated as black bodies) on the flux at 6.7 km are indicated by the ascent curves IIa, IIb, IIIa, and IIIb. A cloud at $z = 5$ km has an upper-surface temperature of 253 K; a cloud at 8.3 km has a lower-surface temperature of 233 K. When a black surface is encountered, the ascent curve goes to $\tilde{m} = \infty$ along the isotherm corresponding to the surface temperature. A large influence of clouds upon radiation fluxes is demonstrated.

6.5. The inverse problem for thermal radiation

6.5.1. *The Kernel functions*

(i) The retrieval equations. All methods of remote sensing from satellites involve observations of electromagnetic radiation scattered by, absorbed by, or emitted from the atmosphere. The most important single application of remote sensing has been the recovery of atmospheric temperature using the 15 μm and the 4.3 μm fundamentals of CO_2 and the 60 GHz line of molecular oxygen. Temperature retrieval provides an exemplar for the entire field and we shall restrict our discussion to this topic. Many other techniques of remote sensing are mentioned in references in the Bibliography.

In the earlier sections of this chapter we discussed how to calculate the radiation field, given data on atmospheric temperatures and absorber concentrations. We are now concerned with the *inverse problem*: Given the radiation field, how well may we infer atmospheric parameters from this information? The theory of this class of problem is *retrieval theory*; atmospheric remote sensing is only one of a number of fields for which retrieval theory is an important topic.

An observed satellite spectrum was shown in Figs. 6.1 and 6.2. This is not the usual way soundings are made; it is more common to record a

number a discrete spectral intervals but the continuous data in these figures are instructive. The spectral resolution of $5\,\mathrm{cm}^{-1}$ smooths out most of the fine structure in the carbon dioxide and ozone bands and some in the water vapor bands. For the most part, remote sensing involves spectral averages over many lines, although in the microwave spectrum monochromatic observations are possible and have some advantages.

Both oxygen and carbon dioxide have almost constant mixing ratios at all levels in the lower and middle atmosphere. Since this is where outgoing thermal radiation originates, the intensities shown in Figs. 6.1 and 6.2 can be expressed as a functional of the atmospheric temperatures, the radiation frequency, and the zenith angle of observation. These latter two are the variables available as a basis for performing temperature inversions.

In terms of Fig. 6.2, the problem is one of finding a temperature distribution that is most consistent with the observed frequency spectrum of CO_2. That such a distribution probably exists is demonstrated by the agreement between observations and the theoretical spectrum shown in that figure. But there are a limited number of completely independent pieces of information in the spectrum and this, together with the existence of instrument noise, gives rise to problems of uniqueness, optimum solutions, and their errors.

The fundamental retrieval equation for monochromatic radiation is (2.98), with $\tau = 0$. For a thermal source function, this may be written

$$I_\nu(\xi) = B_\nu(\theta^*)T_\nu(0, \xi) + \int_0^\infty B_\nu[\theta(z)]\frac{\partial T_\nu(z, \xi)}{\partial z}\,dz, \tag{6.85}$$

where the transmission is

$$T_\nu(z, \xi) = \exp\frac{-\tau_\nu(z)}{\xi}. \tag{6.86}$$

For an optically thick atmosphere, the first term on the right-hand side of (6.85) may be neglected. If not, we may formally regard it as a quantity to be evaluated approximately from measurements made in some other, transparent spectral region, where it will dominate; we may then define a transformed intensity

$$\tilde{I}_\nu(\xi) = I_\nu(\xi) - B_\nu(\theta^*)T_\nu(0, \xi). \tag{6.87}$$

However we may choose to handle this problem, we may adopt as the retrieval equation

$$I_\nu(\xi) = \int_0^\infty B_\nu[\theta(z)]\frac{\partial T_\nu(z, \xi)}{\partial z}\,dz. \tag{6.88}$$

In practice, coordinate transformations, $z \rightarrow p$ or $z \rightarrow \ln p$, where p is the air pressure, are often used.

For a finite spectral interval (i), narrow compared to variations of the Planck function, the expression consistent with the treatment of § 6.1.3 is

$$I_i(\xi) = \int_0^\infty B_i[\theta(z)] \frac{\partial \bar{T}_i(z, \xi)}{\partial z} dz. \tag{6.89}$$

Equations (6.88) and (6.89) are *Fredholm integral equations of the first kind.* The left-hand side, I_i, and the *kernel function,* $\partial T_i(z, \xi)/\partial z$, are known functions. The unknown function to be solved for, $B_i[\theta(z)]$, is under the integral sign. The mathematical difficulties encountered in solving these equations will be discussed in § 6.5.3. Since the difficulties are rooted in the nature of the physics of the problem, no amount of mathematical ingenuity alone can circumvent them. Instead, we must give up the conventional concept of a rigorous and unique solution and be content with a nonunique solution, estimated from the equations, within the errors of measurement, and consistent with our a priori knowledge of the solution. We shall discuss these questions qualitatively at first, followed by a convergent numerical technique that derives from a physical view of the problem (§ 6.5.2) and, finally, in a more general framework provided by linear theories (§ 6.5.3).

Table 6.3 summarizes the most useful spectral features for remote sensing in the earth's atmosphere. As we have discussed, temperature retrievals require an absorber whose concentation is known. Nitrogen cannot be used because it has no convenient spectral signature. Oxygen has absorption lines in the microwave region, while CO_2 has two strong fundamentals in the infrared spectrum. The abundance of species may be determined if we know the temperature. Water vapor soundings are of great importance for meteorology and may be made using the 6.3 μm or the rotation band beyond 10 μm. Ozone is important for the thermal balance of the middle atmosphere. Remote sounding of ozone densities can be performed in the infrared (9.6 μm), the visible, and the ultraviolet spectrum. For the visible and ultraviolet spectrum a scattering source function is necessary. Historically, ozone retrievals in the ultraviolet spectrum were the first remote soundings made in the earth's atmosphere (see the Bibliography). There are many trace constituents, such as N_2O, CH_4, $CFCl_3$, and CF_2Cl_2, some of which have a fairly simply chemistry that can confidently be computed. These species may be important in their own right but they are also of increasing importance as tracers of atmospheric motions.

(ii) Temperature retrieval. The characteristics of atmospheric inversions (or retrievals) are determined by the properties of the kernel function;

Table 6.3. Spectral features for remote sensing

Gas	Wavelength	Transition	Parameter
O_2	5 mm	Magnetic dipole	Atmospheric temperature
CO_2	15 and 4.3 μm	Vibration–rotation	Atmospheric temperature
H_2O	6.3 μm, 10 μm $-\infty$	Vibration–rotation	Gas density
O_3	9.6 μm	Vibration–rotation	Gas density
O_3	UV, visible	Electronic	Gas density
N_2O	7.8 μm	Vibration–rotation	Gas density, tracer
CH_4	7.7 μm	Vibration–rotation	Gas density, tracer
CCl_2F_2	9.1, 8.7, 10.9 μm	Vibration–rotation	Gas density, tracer
CCl_3F	9.2, 11.8 μm	Vibration–rotation	Gas density, tracer
Surface	~10, ~3.8 μm windows	Continua	Surface temperature

Source: After Rodgers (1971).

for (6.88) and (6.89), the kernel function is

$$K(z, \xi) = \frac{\partial T(z, \xi)}{\partial z}, \tag{6.90}$$

where K stands for K_i or for K_ν and T for $\bar{T}_i$ or T_ν. If $K_i(z, \xi)$ were a Dirac δ-function, $\delta(z - z_{i,\xi})$, the solution to the integral equation would be

$$I_i(\xi) = B_i[\theta(z_{i,\xi})]. \tag{6.91}$$

Given that $z_{i,\xi}$ varies with i and ξ, intensities measured at different zenith angles, or in different spectral intervals, lead directly to temperatures at known atmospheric levels.

Quantitative ideas about temperature retrievals can make use of the fact that real kernel functions usually have a single maximum; for both transparent paths ($\bar{T}_i = 1$) and for opaque paths ($\bar{T}_i = 0$) transmission is constant and the kernel function is zero. For monochromatic radiation, the height dependence of the kernel function, based upon (6.86), is identical to the height dependence of solar heating given by (6.50). The

Chapman function, Fig. 6.9, is, with suitable normalization, the kernel function for a gas with constant mixing ratio, constant atmospheric temperature, and an absorption coefficient independent of pressure. The shapes of realistic monochromatic kernel functions are influenced by temperature and pressure as well as by the vertical variation of mixing ratio; nevertheless, the Chapman function displays many of the important characteristics of a realistic kernel function. We have already discussed, in connection with Fig. 6.9, how the peak of the Chapman function moves when the absorption coefficient and the zenith angle vary. If we look at the Chapman function as an approximation to a δ-function, we may now interpret Fig. 6.2. Radiation in the center of the 15 μm band originates from a level of unit optical depth for which the temperature is 220 K. This is the highest level from which temperature information is available. The lowest level is the earth's surface, with a temperature close to 295 K.

Intermediate points on the flank of the 15 μm band can be used to construct the temperature profile between the 220 K and 295 K temperature levels. If the temperature changes monotonically with height, this solution should be unique. However, with a kernel function as wide as the Chapman function, the vertical resolution cannot be much better than one scale height, and details on a smaller scale will be unrecoverable, regardless of the retrieval technique.

For real gases and finite spectral intervals, there are offsetting influences on the kernel function. A finite spectral interval leads to a broadening of the kernel function because the effect is to average over a number of Chapman functions. Pressure broadening, on the other hand, can lead to a narrowing of the kernel function. If absorption were proportional to the pressure, the scaled gas density for use with a scaling approximation would vary as the pressure squared with half the atmospheric scale height, and the kernel function would narrow by the same factor. Temperature can also affect the kernel function and can lead to either broadening or narrowing. The net effect on kernel functions for the 4.3 μm CO_2 band, as shown in Fig. 6.17, is to give a breadth close to that of a Chapman function.

(iii) Angular scans. The peak of the Chapman function moves upward or downward as ξ decreases or increases (i.e., as zenith angle increases or decreases). This change in the kernel function can be used as a basis for temperature retrievals.

For mechanical reasons, it is not easy to perform angular scans from satellites, but the variation of emergent intensity with zenith angle has played an important role in studies of the sun and the planets. Emission from these extended objects appears, to a terrestrial observer, to be approximately symmetrical with respect to the center of the disc ($\xi = 1$), with a uniform change toward the limb ($\xi = 0$), called *limb darkening* or

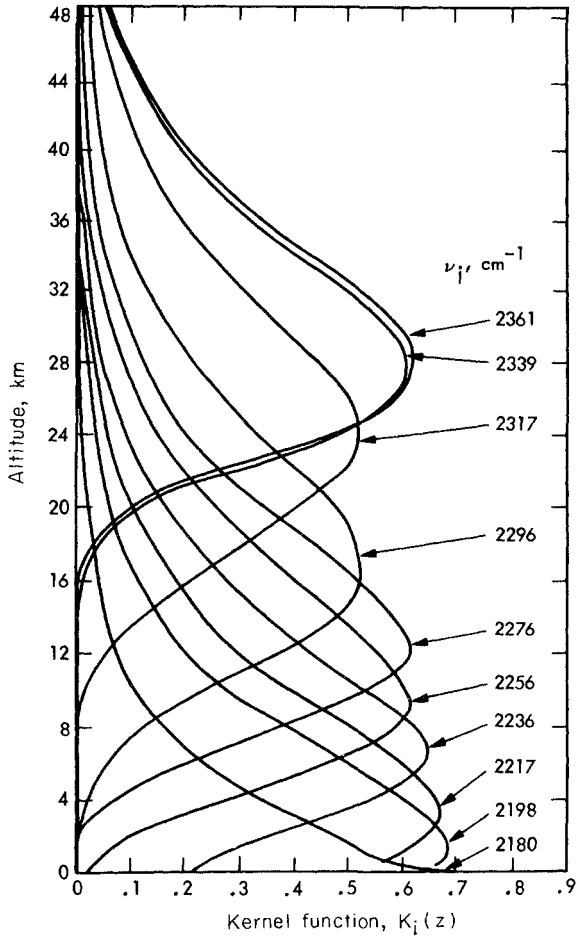


FIG. 6.17. Kernel functions for the 4.3 μm (2349 cm^{-1}) CO_2 band. Values of ν_1 for the spectral intervals used are indicated against each curve. After Chahine (1977a).

limb brightening. The implication is that the vertical profile of temperature is more or less uniform over the disc, and that we are observing the effect of changing zenith angle,

$$\xi = [1 - (r/r_0)^2]^{1/2}, \tag{6.92}$$

where r is the distance from the center and r_0 is the limb radius.

Early measurements on the planet Venus gave limb darkening in the 10 μm atmospheric window of the form

$$I(\xi) = a_0 + a_1\xi + a_2\xi^2 + a_3\xi^3 + \cdots. \tag{6.93}$$

The following solution may be confirmed by substitution in (6.88),

$$B(\tau) = a_0 + a_1\tau + a_2\tau^2/2 + a_3\tau^3/6 + \cdots. \tag{6.94}$$

For uniform limb darkening, B must increase as τ increases, i.e., as height decreases; radiation originates from a region in which temperature decreases with height. For limb brightening, on the other hand, radiation originates from a region of increasing temperature with height.

This is not a sophisticated inversion. The series (6.93) and (6.94) must be terminated after a few terms to avoid numerical instabilities, negative intensities, and other undesirable features. The solution assumes monochromatic absorption, which is probably incorrect, but it illustrates in simple terms the principle of inversions by means of angular scans.

(iv) Limb scans. A collimated radiometer or spectrometer, viewing horizontally, receives radiation from a path for which the closest approach to the surface is the *tangent height*, z_T. Equations (6.88) and (6.89) are valid expressions for the intensity, except that the path is doubled and there is no contribution from below the tangent height; the kernel function is strongly influenced by the curved geometry. Figure 6.18 shows limb kernel functions for a wide spectral interval (585–705 cm^{-1}) in the 15 μm CO_2 band. The kernel functions for tangent

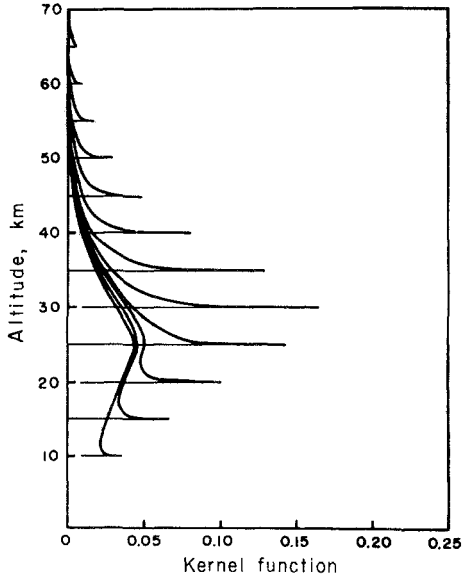


FIG. 6.18. Limb kernel functions. Calculations are for an infinitely narrow instrumental field of view. The lowest level for each curve is the tangent height. The spectral region 585–705 cm^{-1} takes in most of the 15 μm CO_2 band. After Gille and House (1971).

heights of 30 km or more display the advantage of this technique. The optical paths are all small and the kernel functions are similar in shape, with most of the contribution coming from within 3 km of the tangent height. With such a narrow kernel function the inversion is a relatively simple task. We may assume the source function to be constant where the kernel function is significant and the only information required for the retrieval is, from the integration of (6.89), the absorption over the entire tangent path. This path can be very long. For a well-mixed layer of scale height $H = 10$ km and a planet of radius $R = 6000$ km, the effective absorption path length is $(2\pi RH)^{1/2} = 630$ km. Limb scans are sensitive to very small concentrations of absorber in the middle atmosphere.

Vertical profiles are obtained by scanning the radiometer over a very small angle, giving a continuous range of tangent heights. The kernel functions for tangent heights less than 20 km tend toward those from nadir viewing instruments, and are less useful for retrievals. If the temperature is known, limb scans yield absorptions, from which densities can be calculated. If mixing ratios are approximately constant over a scale height, the density inversion is essentially trivial.

Limb scans present many advantages for middle atmosphere observations, both for temperature and for gaseous density soundings. There are, however, some compensating disadvantages in the difficulties involved in construction of the instrument, and of the poor horizontal resolution of the measurements.

6.5.2. *A "physical" approach to retrieval*

Out of a number of inversion techniques in common use, we shall discuss results from only one, that of Chahine. This technique derives from the qualitative discussion in § 6.5.1(i), in which we treated a typical kernel function as an approximation to a Dirac δ-function. Because of this simple idea, the method is easy to visualize and to extend to new circumstances, and it converges rapidly to useful solutions. It does not have a rigorous basis, however, and it must be tested numerically for each set of differing circumstances.

Equation (6.89) may be written

$$I_i = \int_0^\infty K_i(z) B_i[\theta(z)]\, dz. \tag{6.95}$$

Equation (6.95) may be regarded as a mapping transformation relating $\theta(z)$ as a function of z to I_i as a function of ν_i. Chahine's method is based on the observation that $K_i(z)$ usually has a maximum at $z = z_i$, the neighborhood of which gives rise to most of the contribution to the integral in (6.95). The discrete relationship $\nu_i \rightarrow z_i$ is assumed to be the mapping from ν to z; the task is to find the mapping from I_i to $\theta(z_i)$.

It is approximately true that

$$I_i = \int_0^\infty K_i(z) B_i[\theta(z)]\, dz$$
$$\approx C K_i(z_i) B_i[\theta(z_i)], \tag{6.96}$$

where C is an unknown constant. This approximation can be justified by the use of the mean value theorem or may be accepted as physically obvious from the dominance of the kernel function near z_i.

We start from an intelligent guess, $\theta^{(0)}(z_i)$, for the temperatures. From the same reasoning that led to (6.96), we find,

$$I_i^{(0)} = \int_0^\infty K_i(z) B_i[\theta^{(0)}(z)]\, dz$$
$$\approx C K_i(z_i) B_i[\theta^{(0)}(z_i)]. \tag{6.97}$$

In general, $I_i^{(0)} \neq I_i$ because $\theta^{(0)}(z_i)$ is not the exact solution. However, we may use the ratio of (6.96) to (6.97) to obtain an improved estimate,

$$\frac{B_i[\theta^{(1)}(z_i)]}{B_i[\theta^{(0)}(z_i)]} = \frac{I_i}{I_i^{(0)}}. \tag{6.98}$$

In (6.98), $I_i^{(0)}$ was obtained from the integral in (6.97). Given $\theta^{(1)}$, a new integral can be performed and the same procedure used to obtain an improved approximation. This procedure usually converges rapidly upon what we may hope to be the correct solution. A continuous profile is obtained from a suitable interpolation scheme. The success of this last step depends on the smoothness of the profile; the method is incapable of resolving details on a scale less than the separation of the z_i, but this limitation is shared, in some degree, by all inversion techniques.

Chahine's method could, in principle, fail to converge or converge to the wrong solution. Failure to converge will be obvious and the solution can be rejected. Only numerical experiments can show whether the method converges to correct solutions. An example of a numerical retrieval experiment is shown in Fig. 6.19. The kernel functions employed are those for the 4.3 μm CO_2 band, shown in Fig. 6.17. From the continuous "exact profile" in Fig. 6.19, discrete intensities are calculated. A random "error" is added to simulate measurement errors and Chahine's method is used to recover the discrete temperatures shown. The results are strikingly good. On the basis of numerical experiments with 30 temperature profiles, it appeared that a 2% measurement error should give rise to a 1 K error in the retrieved temperature.

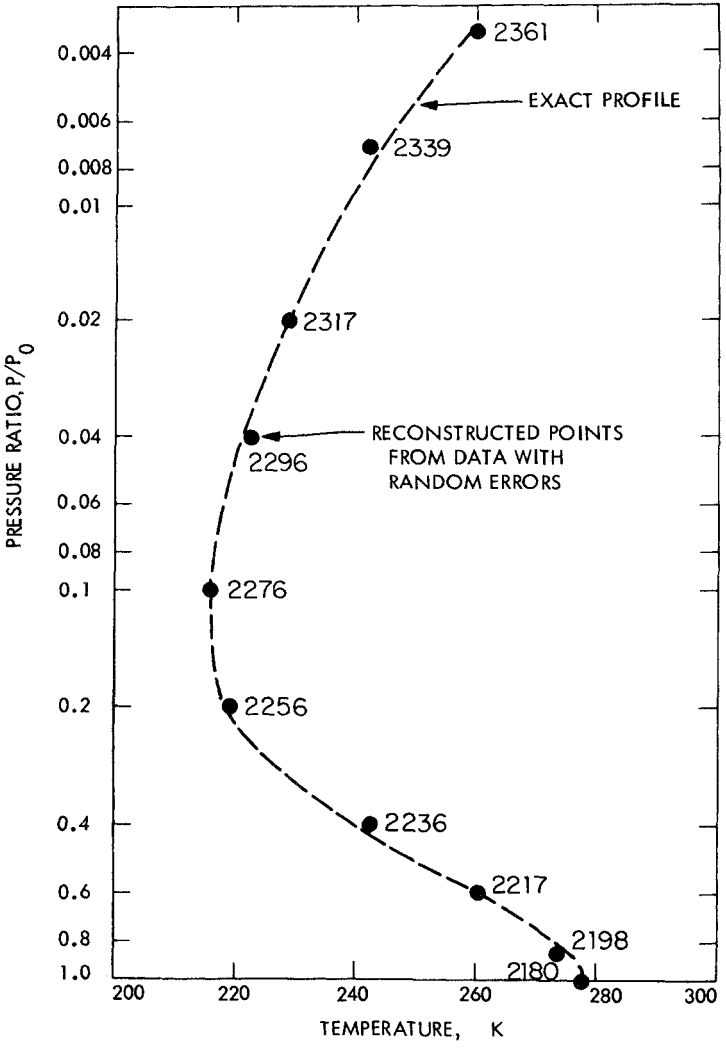


FIG. 6.19. Temperature retrieval using Chahine's method. This is a theoretical investigation. Outgoing intensities were calculated for the kernel functions shown in Fig. 6.17 and the continuous temperature distribution shown by the broken line. These intensities were modified by random errors with amplitudes between 4.8 and 9.6%. The mean difference between retrievals with and without the errors was 1.5 K. The numbers against each curve represent the ν_i (cm^{-1}) that map onto the z_i, given by the altitudes of the plotted points. After Chahine (1977a).

The foregoing discussion was appropriate to a clear sky, but there are often clouds in the field of view of a satellite radiometer. Chahine's method can be readily extended to a partly cloudy field. No assumption is made as to the nature of the clouds, which may be thin or thick, black or gray, and at any level.

We assume that there is a fixed difference in intensity, $G(\nu, p)$, between cloudy and cloud-free sections of the field; p is the pressure of the cloud surface, introduced so that we may consider multiple cloud layers. Consider first a single cloud layer. If $N_1(p)$ is the fraction of clouds at the level, p, in a field designated by the subscript (1), we may write the clear column intensity in the form

$$I_1(\nu) = \tilde{I}_1(\nu) - N_1(p)G(\nu, p), \tag{6.99}$$

where the tilde indicates the measured intensity. We do not know $G(\nu, p)$ but we can eliminate it if we may assume that a neighboring field (subscript 2) differs from the first only in the cloud fraction $N_2(p)$,

$$I_2(\nu) = \tilde{I}_2(\nu) - N_2(p)G(\nu, p). \tag{6.100}$$

We further assume that the clear column intensities are the same in the two fields,

$$I_1(\nu) = I_2(\nu) = \bar{I}(\nu). \tag{6.101}$$

Eliminating I_1, I_2, and G from (6.99), (6.100), and (6.101),

$$\bar{I}(\nu) = \tilde{I}_1(\nu) + \eta[\tilde{I}_1(\nu) - \tilde{I}_2(\nu)], \tag{6.102}$$

where

$$\eta = \frac{N_1}{N_1 - N_2}. \tag{6.103}$$

For this procedure to be useful, the cloud fractions in the two fields must differ, the more the better.

Since N_1 and N_2 should not depend upon the frequency, η can be determined from measurements at another frequency ν',

$$\eta = \frac{\bar{I}(\nu') - \tilde{I}_1(\nu')}{\tilde{I}_1(\nu') - \tilde{I}_2(\nu')}. \tag{6.104}$$

$\theta(p)$ and η can now be obtained simultaneously from a coupled iterative scheme involving the two frequencies and the results from two fields.

The method outlined above can be extended to deal with multiple

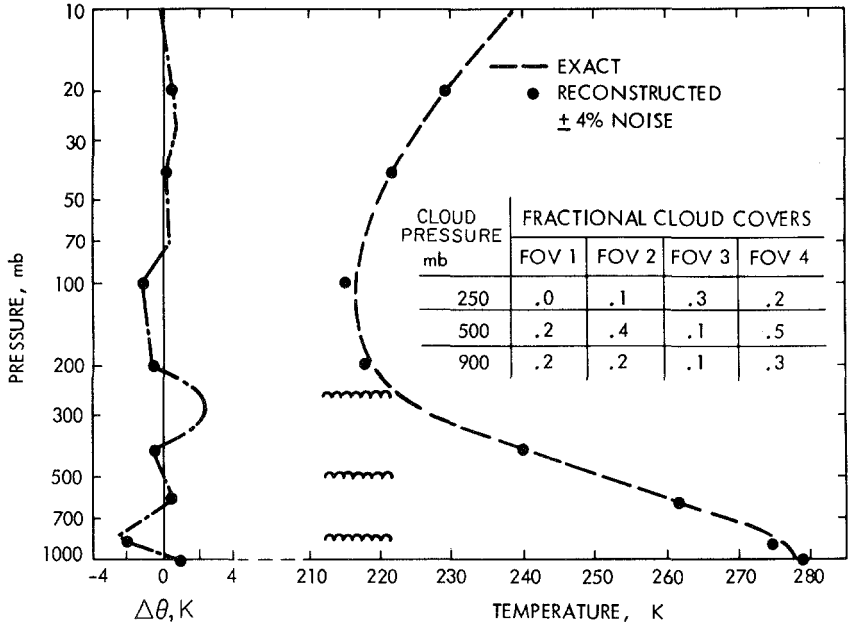


FIG. 6.20. Temperature retrieval in the presence of cloud. This is a theoretical reconstruction. The broken line is the assumed temperature profile. Intensities were calculated for 10 frequencies in the 4.3 μm band of CO_2 and for three in the 15 μm band. Four percent random errors were added to the calculated intensities. The fractional cloud amounts in four fields of view (FOV) are given. The method described in the text was used to obtain the clear column intensities and retrieved temperatures are shown by the points. Average errors in the temperature retrieval are 0.80 K. After Chahine (1977b).

cloud layers. A numerical illustration of the success of the method using synthetic data is shown in Fig. 6.20. The temperature field is retrieved in the presence of four cloud decks. Data were employed from four adjacent fields, and from both the 4.3 μm and the 15 μm CO_2 bands.

Despite these striking results, it must again be emphasized that success with Chahine's method is not guaranteed. The rapidity of the convergence depends greatly on the sharpness of the maximum of the kernel function; without a single maximum in the kernel function, the method may not converge or converge to an incorrect solution.

6.5.3. *Linear analysis*

Chahine's method introduces no approximations into the retrieval equation, (6.89). If we accept two reasonable approximations, however, the equation can be linearized and the retrieval problem can be posed in a more rigorous framework. The required approximations are that the Planck function be independent of the frequency and that the kernel function be independent of temperature. The former is satisfactory for

most single bands but not for the water vapor bands nor for coupled schemes that use information from more than one band; the latter approximation is less acceptable, but may be improved by iterating the solution with improved temperatures in the calculation of the transmission.

Unlike Chahine's method, a linear solution gives a definite result, but that result may be physically unacceptable; Chahine's method is more constrained and there is less possibility of unreasonable solutions. Both approaches have strong supporters.

(i) Noise amplification. Measurements are made at discrete frequencies $\nu_1, \nu_2, \ldots \nu_i, \ldots \nu_M$. We assume $B_i = \bar{B}$ and that the kernel functions, $K_i(z) = \partial \bar{T}_i(z)/\partial z$, are independent of temperature. We then have the linear relation,

$$y_i = I_i = \int_0^\infty K_i(z)\bar{B}(z)\, dz \qquad i = 1, 2, \ldots M. \tag{6.105}$$

It is convenient to express $\bar{B}(z)$ in a series,

$$\bar{B}(z) = \sum_{i=1}^{M} x_i b_i(z), \tag{6.106}$$

where $b_i(z)$ are suitable basis functions, such as polynomials, harmonic functions, or empirical functions, and x_i are unknown coefficients. Substituting (6.106) into (6.105) we have

$$y_i = \sum_{j=1}^{M} A_{ij} x_j, \qquad i = 1, 2, \ldots M, \tag{6.107}$$

where

$$A_{ij} = \int_0^\infty K_i(z) b_j(z)\, dz. \tag{6.108}$$

Inversion of (6.107) yields x_i,

$$x_i = \sum_{j=1}^{M} A_{ij}^{-1} y_j \qquad i = 1, 2, \ldots M, \tag{6.109}$$

where $\mathbf{A}^{-1}$ is the inverse of the matrix $\mathbf{A}$, the elements of which are given by (6.108). Substituting (6.108) into (6.106), we have

$$\bar{B}(z) = \sum_{i=1}^{M} D_i(z) y_i, \tag{6.110}$$

where

$$D_i(z) = \sum_{j=1}^{M} A_{ji}^{-1} b_j(z) \qquad i = 1, 2, \ldots M \tag{6.111}$$

is the *contribution function.* $D_i(z)$ weighs the contribution of the specific intensity y_i to the solution $\bar{B}(z)$. From (6.105) and (6.110), the contribution function must satisfy the relation

$$\int_0^\infty K_i(z) D_j(z)\, dz = \delta_{ij}, \tag{6.112}$$

where δ_{ij} is the Kronecker δ-function.

In subsequent sections we will consider what needs to be done to avoid the possibility that A_{ij} is ill conditioned. An error ϵ_i in the measurement y_i will introduce an error $D_i(z)\epsilon_i$ into the solution. If the measurements, y_i, have associated noise variances σ_i^2, the noise variance σ_B^2 of the solution $B(z)$ is

$$\sigma_B^2 = \sum_{i=1}^{M} D_i^2(z)\sigma_i^2. \tag{6.113}$$

If $\sigma_i^2 = \sigma^2$, we have

$$\sigma_B^2(z) = D^2(z)\sigma^2, \tag{6.114}$$

where

$$D^2(z) = \sum_{i=1}^{M} D_i^2(z) \tag{6.115}$$

is a *noise amplification factor.*

As an illustration of the behavior of the linear equations we consider a numerical example that simulates the essential features of remote temperature soundings. Eight idealized kernel functions are shown in Fig. 6.21. With polynomials as the basis functions, $b_i(z)$, we may compute the contribution functions, $D_i(z)$ using (6.108) and (6.110). The results are displayed in Fig. 6.22. According to (6.113), these contribution functions imply very high sensitivity of the solution to noise.

It can be argued that the choice of polynomials as basis functions is arbitrary and that a more judicious choice would lead to less noise amplification. This is certainly the case. Formally, it is possible to ask what representation of $\bar{B}(z)$ would lead to a minimum in the average value of the noise amplification factor, subject to the constraint (6.112).

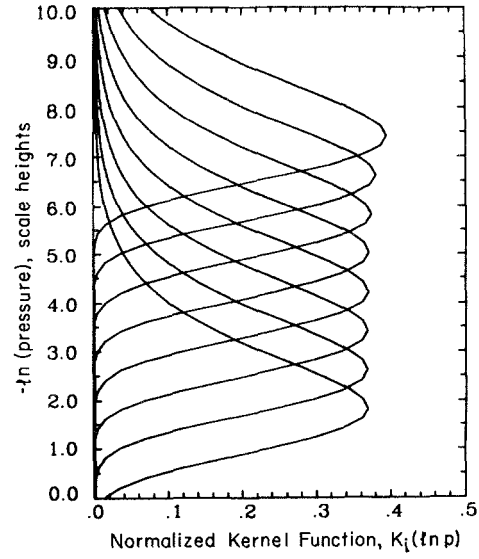


FIG. 6.21. Idealized kernel functions. The kernel functions are Chapman functions, based upon ln(pressure) as the independent variable. After Houghton et al. (1984).

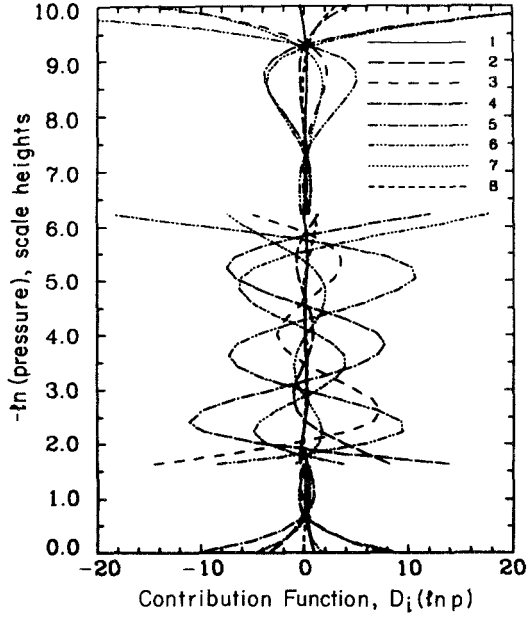


FIG. 6.22. Contribution functions. The contribution functions are computed for the kernel functions shown in Fig. 6.21 with polynomials as basis functions (curve 1 is the uppermost kernel function). The scale of the abscissa corresponds to the middle section. For the upper and lower sections the amplitude of the contribution function is 1000. After Houghton et al. (1984).

The minimum contribution functions can be shown to be

$$D_i(z) = \sum_{j=1}^{M} L_{ij}^{-1} K_j(z), \tag{6.116}$$

where L_{ij}^{-1} is the inverse of the matrix

$$L_{ij} = \int_0^\infty K_j(z) K_i(z)\, dz. \tag{6.117}$$

Optimum contribution functions for the kernel functions in Fig. 6.21 are shown in Fig. 6.23. They are a marked improvement over the results shown in Fig. 6.22, but the noise sensitivity is still poor. An error of ϵ in an intensity measurement can lead to an error as large as 15ϵ in the retrieved Planck function.

(ii) A least-squares solution. It is possible to reduce the noise amplification factor by choosing a representation of the source function with fewer terms than the number of measurements. In place of (6.106) we write

$$\bar{B}(z) = \sum_{\alpha=1}^{N} x_\alpha b_\alpha(z) \qquad N < M. \tag{6.118}$$

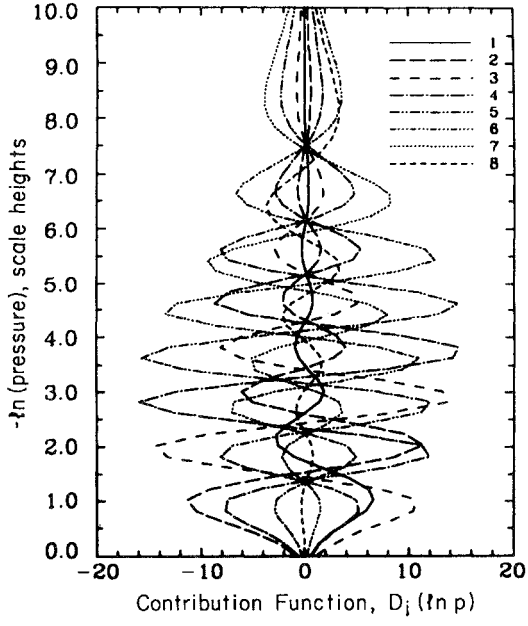


FIG. 6.23. Optimum contribution functions. The kernel functions in Fig. 6.21 are used with (6.116) and (6.117). After Houghton et al. (1984).

We observe the convention that latin indices run from 1 to M (the number of measurements) while greek indices run from 1 to N (the number of terms in the solution).

Using (6.105), we obtain the analog of (6.107)

$$y_i = \sum_{\alpha=1}^{N} A_{i\alpha} x_\alpha \qquad i = 1, 2, \ldots M. \tag{6.119}$$

Equation (6.119) cannot be inverted directly because there are more equations than there are unknowns, i.e., the system is overdetermined. The standard approach to such problems is the method of least squares. The difference between the left-hand side and the right-hand side of (6.119) is

$$\Delta_i = y_i - \sum_{\alpha=1}^{N} A_{i\alpha} x_\alpha \qquad i = 1, 2, \ldots M. \tag{6.120}$$

The least-squares solution minimizes the quantity

$$\begin{aligned} \Delta^2 &= \sum_{i=1}^{N} \Delta_i^2 \\ &= A_{i\alpha} A_{i\beta} x_\alpha x_\beta - 2A_{i\alpha} y_i x_\alpha + y_i y_i, \end{aligned} \tag{6.121}$$

where summation over repeated indices is implied.

Δ^2 is always positive or zero; in the latter case the solution is exact. The minimum condition on (6.121) is obtained by setting the partial derivatives of Δ^2 with respect to x_α ($\alpha = 1, 2, \ldots N$) equal to zero. After some straightforward algebra, we find

$$A_{i\alpha} A_{i\beta} x_\beta = A_{i\alpha} y_i, \tag{6.122}$$

or, in vector notation,

$$\mathbf{A}^{\mathbf{T}} \mathbf{A} \mathbf{x} = \mathbf{A}^{\mathbf{T}} \mathbf{y}, \tag{6.123}$$

where $\mathbf{x} = (x_1, x_2, \ldots x_N)$, $\mathbf{y} = (y_1, y_2, \ldots y_M)$, $\mathbf{A}$ is an $M \times N$ matrix whose elements are given by (6.108), and $\mathbf{A}^{\mathbf{T}}$, the transpose of $\mathbf{A}$, is an $N \times M$ matrix. Equation (6.123) contains the same number of equations as unknowns and can be inverted for $\mathbf{x}$

$$\mathbf{x} = (\mathbf{A}^{\mathbf{T}} \mathbf{A})^{-1} \mathbf{A}^{\mathbf{T}} \mathbf{y}. \tag{6.124}$$

This is the *least-squares solution* or the *Penrose pseudoinverse*. It is the same as (6.108) for $N = M$. With a suitable choice of N and of $b_i(z)$, the

least-squares solution can be much less sensitive to noise than are direct solutions.

(iii) Smooth solutions. The discussion in the previous two sections illustrates the difficulties in obtaining a good temperature inversion from the measurements alone. Additional information, based upon a priori knowledge, is, however, available and should be incorporated into the solution. For example, the basis functions, $b_i(z)$, may be chosen to be empirical orthogonal functions of the climatological data. The technique of Twomey concentrates, instead, upon the smoothness of the temperature field; whether or not the temperature is smooth in the sense that is imposed on the solution, the use made of the temperature field in numerical weather forecasts does involve a smoothness assumption. Twomey defines "smoothness" in a number of different ways; we consider one of them.

We construct a quantity corresponding to the second derivative of the unknown function, $\mathbf{x}$,

$$\boldsymbol{\delta}(\mathbf{x}) = (x_3 - 2x_2 + x_1, x_4 - 2x_3 + x_2, \ldots, x_N - 2x_{N-1} + x_{N-2}). \quad (6.125)$$

A squared form of this "derivative" can be shown to be $\mathbf{x}^{\mathbf{T}}\mathbf{H}\mathbf{x}$, where

$$\mathbf{H} = \begin{pmatrix} 1 & -2 & 1 & . & . & . & 0 & 0 & 0 \\ -2 & 5 & -4 & 1 & . & . & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & . & . & . & 1 & -2 & 1 \end{pmatrix} \quad (6.126)$$

$\mathbf{H}$ is a symmetric $N \times N$ matrix; $\mathbf{x}^{\mathbf{T}}$ is the transpose of $\mathbf{x}$. One way to ensure that the solution is smooth is to require that this quadratic quantity be a minimum. This condition can be combined with (6.121). Consider the functional

$$F(\mathbf{x}) = A_{i\alpha}A_{i\beta}x_\alpha x_\beta - 2A_{i\alpha}y_i x_\alpha + y_i y_i + \gamma H_{\alpha\beta}x_\alpha x_\beta, \quad (6.127)$$

where γ is an arbitrary smoothing constant. Requiring the functional to be stationary simultaneously solves (6.119) in the least-squares sense and guarantees that the solution is smooth. The derivation of the solution follows that from (6.121) to (6.124) and the result is

$$\mathbf{x} = (\mathbf{A}^{\mathbf{T}}\mathbf{A} + \gamma\mathbf{H})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{y}. \quad (6.128)$$

For $\gamma = 0$, (6.128) reduces to (6.124). The smoothness requirement ($\gamma \neq 0$) makes the matrix $(\mathbf{A}^{\mathbf{T}}\mathbf{A} + \gamma\mathbf{H})$ more diagonally dominant.

(iv) Statistical solutions. Both the measured intensities and the derived source functions are members of statistical ensembles; statistical methods are, therefore, appropriate to their study. In addition, information theory may be used with the statistical data to give objective measures of the value of different protocols.

The coefficients x_α in (6.107) are members of an ensemble whose mean value, $\bar{x}_\alpha$, is the climatological mean for the specified observational period. If each x_α is normally distributed with a variance σ^2 (for simplicity, we do not distinguish between different σ_α^2), the joint a priori probability density of N values of x_α is

$$P_0(\mathbf{x}) = \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{1/2}\sigma} \exp \frac{-(x_\alpha - \bar{x}_\alpha)^2}{2\sigma^2}. \tag{6.129}$$

The measured quantities, y_i ($i = 1$ to M), are subject to errors ϵ_i, also assumed to be normally distributed about a zero mean, with a single variance s^2. The joint probability density of the errors is

$$P_e(\boldsymbol{\epsilon}) = \prod_{i=1}^{M} \frac{1}{(2\pi)^{1/2}\sigma} \exp \frac{-\epsilon_i^2}{2s^2}. \tag{6.130}$$

Since y_i is subject to the error ϵ_i, we must replace (6.107) by

$$\sum_{\alpha=1}^{N} A_{i\alpha} x_\alpha = y_i + \epsilon_i \qquad i = 1, 2, \ldots M. \tag{6.131}$$

The probability density for the measurement $\mathbf{y}$ is now the joint probability of $\mathbf{x}$ and $\boldsymbol{\epsilon}$. After some algebra this can be shown to be proportional to $\exp[-F(\mathbf{x})/2]$, where

$$F(\mathbf{x}) = \frac{(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}})}{\sigma^2} + \frac{(\mathbf{A}\mathbf{x} - \mathbf{y})^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y})}{s^2}. \tag{6.132}$$

We now seek the distribution, $\mathbf{x}_P$, that leads to the most probable distribution of $\mathbf{y}$, i.e., that for which (6.132) is a minimum. The result can be shown to be

$$\mathbf{x}_P = [\mathbf{A}^{\mathrm{T}}\mathbf{A} + (s^2/\sigma^2)\mathbf{I}]^{-1}[\mathbf{A}^{\mathrm{T}}\mathbf{y} + (s^2/\sigma^2)\mathbf{x}], \tag{6.133}$$

where $\mathbf{I}$ is the identity matrix. In the limit $s^2/\sigma^2 \to 0$, this solution is the same as that obtained using Twomey's method (6.128), with $\mathbf{H} = (s^2/\sigma^2)\mathbf{I}$.

Many of the eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ may be very small. For example, the eigenvalue is identically zero for redundant measurements. This is the

reason for the unsatisfactory noise amplification of the least-squares solution, since small eigenvalues may occur in the denominator of (6.124). The term (s^2/σ^2) in (6.133) suppresses the influence of these small eigenvalues. This term serves the same function as does Twomey's empirical smoothing term, but is derived from a more secure foundation of climatological and measurement statistics.

We now turn to the information content of the measurements. For a statistical variable x_j with a probability density $P(x_j)$, the amount of information is usually measured by the change in the *entropy*,

$$S_j(P) = -\int_{-\infty}^{+\infty} P(x_j) \log_2 P(x_j)\, dx, \tag{6.134}$$

resulting from the performance of the measurement. The entropy may be regarded as the number of bits required to represent the number of distinct measurements that could have been made leading to the same x_j. For a gaussian distribution of a single variable [a single term in (6.129)],

$$S(P) = -\log_2[(2\pi e)^{1/2}\sigma]. \tag{6.135}$$

The definition of entropy can be extended to many normally distributed variables,

$$S(P) = \sum_j S_j(P_j). \tag{6.136}$$

For the most probable solution, (6.133), the increase in information over that which previously existed in the climatological data can be shown to be

$$\Delta S = (1/2) \sum_{\alpha=1}^{N} \log_2(1 + \sigma^2\lambda_\alpha), \tag{6.137}$$

where λ_α is an eigenvalue of $\mathbf{A}^T\mathbf{A}$.

From (6.137), we see that the information content is not necessarily increased by having a large number of measurements or discrete levels for the inverted data. Increasing the size of the matrix $\mathbf{A}$, and thereby increasing the number of eigenvalues, may have little influence upon the information available if many of the eigenvalues, λ_α, are small. In fact, given the measurement errors and the kernel functions, it is not possible to increase the information available over that given by the few largest eigenvalues. We may regard Chahine's method of inversion as an ingenious way to construct $\mathbf{A}$ such that most of the information available can be easily extracted. By choosing to invert the unknown quantities at

the maxima of the kernel functions, we are guaranteed a diagonally dominant matrix. The problem of redundancy is avoided and that of the mixing of unknowns is minimized. The solution is, therefore, stable.

BIBLIOGRAPHY

6.1. Introduction

6.1.1. *Line-by-line calculations*
Line-by-line calculations serve as a standards against which to compare approximate algorithms. Examples are

Scott, N. A., and Chedin, A., 1981, "A fast line-by-line method for atmospheric absorption computations: The automatized atmospheric absorption atlas," *J. Appl. Meteorol.* **20,** 802.

Fels, S. B., and Schwarzkopf, M. D., 1981, "An efficient, accurate algorithm for calculating CO_2 15 μm band cooling rates," *J. Geophys. Res.* **86,** 1205.

The calculations in Figs. 6.1 and 6.2 are from

Conrath, B. J., Hanel, R. A., Kunde, V. G., and Prabhakara, C. 1970, "The infrared interferometer experiment on Nimbus 3," *J. Geophys. Res.* **75,** 5831.

6.1.2. *The angular integration*
The diffusivity approximation has received attention out of proportion to its importance. The two-stream approximations (§ 2.4.5) are of this genre, and go back to the beginning of the century. The widely favored value $r = 1.66$ was first suggested by Elsasser (1943), see § 4.5. Godson (1954), see § 4.3, discussed the factor for both regular and random models. More recent papers are by

Armstrong, B. H., 1968, "Theory of the diffusivity factor for atmospheric radiation," *J. Quant. Spectrosc. Radiat. Transfer* **8,** 1577.

——— 1969, "The radiative diffusivity factor for the random Malkmus band," *J. Atmos. Sci.* **26,** 741.

6.2. Transmission through a nonhomogeneous atmosphere

6.2.1. *Exact solutions for constant mixing ratio*
The solution, (6.20), was first given by

Pedersen, F., 1942, "On the temperature-pressure effect on absorption of long-wave radiation by water vapor," *Meteorol. Ann.* (*Norwegian Meteorol. Inst.*) **1**(6).

Strong, J., and Plass, G. N., 1950, "The effect of pressure broadening of spectral lines on atmospheric temperature," *Astrophys. J.* **112,** 365.

The integral, (6.21), has been evaluated for discrete values of y by

Kaplan, L. D., 1952, "On the pressure dependence of radiative heat transfer in the atmosphere," *J. Meteorol.* **9,** 1.

Solutions to (6.19) and (6.21) in terms of hypergeometric functions have been given by

Yamamoto, G., and Aida, M., 1970, "Transmission in a non-homogeneous atmosphere with an absorbing gas of constant mixing ratio," *J. Quant. Spectrosc. Radiat. Transfer* **10,** 593.

6.2.2. *Scaling approximations*

The origin of the scaling approximation is obscure. Prior to the availability of computers it was the only feasible approach to nonhomogeneous atmospheric paths and was rediscovered many times.

The accuracy of the strong-line approximation is discussed by

Wu, M-L. C., 1980, "The exchange of infrared energy in the troposphere," *J. Geophys. Res.* **85,** 4084.

Chou, M-D., and Arking, A., 1980, "Computation of infrared cooling rates in the water vapor bands," *J. Atmos. Sci.,* **37,** 855.

Rodgers, C. D., and Walshaw, C. D., 1966, "The computation of infra-red cooling rate in planetary atmospheres," *Quart. J. Roy. Meteorol. Soc.* **92,** 67.

6.2.3. *The H–C–G approximation*

The H–C–G approximation was first proposed in a little known paper by

van de Hulst, H. C., 1945, "Theory of absorption lines in the atmosphere of the Earth," *Ann. Astrophys.* **8,** 1.

van de Hulst obtained his result both by the method given here and also by means of an ingenious expansion of the fourier transform of the optical depth. The fourier technique was used to give higher order approximations by

Goody, R. M., 1964, "The transmission of radiation through an inhomogeneous atmosphere," *J. Atmos. Sci.* **21,** 575.

The H–C–G approximation was later proposed for an isothermal atmosphere by

Curtis, A. R., 1952, "Contribution to a discussion of 'A statistical model for water vapour absorption', by R. M. Goody," *Quart. J. Roy. Meteorol. Soc.* **78,** 638.

The appropriate form for a nonisothermal atmosphere was given by

Godson, W. L., 1953, "The evaluation of infra-red radiative fluxes due to atmospheric water vapour," *Quart. J. Roy. Meteorol. Soc.* **79,** 367.

Errors arising from the H–C–G approximation have been studied by Goody (1964) and by

Walshaw, C. D., and Rodgers, C. D., 1963, "The effect of the Curtis–Godson approximation on the accuracy of radiative heating-rate calculations," *Quart. J. Roy. Meteorol. Soc.,* **89,** 122.

Approximations of higher order than H–C–G have been proposed by Goody (1964) and by

Armstrong, B. H., 1968, "Analysis of the Curtis–Godson approximation and radiation transmission through inhomogeneous atmospheres," *J. Atmos. Sci.* **25,** 312.

Yamamoto, G., Aida, M., and Yamamoto, S., 1972, "Improved Curtis–Godson approximation in a non-homogeneous atmosphere," *J. Atmos. Sci.* **29,** 1150.

Kurian, J. G., Mitra, S. K., and Ueyoshi, K., 1978, "Heating due to the 9.6 μm ozone band in an inhomogeneous atmosphere—a new approximation," *J. Quant. Spectrosc. Radiat. Transfer* **20,** 245.

Lindquist, G. H., and Simmons, F. S., 1972, "A band model formulation for very nonuniform paths," *J. Quant. Spectrosc. Radiat. Transfer* **12,** 807.

Young, S. J., 1975, "Band model formulation for inhomogeneous optical paths," *J. Quant. Spectrosc. Radiat. Transfer* **15,** 483.

The final two papers are concerned with an engineering problem involving very large temperature differences along an absorbing path.

The paper by Armstrong (1968) derives the H–C–G approximation from a first-order Gaussian quadrature to a pressure integral. This method, and van de Hulst's fourier transform technique, can both be taken to higher orders of approximation.

The calculations in Fig. 6.4 are from

Ellingson, R. G., and Gille, J. C., 1978, "An infrared radiative transfer model. Part 1: Model description and comparison of observations with calculations," *J. Atmos. Sci.* **35,** 523.

6.2.4. *Correlated k*

Although this method has been used for some years by the Institute of Space Studies, New York, very few details have been published. Brief mention is to be found in

Hansen, J., Russell, G. G., Rind, D., Stone, P., Lacis, A., Lebedeff, S., Ruedy, R., and Travis, L., 1983, "Efficient three-dimensional models for global climate studies: Models I and II," *Mon. Wea. Rev.* **111,** 609.

A more detailed treatment by Lacis and Oinas (1986, § 4.8) was available in draft form when Chapter 6 was revised.

6.3. Topics concerning heating rates

For many years the following paper provided a standard against which to test approximations to radiation algorithms:

Rodgers, C. D., and Walshaw, C. D., 1966, "The computation of infra-red cooling rate in planetary atmospheres," *Quart. J. Roy. Meteorol. Soc.* **92,** 67.

A recent survey of six major numerical models is by

Stephens, G. L., 1984, "The parameterization of radiation for numerical weather prediction and climate models," *Mon. Wea. Rev.* **112,** 826.

This paper has an extensive and valuable bibliography.

6.3.1. *The Chapman layer*
The theory was first advanced to explain the nature of ionized layers in the atmosphere in

Chapman, S., 1939, "The atmospheric height distribution of band-absorbed solar radiation," *Proc. Phys. Soc. London* **51,** 93.

6.3.2. *The Curtis matrix*
First used by

Curtis, A. R., 1956, "The computation of radiative heating rates in the atmosphere," *Proc. Roy. Soc. London Ser. A* **236,** 156.

The following papers develop a perturbation scheme for the effect of temperature upon the diffuse transmission of radiation by carbon dioxide between given pressure levels, from which the Curtis matrix elements can be obtained:

Fels, S. B., and Schwarzkopf, M. D., 1981, "An efficient, accurate algorithm for calculating CO_2 15 μm band cooling rates," *J. Geophys. Res.* **86,** 1205.

Schwarzkopf, M. D., and Fels, S. B., 1985, "Improvements to the algorithm for computing CO_2 transmissivities and cooling rates," *J. Geophys. Res.* **90,** 10,541.

The calculations in Fig. 6.10 are by Williams (1971) (§ 2.2).

6.3.3. *Calculations for the middle atmosphere*
The data in Table 6.2 are from Curtis and Goody (1956), (§ 2.2). The remainder of this section follows the work of Lopez–Puertas et al. (1986), (§ 2.2), and

Lopez–Puertas, M., Rodrigo, R., Lopez–Moreno, J. J., and Taylor, F. W., 1986, "A non-LTE radiative transfer model for infrared bands in the middle atmosphere. II. CO_2 (2.7 μm and 4.3 μm) and water vapour (6.3 μm) and $N_2(1)$ and $O_2(1)$ vibrational levels," *J. Atmos. Terr. Phys.* **48,** 749.

Other important papers are by Dickinson (1984), Kuhn and London (1969) (see § 2.2 for both), and

Kutepov, A. A. and Shved, G. M., 1978, "Radiative transfer in the 15 μm CO_2 band with non-LTE in the earth's atmosphere," *Atmos. Ocean. Phys.* **14,** 28.

Houghton, J. T., 1969, "Absorption and emission by carbon-dioxide in the mesosphere," *Quart. J. Roy. Meteorol. Soc.* **95,** 1.

Wehrbein, W. M., and Leovy, C. B., 1983, "An accurate radiative heating and cooling algorithm for use in a dynamical model of the middle atmosphere," *J. Atmos. Sci.*, **39,** 1532.

Apruzese, J. P., Strobel, D. F., and Schoeberl, M. R., 1984, "Parameterization

of IR cooling in a middle atmosphere dynamics model, 2, non-LTE radiative transfer and globally averaged temperature of the mesosphere and lower thermosphere," *J. Geophys. Res.* **89,** 4917.

6.4. Approximate methods

6.4.1. *Exchange of radiation with the boundaries*

The first use of radiation-to-space as an approximation was by Curtis (1956), (§ 6.3.2). The importance of the approximation for all atmospheric gases and at all atmospheric levels was demonstrated by Rodgers and Walshaw (1966), (§ 6.2.2).

The following paper draws a formal distinction between exchange of radiation with the boundaries and exchange with other layers of the atmosphere:

Green, J. S. A., 1967, "Division of radiative streams into internal transfer and cooling to space," *Quart. J. Roy. Meteorol. Soc.* **93,** 371.

An assessment of the importance of both boundary exchange terms has been made by

Fels, S. B., and Schwarzkopf, M. D., 1975, "The simplified exchange approximation: A new method for radiative transfer calculations," *J. Atmos. Sci.* **32,** 1475.

Boundary exchange approximations can be improved, without a serious increase in complexity, by employing a trapezoidal approximation for the integral over atmospheric layers:

Gierasch, P., and Goody, R., 1967, "An approximate calculation of radiative heating and radiative equilibrium in the martian atmosphere," *Planet. Space Sci.* **15,** 1465.

6.4.2. *Use of emissivities*

Before the advent of large computers all investigators made some use of emissitivies. Elsasser (1943), (§ 4.5), tabulates ϵ^*. Improved data were given by

Elsasser, W. M., and Culbertson, M. F., 1961, *Atmospheric radiation tables.* Boston, Mass.: American Meteorological Society, Meteorological Monographs **4,** No. 23.

A careful study of the relationship between ϵ, ϵ^*, and empirically derived emissitivities is by

Rodgers, C. D., 1967, "The use of emissivity in atmospheric radiation calculations," *Quart. J. Roy. Meteorol. Soc.* **93,** 43.

Different calculations of ϵ and ϵ^* are compared by

Sasamori, T., 1986, "The radiative cooling calculation for application to general circulation experiments," *J. Appl. Meteorol.* **7,** 721.

Tabulations of emissivities for water vapor, carbon dioxide, and ozone, together with overlap corrections, are given by

Staley, D. O., and Jurica, G. M., 1970, "Flux emissivity tables for water vapor, carbon dioxide and ozone," *J. Appl. Meteorol.* **9,** 365.

For the use of emissivities in numerical radiation models, see Stephens (1984), (§ 6.3). One example is

Ramanathan, V., Pitcher, E. J., Malone, R. C., and Blackmon, M. C., 1983, "The response of a spectral general circulation model to refinements in radiative processes," *J. Atmos. Sci.*, **40,** 605.

The importance of nonisothermal emissivities is discussed by

Ramanathan, V., and Downey, P., 1986, "A non-isothermal emissivity and absorptivity formulation for water vapor," *J. Geophys. Res.* **91,** 8649.

The data for Fig. 6.15 are from Fels and Schwarzkopf (1975), (§ 6.4.1).

6.4.3. *Radiation charts*
The first radiation chart was proposed by

Mügge, R., and Möller, F., 1932, "Zur Berechnung von Strahlungsströmen und Temperaturänderungen in Atmosphären von beliebigem Aufbau," *Z. Geophys.* **8,** 53.

It was put into a practical form by

Möller, F., 1943, *Das Strahlungsdiagram.* Reichsampt für Wetterdienst (Luftwaffe).

The first chart available in the English language was by Elsasser (1943), (§ 4.3). Details of an earlier soviet chart, prepared in 1940 by Dimitriev, are not available.

Yamamoto's chart is described by

Yamamoto, G., 1952, "On a radiation chart," *Sci. Rep. Tohoku Univ., Ser. 5, Geophys.* **4**(1), 9.

An example of a heating rate chart is

Yamamoto, G., and Onishi, G., 1953, "A chart for the calculation of radiative temperature changes," *Sci. Rep. Tohoku Univ., Ser. 5, Geophys.* **4**(3), 108.

The simplest chart on record (the Kew chart) uses emissivities that are independent of temperature and, consequently, has linear isopleths:

Robinson, G. D., 1950, "Notes on the measurement and estimation of atmospheric radiation," *Quart. J. Roy. Meteorol. Soc.* **76,** 37.

6.5. The inverse problem for thermal radiation

6.5.1. *Kernel functions*
The fundamental ideas in remote sensing were developed by solar astronomers during the 1940s and 1950s. Conclusions were drawn from solar limb darkening and, when the absorption spectrum of negative hydrogen became known, spectral methods could also be used. The solar problem is more difficult than that discussed in this chapter because absorber concentrations are not known a priori, and must be calculated self-consistently, and also because thermodynamic

equilibrium cannot be assumed. For references, see any book on solar physics, e.g.,

Athay, R. G., 1976, *The Solar Chromosphere and Corona.* Dordrecht: Reidel.

The first systematic attempts to perform remote sensing in the earth's atmosphere were for the purpose of obtaining the vertical distribution of ozone from twilight scattering measurements (Gotz's *Umkehr* method). For references, see

Goody, R. M., 1954, *The physics of the stratosphere.* London: Cambridge University Press.

The first attempts to perform inversions on the atmospheric thermal spectrum were also for the purpose of measuring atmospheric ozone:

Epstein, E. S., Osterberg, C., and Adel, A., 1956, "A new method for the determination of the vertical distribution of atmospheric ozone from a ground station," *J. Meteorol.* **13,** 319.

Goody, R. M., and Roach, W. T., 1956, "The determination of the vertical distribution of ozone from emission spectra," *Quart. J. Roy. Meteorol. Soc.* **82,** 217.

Walshaw, C. D., 1960, "The accuracy of determination of the vertical distribution of atmospheric ozone from emission spectrophotometry in the 1043 cm^{-1} band at high resolution," *Quart. J. R. Meteorol. Soc.* **86,** 519.

At the same time, discussion began on the possibility of remote soundings from satellites. The following book contains two important articles, the first by S. F. Singer on ozone soundings in the ultraviolet spectrum and the second by J. I. F. King on the use of angular scans in the thermal spectrum:

van Allen, J. A. (Ed.), 1956, *Scientific uses of earth satellites.* Ann Arbor: University of Michigan Press.

The first proposal to use the spectral properties of the 15 μm CO_2 band, the method that forms the basis of most meteorological soundings, was by

Kaplan, L. D., 1959, "Inference of atmospheric structure from remote radiation measurements," *J. Opt. Soc. Am.* **49,** 1004.

The following three books treat the subject of remote sensing in detail:

Houghton, J. T., Taylor, F. W., and Rodgers, C. D., 1984, *Remote sounding of atmospheres.* London: Cambridge University Press.

Deepak, A. (Ed.), 1977, *Inversion methods in remote atmospheric sounding.* New York: Academic Press.

Twomey, S., 1977, *Introduction to the mathematics of inversion in remote sensing and indirect measurements.* Amsterdam: Elsevier.

Table 6.3 follows

Rodgers, C. D., 1971, "Some theoretical aspects of remote sounding in the earth's atmosphere," *J. Quant. Spectrosc. Radiat. Transfer* **11,** 767.

The discussion of angular scans on Venus is from

Goody, R., 1965, "The structure of the Venus cloud veil," *J. Geophys. Res.* **70,** 5471.

Limb scans are discussed by

Gille, J. C., and Bailey, P. L., 1977, "Inversion of infrared limb emission measurements for temperature and trace gas concentrations," Deepak (1977), § 6.5.1, 195.

Gille, J. C., and House, F. B., 1971, "On the inversion of limb radiance measurements: I. Temperature and thickness," *J. Atmos. Sci.* **28,** 1427.

6.5.2. *A "physical" approach to retrieval*
Chahine's method was developed in a series of articles

Chahine, M. T., 1968, "Determination of the temperature profile in an atmosphere from its outgoing radiance," *J. Opt. Soc. Am.* **58,** 1634.

———, 1970, "Inverse problems in radiative transfer: Determination of atmospheric parameters," *J. Atmos. Sci.* **27,** 960.

———, 1972, "A general relaxation method for inverse solution of the full radiative transfer equation," *J. Atmos. Sci.* **29,** 741

Barcilon, V., 1975, "On Chahine's relaxation method for the radiative transfer equation," *J. Atmos. Sci.* **32,** 1626.

Figures 6.17 and 6.19 are from

Chahine, M. T., 1977a, "Generalization of the relaxation method for the inverse solution of nonlinear and linear transfer equations," Deepak (1977), § 6.5.1, 67.

The theory of remote sensing in the presence of clouds was developed by M. T. Chahine in another series of papers:

Chahine, M. T., 1977b, "Remote sounding of cloudy atmospheres. II. Multiple cloud formations," *J. Atmos. Sci.* **34,** 744.

Aumann, H. H., and Chahine, M. T., 1976, "An infrared multidetector spectrometer for remote sensing of temperature profiles in the presence of clouds," *Appl. Opt.* **15,** 2091.

Chahine, M. T., Aumann, H. H., and Taylor, F. W., 1977, "Remote sounding of cloudy atmospheres. III. Experimental verifications," *J. Atmos. Sci.* **34,** 758.

6.5.3. *Linear analysis*
The presentation closely follows a review article by

Rodgers, C. D., 1975, "Retrieval of atmospheric temperature and composition

from remote measurements of thermal radiation," *Rev. Geophys. Space Phys.* **14,** 609.

Figures 6.21, 6.22, and 6.23 are all from Houghton et al. (1984), § 6.5.1.
The method of least squares for ill-conditioned matrices is discussed by

Penrose, R., 1955, "A generalized inverse for matrices," *Proc. Camb. Phil. Soc.* **51,** 406.

Twomey's method was developed by

Twomey, S., 1963, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," *J. Assoc. Comput. Mach.* **10,** 97.

———, 1965, "The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements," *J. Franklin Inst.* **279,** 95.

Tikhonov, A. N., 1963, "On the solution of incorrectly stated problems and a method of regularization," *Dokl. Akad. Nauk. USSR* **15,** 501 (p. 1035 in the English edition).

See also Twomey's book, § 6.5.1.
The concept and the implemenation of the statistical method are by

Westwater, E. R., and Strand, O. N., 1968, "Statistical information content of radiation measurements used in indirect sensing," *J. Atmos. Sci.* **25,** 750.

Turchin, V. F., and Nozik, V. Z., 1969, "Statistical regularization of the solution of incorrectly posed problems," *Atmos. Ocean. Phys.* **5,** 14.

Rodgers, C. D., 1970, "Remote sounding of the atmospheric temperature profiles in the presence of cloud," *Quart. J. Roy. Meteorol. Soc.* **96,** 654.

DeLuisi, J. J., and Mateer, C. L., 1971, "On the application of the optimum statistical inversion technique to the evaluation of Umkehr observations," *J. Appl. Meteorol.* **10,** 328.

See also, Rodgers (1977), § 6.5.1.
The application of information theory to measurements, as described in this section, follows

Wiener, N., 1948, *Cybernetics.* Cambridge, Mass.: MIT Press.

7

EXTINCTION BY MOLECULES AND DROPLETS

7.1. The problem in terms of the electromagnetic theory

The formal theory developed in Chapter 2 assumed the Stokes parameters to be additive. The sufficient condition for additivity is that the radiation fluxes in the atmosphere shall have no *phase coherence.* Thermal emission from independently excited molecules is necessarily incoherent with respect to phase. Atmospheric scattering centers are widely and randomly spaced, and they can be treated as independent and incoherent scatterers. The situation differs, however, when we consider details of the scattering process within a single particle, and in order to derive the extinction coefficient and the scattering matrix (see § 2.1.3) we must make use of a theoretical framework that involves the phase explicitly.

The problem of the interaction between an electromagnetic wave and a dielectric particle can be precisely formulated using Maxwell's equations. For a plane wave and a spherical particle, *Mie's theory* provides a complete solution (see § 7.6). But the general problem is complicated and our understanding is rendered more difficult by preconceptions based on the approximations of elementary optics. This chapter provides a brief survey of the important results and the underlying concepts.

The geometry of the problem is illustrated in Fig. 7.1. An isolated particle is irradiated by an incident, plane electromagnetic wave. The plane wave preserves its character only if it propagates through a homogeneous medium; the presence of the scattering particle, with electric and magnetic properties differing from those of the surrounding medium, distorts the wave front. The disturbance has two aspects: first, the plane wave is diminished in amplitude; second, at distances from the particle that are large compared with the wavelength and particle size, there is an additional, outward-traveling spherical wave. The energy carried by this spherical wave is the *scattered* energy; the total energy lost by the plane wave corresponds to *extinction*; the difference is the *absorption.*

The properties of the spherical wave in one particular direction (the *line of sight*) will be considered. This direction can be specified by the

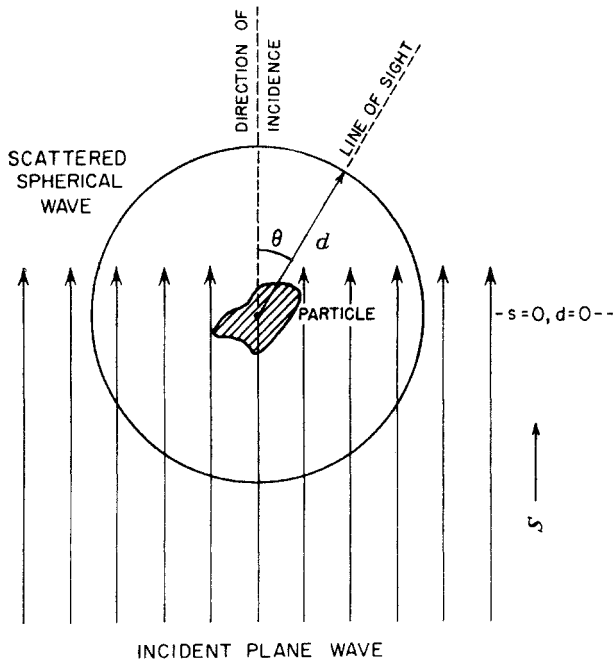


FIG. 7.1. Definition of scattering angle θ.

scattering angle θ (see Fig. 7.1) in a plane containing both the incident and scattered wave normals (*the plane of reference*), and the azimuth angle ϕ between the plane of reference and a plane fixed in space. For spherical particles, from symmetry considerations, scattering will be independent of the azimuth angle.

An electromagnetic wave is characterized by electric and magnetic vectors **E** and **H** that form an orthogonal set with the direction of propagation of the wave (i.e., the direction of the wave normal). In any one medium $|\mathbf{E}|$ and $|\mathbf{H}|$ are related and, since we will examine the properties of the radiation in air surrounding a scattering particle, we may, without loss of generality, speak in terms of the electric vector only. The *direction of polarization* is defined as the direction of the electric vector.

Let **l** and **r** be two unit vectors that form an orthogonal set with the direction of propagation, respectively, parallel to and perpendicular to the *plane of reference.*[1] Note that this defines two directions *that depend upon the direction of the line of sight and are not fixed in space.* If $E^{(l)}$ and $E^{(r)}$ are the complex amplitudes of the parallel and perpendicular

[1] *l* and *r* are the last letters of the words parall*el* and perpendicul*ar*, respectively.

components then

$$\mathbf{E} = E^{(l)}\mathbf{l} + E^{(r)}\mathbf{r}, \tag{7.1}$$

and, for a general (elliptically polarized) wave,

$$\begin{aligned} E^{(l)} &= |E^{(l)}|\, e^{-i\delta} e^{-2\pi i s/\lambda + 2\pi i \nu t}, \\ E^{(r)} &= |E^{(r)}|\, e^{-2\pi i s/\lambda + 2\pi i \nu t}, \end{aligned} \tag{7.2}$$

where δ is the phase difference between the two components and s is the distance along the direction of propagation. λ varies from medium to medium and may be written $\lambda_0/\tilde{m}$, where $\tilde{m}$ is the refractive index.[2] In general $\tilde{m}$ is complex,

$$\tilde{m} = \tilde{n} - i\tilde{n}', \tag{7.3}$$

where $\tilde{n}$ and $\tilde{n}'$ are both real. From (7.3) and (7.2), the complex index leads to an attenuation factor $e^{-2\pi\tilde{n}'s/\lambda_0}$ in the amplitude of the electric vector.

The energy carried by an electromagnetic wave is measured by the Poynting vector, directed along the wave normal, whose magnitude is

$$\begin{aligned} N &= |\mathbf{E}|^2 \\ &= |E^{(l)}|^2 + |E^{(r)}|^2. \end{aligned} \tag{7.4}$$

Since it is irrelevant to our discussion, no constant of proportionality is included in (7.4), thus avoiding questions associated with the choice of electromagnetic units.

It is a straightforward matter to show that there is a vector $(N^{(1)}, N^{(2)}, N^{(3)}, N^{(4)})$ corresponding to $(I, Q, U, V)\, d\omega$, with components

$$\begin{aligned} N^{(1)} &= E^{(l)}E^{(l)*} + E^{(r)}E^{(r)*}, \\ N^{(2)} &= E^{(l)}E^{(l)*} - E^{(r)}E^{(r)*}, \\ N^{(3)} &= E^{(l)}E^{(r)*} + E^{(r)}E^{(l)*}, \\ N^{(4)} &= i(E^{(l)}E^{(r)*} - E^{(r)}E^{(l)*}), \end{aligned} \tag{7.5}$$

where the asterisk denotes the complex conjugate.

If we apply these definitions to the elliptically polarized wave (7.2),

[2] The tilde denotes the optical properties of the particle. Properties of the entire medium, consisting of small particles in space or suspended in a gas, will be designated m, n, and n'.

we obtain

$$\begin{aligned} N^{(1)} &= |E^{(l)}|^2 + |E^{(r)}|^2, \\ N^{(2)} &= |E^{(l)}|^2 - |E^{(r)}|^2, \\ N^{(3)} &= 2\,|E^{(l)}|\,|E^{(r)}|\cos\delta, \\ N^{(4)} &= 2\,|E^{(l)}|\,|E^{(r)}|\sin\delta. \end{aligned} \tag{7.6}$$

If the direction of polarization happens to be **r** or **l**, then $|E^{(l)}|$ or $|E^{(r)}|$ is zero, and $N^{(3)} = N^{(4)} = 0$. Natural or unpolarized light can be looked upon as an incoherent sum of two beams polarized at right angles. We may, for convenience, take these two beams to be polarized in the **l** and **r** directions. If there is no phase coherence, the Stokes parameters add, and $N^{(2)}$, $N^{(3)}$, and $N^{(4)}$, are, therefore, all zero for natural light.

7.2. Scattering functions

Let the components of the electric vector be $E_0^{(l)}$ and $E_0^{(r)}$ for the incident wave and $E^{(l)}$ and $E^{(r)}$ for the scattered spherical wave. For distances (d), large compared with the wavelength and the particle size, $E^{(l)}$ and $E^{(r)}$ fall off as d^{-1}. Moreover, the scattered wave will possess a phase difference $2\pi i(d-s)/\lambda$ from the incident wave (see Fig. 7.1). When writing a formal relationship between incident and scattered amplitudes it is convenient to take out a factor

$$\left(\frac{2\pi i d}{\lambda_0}\right)^{-1} e^{-2\pi i(d-s)/\lambda_0}.$$

According to Maxwell's equations, a linear relationship exists between scattered and incident amplitudes. We can write in matrix form

$$\begin{pmatrix} E^{(l)} \\ E^{(r)} \end{pmatrix} = \left(\frac{2\pi i d}{\lambda_0}\right)^{-1} e^{-2\pi i(d-s)/\lambda_0} \begin{pmatrix} S_2 & S_3 \\ S_4 & S_1 \end{pmatrix} \begin{pmatrix} E_0^{(l)} \\ E_0^{(r)} \end{pmatrix}. \tag{7.7}$$

The matrix $\begin{pmatrix} S_2 & S_3 \\ S_4 & S_1 \end{pmatrix}$ is known as the *amplitude scattering matrix*; the unusual numbering of elements follows established convention. For homogeneous, spherical scatterers (the only case that we will consider in detail) S_3 and S_4 vanish. Our discussion will therefore be limited to S_2 and S_1; for spherical scatterers these are functions of scattering angle (θ) only.

A particularly important role is played in the theory by the matrix coefficients for $\theta = 0$. Consider a thin slab of material of thickness ds

containing N scattering centers per unit volume, and a wave incident from one side. The electric field on the far side can be found by compounding the incident wave with all scattered waves, taking due account of the phase. Consider a scalar wave, with a one-component scattering matrix. The resultant amplitude E_0' can be shown to be

$$E_0' = E_0\left[1 - \frac{\lambda_0^2 N\, ds S(0)}{2\pi}\right]. \tag{7.8}$$

Looking at the problem from a different viewpoint we may now suppose the slab to have a complex refractive index $m = n - in'$. The amplitude can be written (7.2)

$$\frac{E_0'}{E_0} = e^{-2\pi i ds(m-1)/\lambda_0} = 1 - \frac{2\pi i\, ds(m-1)}{\lambda_0}, \tag{7.9}$$

since ds is infinitesimally small. Comparing (7.8), (7.9), and (7.3), we find

$$n = 1 + 2\pi N(2\pi/\lambda_0)^{-3}\mathcal{I}\{S(0)\}, \tag{7.10}$$

$$n' = 2\pi N(2\pi/\lambda_0)^{-3}\mathcal{R}\{S(0)\}, \tag{7.11}$$

where $\mathcal{R}$ and $\mathcal{I}$ denote real and imaginary components. Relations similar to (7.10) and (7.11) can be shown to hold for vector waves, but, in the case of spherical particles, it is clear from symmetry considerations that for $\theta = 0$ both states of polarization will be similarly affected, and therefore $S_1(0) = S_2(0)$. Equations (7.10) and (7.11), therefore, also apply to spherical particles. Since intensity is proportional to the square of the amplitude, it involves an attenuating factor $e^{-4\pi n'\, ds/\lambda_0}$. From the definition of extinction coefficient, (2.15), it follows that

$$e_v = \frac{4\pi n'}{\lambda_0}. \tag{7.12}$$

Since the theory of Chapter 2 is given in terms of the phase matrix, we must show how this quantity can be related to the amplitude scattering matrix. From (7.5) and (7.7) we can derive a linear relation between scattered and incident Poynting vectors. Extracting a factor $(\lambda_0/2\pi d)^2$, we may write

$$N^{(j)} = \left(\frac{\lambda_0}{2\pi d}\right)^2 F_{ij}(\theta, \phi) N_0^{(i)}, \tag{7.13}$$

where the sum convention for repeated indices is employed. Equation (7.13) can be compared with (2.29) and (2.30) if we bear in mind that

(7.13) applies to a single particle only. Make the transformation

$$N \rightarrow I\, d\omega, \tag{7.14}$$

and take account of the spherical nature of the scattered wave by writing $d\omega_d = d^{-2}$. There results

$$\left(\frac{\lambda_0}{2\pi}\right)^2 F_{ij}(\theta, \phi) \equiv \frac{s_n}{4\pi} P_{ij}(\mathbf{s}, \mathbf{d}), \tag{7.15}$$

where s_n is the scattering coefficient per particle (see Appendix 2).

The scattering coefficient can be evaluated independently by applying the first law of thermodynamics to the intensity. Let us suppose that the incident light is unpolarized $\{(N_0^{(2)}, N_0^{(3)}, N_0^{(4)}) = 0\}$ and inquire about the scattered *intensity* $(N^{(1)})$. From (7.13)

$$N^{(1)} = \left(\frac{\lambda_0}{2\pi d}\right)^2 F_{11}(\theta, \phi) N_0^{(1)}.$$

Now integrate over the surface of a sphere of radius d to discover the total scattered component of the Poynting vector. The fraction scattered by a single particle can be equated to s_n, and hence

$$s_n = \int \left(\frac{\lambda_0}{2\pi}\right)^2 F_{11}(\theta, \phi)\, d\omega. \tag{7.16}$$

s_n has the dimensions of an area, and is conveniently made nondimensional by dividing by the cross-sectional area of the particle (πr^2, where r = radius of the particle). The result is a *scattering efficiency factor*

$$Q_s = \int \frac{F_{11}(\theta, \phi)\, d\omega}{\pi x^2}, \tag{7.17}$$

where $x = 2\pi r/\lambda_0$. Similarly, from (7.12) and (7.11), we can define an *extinction efficiency factor*

$$Q_e = \frac{e_n}{\pi r^2} = \frac{e_v}{N\pi r^2} = \frac{4}{x^2} \mathscr{R}\{S(0)\}. \tag{7.18}$$

The difference,

$$Q_a = Q_e - Q_s, \tag{7.19}$$

is the *absorption efficiency factor.*

Now let

$$i_1 = |S_1|^2,$$
$$i_2 = |S_2|^2, \tag{7.20}$$

and let δ be the phase difference between S_1 and S_2 (both complex in the general case). Substituting (7.7) in (7.5), with $S_3 = S_4 = 0$, we find

$$\begin{aligned}
N^{(1)} &= \left(\frac{\lambda_0}{2\pi d}\right)^2 [\tfrac{1}{2}(i_1 + i_2)N_0^{(1)} + \tfrac{1}{2}(i_2 - i_1)N_0^{(2)}], \\
N^{(2)} &= \left(\frac{\lambda_0}{2\pi d}\right)^2 [\tfrac{1}{2}(i_2 - i_1)N_0^{(1)} + \tfrac{1}{2}(i_1 + i_2)N_0^{(2)}], \\
N^{(3)} &= \frac{\sqrt{(i_1 i_2)}}{(\lambda_0/2\pi d)^2}(N_0^{(3)} \cos\delta - N_0^{(4)} \sin\delta), \\
N^{(4)} &= \frac{\sqrt{(i_1 i_2)}}{(\lambda_0/2\pi d)^2}(N_0^{(3)} \sin\delta + N_0^{(4)} \cos\delta).
\end{aligned} \tag{7.21}$$

The transformation matrix is, therefore,

$$F_{ij} = \begin{pmatrix} \tfrac{1}{2}(i_1 + i_2) & \tfrac{1}{2}(i_2 - i_1) & 0 & 0 \\ \tfrac{1}{2}(i_2 - i_1) & \tfrac{1}{2}(i_1 + i_2) & 0 & 0 \\ 0 & 0 & \sqrt{(i_1 i_2)} \cos\delta & -\sqrt{(i_1 i_2)} \sin\delta \\ 0 & 0 & \sqrt{(i_1 i_2)} \sin\delta & \sqrt{(i_1 i_2)} \cos\delta \end{pmatrix}. \tag{7.22}$$

7.3. Rayleigh's solution for small particles

One confusing aspect of scattering theory is that a complete formal solution (Mie's theory) exists for homogeneous spheres, which sometimes seems to differ from approximate solutions applicable in certain limiting cases. Examples of such limiting cases are ray optics, Huygens' principle, Fresnel's theory of diffraction, and Rayleigh's theory of molecular scattering. Mie's theory contains all the diverse phenomena of classical optics, and is difficult to comprehend in simple terms. Despite its generality, the complete theory cannot always replace simple asymptotic forms. We first consider the limit of very small particles.

The phase change along the radius of a sphere is $2\pi a\,|\tilde{m}|/\lambda_0$. If this is small, i.e., if $|\tilde{m}x| \ll 1$, then the impressed electric field is constant throughout the particle and equal to $\mathbf{E}_0$. If the polarizability tensor is $\boldsymbol{\alpha}$, then the induced dipole moment is, by definition,

$$\mathbf{M} = \boldsymbol{\alpha}\mathbf{E}_0. \tag{7.23}$$

We now assume that the scattered radiation corresponds to that emitted by the dipole. This neglects interactions between the dipole and the radiation field, with consequences that appear later. The classical solution by Hertz for the wave propagating in space for $d \gg \lambda_0$ is

$$E^{(l),(r)} = \frac{1}{d\mathbf{c}^2}\frac{\partial^2 M^{(l),(r)}}{\partial t^2}\sin\gamma^{(l),(r)}$$

$$= (2\pi/\lambda_0)^2\frac{M^{(l),(r)}}{d}\sin\gamma^{(l),(r)}, \tag{7.24}$$

where γ is the angle between $\mathbf{M}$ and the direction of observation, and we have replaced the operator $\partial/\partial t$ by $2\pi i\nu = 2\pi i\mathbf{c}/\lambda_0$.

Rayleigh's original theory assumed a scalar polarizability (α), which is correct for a sphere. Then $\mathbf{M}$ and $\mathbf{E}_0$ are parallel and the cross terms in the scattering matrix are zero. Since $\mathbf{r}$ is, by definition, perpendicular to the plane of reference it follows that this component corresponds to viewing the induced dipole sideways (see Fig. 7.2) and $\sin\gamma^{(r)} = 1$. For parallel polarization on the other hand

$$\sin\gamma^{(l)} = \cos\theta.$$

Combining (7.23) and (7.24)

$$E^{(r)} = \left(\frac{2\pi}{\lambda_0}\right)^2\frac{\alpha}{d}E_0^{(r)},$$

$$E^{(l)} = \left(\frac{2\pi}{\lambda_0}\right)^2\frac{\alpha}{d}E_0^{(l)}\cos\theta, \tag{7.25}$$

giving a scattering matrix

$$\mathbf{S} = i\left(\frac{2\pi}{\lambda_0}\right)^3\alpha\begin{pmatrix}\cos\theta & 0\\ 0 & 1\end{pmatrix}, \tag{7.26}$$

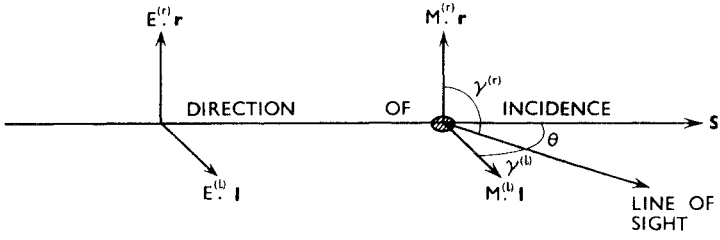


FIG. 7.2. Dipole scattering. By definition, $\mathbf{l}$, $\mathbf{s}$, and the line of sight all lie in the plane of reference. From the definition of scattering angle, $\theta + \gamma^{(l)} = \frac{1}{2}\pi$ and $\gamma^{(r)} = \frac{1}{2}\pi$.

and a transformation matrix

$$F_{ij} = \left(\frac{2\pi}{\lambda_0}\right)^6 |\alpha|^2 \begin{pmatrix} \frac{1}{2}(1+\cos^2\theta) & -\frac{1}{2}\sin^2\theta & 0 & 0 \\ -\frac{1}{2}\sin^2\theta & \frac{1}{2}(1+\cos^2\theta) & 0 & 0 \\ 0 & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & \cos\theta \end{pmatrix}. \quad (7.27)$$

An alternative form of the transformation matrix, often quoted in the literature, is

$$F'_{ij} = \left(\frac{2\pi}{\lambda_0}\right)^6 |\alpha|^2 \begin{pmatrix} \cos^2\theta & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & \cos\theta \end{pmatrix}. \quad (7.28)$$

This is the appropriate form if $(I^{(l)}, I^{(r)}, U, V)$ are chosen in place of (I, Q, U, V) as Stokes parameters. The apparent advantage of simplicity is illusory, but it makes more obvious the complete perpendicular polarization of scattered radiation for $\theta = \frac{1}{2}\pi$.

In Fig. 7.3, the results are shown graphically on a *scattering diagram.* Incidence is from left to right. The length of a radius vector at the scattering angle θ from the central point gives the scattered intensity. The scattering depends upon the polarization of the incident radiation. Three possibilities are shown: for polarization in the r or l directions, and for natural or unpolarized radiation. Note that the scale is not the same in each case. According to (7.27), the only relevant matrix element for natural radiation ($Q = U = V = 0$) is $(2\pi/\lambda_0)^6 |\alpha|^2 \frac{1}{2}(1+\cos^2\theta)$, while, according to (7.28), the matrix elements for parallel and perpendicular polarization are $(2\pi/\lambda_0)^6 |\alpha|^2 \cos^2\theta$ and $(2\pi/\lambda_0)^6 |\alpha|^2$, respectively. The factor 1/2 means that, for natural light, half of the intensity is to be attributed to each state of polarization.

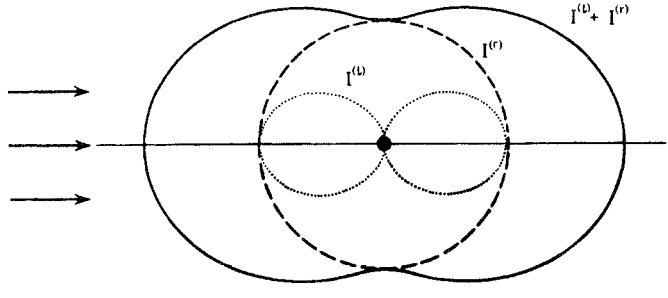


FIG. 7.3. Scattering diagram for small particles. Solid line = $I^{(r)} + I^{(l)}$; broken line = $I^{(r)}$; dotted line = $I^{(l)}$.

Using a prime to denote scattering of natural light

$$s_n \frac{P'(\cos\theta)}{4\pi} = \left(\frac{\lambda_0}{2\pi}\right)^2 F(\theta) = \left(\frac{2\pi}{\lambda_0}\right)^4 |\alpha|^2 \tfrac{1}{2}(1+\cos^2\theta). \tag{7.29}$$

This is all that is required for primary scattering of solar radiation, multiplying (7.29) by $d\omega$ and integrating over all solid angles

$$\begin{aligned} s_n &= \frac{1}{2}\left(\frac{2\pi}{\lambda_0}\right)^4 |\alpha|^2 \int (1+\cos^2\theta)\, d\omega \\ &= \frac{8\pi}{3}\left(\frac{2\pi}{\lambda_0}\right)^4 |\alpha|^2. \end{aligned} \tag{7.30}$$

The scalar phase function for incident unpolarized radiation is, therefore,

$$P'(\cos\theta) = \tfrac{3}{4}(1+\cos^2\theta). \tag{7.31}$$

According to (7.26) **S** is imaginary, and according to (7.11) m is real. Eliminating $S_1(0)$ or $S_2(0)$ between (7.10) and (7.26), we find

$$\alpha = \frac{m-1}{2\pi N}, \tag{7.32}$$

and, for m close to unity, as for a gas

$$s_n = \frac{32\pi^3}{3\lambda_0^4}\frac{(m-1)^2}{N^2} \simeq \frac{8\pi^3}{3\lambda_0^4}\frac{(m^2-1)^2}{N^2}. \tag{7.33}$$

This is the usual form of Rayleigh's inverse fourth-power law of scattering. Its important application is to molecular scattering. Only spherical top molecules have a scalar polarizability, however, and a small, but significant, correction term must therefore be included for nonspherical molecules. Let the tensor components of $\boldsymbol{\alpha}$, as referred to the three principal axes, be α_1, α_2, and α_3. We define

$$\begin{aligned} A &= \tfrac{1}{15}(\alpha_1\alpha_1^* + \alpha_2\alpha_2^* + \alpha_3\alpha_3^*), \\ B &= \tfrac{1}{30}(\alpha_1\alpha_2^* + \alpha_2\alpha_3^* + \alpha_3\alpha_1^* + \alpha_2\alpha_1^* + \alpha_3\alpha_2^* + \alpha_1\alpha_3^*). \end{aligned} \tag{7.34}$$

Since in meteorological problems we are interested in the combined effect of many randomly oriented particles, it is permissible to take suitable averages of the transformation matrix. The matrix F'_{ij} (7.28)

becomes

$$F'_{ij} = \left(\frac{2\pi}{\lambda_0}\right)^6 \times \begin{pmatrix} (2A+3B)\cos^2\theta + A - B & A-B & 0 & 0 \\ A-B & 3A+2B & 0 & 0 \\ 0 & 0 & (2A+3B)\cos\theta & 0 \\ 0 & 0 & 0 & 5B\cos\theta \end{pmatrix}. \tag{7.35}$$

The most important difference between (7.28) and (7.35) lies in the polarization at a scattering angle $\theta = \frac{1}{2}\pi$. For incident natural light ($I_0^{(r)} = I_0^{(l)}$), (7.35) gives

$$\Delta = \frac{I^{(l)}(\theta = \frac{1}{2}\pi)}{I^{(r)}(\theta = \frac{1}{2}\pi)} = \frac{2A - 2B}{4A + B}, \tag{7.36}$$

where Δ is the *depolarization factor*, a quantity that lends itself relatively easily to measurement in the laboratory.

For natural light (but not for other polarizations), the scattering coefficient and phase function can be specified in terms of Δ alone,

$$s_n = \frac{8\pi^3}{3\lambda_0^4}\frac{(m^2-1)^2}{N^2}\frac{3(2+\Delta)}{(6-7\Delta)}, \tag{7.37}$$

$$P'(\cos\theta) = \frac{3}{2(2+\Delta)}\{1 + \Delta + (1-\Delta)\cos^2\theta\}. \tag{7.38}$$

Depolarization factors of atmospheric constituents are as follows: $\Delta(O_2) = 0.054$, $\Delta(N_2) = 0.0305$, $\Delta(CO_2) = 0.0805$, and $\Delta(A) = 0$, with an effective mean of $\Delta(\text{air}) = 0.0350$. At s.t.p. (7.37) and (7.38) give

$$s_n(\text{s.t.p.}) = 1.215 \times 10^{-37}\frac{(m^2-1)^2}{\lambda_0^4}, \tag{7.39}$$

$$P'(\cos\theta) = 0.7629(1 + 0.932\cos^2\theta). \tag{7.40}$$

The refractive index of air depends upon wavelength, varying from $m - 1 = 3.4187 \times 10^{-4}$ at $\lambda_0 = 0.2\ \mu$m and 0°C to $m - 1 = 2.8757 \times 10^{-4}$ at $\lambda_0 = 20\ \mu$m and 0°C. This gives rise to slight departures from a simple inverse fourth-power law for the wavelength dependence of the Rayleigh scattering coefficient.

In the following sections we will discuss the properties of single

particles rather than an assembly of particles and the equations must be rewritten in terms of the optical properties of the particle itself. For a sphere, the Lorentz relation states

$$\alpha = \frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} r^3, \tag{7.41}$$

where r is the drop radius, and $\tilde{m}$ is complex and not necessarily close to unity. There results from (7.30)

$$Q_s = \frac{s_n}{\pi r^2} = \tfrac{8}{3} x^4 \left| \frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right|. \tag{7.42}$$

From (7.18) and (7.26) we have

$$\begin{aligned} Q_e &= \frac{4}{x^2} \mathscr{R}\left[i\alpha \left(\frac{2\pi}{\lambda_0} \right)^3 \right] \\ &= -4x \mathscr{I}\left(\frac{\tilde{m}^2 - 1}{\tilde{m}^2 + 2} \right). \end{aligned} \tag{7.43}$$

Equation (7.43) is paradoxical. It implies that if α is real, $Q_e = 0$. But we expect $Q_e = Q_s$ if α is real and, from (7.42), both must be nonzero. Radiation reaction is neglected in the Rayleigh theory and because of this the phase of the scattered wave is incorrect; as a result only the absorbed component is properly accounted for, and it can be shown that (7.43) gives Q_a—the absorption efficiency—and not the extinction efficiency.

Under certain conditions small elements of a large particle of arbitrary shape can be treated as independent Rayleigh scatterers whose amplitudes can be summed on the scattered wave front, provided due account is taken of the phase. van de Hulst calls this the Rayleigh–Gans approximation and shows the appropriate conditions to be

$$\begin{aligned} &|\tilde{m} - 1| \ll 1, \\ \text{and} \qquad &\\ &2x\,|\tilde{m} - 1| \ll 1. \end{aligned} \tag{7.44}$$

These conditions can be satisfied in the X-ray spectrum, but not in the visible spectrum for any circumstances encountered in the earth's atmosphere. While the approximation has no applications, it is interesting to note some of the properties of Rayleigh–Gans scattering by a sphere, because they are typical of large-particle scattering in general.

First, the scattering efficiency factor tends to (7.42) in the limit $x \ll 1$, as is expected. For $x \gg 1$, however, we have

$$Q_s = 2x^2\,|\tilde{m} - 1|. \tag{7.45}$$

In place of Rayleigh's fourth-power law, scattering now varies as λ_0^{-2}, i.e., the scattering is "whiter."

Second, owing to destructive interference, the phase matrix can have zeros for certain values of the scattering angle. The scattering diagram can therefore exhibit *lobes*.

Third, for $\theta = 0$ the scattering is the same as for the Rayleigh case but, for all other scattering angles, the intensity is diminished. The scattering has a strongly forward component, and the scattering coefficient is less than predicted by the Rayleigh theory.

In subsequent sections we will find all three properties to be typical of scattering by large, dielectric spheres in contrast to Rayleigh's theory.

7.4. Large particles as $|\tilde{m}| \to 1$

The condition $x \gg 1$ does not, by itself, uniquely define an important asymptotic form of the exact electromagnetic theory. Clearly, there some connection with the laws of refraction and reflection of geometric optics, since these are usually effective for large surfaces. Also, the concept of a *ray* becomes meaningful. If the particle is very much larger than the first few Fresnel zones, then we may look upon a ray as *localized* in these central zones. A complete wave description then ceases to be necessary, although this does not mean that the wave character (i.e., the properties of amplitude and phase) can be neglected.

The concept of localization enables us to make a useful distinction between *diffraction* on the one hand and *reflection* and *refraction* on the other. In the context of a complete electromagnetic theory the distinction is not meaningful, all phenomena being aspects of a single solution of Maxwell's equations. However, if we recognize the localization principle, we can refer to the *shadow area* of an obstacle and consider rays outside it to be *diffracted* while those inside are *reflected* or *refracted*. The distinction is, however, fraught with paradoxes. For example, for absorbing spheres with moderate values of x, Q_a can exceed unity. Now, since absorption can occur only if a photon actually strikes the scattering particle, we appear to have the situation whereby more quanta strike the particle than pass through the shadow area. But the complete theory is so complex that despite such difficulties we cannot pass over the opportunity of utilizing the large body of optical theory based on the localization principle, even if some risk of confusion is entailed.

We know from theory and observation that the *diffracted* light can be accounted for approximately on the basis of Fresnel's theory. Narrow diffraction rings are observed, which become narrower as the particle size is increased. On a scattering diagram, these diffraction rings appear as strong, forward lobes, but they will not necessarily be detectable experimentally. If the scattered light is viewed with an instrument of large angular dispersive power, the diffraction rings will be resolved and distinguished from the incident beam. In this case a measurement of the

extinction will include the diffraction term. If, however, we picture a small water drop close to, and in front of a pyrheliometer or other non-image-forming device, it is obvious that the diffracted radiation cannot be distinguished from undisturbed incident radiation, and the Fresnel diffraction should not be included in an extinction computation (see § 2.4.7).

A relationship between the two components is provided by Babinet's principle. Consider a beam of light falling on an opaque obstacle of shadow area G, and also consider the complementary experiment whereby all of the wave front is obscured *except* the area G. The amplitudes in the two cases must add up to the original, undisturbed wave front. It follows that the diffracted *wave amplitudes* are equal and opposite in the two cases; since intensity is proportional to the square of the amplitude, diffracted *intensities* are, therefore, the same in the two experiments. Now, in the case of the opening G in the otherwise opaque wave front, we know that all of the radiation is disturbed by diffraction to a greater or lesser extent. The same, therefore, applies to the opaque obstacle of area G, which must therefore have a diffraction cross section G. Since it also has a cross section G for interception of photons in the shadow zone, it follows that the total cross section is $2G$. As has already been shown, the result of an experiment with limited angular resolution may be to measure G only, but the expected theoretical result is $Q_e \rightarrow 2$ as $x \rightarrow \infty$.

We shall now state the familiar theory of Fresnel diffraction at an aperture (Fig. 7.4) in the formal language adopted in this chapter, so that it can later be applied to the problem of a sphere with refractive index

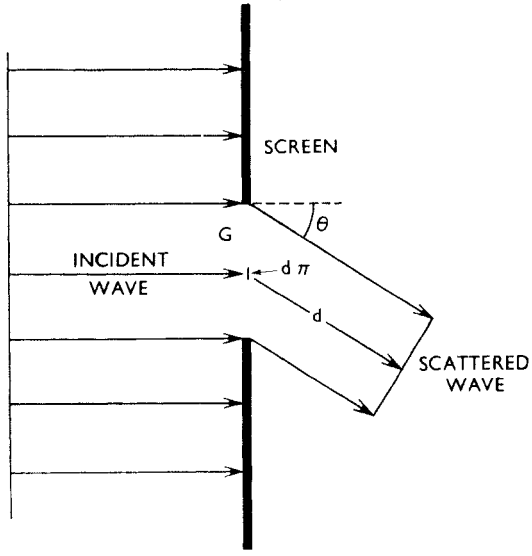


FIG. 7.4. Fresnel diffraction.

($\tilde{m}$) close to unity. Fresnel diffraction involves no phase change and no change in the state of polarization. It follows that $S_1 = S_2$ and that both are real quantities. Moreover, foreseeing the result that diffraction effects are important only for small scattering angles, and considering distances large compared to the aperture, we may regard each element of the aperture G as an isotropic scatterer, contributing an intensity proportional only to its area. The only factor distinguishing the scattering in different directions is that caused by destructive interference. Thus,

$$\frac{S_{1,2}(\theta, \phi)}{S(0)} = \left| \frac{\int_G e^{-2\pi i d/\lambda_0} \, d\pi}{\int_G d\pi} \right| = D(\theta, \phi). \tag{7.46}$$

A little consideration shows that $D(\theta, \phi)$ is a function only of θ, ϕ, and G and not of d. For real $S(0)$, (7.18) can be written

$$Q_e = \frac{\lambda_0^2}{\pi G} S(0), \tag{7.47}$$

and we have already shown that $Q_e = 2$ for a large obstacle. By Babinet's principle, the *ratio* (7.46) is the same for obstacle and aperture; consequently for an obstacle

$$S_{1,2}(\theta, \phi) = \frac{2\pi G}{\lambda_0^2} D(\theta, \phi). \tag{7.48}$$

If G is circular, as for a sphere, (7.46) can be evaluated in terms of Bessel functions of integral order

$$S_{1,2}(\theta) = x^2 \frac{J_1(x \sin \theta)}{x \sin \theta}. \tag{7.49}$$

Let us briefly review the meaning of (7.49). It refers to the *diffraction component* of the amplitude scattering matrix ($\mathbf{S}_d$) for a very large sphere, whether opaque or transparent. To obtain the complete matrix it must be added (taking account of phase) to the matrix resulting from rays in the shadow zone ($\mathbf{S}_s$). This result has been obtained by a very loose argument based on Babinet's principle, together with the conclusion that $Q_e = 2$ for an *opaque* aperture or screen. However, the localization principle permits us to distinguish between $\mathbf{S}_d$ and $\mathbf{S}_s$, and therefore the result should be true for any kind of obstacle. For a translucent obstacle, however, we need not expect to find $Q_e = 2$; indeed, in the limit $\tilde{m} = 1$ we have no screen at all and therefore Q_e must

obviously be zero. For values of $\tilde{m}$ between 1 and ∞, some light will penetrate the sphere and interfere with the diffracted component, leading to a phenomenon called *anomalous diffraction* by van de Hulst.

In one circumstance, namely $|\tilde{m}| \to 1$, the problem of anomalous diffraction can be treated very simply. Since we now have two parameters (viz. x and $\tilde{m}$) going to limits, we must be careful to define their mutual behavior. Let us assume $\tilde{m}$ to be real. The important parameter is then

$$\rho = 2x(\tilde{m} - 1), \tag{7.50}$$

the phase lag of a ray passing through the centre of the sphere. We require that this parameter remain finite but it need no longer be small, as in the Rayleigh–Gans case. The importance of this condition is that the ray suffers a phase change as it passes through the sphere, but (since $\tilde{m} - 1$ is small and real) is not reflected, refracted, or absorbed. The amplitude of a wave reaching $d\pi$ (Fig. 7.5) is the same as if the sphere were not there, but the phase is shifted by $\rho \sin \tau$.

From (7.47), we can write the diffraction component of the amplitude matrix in the form

$$S_d(0) = \frac{2\pi G}{\lambda_0^2} = \frac{2\pi}{\lambda_0^2} \int_G d\pi. \tag{7.51}$$

By Babinet's principle

$$S_s(0) = -\frac{2\pi}{\lambda_0^2} \int_G d\pi \tag{7.52}$$

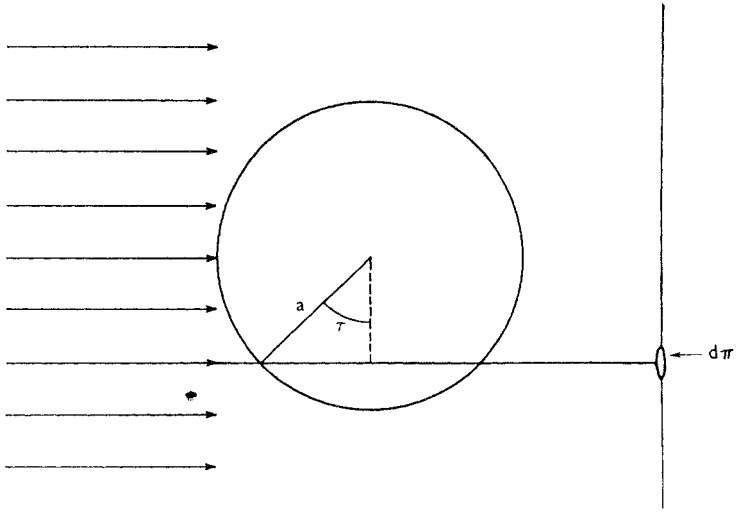


FIG. 7.5. Geometry of a sphere as $|\tilde{m}| \to 1$.

is the amplitude factor for the transmitted ray through a sphere with $\tilde{m} - 1 = 0$ (i.e., an aperture). The sphere depicted in Fig. 7.5 introduces a phase shift $\rho \sin \tau$ for the ray passing through $d\pi$, and therefore we have to modify (7.52) to

$$S_s(0) = -\frac{2\pi}{\lambda_0^2} \int_G e^{-i\rho \sin \tau} \, d\pi. \tag{7.53}$$

Adding (7.52) and (7.53), setting $d\pi = 2G \cos \tau \sin \tau \, d\tau$, and performing the integration, we find

$$S(0) = x^2 K(i\rho), \tag{7.54}$$

where

$$K(w) = \frac{1}{2} + \frac{e^{-w}}{w} + \frac{e^{-w} - 1}{w^2}.$$

Hence,

$$Q_e = \frac{4}{x^2} \mathcal{R}[S(0)] = 4\mathcal{R}[K(i\rho)]. \tag{7.55}$$

For real $\tilde{m}$, (7.55) is

$$Q_e = 2 - \frac{4}{\rho} \sin \rho + \frac{4}{\rho^2} (1 - \cos \rho). \tag{7.56}$$

Figure 7.6 shows a comparison between this equation and exact computations for $\tilde{m} = 0.8$, 0.93, 1.33, and 1.5 (see § 7.6). The agreement is remarkably good, even though $\tilde{m}$ is far from unity. Maxima and minima appear at the predicted values of ρ, indicating that these features of the extinction curve are the result of interference between the transmitted and diffracted rays. If $\tilde{m} \neq 1$ the curves have a fine structure, which is clearly not of fundamental importance in natural systems, for the inevitable mixture of particle sizes will blur over any such detail. There is a systematic increase in the heights of the maxima and minima as $\tilde{m}$ increases that is not given by the approximate theory. It has been proposed, as an empirical correction, that approximate extinction efficiencies should be multiplied by a factor

$$F(\tilde{m}, x) = 1 + \frac{3\tilde{m} + 1}{2} x^{-2/3}. \tag{7.57}$$

We have already seen from the case of Rayleigh–Gans scattering

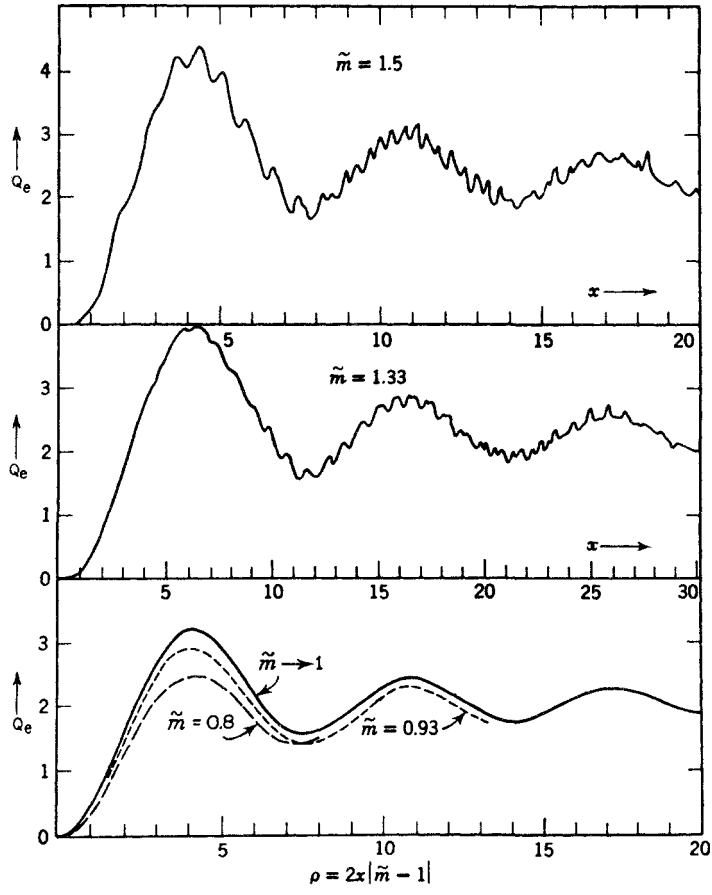


FIG. 7.6. Extinction by large dielectric spheres. The solid line in the lower panel follows the approximate equation (7.56). The other four curves are computed from Mie's theory (§ 7.6) for the values of $\tilde{m}$ shown. The accuracy of computation for $\tilde{m} = 0.8$ and 0.93 is somewhat lower than for $\tilde{m} = 1.5$ and $\tilde{m} = 1.33$. After van de Hulst (1957).

how increase of particle size leads to more neutral scattering. Figure 7.6 shows how this trend continues, and how Q_e hardly varies with λ for $\rho > 10$. Of interest for atmospheric optical phenomena are the portions of the efficiency curve that slope downward to the right, particularly that section for $4.09 < \rho < 7.63$. In this region, close to an octave in extent, long wavelengths are scattered more strongly than short wavelengths. Such behavior is occasionally observed for natural aerosols.

For spheres with a complex refractive index, we may redefine ρ using only the real part of the refractive index and introduce the new

variable

$$\tan\beta = \frac{\tilde{n}'}{\tilde{n}-1}. \tag{7.58}$$

Thus, $\beta = 0$ denotes a nonabsorbing medium and $\beta = \frac{1}{4}\pi$ a *black* medium. Equation (7.56) becomes

$$Q_e = 2 - 4e^{-\rho\tan\beta}\frac{\cos\beta}{\rho}\sin(\rho-\beta) - 4e^{-\rho\tan\beta}\left(\frac{\cos\beta}{\rho}\right)^2\cos(\rho-2\beta) + 4\left(\frac{\cos\beta}{\rho}\right)^2\cos 2\beta. \tag{7.59}$$

Accepting the localization principle, we know the ray paths through the drop, and can calculate the absorption from first principles. We find

$$Q_a = 2K(2\rho\tan\beta). \tag{7.60}$$

Equations (7.59) and (7.60) are illustrated in Fig. 7.7. We have seen that the maxima and minima on the curve $\beta = 0°$ owe their existence to

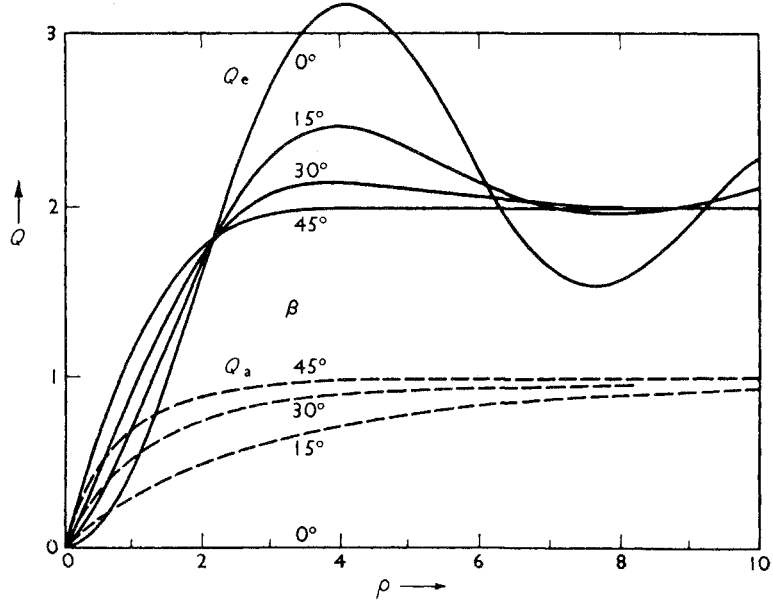


FIG. 7.7. Extinction and absorption efficiencies for a medium with refractive index $\tilde{m} = 1 + \epsilon - i\epsilon\tan\beta$ ($\epsilon \ll 1$). After van de Hulst (1957).

interference between rays penetrating and passing the sphere. As the absorption coefficient increases, this interference decreases and the wave structure disappears, while preserving the asymptotic value $Q_e = 2$ as $\rho \to \infty$.

Finally, we wish to know the angular distribution of the scattered light. This involves a straightforward application of the principles that we have already established. The integrals in (7.51) and (7.53) give the amplitude of the wave front in the area $d\pi$ of Fig. 7.5. The amplitude

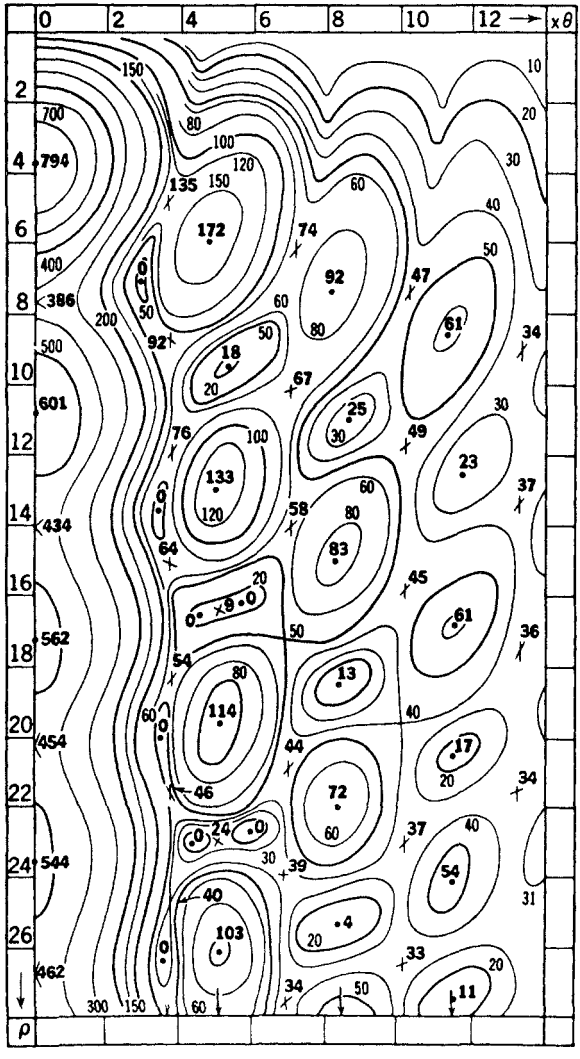


FIG. 7.8. Altitude chart of the amplitude function $|A| = |x^{-2}S|$. after van de Hulst (1957).

factor in a direction θ can now be found by vectorial addition, as for (7.46).

Integrating over the wave front for a sphere with real refractive index, we find in a straightforward manner

$$S_1(\theta) = S_2(\theta) = x^2 \int_0^{\pi/2} (1 - e^{i\rho \sin \tau}) J_0(x\theta \cos \tau) \cos \tau \sin \tau \, d\tau$$

$$= x^2 A(\rho, x\theta). \tag{7.61}$$

This integral has to be computed by numerical quadrature, and the results are displayed in the form of an altitude chart in Fig. 7.8. In Fig.

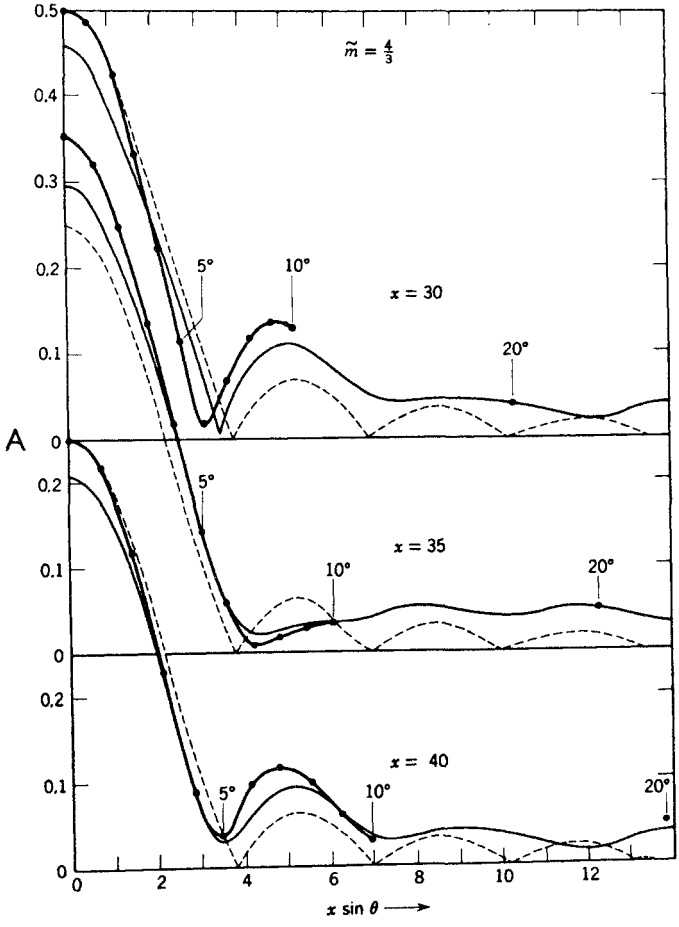


FIG. 7.9. Amplitude function $A = x^{-2}S$ for $\tilde{m} = 4/3$. The points and heavy solid lines are exact calculations. The less heavy solid line is the approximation $\tilde{m} \to 1$. The broken line is the Fraunhofer diffraction pattern. Values of θ are given against some points. After van de Hulst (1957).

7.9, a cross section of this diagram, appropriate to water drops, is compared with exact computations, and also with the Fraunhofer diffraction pattern of an aperture.

The agreement between the approximation and exact computations shown in Fig. 7.9 is good both qualitatively and quantitatively, except that the approximation predicts no polarization effects at all. As regards the results displayed in Fig. 7.9, this is a good prediction, for the difference between S_1 and S_2 would not be visible on this scale. This is because the computations are restricted to scattering angles of 20° and less, but for larger scattering angles polarization is an important feature of the scattering pattern.

A solution has been proposed that reduces to that discussed in this section for $\theta = 0$, but that permits polarization effects at other angles. This is a mathematical approximation to the complete Mie theory (see § 7.6), which replaces the internal electromagnetic field in the droplet by the WKBJ approximation, and proceeds analytically from this point. The degree of polarization for large scattering angles is predicted with reasonable accuracy, but the absolute magnitude of the scattered intensities is less satisfactory.

7.5. Geometric optics

The elementary rainbow theory of Descartes and other early theories of droplet scattering are based upon the principles of *geometric optics*. This is an inclusive term for a number of assumptions of which the localization principle is one, but not the only one. It is assumed in addition that the disturbance is scalar and that the energy flux is proportional to the density of rays. Thus, if the incident wave front is represented by an equidistant set of rays, the scattering matrix is proportional to the angular density of rays leaving the scattering centre, the path of each having been traced by the laws of reflection and refraction (an attenuating factor, dependent upon the path length, can be added if absorption has to be taken into account). As we have already discussed, the concept of a ray (i.e., the localization principle) is sound, but rays can be considered to be independent only as long as two or more do not meet, when, according to the assumptions of geometric optics, the intensity is infinite. Since the waves are now in a position to interfere the concept of a scalar ray is inadequate, and phase must be taken into account. At these singular points Fresnel's theory can be used to modify the simple concepts of geometric optics. Applying the theory to the wave front before a focal point, a system of diffraction rings results. The focal point of principal interest is the final image, which is formed at infinity from an emergent parallel bundle of rays. Interest in intermediate foci is restricted to the phase changes that may take place; the diffraction blurring is too small to matter. It is a well-known result of Fresnel's theory that on passing

through a focus there is a phase advance of $\frac{1}{2}\pi$. At a reflection there can be shown to be a phase shift of π, while refraction involves no phase shift. With these results we can determine the phase at any point on an emergent ray.

There is a variety of different situations in which the phase is important. Figure 7.10 shows a familiar diagram illustrating the formation of the first rainbow. Because there is a ray whose deviation is a minimum (ray 3), there must be flanking rays that are parallel. These can interfere at the image formed in the eye or telescope. The nature of the image in the vicinity of this minimum deviation can be estimated by constructing the wave front B and applying Fresnel's theory. Other less obvious singular points are when a ray intersects the line XY or runs parallel to it. Since the incident wave front is unlimited in size, the net effect of wave fronts such as A is obtained by rotating the figure about the axis XY. A ray intersecting this axis will therefore intersect others, and a ray parallel to the axis ($\theta = 0$ or π) will form a parallel bundle.

It is scarcely necessary to add that geometric optics takes no account of diffraction of light that is not incident upon the scatterer. For large particles, this diffracted component is all in a strongly forward direction. If it is not to interfere with transmitted rays these latter must be strongly

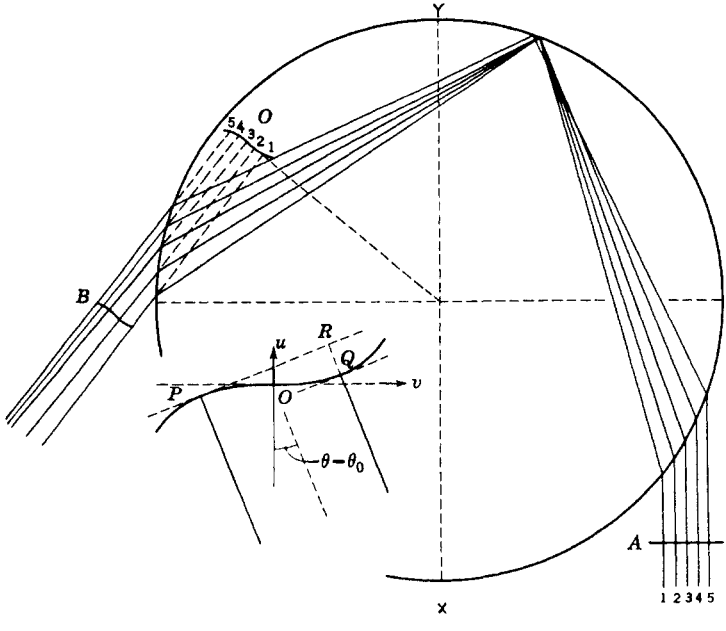


FIG. 7.10. Drawing in correct proportion of five equidistant rays that contribute to the first rainbow ($\tilde{m} = 4/3$). Inset is a schematic drawing of the cubic wave front at O. After van de Hulst (1957).

deflected. The requirement is that $\tilde{m} - 1$ shall *not* be vanishingly small. Then we can consider the diffracted component separately and, if the particle is large enough, neglect its deviation entirely.

The most important and successful application of geometric optics is to the problem of the rainbow, i.e., to scattering by a raindrop, which typically has a value of $x = 5 \times 10^3$ in the visible spectrum. Although it is possible to sum the Mie series numerically to obtain an exact result, only geometric optics, with all the associated doubts as to its validity, can provide an intuitive understanding.

Except for the singular directions at $\theta = 0$ and π, and the rainbow angles (at minimum or maximum deviation), geometric optics should be satisfactory. Ignoring diffraction effects entirely, we obtain $Q_s = 1$ and a phase function that does not depend upon particle size. This phase function was first computed by Wiener in 1909 and is named after him.

A number of different rays have to be distinguished, identified by the ordinal numbers $p = 0, 1, 2, 3, \cdots$. $p = 0$ is the externally reflected ray, which does not enter the drop. $p = 1$ is the ray that leaves the drop after two refractions, but no reflections. $p = 2$ is the ray that suffers one internal reflection, $p = 3$ suffers two, and so on. In terms of the angles θ and τ (defined in Fig. 7.5), it is straightforward to show that

$$N^{(r)}(p, \tau) = \frac{a^2}{d^2} N_0^{(r)} \epsilon_r^2 \frac{\sin \tau \cos \tau}{\sin \theta \, |d\theta'/d\tau|}, \tag{7.62}$$

where

$$\begin{aligned} \epsilon_r &= q_r \qquad \text{for } p = 0, \\ &= (1 - q_r^2)(-q_r)^{p-1} \qquad \text{for } p = 1, 2, 3, \end{aligned}$$

and other symbols are defined below.

In (7.62) the factor d^{-2} accounts for the divergence of light leaving the raindrop. The factor ϵ_r, involving the Fresnel reflection coefficient q_r, accounts for the loss of light at the reflections and refractions that take place along the path. The index r refers to perpendicular polarization; the same expression with appropriate indices holds for the parallel case. Absorption in the drop has been neglected. Finally, the total deflection θ' is defined to take account of the fact that the outgoing ray may have rotated through one or more multiples of 2π by means of internal reflections.

It is convenient to write

$$G^{(r)} = 4\epsilon_r^2 \frac{\sin \tau \cos \tau}{\sin \theta \, |d\theta'/d\tau|}, \tag{7.63}$$

Table 7.1. Rainbow angles and degree of polarization

		$\tilde{m}$					
p		**1.3200**	**1.3250**	**1.3300**	**1.3350**	**1.3400**	**1.3450**
2	θ	135°59′36″	136°44′40″	137°29′00″	138°12′24″	138°55′46″	139°38′08″
	P	0.8995	0.9087	0.9185	0.9277	0.9364	0.9446
3	θ	132°34′38″	131°13′44″	129°53′58″	128°34′12″	127°17′30″	126°00′48″
	P	0.7832	0.7929	0.8017	0.8108	0.8193	0.8277
4	θ	46°35′58″	44°42′34″	42°50′40″	41°00′12″	39°09′48″	37°23′28″
	P	0.7573	0.7658	0.7752	0.7838	0.7922	0.8004

Source: After Shifrin and Rabinovich (1957).

where $G^{(r)}$ is the gain relative to isotropic scattering. It is a well-known property of a spherical drop that for $p \geq 2$, $d\theta'/d\tau = 0$ at certain angles, known as the *rainbow angles*. Thus, geometric optics gives $G^{(r)} = \infty$ at the rainbow, but if the gain factor is averaged over a small but finite angle, the result is finite. Table 7.1 shows rainbow angles and degree of polarization $[P = (G^{(r)} - G^{(l)})/(G^{(r)} + G^{(l)})]$ for the first, second, and third rainbows ($p = 2$, 3, and 4, respectively; $p = 1$ gives no rainbow). The first rainbow has a minimum deviation, while the second and third rainbows are maxima. The dispersion of $\tilde{m}$ gives different edges for each color, leading to the familiar rainbow effect. The color sequence in the second and third rainbows reverses that of the first. An extract from an extensive computation of gain factors is shown in Table 7.2. Also shown in Table 7.2 are the degree of polarization (P_N) and gain factor (G_N) for incident natural (unpolarized) light. Calculations are made at integral numbers of degrees and not at the actual rainbow angles; therefore there are no infinities except for $\theta = 0$, $p = 4$. Averaged over a degree, the contribution from this intensity is negligible and it is not included in G_N.

We now have to consider the singularities in the light of Fresnel's theory of diffraction. At $\theta = 0$ we are concerned with the diffraction rings for rays passing close to, but not through, the raindrop. The theory for large x has been given in § 7.4 and the result for a sphere is given by (7.49). The Bessel function has zeros at $x \sin \theta = 3.83$, 7.02, 13.32, etc., so that the diffraction rings close in as x increases. For $x = 10^4$ (a raindrop) the first zero is at $\theta = 1.4'$, and no distinct rings will be seen around a source as large as the sun. For mists of very small droplets the rings may sometimes be sufficiently far apart to be be seen around either the sun or moon.

Near the rainbow angles a full theory involves asymptotic solutions to the Mie theory for large x ($x \gg 1000$). There has been considerable progress in this direction (see Bibliography). However, in this section we shall restrict our discussion to the Airy theory, which applies Fresnel's theory to the outgoing wave front, assuming it to have a cubic form.

Table 7.2. Gain factors and polarization for water drops ($n = 1.3350$)

	$p=0$		$p=1$		$p=2$		$p=3$		$p=4$		Total	
θ	$G^{(r)}$	$G^{(l)}$	$G^{(r)}$	$G^{(l)}$	$G^{(r)}$	$G^{(l)}$	$G^{(r)}$	$G^{(l)}$	$G^{(r)}$	$G^{(l)}$	$2G_N$	P_N
0	1.0000	1.0000	15.2339	15.2339	—	—	0.0003	0.0003	(∞)	(∞)	32.4684	0.0000
2	0.9246	0.8696	15.1147	15.0947	—	—	0.0003	0.0003	0.0102	—	32.0144	0.00265
5	0.8211	0.7034	14.5377	14.6077	—	—	0.0003	0.0003	0.0052	—	30.6757	0.00172
10	0.6746	0.4931	12.7009	12.9462	—	—	0.0003	0.0003	0.0028	—	26.8182	−0.00227
20	0.4582	0.2389	7.9051	8.4030	—	—	0.0003	0.0002	0.0026	—	17.0083	−0.0162
30	0.3152	0.1108	4.1517	4.7692	—	—	0.0003	0.0002	0.0055	—	9.3529	−0.0436
40	0.2208	0.0473	1.9692	2.5256	—	—	0.0004	0.0002	0.0933	0.0184	4.8752	−0.0631
50	0.1580	0.0171	0.8312	1.2320	—	—	0.0004	0.0002	—	—	2.2389	−0.1606
60	0.1158	0.0043	0.2787	0.4955	—	—	0.0005	0.0002	—	—	0.8950	−0.1173
70	0.0873	0.0002	0.0527	0.1169	—	—	0.0007	0.0001	—	—	0.2579	0.0911
80	0.0674	0.0005	0.0007	0.0016	—	—	0.0010	0.0001	—	—	0.0713	0.9383
90	0.0534	0.0028	—	—	—	—	0.0016	—	—	—	0.0578	0.9031
100	0.0435	0.0059	—	—	—	—	0.0025	—	—	—	0.0519	0.7726
110	0.0364	0.0091	—	—	—	—	0.0053	0.0000	—	—	0.0508	0.6417
120	0.0312	0.0121	—	—	—	—	0.0310	0.0192	—	—	0.0935	0.3305
130	0.0275	0.0146	—	—	—	—		—	—	—	0.0421	0.3064
140	0.0247	0.0168	—	—	1.1239	0.1545		—	—	—	1.3199	0.7404
150	0.0228	0.0185	—	—	0.2983	0.2021		—	—	—	0.5417	0.1855
160	0.0217	0.0197	—	—	0.1133	0.0995		—	—	—	0.2542	0.0622
170	0.0208	0.0214	—	—	0.0839	0.0768		—	—	—	0.2019	0.0371
180	0.0206	0.0206	—	—	0.0795	0.0795		—	—	—	0.2002	0.0000

Source: Extracted from the tables of Shifrin and Rabinovich (1957).

Airy's theory is illustrated in Fig. 7.10. The insert shows details of the virtual wave front at O, obtained by considering path lengths along the various rays. The fundamental assumption is that the equation of the wave front is approximately

$$u = \frac{hv^3}{3a^2}, \tag{7.64}$$

where the constant h is 4.89 for the first rainbow and 27.86 for the second, if $\tilde{m} = 1.3333$. Applying Fresnel's principle at an angle $(\theta - \theta_0)$ to the rainbow direction gives an amplitude factor proportional to

$$\int_{-\infty}^{+\infty} e^{-2\pi iv(\theta-\theta_0)/\lambda + 2\pi ihv^3/3a^2\lambda}\, dv,$$

which can be written in the form of the Airy integral,

$$f(Z) = \int_0^{\infty} \cos \tfrac{1}{2}\pi(Zt - t^3)\, dt, \tag{7.65}$$

$$Z = (12/h\pi^2)^{1/2} x^{2/3}(\theta - \theta_0),$$

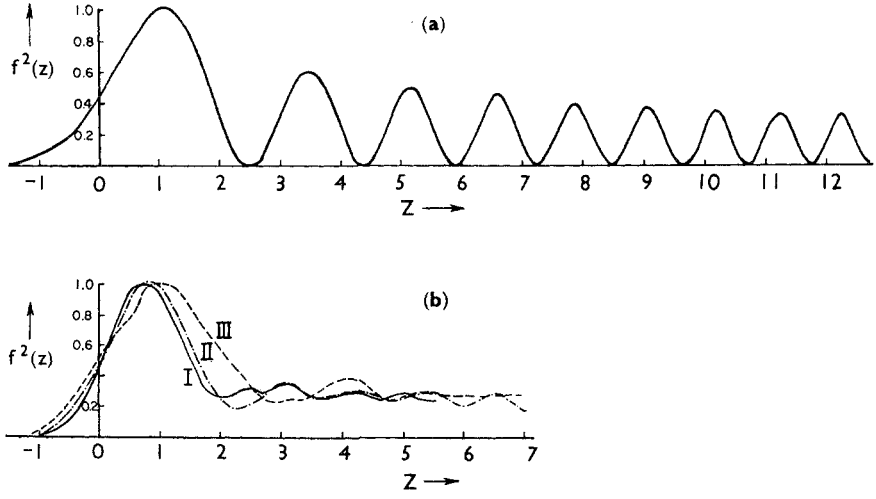


Fig. 7.11. The Airy rainbow integral for $\tilde{m} = 4/3$. (a) $f^2(Z)$ from (7.65). (b) Averaged over naturally occurring drop-size distributions. I, light shower rain; II, warm front rain; III, drizzle. Z is computed for a mean droplet radius of 0.5 mm, since the maximum contribution to the rainbow intensity comes from drops of approximately this size. After Volz (1961).

where θ_0 is the rainbow angle. The integral has been computed numerically and its square (proportional to the scattered intensity) is shown in Fig. 7.11. It has a series of maxima that, under ideal conditions, can be seen inside the first rainbow and outside the second.

Rain has a wide spectrum of drop sizes, which smooths out the diffraction pattern. Figure 7.11b illustrates three typical cases. The finite angular diameter of the sun leads to further averaging, and as a result it is rare that the diffraction rings (*supernumerary bows*) are seen in nature.

7.6. The Mie theory

The Mie theory is a complete, formal theory of the interaction of a plane wave with a dielectric sphere. Its results can be given in the form of an infinite series, but the convergence is slow for large values of x. It can be shown that the minimum number of terms that must be retained in the series to ensure reasonable accuracy is about x. However, with modern computers, this is not a serious defect and a large number of computations of scattering functions has now been made for both real and complex $\tilde{m}$. In addition, asymptotic series for large x are now available, and this greatly extends the domain of practical numerical computations.

The derivation of the solution is a straightforward application of classical electrodynamics, and here we shall quote only the results (see Bibliography for further references). The two amplitude functions have the symmetrical form

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [a_n \pi_n(\cos\theta) + b_n \tau_n(\cos\theta)],$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [b_n \pi_n(\cos\theta) + a_n \tau_n(\cos\theta)], \tag{7.66}$$

where

$$\pi_n(\cos\theta) = \frac{1}{\sin\theta} P_n^1(\cos\theta),$$

$$\tau_n(\cos\theta) = \frac{d}{d\theta} P_n^1(\cos\theta),$$

P_n^1 is an associated Legendre polynomial, and the coefficients a_n and b_n are given by

$$a_n = \frac{\psi_n'(\tilde{m}x)\psi_n(x) - \tilde{m}\psi_n(\tilde{m}x)\psi_n'(x)}{\psi_n'(\tilde{m}x)\zeta_n(x) - \tilde{m}\psi_n(\tilde{m}x)\zeta_n'(x)},$$

$$b_n = \frac{\tilde{m}\psi_n'(\tilde{m}x)\psi_n(x) - \psi_n(\tilde{m}x)\psi_n'(x)}{\tilde{m}\psi_n'(\tilde{m}x)\zeta_n(x) - \psi_n(\tilde{m}x)\zeta_n'(x)},$$

where

$$\psi_n(x) = (\tfrac{1}{2}\pi x)^{1/2} J_{n+1/2}(x),$$
$$\zeta_n(x) = (\tfrac{1}{2}\pi x)^{1/2} H^{(2)}_{n+1/2}(x),$$

are Riccati–Bessel functions and $J_{n+1/2}$ and $H^{(2)}_{n+1/2}$ are spherical Bessel functions. The Bessel functions have zeros that increase in number with the size of the argument, with the result that S_1 and S_2 can change rapidly for very small variations of x.

If extinction and scattering efficiencies alone are required these can be obtained from the expressions

$$Q_e = \frac{2}{x^2} \sum_{n=1}^{\infty} (2n+1)(a_n + b_n),$$
$$Q_s = \frac{2}{x^2} \sum_{n=1}^{\infty} (2n+1)(|a_n|^2 + |b_n|^2). \tag{7.67}$$

van de Hulst has attempted to list all the calculations prior to 1957 in his monograph; computations that have appeared since that date, and some of particular importance for water droplets in the visible and infrared spectrum, are noted in the Bibliography. Some results for real

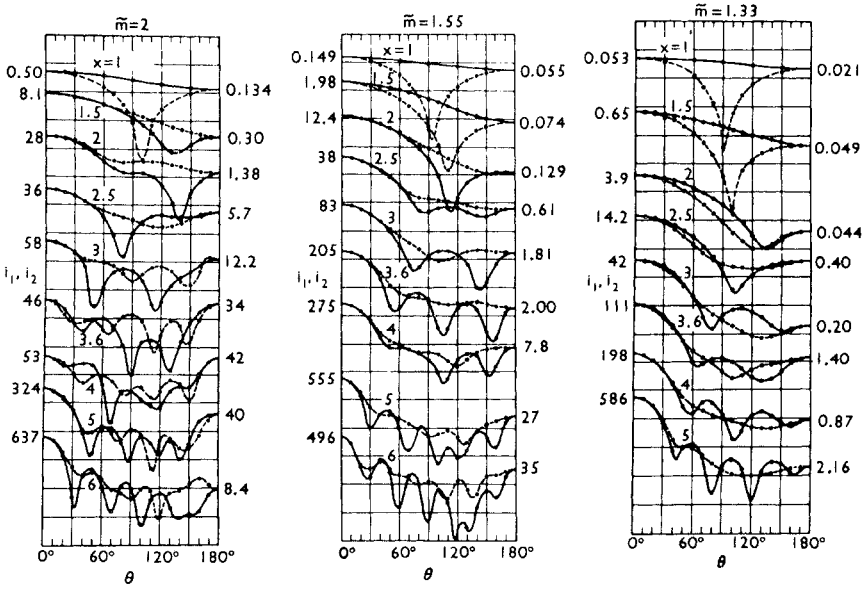


FIG. 7.12. Scattering diagrams from Mie theory. The solid curves are for $i_1 = |S_1|^2$ and the broken curves for $i_2 = |S_2|^2$. The vertical scale is logarithmic, with 1 div = a factor 10. Values of i_1 and i_2 at 0° and 180° (where they are equal) are given beside the diagram. The results are taken from the tables of Lowan (1949). After van de Hulst (1957).

refractive indices, close to that of water, have already been given in Figs. 7.6 and 7.9. Further results for all scattering angles and $\tilde{m} = 1.33$, 1.5, and 2.0 are shown in Fig. 7.12. These do not lend themselves to any simple discussion. The curves for $x = 1$ show some of the features of Rayleigh scattering, with large, positive polarization (i.e., $i_1 > i_2$) near 90°. For $x > 2$, however, both positive and negative polarizations occur, with changes from one to another occurring more frequently as x increases. For large particles, the forward lobe often has a net negative polarization.

Some results for complex refractive indices are shown in Figs. 7.13 and 7.14. Figure 7.13 shows the qualitative features of the approximate

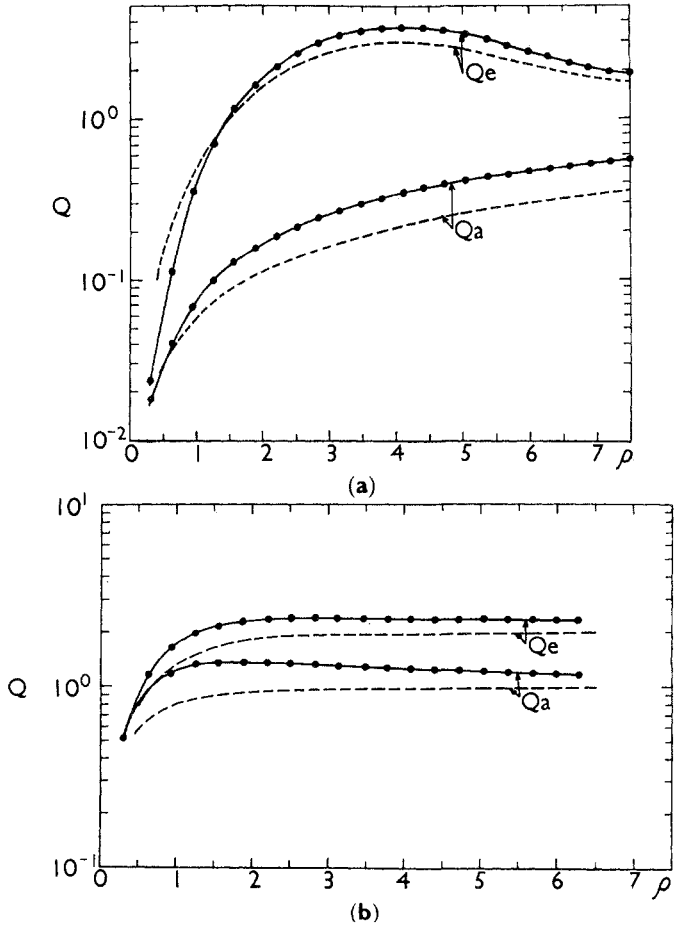


FIG. 7.13. Q_e and Q_a for $\tilde{n} = 1.315$ according to the Mie theory (points and solid line) and the approximate theory for $|\tilde{m}| \to 1$ (broken line). (a) $\tilde{n}' = 0.0143i$, (b) $\tilde{n}' = 0.4298i$. After Deirmendjian et al. (1961).

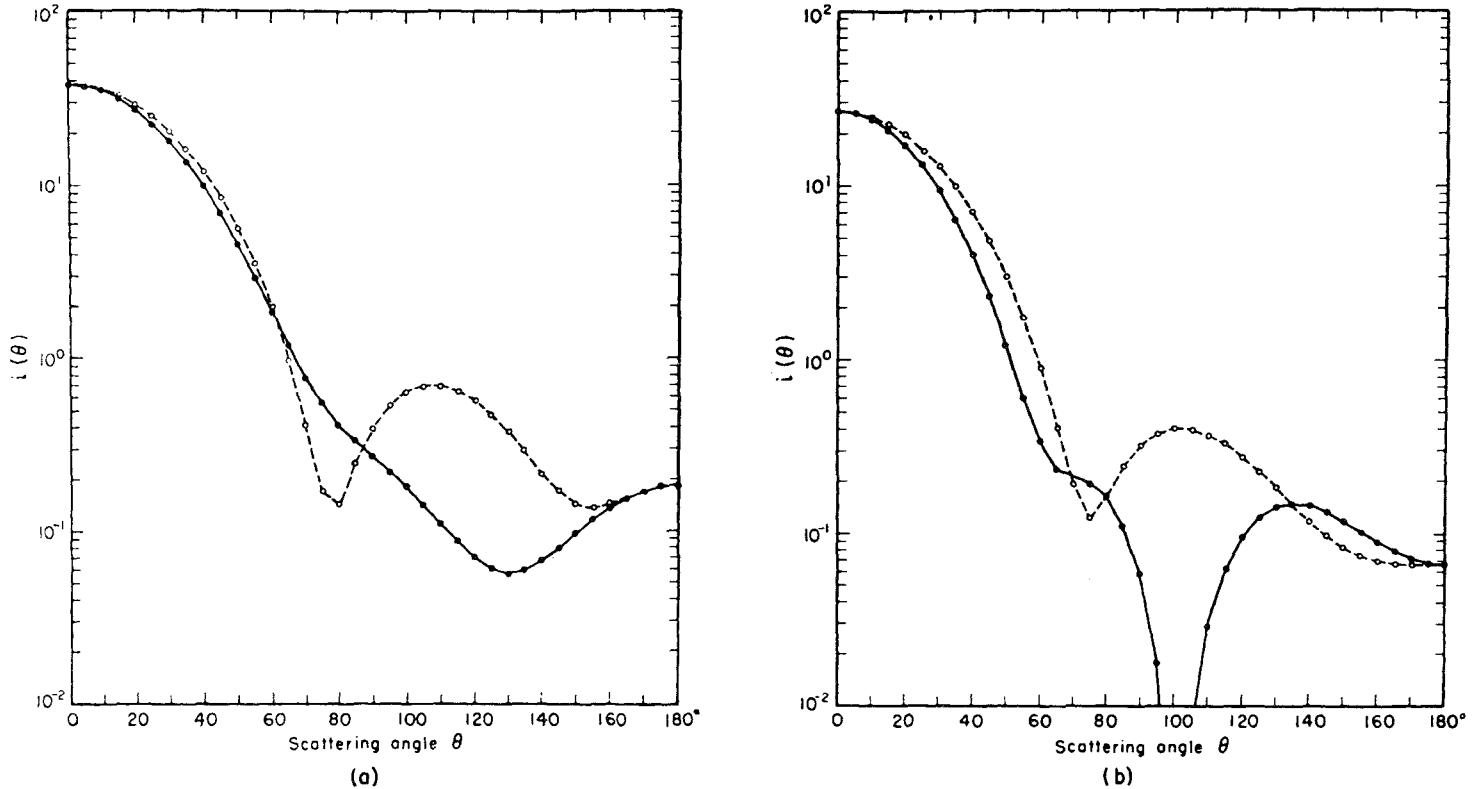


FIG. 7.14. $i_1(\theta)$ (broken line) and $i_2(\theta)$ (solid line) for $x = 3$, $\tilde{n} = 1.315$. (a) $\tilde{n}' = 0.0$, (b) $\tilde{n}' = 0.4298i$. After Deirmendjian et al. (1961).

theory for $|\tilde{m}| \to 1$; in fact, the quantitative agreement between comparable curves is also good if the empirical factor (7.57) is applied. The extinction maximum near $\rho = 4$ is damped out as the absorption increases and the asymptotic limits as $\rho \to \infty$ ($Q_e \to 2$, $Q_a \to 1$) are approached more rapidly. A weak maximum near $\rho = 1$ appears when absorption is large; this feature does not appear in the approximate theory. Comparison with the curves in Fig. 7.6 shows that the fine structure is completely eliminated, even for small absorptions. The curves in Fig. 7.14 illustrate the large changes that result from a variation in the complex component of the refractive index.

Mie theory results, with $\tilde{m} = 1.33$ and 1.50, for intermediate and large spheres, are shown in Fig. 7.15. The upper figure gives the phase function (G_N) and the lower figure gives the polarization (P_N) for scattering of natural light. These computations are averaged over a particle size distribution

$$n(x) = x^6 \exp(-6x/x_m)$$

where $n(x)\,dx$ is the number of particles with size parameter between x and $x + dx$, and x_m is an adjustable parameter. The effect of this averaging is to wash out the fine structure due to resonance and interference and to facilitate comparison with geometric optics (see Table 7.2). Geometric optics (dotted lines) predicts that the primary and the secondary rainbows should occur at scattering angles 137.5° and 157.2°, respectively, for $\tilde{m} = 1.33$, and at 129.9° and 93.1° for $\tilde{m} = 1.5$. The Mie calculations for large spheres ($x_m = 400$) reproduce these features and the polarizations quite well.

Referring to Fig. 7.15 we note the sharp increase in scattering (G_N) near the backward direction ($\theta = 180°$) for large spheres with $\tilde{m} = 1.33$. This strong enhancement in intensity, often accompanied by colored rings, is the *glory*. The glory cannot be explained by a combination of geometric optics and Fresnel theory. One major contributor to glory scattering comes from tangential rays (such as *IT* in Fig. 7.16). These give rise to a surface wave along *TA* critically refracted to the inside at *A*, "totally reflected" at *B*, and critically refracted to the outside at *C*, traveling as a surface wave *CT'* and finally emerging as the scattered tangential ray *T'S*. Mie theory fully accounts for the "surface waves," *TA* and *CT'*, although they have no place in geometric optics.

In the discussion of stratified atmospheres in § 2.4.5, the relevant properties of scatterers could be summed up in terms of the single scattering albedo ($a = s/e$) and the asymmetry factor (g). In terms of the symbols Q_s, Q_e, $G^{(r)}$, and $G^{(l)}$, they may be written as

$$a = \frac{Q_s}{Q_e}, \tag{7.68}$$

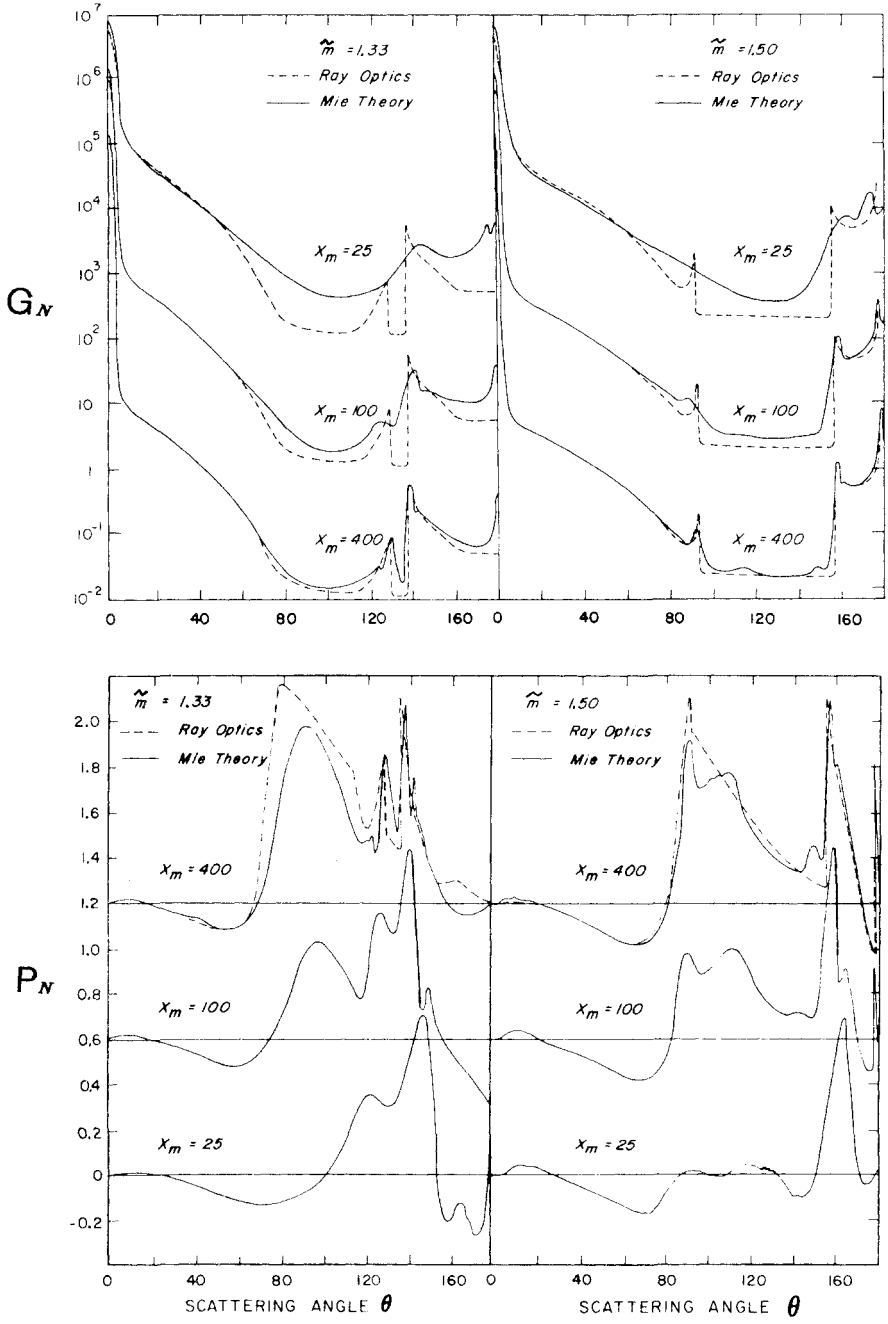


FIG. 7.15. Comparison between geometric optics and Mie theory. The phase function (G_N, upper figure) and polarization (P_N, lower figure) are for single scattering of unpolarized light by spheres. Results are shown for two real refractive indices and three values of x_m, which is the effective size parameter. After Hansen and Travis (1974).

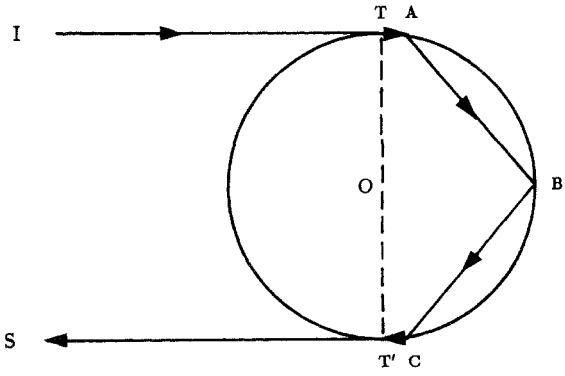


FIG. 7.16. Formation of the glory. "Surface" waves are a major contributor to the glory. After Nussenzveig (1979).

$$g = \frac{\int [G^{(r)}(\theta) + G^{(i)}(\theta)] \cos\theta \, d\omega}{\int [G^{(r)}(\theta) + G^{(i)}(\theta)] \, d\omega}. \tag{7.69}$$

Using (7.66), we can express (7.69) in a series

$$g = \frac{4}{x^2 Q_s} \sum_{n=1}^{\infty} \frac{n(n+2)}{n+1} \mathcal{R}(a_n a_{n+1}^* + b_n b_{n+1}^*) + \frac{2n+1}{n(n+1)} \mathcal{R}(a_n b_n^*). \tag{7.70}$$

For isotropic and Rayleigh scattering, $g = 0$; g is positive or negative according as the particle scatters more energy into the forward or backward direction. a is unity for a real index of refraction and less than unity for complex $\tilde{m}$. Numerical values for g and a are presented in Figs. 7.17 and 7.18, respectively. It is clear from Fig. 7.17 that $g \approx 0.8$ is a good approximation for cloud droplets, and, from Fig. 7.18, that a large sphere ($x = 1000$) can be quite dark for $\tilde{n}'$ as small as 10^{-3}. The asymptotic behavior of a as $x \to \infty$ in Fig. 7.18 can be understood if we recall that a metallic sphere is totally reflecting.

It is now possible to compute scattering functions using (7.66) for spheres with size parameter x as large as 1000. But such efforts do not reveal how to recover analytically the geometric optics limit from (7.66). Recent work indicates that this goal can be accomplished.

The series in (7.66) are both of the form

$$S = \sum_{n=1}^{\infty} \Phi(n). \tag{7.71}$$

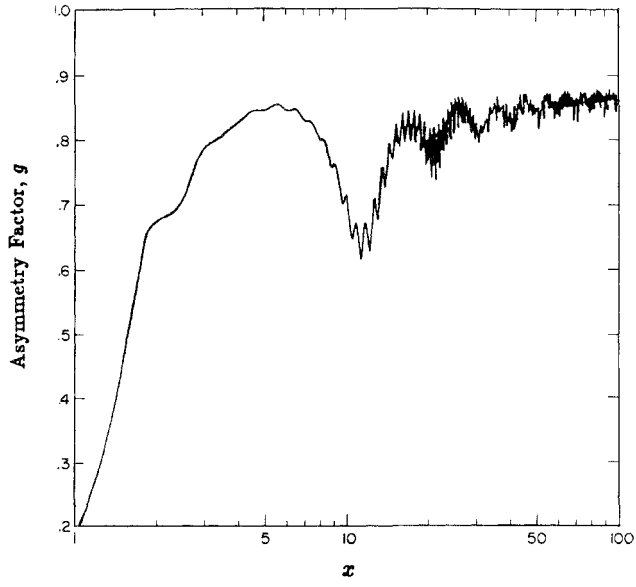


FIG. 7.17. Asymmetry factor. g is plotted as a function of x for $\tilde{m} = 1.33$. After Hansen and Travis (1974).

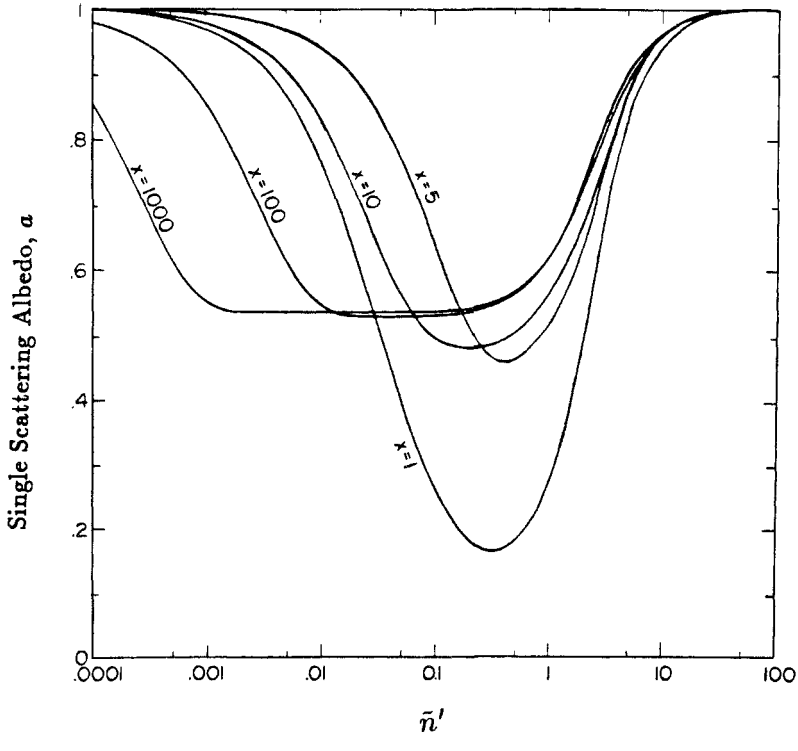


FIG. 7.18. Single scattering albedo. a is plotted as a function of the imaginary part of the index of refraction $\tilde{m} = \tilde{n} - i\tilde{n}'$, with $\tilde{n} = 1.33$. After Hansen and Travis (1974).

Applying the Poisson sum formula to (7.71), we have

$$S = \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} \Phi(\Lambda) \exp(2im\pi\Lambda)\, d\Lambda. \tag{7.72}$$

Note that the $m = 0$ term corresponds to approximating the discrete sum (7.71) by an integral. Using the Watson transformation, the integral over the whole real axis can now be replaced by a contour integral in the complex Λ plane.

For a large dielectric sphere, $\Phi(\Lambda)$ can be expanded into a Debye series of surface waves and their multiple reflections. It turns out that the combination of these two transformations, together with the proper choice of the path of integration, is sufficient to produce a rapidly convergent series whose leading terms are the same as those given by geometric optics and the Airy theory. Furthermore, the rays with complex angular momentum (Λ) can tunnel into the geometrically forbidden regions and these rays give rise to the glory. Thus, a satisfactory connection between Mie theory and the phenomena of the rainbow and the glory scattering has been established.

For practical computations in radiative transfer it is often convenient to approximate the Mie phase function (G_N) by the Henyey–Greenstein phase function

$$P_{HG}(\cos\theta; g) = \frac{1 - g^2}{(1 + g^2 - 2g\cos\theta)^{3/2}} = \sum_{l=0}^{\infty} (2l + 1) g^l P_l(\cos\theta), \tag{7.73}$$

where θ = scattering angle and $P_l(x)$ is a Legendre polynomial. The asymmetry factor of this phase function is g. Figure 7.19 shows a comparison between Mie phase functions and Henyey–Greenstein phase functions for particles with $\tilde{m} = 1.33$. While the Henyey–Greenstein phase functions reproduce the forward peak of Mie scattering quite well, they fail to reproduce the backscattering behavior. This situation can be readily remedied by using a double Henyey–Greenstein phase function,

$$P(\cos\theta) = bP_{HG}(\cos\theta; g_1) + (1 - b)P_{HG}(\cos\theta; g_2), \tag{7.74}$$

where b is a positive fraction, and g_2 can be assigned a negative value to account for the backscattering peak. For instance, the choice $g_1 = 0.824$, $g_2 = -0.55$, and $b = 0.9724$ provides a realistic simulation of scattering by maritime haze at 0.7 μm.

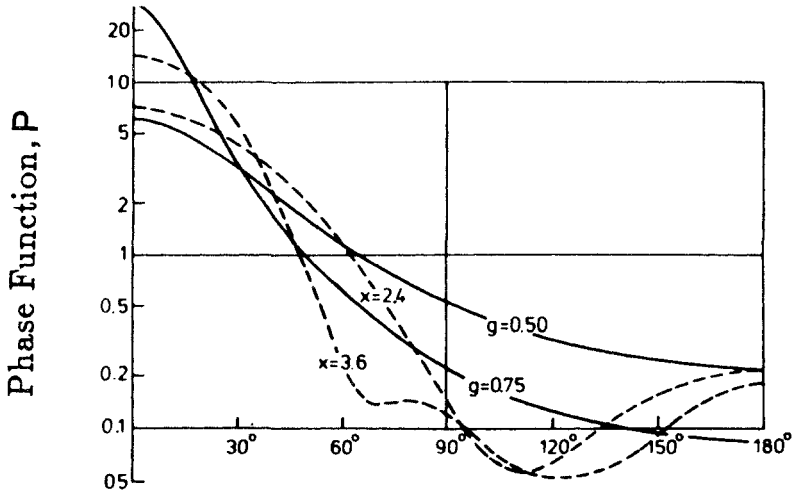


FIG. 7.19. Mie and Henyey–Greenstein phase functions. Broken lines are Mie phase functions for $\tilde{m} = 1.33$ and $x = 2.4$ and 3.6. Solid lines are Henyey–Greenstein phase functions for $g = 0.50$ and 0.75. After van de Hulst (1980).

7.7. Nonspherical particles

The Mie theory for dielectric spheres (§ 7.6) can be generalized to spheroids. Figure 7.20 shows the scattering efficiency Q_s, defined with respect to the geometrical shadow area, for an incident beam parallel to the symmetry axis. The results are similar to those obtained for a sphere. In the limit of a small spheroid, the behavior corresponds to Rayleigh scattering. For larger spheroids, characteristic features arising from interference and resonance effects become evident. The sharp increase in Q_s for highly prolate spheroids is attributed to surface phenomena.

The general problem of scattering of light by an object of arbitrary shape can be formulated as a Fredholm integral equation

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \int_V \mathbf{G}(\mathbf{r}, \mathbf{r}')\gamma(\mathbf{r}')\mathbf{E}(\mathbf{r}')\, d\mathbf{r}', \tag{7.75}$$

where $\mathbf{E}_0(\mathbf{r})$ is the incident plane wave, $\mathbf{G}(\mathbf{r}, \mathbf{r}')$ is the dyadic Green's function given by

$$\mathbf{G}(\mathbf{r}, \mathbf{r}') = \left(\mathbf{I} + \frac{\lambda_0^2}{2\pi} \nabla\nabla\right) \frac{\exp(2\pi i\, |\mathbf{r} - \mathbf{r}'|/\lambda_0)}{4\pi\, |\mathbf{r} - \mathbf{r}'|}, \tag{7.76}$$

$\mathbf{I}$ is the identity operator, and $\gamma(\mathbf{r})$ is related to the index of refraction

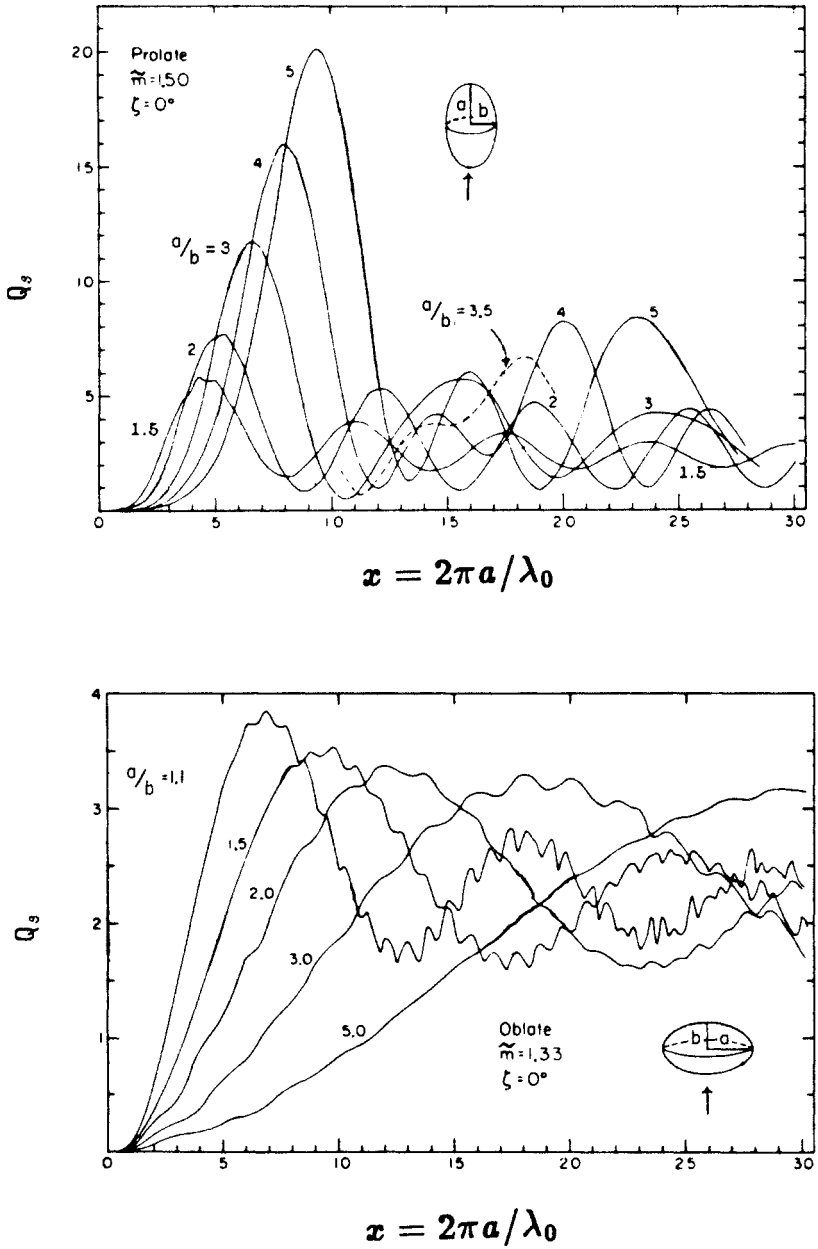


FIG. 7.20. Extinction by spheroids. The incident light is parallel to the symmetry axis. The projected area relative to which Q_s is defined is πb^2 or πa^2, respectively, for prolate and oblate spheres. After Asano (1979).

(which can be a function of **r**)

$$\gamma(\mathbf{r}) = \left(\frac{2\pi}{\lambda_0}\right)^2 [\tilde{m}^2(\mathbf{r}) - 1]. \tag{7.77}$$

The integration in (7.75) is restricted to the volume of the object V. Solving (7.75), usually by numerical inversion, yields the electric field inside V. All the important physical quantities are simply related to the integral

$$\mathbf{q}(\mathbf{s}) = \int_V \exp\left(-\frac{2\pi i}{\lambda_0}\mathbf{s}\cdot\mathbf{r}\right)\gamma(\mathbf{r})\mathbf{E}(\mathbf{r})\, d\mathbf{r}, \tag{7.78}$$

where **s** is a unit vector.

The main advantage of (7.75) versus Mie theory is that there are no explicit boundary conditions; all the effects of arbitrary geometry are absorbed by integration over V. The chief disadvantage of (7.75) is that its solution is a formidable task, although considerable progress has been made by approximating the integral by a large number of discrete dipole elements.

BIBLIOGRAPHY

This chapter follows closely the excellent text by

van de Hulst, H. C., 1957, *Light scattering by small particles*. New York: Wiley.

This contains a wealth of references. A modern text is by

Bohren, C. F., and Huffman, D. R., 1983, *Absorption and scattering of light by small particles*. New York: Wiley.

7.1. The problem in terms of the electromagnetic theory
Two textbooks on general electromagnetic theory are by

Stratton, J. A., 1941, *Electromagnetic theory*. New York: McGraw-Hill.

Jackson, J. D., 1975, *Classical electrodynamics*. New York: Wiley.

An historical account of various approximate theories by Kirchhoff, Fresnel, Rayleigh, and Sommerfeld can be found in

Goodman, J. W., 1968, *Introduction to Fourier optics*. New York: McGraw-Hill.

7.2. Scattering Functions
Numerical values of scattering functions are to be found in

Penndorf, R. B., 1957a, "Tables of the refractive index for standard air and the Rayleigh scattering coefficient for the spectral region between 0.2μ and 20.0μ and their application to atmospheric optics," *J. Opt. Soc. Am.* **47,** 176.

———, 1957b, "Total Mie scattering coefficients for spherical particles of refractive index $n \sim 1.0$," *J. Opt. Soc. Am.* **47,** 603.

———, 1957c, "New tables of total Mie scattering coefficients for spherical particles of real refractive indexes ($1.33 \leq n \leq 1.50$)," *J. Opt. Soc. Am.* **47,** 1010.

Hansen, J. E., and Travis, L. D., 1974, "Light scattering in planetary atmospheres," *Space Sci. Rev.* **16,** 527.

Liou, K. N., and Hansen, J. E., 1971, "Intensity and polarization for single scattering by polydisperse spheres: A comparison for ray optics and Mie theory," *J. Atmos. Sci.* **28,** 995.

Wickramasinghe, N. C., 1972, *Light scattering functions for small particles with applications to astronomy.* New York: Wiley; the g values computed in this reference are incorrect (1986, private communication).

7.3. Rayleigh's solution for small particles

The best reference is van de Hulst (1957), cited earlier.

7.4. Large particles as $|\tilde{m}| \to 1$

The empirical correction factor in (7.57) and a description of the WKBJ theory (this theory was developed by D. S. Saxon but remains unpublished) are given by

Deirmendjian, D., 1957, "Theory of the solar aureole. Part I. Scattering and radiative transfer," *Ann. Géophys.* **13,** 286.

7.5. Geometric optics

The scattering by raindrops is reviewed, with a copious bibliography, by

Volz, F., 1961, "Der Regenbogen," *Handbuch der Geophysik,* Band viii, Berlin: Borntraeger, p. 943.

Meyer, R., 1956, "Kränze, Glorien and verwandte Erscheinungen," *Handbuch der Geophysick,* Band viii. Berlin: Bortraeger, p. 898.

Two early books of great interest are

Minneart, M., 1940, *Light and colour in the open air.* London: Bell.

Pernter, J. M., and Exner, F. M., 1922, *Meteorologische Optik.* Vienna: Braumüller.

Tables 7.1 and 7.2 follow

Shifrin K. S., and Rabinovich, Yu. I., 1957, "The spectral indicatrices of large water drops and the spectral polarisation of rainbows," *Bull. Acad. Sci. U.S.S.R. Geophys.* **12,** 173.

7.6. The Mie theory

The expressions (7.66) and (7.67) are obtained in Stratton (1941, § 7.1)

The Mie theory of the raindrop was developed in a series of papers by

Nussenzveig and collaborators:

Nussenzveig, H. M., 1969a, "High-frequency scattering by a transparent sphere. I. Direct reflection and transmission," *J. Math. Phys.* **10,** 82.

———, 1969b, "High-frequency scattering by a transparent sphere. II. Theory of the rainbow and the glory," *J. Math. Phys.* **10,** 125.

———, 1979, "Complex angular momentum theory of the rainbow and the glory," *J. Opt. Soc. Am.* **69,** 1068.

Khare, V., and Nussenzveig, H. M., 1974, "Theory of the rainbow," *Phys. Rev. Lett.* **33,** 976.

———, 1977, "Theory of the glory," *Phys. Rev. Lett.* **38,** 1279.

The results shown in Fig. 7.13 are taken from the tables of

Lowan, A. N., 1949, *Tables of scattering functions for spherical particles.* Nat. Bur. of Standards. Appl. Math. Ser. **4.** Washington, D.C.: Government Printing Office.

Scattering diagrams for complex refractive indices of water drops in the infrared spectrum have been computed by

Deirmendjian, D., Clasen, R., and Viezee, W., 1961, "Mie scattering with complex index of refraction," *J. Opt. Soc. Am.* **51,** 620.

Stevens, J. J., and Gerhardt, J. R., 1961, "Absorption cross-sections of water drops for infrared radiation," *J. Meteorol.* **18,** 818.

———, 1961, "Spectrally averaged total attenuation, scattering and absorption cross-sections for infrared radiation," *J. Meteorol.* **18,** 822.

Herman, B. M., 1962, "Infra-red absorption, scattering, and total attenuation cross-sections for water spheres," *Quart. J. R. Meteorol. Soc.* **88,** 143.

Mie computations for spheroids are taken from

Asano, S., 1979, "Light scattering properties of spheroidal particles," *Appl. Opt.* **18,** 712.

The Henyey–Greenstein phase function was first proposed by

Henyey, L. C., and Greenstein, J. L., 1941, "Diffuse radiation in the galaxy," *Astrophys. J.* **93,** 70.

Comparison between Mie and Henyey–Greenstein phase functions is taken from

van de Hulst, H. C., 1980, *Multiple light scattering.* New York: Academic Press, p. 321.

The double Henyey–Greenstein phase function seems to have been first used by

Irvine, W. M., 1968, "Multiple scattering by large particles. II. Optically thick layers," *Astrophys. J.* **152,** 823.

7.7. Nonspherical particles

The integral equation approach to scattering (§ 7.7) was developed by

Levine, H., and Schwinger, J., 1950, "On the theory of electromagnetic wave diffraction by an aperture in an infinite plane conducting screen," *Comm. Pure Appl. Math.* **3,** 355.

Purcell, E. M., and Pennypacker, C. R., 1973, "Scattering and absorption of light by non-spherical dielectric grains," *Astrophys. J.* **186,** 705.

Holt, A. R., Uzunoglu, N. K., and Evans, B. G., 1978, "An integral equation solution to the scattering of electromagnetic radiation by dielectric spheroids and ellipsoids," *IEEE Trans. Antenna Propagation* **26,** 706.

8

RADIATIVE TRANSFER IN A SCATTERING ATMOSPHERE

8.1. Introduction

The source function for scattering, (2.32), is more complicated than a thermal source function on two accounts: it is not a function of local conditions alone, but involves conditions throughout the atmosphere, through the local radiation field, and the phase function, $P_{ij}(\mathbf{s}, \mathbf{d})$, may be an extremely complex function of the directions, **s** and **d**, and the states of polarization, i and j. The general solution, (2.87), is still valid, but it is now an integral equation, involving the intensity both on the left-hand side and under the integral on the right-hand side. Successive approximations, starting with the first-order scattering term [third term on the right-hand side of (2.116)], are an obvious approach, and would lead to a solution, but there are more efficient and more accurate ways to solve the problem.

Many methods are available because their fundamental theory has proved to be mathematically interesting and because there are important applications in neutron diffusion theory and astrophysics. These motivations are extraneous to atmospheric science, but the availability of the methodology has led to its adoption and extension to atmospheric problems.

Solutions to scattering problems can be elaborate and mathematically elegant; they can also be numerically onerous but, with access to modern computers, "exact" solutions are feasible, given the input parameters τ_ν, a_ν $(=s_\nu/e_\nu)$, and P_{ij}. For monochromatic calculations with simple phase functions, numerical solutions present few difficulties. Nevertheless, the combination of unfamiliar formalism with inaccessible and undocumented algorithms can be daunting for those with only a peripheral interest in radiation calculations. It is, therefore, relevant to note that available data are imprecise and virtually never require the accuracy available from exact methods. Easily visualized two-stream approximations, combined with similarity relations to handle complex phase functions (see §§ 8.4.4 and 8.5.6), are often more than adequate, and some angular information can be added, if required, from the use of Eddington's second approximation (§ 2.4.5).

It is not the purpose of this book to discuss observations or numerical calculations, but in § 8.6 we briefly summarize some of the observational data and uses that have been made of scattering theory. The purpose of these examples is to motivate studies of atmospheric scattering and to provide some perspective on the awkward conjunction between elegant theory and imprecise data. We leave this until the end of the chapter so that some theoretical results are available, but the reader may benefit from treating it as introductory material.

The methods described in this chapter fall into two broad categories. The first involves formal solutions to the integrodifferential equation of transfer (§ 2.3.3). The alternative approach seeks simple relations for the radiation field, not derived directly from the equation of transfer, but of equal physical validity. It leads to the *doubling and adding method*, the most efficient and accurate method capable of dealing with realistic atmospheric problems.

Polarization is a subject of interest for some special problems, but to alleviate the burden of too many indices we shall make no formal reference to it. All results derived in this chapter can be extended to include polarization by using Stokes' parameters and scattering matrices.

8.2. Integrodifferential equation

8.2.1. *Fourier series expansion*

The integrodifferential equation for the scattered component of non-polarized monochromatic radiation in a plane-parallel atmosphere is, from (2.113),

$$\xi \frac{dI(\tau; \xi, \phi)}{d\tau} = I(\tau; \xi, \phi) - a(\tau) \int P(\tau; \xi, \phi; \xi', \phi') I(\tau; \xi', \phi') \frac{d\omega'}{4\pi} - \Sigma(\tau; \xi, \phi), \tag{8.1}$$

where $a(\tau) = s(\tau)/e(\tau)$ is the *albedo for single scattering* and $\Sigma(\tau; \xi, \phi)$ is the primary source of radiation, such as the third term on the right-hand side of (2.116). Equation (8.1) does not include a term for the direct solar beam. Following the treatment of § 2.3.4 the direct beam is estimated independently, (2.115).

The phase function $P(\tau; \xi, \phi; \xi', \phi')$ is generally assumed to be a function of the scattering angle θ only. The latter is related to the incident direction (ξ', ϕ') and the emergent direction (ξ, ϕ) by the expression

$$\cos\theta = \xi\xi' + (1 - \xi^2)^{1/2}(1 - \xi'^2)^{1/2} \cos(\phi - \phi'). \tag{8.2}$$

With this restriction on the functional form, the phase function admits a

series expansion,

$$a(\tau)P(\tau;\xi,\phi;\xi',\phi') = a(\tau)P(\tau;\cos\theta),$$
$$= \sum_{l=0}^{N} \alpha_l(\tau)P_l(\cos\theta), \tag{8.3}$$

where $P_l(x)$ are Legendre polynomials and $\alpha_l(\tau)$ are expansion coefficients. Note that (8.3) is exact in the limit $N\to\infty$. In practice, N is taken to be a finite number, determined by the anisotropy of the phase function (see Chapter 7). From the addition theorem for spherical harmonics, the Legendre polynomials in (8.3) can be expanded in a Fourier series in the azimuthal variable

$$P_l(\cos\theta) = P_l(\xi)P_l(\xi') + 2\sum_{m=1}^{l}\frac{(l-m)!}{(l+m)!}P_l^m(\xi)P_l^m(\xi')\cos m(\phi-\phi'), \tag{8.4}$$

where $P_l^m(x)$ are associated Legendre polynomials. Substituting (8.4) into (8.3), we have

$$a(\tau)P(\tau;\xi,\phi;\xi',\phi') = \sum_{m=0}^{N}\rho^m(\tau;\xi,\xi')\cos m(\phi-\phi'), \tag{8.5}$$

where

$$\rho^m(\tau;\xi,\xi') = \sum_{l=m}^{N}(2-\delta_{0m})\alpha_l(\tau)\frac{(l-m)!}{(l+m)!}P_l^m(\xi)P_l^m(\xi') \tag{8.6}$$

and δ_{0m}, is the Kronecker δ-function. From the fact that

$$P_l^m(-\xi) = (-1)^{l+m}P_l^m(\xi), \tag{8.7}$$

we can derive two important symmetry properties for $\rho^m(\tau;\xi,\xi')$:

$$\rho^m(\tau;\xi,-\xi') = \rho^m(\tau;-\xi,\xi'), \tag{8.8}$$

$$\rho^m(\tau;-\xi,-\xi') = \rho^m(\tau;\xi,\xi'). \tag{8.9}$$

In view of (8.5) we may expand the other functions in (8.1) as Fourier series in the azimuthal variable

$$I(\tau;\xi,\phi) = \sum_{m=0}^{N} I^m(\tau,\xi)\cos m(\phi-\phi_0), \tag{8.10}$$

$$\Sigma(\tau;\xi,\phi) = \sum_{m=0}^{N}\Sigma^m(\tau,\xi)\cos m(\phi-\phi_0), \tag{8.11}$$

where ϕ_0 is a reference angle. Upon substituting (8.5), (8.10), and (8.11) into (8.1), and using the orthogonality of Fourier series, (8.1) splits up into $N+1$ independent equations:

$$\xi \frac{dI^m(\tau, \xi)}{d\tau} = I^m(\tau, \xi) - \tfrac{1}{4}(1+\delta_{0m}) \int_{-1}^{1} \rho^m(\tau; \xi, \xi') I^m(\tau, \xi')\, d\xi' - \Sigma^m(\tau, \xi) \qquad (m = 0, 1, 2, \ldots N). \tag{8.12}$$

We have succeeded in reducing an integrodifferential equation in three variables (τ, ξ, ϕ) to a set of uncoupled integrodifferential equations in two variables (τ, ξ). Since (8.12) is to be solved separately for each m, no confusion should arise by dropping the index m in subsequent discussion, and we shall consider an equation of the type

$$\xi \frac{dI(\tau, \xi)}{d\tau} = I(\tau, \xi) - \tfrac{1}{2}\gamma \int_{-1}^{1} \rho(\tau; \xi, \xi') I(\tau, \xi')\, d\xi' - \Sigma(\tau, \xi), \tag{8.13}$$

where $\gamma = \frac{1}{2}(1+\delta_{0m})$.

8.2.2. *Discrete ordinates*

This method is a natural generalization of the two-stream approximation discussed in § 2.4.8. Starting with (8.10), we keep τ as a continuous variable, but approximate the angle variable ξ $(-1 \leqslant \xi \leqslant 1)$ by $2n$ values $\xi_{\pm i}$ $(i = 1, 2, \ldots n)$, where $\xi_i > 0$ and $\xi_{-i} = -\xi_i$. The usual choice for $\xi_{\pm i}$ is the $2n$ roots of the Legendre polynomial $P_{2n}(\xi)$. With this choice we can replace the integral over ξ' in (8.13) by the extremely accurate Gaussian quadrature formula and obtain a discretized version of (8.13),

$$\xi_{\pm i} \frac{dI(\tau, \xi_{\pm i})}{d\tau} = I(\tau, \xi_{\pm i}) - \tfrac{1}{2}\gamma \sum_{j=1}^{n} a_j \rho(\tau; \xi_{\pm i}, \xi_j) I(\tau, \xi_j) - \tfrac{1}{2}\gamma \sum_{j=1}^{n} a_j \rho(\tau; \xi_{\pm i}, \xi_{-j}) I(\tau, \xi_{-j}) - \Sigma(\tau, \xi_{\pm i}) \qquad (i = 1, 2, \ldots n), \tag{8.14}$$

where a_j are Gaussian weights (a Gaussian quadrature using $2n$ points evaluates the integral exactly for all polynomials of degree less than $4n$). It should be noted that in order to achieve the high accuracy of the Gaussian quadrature formula the total number of Gaussian divisions, $2n$, must exceed the number of terms in the expansion of the phase function using Legendre polynomials (8.3).

Equation (8.14) is a system of $2n$ first-order nonhomogeneous differential equations, and can be written more compactly in vector

notation

$$\frac{d}{d\tau}\mathbf{I}^{+}(\tau) = \mathbf{M}^{+}(\tau)\mathbf{I}^{+}(\tau) + \mathbf{M}^{-}(\tau)\mathbf{I}^{-}(\tau) - \boldsymbol{\sigma}^{+}(\tau), \tag{8.15}$$

$$\frac{d}{d\tau}\mathbf{I}^{-}(\tau) = -\mathbf{M}^{-}(\tau)I^{+}(\tau) - \mathbf{M}^{+}(\tau)\mathbf{I}^{-}(\tau) - \boldsymbol{\sigma}^{-}(\tau), \tag{8.16}$$

where

$$\mathbf{I}^{\pm}(\tau) = [I(\tau, \pm\xi_1), I(\tau, \pm\xi_2), \ldots I(\tau, \pm\xi_n)], \tag{8.17}$$

$$\boldsymbol{\sigma}^{\pm}(\tau) = \left[\pm\frac{\Sigma(\tau, \pm\xi_1)}{\xi_1}, \pm\frac{\Sigma(\tau, \pm\xi_2)}{\xi_2}, \ldots \pm\frac{\Sigma(\tau, \pm\xi_n)}{\xi_n}\right], \tag{8.18}$$

and the matrix elements of the $n \times n$ matrices are given by

$$M_{ij}^{+}(\tau) = \frac{1}{\xi_i}[\delta_{ij} - \tfrac{1}{2}\gamma a_j \rho(\tau; \xi_i, \xi_j)], \tag{8.19}$$

$$M_{ij}^{-}(\tau) = -\frac{1}{2\xi_i}\gamma a_j \rho(\tau; \xi_i, -\xi_j). \tag{8.20}$$

The symmetry relations (8.8) and (8.9) have been used in deriving these equations. Physically, $\mathbf{I}^{+}(\tau)$ and $\mathbf{I}^{-}(\tau)$ represent upward traveling and downward traveling beams, respectively.

A simplification of (8.14) occurs if the atmosphere is vertically homogeneous [a and P independent of τ, so that α_l is also independent of τ from (8.3)]; from (8.6) $\rho(\tau; \xi_i, \xi_j) = \rho(\xi_i, \xi_j)$. In this case (8.14) becomes $2n$ first-order differential equations with constant coefficients plus nonhomogeneous terms. Let us first consider the homogeneous equations

$$\frac{d\mathbf{I}^{+}(\tau)}{d\tau} = \mathbf{M}^{+}\mathbf{I}^{+}(\tau) + \mathbf{M}^{-}\mathbf{I}^{-}(\tau), \tag{8.21}$$

$$\frac{d\mathbf{I}^{-}(\tau)}{d\tau} = -\mathbf{M}^{-}\mathbf{I}^{+}(\tau) - \mathbf{M}^{+}\mathbf{I}^{-}(\tau), \tag{8.22}$$

where the matrices $\mathbf{M}^{\pm}$ are now independent of τ. Let the solutions be of the form $\mathbf{I}^{\pm} = \mathbf{g}^{\pm}e^{k\tau}$. We have, from (8.21) and (8.22),

$$\begin{vmatrix} \mathbf{M}^{+} & \mathbf{M}^{-} \\ -\mathbf{M}^{-} & -\mathbf{M}^{+} \end{vmatrix} \begin{vmatrix} \mathbf{g}^{+} \\ \mathbf{g}^{-} \end{vmatrix} = k \begin{vmatrix} \mathbf{g}^{+} \\ \mathbf{g}^{-} \end{vmatrix}, \tag{8.23}$$

which may be solved as a standard eigenvalue problem. It has been

shown that all the eigenvalues are real and that they occur in pairs $(\pm k)$. The latter property enables us to transform the eigenvalue problem (8.23) into an equivalent one in k^2 and thus reduce the order of the problem by 2. This can be accomplished as follows. From (8.23) we have

$$\mathbf{M}^+\mathbf{g}^+ + \mathbf{M}^-\mathbf{g}^- = k\mathbf{g}^+, \tag{8.24}$$

$$-\mathbf{M}^-\mathbf{g}^+ - \mathbf{M}^+\mathbf{g}^- = k\mathbf{g}^-. \tag{8.25}$$

Add and subtract (8.24) and (8.25), and we have

$$(\mathbf{M}^+ - \mathbf{M}^-)(\mathbf{g}^+ - \mathbf{g}^-) = k(\mathbf{g}^+ + \mathbf{g}^-), \tag{8.26}$$

$$(\mathbf{M}^+ + \mathbf{M}^-)(\mathbf{g}^+ + \mathbf{g}^-) = k(\mathbf{g}^+ - \mathbf{g}^-). \tag{8.27}$$

Substituting (8.27) into (8.26) yields

$$(\mathbf{M}^+ - \mathbf{M}^-)(\mathbf{M}^+ + \mathbf{M}^-)(\mathbf{g}^+ + \mathbf{g}^-) = k^2(\mathbf{g}^+ + \mathbf{g}^-). \tag{8.28}$$

Equation (8.28) is an eigenvalue problem of order n. Once we obtain the eigenvectors $\mathbf{g}^+ + \mathbf{g}^-$ from (8.28), we can use (8.27) to find the vectors $\mathbf{g}^+ - \mathbf{g}^-$ and thus the eigenvectors of the original system $\mathbf{g}^\pm$ can be obtained from those of the reduced system.

The homogeneous solution can now be written

$$I(\tau, \xi_i) = \sum_{j=-n}^{n} L_j g_j(\xi_i) e^{-k_j\tau}, \tag{8.29}$$

where k_j and $g_j(\xi)$ are the eigenvalues and eigenfunctions, respectively, of (8.23), and the L_j are constants. Let $I_p(\tau, \xi)$ be the particular solution. Then the general solution is

$$I(\tau, \xi_i) = \sum_{j=-n}^{n} L_j g_j(\xi_i) e^{-k_j\tau} + I_p(\tau, \xi_i). \tag{8.30}$$

The $2n$ constants L_j are to be determined by $2n$ boundary conditions imposed on $\mathbf{I}^\pm$ at the upper and lower boundaries.

We will give two specific solutions as illustrations of the discrete ordinate method. The solution for a conservative, isotropically scattering infinite atmosphere ($a = 1$, $P = 1$) with constant flux, F, is

$$I(\tau, \xi_i) = \frac{3F}{4\pi}\left(\tau + Q + \sum_{j=1}^{n-1} L_j \frac{e^{-k_j\tau}}{1 + k_j\xi_i} + L_{-j}\frac{e^{k_j\tau}}{1 - k_j\xi_i}\right), \tag{8.31}$$

where Q and $L_{\pm j}$ are constants of integration. The source function for

this solution is

$$J(\tau) = \frac{3F}{4\pi}\left(\tau + Q + \sum_{j=1}^{n-1} L_j e^{-k_j\tau} + L_{-j}e^{k_j\tau}\right)$$
$$= \frac{3F}{4\pi}[\tau + q(\tau)], \tag{8.32}$$

where $q(\tau)$ is *Hopf's function.* In a nonconservative, isotropically scattering atmosphere illuminated by sunlight, we have a primary source, (2.116),

$$\Sigma(\tau, \xi_i) = \frac{a}{4\pi} f(0) e^{\tau/\xi_\odot}. \tag{8.33}$$

The solution in this case is

$$I(\tau, \xi_i) = \frac{af(0)H(\xi_\odot)H(-\xi_\odot)}{4(1 - \xi_i/\xi_\odot)} e^{\tau/\xi_\odot} + \sum_{j=-n}^{n} L_j \frac{e^{-k_j\tau}}{1 + k_j\xi_i}, \tag{8.34}$$

where $a \neq 1$, L_j are constants of integration, and H is *Chandrasekhar's H function*; in the *nth* approximation H is

$$H(\xi) = \prod_{j=1}^{n} \frac{(\xi + \xi_j)}{\xi_j(1 + k_j\xi)}. \tag{8.35}$$

The primary advantages of the discrete ordinate method are as follows:

1. The solution for specific intensity can be derived explicitly, and includes both the emergent as well as the internal radiation field.
2. The principal concepts such as streams and eigenvalues can be grasped intuitively. An eigenvalue k_i may be interpreted as a diffusivity factor such that $k_i\tau$ represents an effective optical path for a stream in the ξ_i direction.
3. The smallest eigenvalue corresponds to the dominant mode of propagation, and this is called the diffusion mode (§ 8.5.1). Partly because of this, the low-order approximations such as the two-stream and other multistream solutions, which can be derived in closed forms, are quite accurate (§ 8.5.6) and particularly useful for radiation flux calculations.

Closely related to the discrete ordinate method is the moment method, which uses a sequence of moments,

$$M_l(\tau) = \frac{1}{2}\int_{-1}^{1} \xi^l I(\tau, \xi)\, d\xi, \tag{8.36}$$

and the spherical harmonic method, which employs the series expansion,

$$I(\tau, \xi) = \sum_{l=0}^{\infty} I_l(\tau) P_l(\xi), \tag{8.37}$$

where $P_l(\xi)$ are Legendre polynomials. However, the boundary conditions for these two methods are more complicated.

The primary shortcoming of the discrete ordinate eigenfunction method is the requirement of vertical homogeneity. This restriction can be overcome by assuming that a nonhomogeneous atmosphere is composed of a series of adjacent homogeneous layers in which the scattering and absorption properties are allowed to vary from layer to layer. In the following section we shall describe a method that explicitly takes into account the vertical nonhomogeneity of the atmosphere.

8.2.3. *Feautrier method*

The two first-order equations (8.15) and (8.16) can be converted into a second-order equation in τ as follows. Adding and subtracting (8.15) and (8.16) lead to

$$\frac{d\mathbf{u}(\tau)}{d\tau} = \mathbf{A}(\tau)\mathbf{v}(\tau) - \mathbf{v}_0(\tau), \tag{8.38}$$

$$\frac{d\mathbf{v}(\tau)}{d\tau} = \mathbf{B}(\tau)\mathbf{u}(\tau) - \mathbf{u}_0(\tau), \tag{8.39}$$

where

$$\mathbf{u}(\tau) = \tfrac{1}{2}[\mathbf{I}^+(\tau) + \mathbf{I}^-(\tau)], \tag{8.40}$$

$$\mathbf{v}(\tau) = \tfrac{1}{2}[\mathbf{I}^+(\tau) - \mathbf{I}^-(\tau)], \tag{8.41}$$

$$\mathbf{u}_0(\tau) = \tfrac{1}{2}[\boldsymbol{\sigma}^+(\tau) - \boldsymbol{\sigma}^-(\tau)], \tag{8.42}$$

$$\mathbf{v}_0(\tau) = \tfrac{1}{2}[\boldsymbol{\sigma}^+(\tau) + \boldsymbol{\sigma}^-(\tau)], \tag{8.43}$$

and the elements of the $n \times n$ matrices are [see (8.6), (8.19) and (8.20) for definitions]

$$\begin{aligned} A_{ij}(\tau) &= M_{ij}^+(\tau) - M_{ij}^-(\tau) \\ &= \frac{1}{\xi_i}\{\delta_{ij} - \tfrac{1}{2}\gamma a_j[\rho(\tau; \xi_i, \xi_j) - \rho(\tau; \xi_i, -\xi_j)]\}, \end{aligned} \tag{8.44}$$

$$\begin{aligned} B_{ij}(\tau) &= M_{ij}^+(\tau) + M_{ij}^-(\tau) \\ &= \frac{1}{\xi_i}\{\delta_{ij} - \tfrac{1}{2}\gamma a_j[\rho(\tau; \xi_i, \xi_j) + \rho(\tau; \xi_i, -\xi_j)]\}. \end{aligned} \tag{8.45}$$

Note the similarity between $\mathbf{u}(\tau)$ and $\mathbf{v}(\tau)$ and $\bar{I}$ and F as given by (2.143) and (2.142). Indeed we may regard the Feautrier method as the multistream generalization of (2.140).

From (8.38) we have

$$\mathbf{v}(\tau) = \mathbf{A}^{-1}(\tau)\left[\frac{d\mathbf{u}(\tau)}{d\tau} + \mathbf{v}_0(\tau)\right]; \tag{8.46}$$

substituting (8.46) into (8.39) yields

$$\frac{d}{d\tau}\left[\mathbf{A}^{-1}(\tau)\frac{d\mathbf{u}(\tau)}{d\tau}\right] - \mathbf{B}(\tau)\mathbf{u}(\tau) = -\frac{d}{d\tau}[\mathbf{A}^{-1}(\tau)\mathbf{v}_0(\tau)] - \mathbf{u}_0(\tau). \tag{8.47}$$

Since $\mathbf{A}(\tau)$ and $\mathbf{B}(\tau)$ are positive definite matrices, (8.47) can be interpreted as the multicomponent generalization of the steady-state diffusion equation. For isotropic scattering in an atmosphere illuminated by sunlight, we have

$$\begin{aligned} A_{ij}^{-1}(\tau) &= \xi_i \delta_{ij}, \\ B_{ij}(\tau) &= \frac{1}{\xi_i}[\delta_{ij} - a(\tau)a_j], \\ u_0(\tau, \xi_i) &= \frac{a(\tau)}{4\pi\xi_i} f(0) e^{\tau/\xi_\odot}, \\ v_0(\tau, \xi_i) &= 0, \end{aligned} \tag{8.48}$$

and (8.47) becomes

$$\begin{aligned} \xi_i^2 \frac{d^2 u(\tau, \xi_i)}{d\tau^2} &= u(\tau, \xi_i) - a(\tau) \sum_{j=1}^{n} a_j u(\tau, \xi_j) \\ &\quad - \frac{a(\tau)}{4\pi} f(0) e^{\tau/\xi_\odot}, \end{aligned} \tag{8.49}$$

which can be recognized as a generalized steady-state diffusion equation with a source (see § 2.4.4 and § 2.4.5).

The analogy between radiative transfer and diffusion is not only mathematical. There is a deeper physical reason. We can intuitively regard $\mathbf{u}(\tau) = [\mathbf{I}^+(\tau) + \mathbf{I}^-(\tau)]/2$ as the "density" of radiation, and $\mathbf{v}(\tau) = [\mathbf{I}^+(\tau) - \mathbf{I}^-(\tau)]/2$ as the "flux". From (8.46), the "flux" is proportional to the gradient of the "density" [let us ignore for the moment the source term $\mathbf{v}_0(\tau)$]. This result is what we would expect if the photons perform random motion back and forth as molecules do in kinetic theory of gases,

and this type of transport is known as diffusion. Equation (8.39) states that the "density" is controlled by the divergence of the "flux" and a source term, an alternative form of the continuity equation.

All the usual mathematical techniques for solving the steady-state diffusion equation can now be applied to (8.47). A convenient method is discretization in τ. The resulting equation may be written in block tridiagonal form and can be easily and accurately inverted. The chief advantage of this method is the ability to deal with nonhomogeneous atmospheres. The main restriction of this method is that the atmosphere must not vary too rapidly in the vertical because of the occurrence of the term $\mathbf{A}^{-1}(\tau)$ under the derivative sign in (8.47).

8.3. Interaction principle

The integrodifferential equation approach described in the previous sections may be regarded as a "microscopic" view of radiative transfer. The master equation (8.1) relates the radiation field at τ to that at $\tau + d\tau$, where $d\tau$ is infinitesimally small. There is, however, an alternative approach to radiative transfer, which may be regarded as the "macroscopic" approach. The key relationship, the equivalent to (8.1), is the *interaction principle.* This is a conservation relationship that equates the radiation emerging from an arbitrary layer to the incident fluxes on the boundary and the source distribution within the layer. Thus, we are interested only in the overall transfer properties of the layer. It can be shown that these two different approaches to radiative transfer are, in fact, equivalent.

In a plane parallel atmosphere, the physical properties depend upon a single variable τ. At any level we may define the upward and downward directed specific intensities $\mathbf{I}^+(\tau)$ and $\mathbf{I}^-(\tau)$ as given by (8.17). Consider a layer of optical depth, τ_a, bounded by surfaces $\tau = \tau_1$ and $\tau = \tau_2$ ($\tau_a = \tau_2 - \tau_1$, see Fig. 8.1). The interaction principle states that there is a linear conservation principle that relates the radiation emerging from the layer, $\mathbf{I}^+(\tau_1)$ and $\mathbf{I}^-(\tau_2)$, to the incident radiation, $\mathbf{I}^-(\tau_1)$ and $\mathbf{I}^+(\tau_2)$, and to the source

$$\mathbf{I}^+(\tau_1) = \mathbf{R}_a\mathbf{I}^-(\tau_1) + \mathbf{T}_a^*\mathbf{I}^+(\tau_2) + \mathbf{\Sigma}_a^+, \tag{8.50}$$

$$\mathbf{I}^-(\tau_2) = \mathbf{T}_a\mathbf{I}^-(\tau_1) + \mathbf{R}_a^*\mathbf{I}^+(\tau_2) + \mathbf{\Sigma}_a^-, \tag{8.51}$$

where $\mathbf{R}_a$ and $\mathbf{T}_a$ are reflection and transmission operators for the downward directed incident stream of radiation, $\mathbf{R}_a^*$ and $\mathbf{T}_a^*$ are the corresponding operators for the upward directed incident stream, and $\mathbf{\Sigma}_a^{\pm}$ represent emission of radiation originating in the layer. The discrete ordinate representations of the reflection and transmission operators remain to be specified (see § 8.3.3). Note that in general $\mathbf{R}_a \neq \mathbf{R}_a^*$ and

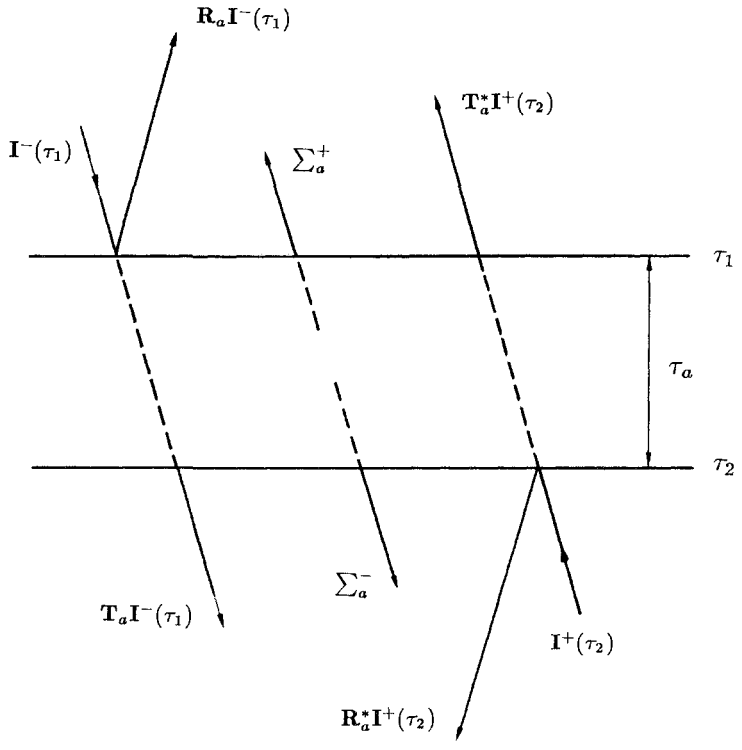


FIG. 8.1. The interaction principle. The figure illustrates the quantities and definitions introduced in the text.

$\mathbf{T}_a \neq \mathbf{T}_a^*$. Equations (8.50) and (8.51) may be written compactly as

$$\begin{vmatrix} \mathbf{I}^+(\tau_1) \\ \mathbf{I}^-(\tau_2) \end{vmatrix} = \mathbf{S}(a) \begin{vmatrix} \mathbf{I}^-(\tau_1) \\ \mathbf{I}^+(\tau_2) \end{vmatrix} + \begin{vmatrix} \mathbf{\Sigma}_a^+ \\ \mathbf{\Sigma}_a^- \end{vmatrix}, \tag{8.52}$$

where,

$$\mathbf{S}(a) = \begin{vmatrix} \mathbf{R}_a & \mathbf{T}_a^* \\ \mathbf{T}_a & \mathbf{R}_a^* \end{vmatrix}. \tag{8.53}$$

8.3.1. *Adding two layers*

Repeated applications of the interaction principle enables us to find the rule for combining two layers. This can be shown as follows. Figure 8.2 shows a schematic of two adjacent layers of optical depths τ_a and τ_b, respectively, bounded by the planes $\tau = \tau_1$, τ_2 and τ_3. For the first layer, the interaction principle gives (8.52). A similar relation is obtained for

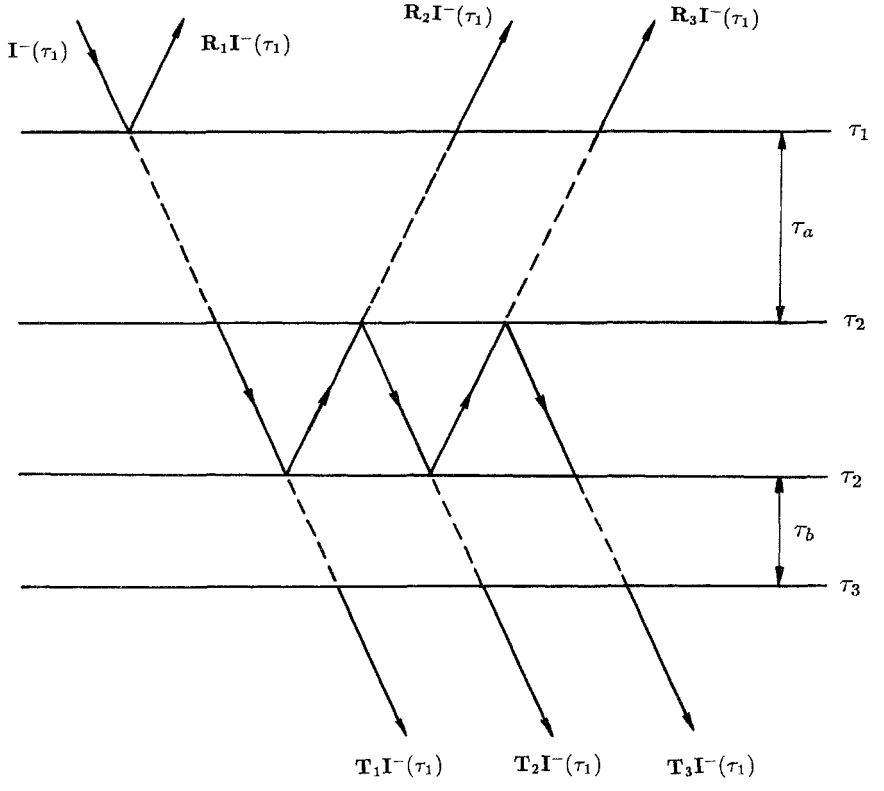


FIG. 8.2. Combining two layers. The two layers, a and b, can together be treated as a single layer, c.

the second layer,

$$\begin{vmatrix} \mathbf{I}^+(\tau_2) \\ \mathbf{I}^-(\tau_3) \end{vmatrix} = \mathbf{S}(b) \begin{vmatrix} \mathbf{I}^-(\tau_2) \\ \mathbf{I}^+(\tau_3) \end{vmatrix} + \begin{vmatrix} \mathbf{\Sigma}_b^+ \\ \mathbf{\Sigma}_b^- \end{vmatrix}, \tag{8.54}$$

where

$$\mathbf{S}(b) = \begin{vmatrix} \mathbf{R}_b & \mathbf{T}_b^* \\ \mathbf{T}_b & \mathbf{R}_b^* \end{vmatrix}. \tag{8.55}$$

We can eliminate $\mathbf{I}^{\pm}(\tau_2)$ from (8.52) and (8.54), and obtain a relation for the combined layer of optical depth $\tau_a + \tau_b$,

$$\begin{vmatrix} \mathbf{I}^+(\tau_1) \\ \mathbf{I}^-(\tau_3) \end{vmatrix} = \mathbf{S}(a+b) \begin{vmatrix} \mathbf{I}^-(\tau_1) \\ \mathbf{I}^+(\tau_3) \end{vmatrix} + \begin{vmatrix} \mathbf{\Sigma}_{a+b}^+ \\ \mathbf{\Sigma}_{a+b}^- \end{vmatrix}, \tag{8.56}$$

where

$$\mathbf{S}(a+b) = \begin{vmatrix} \mathbf{R}_a + \mathbf{T}_a^*(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1}\mathbf{R}_b\mathbf{T}_a & \mathbf{T}_a^*(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1}\mathbf{T}_b^* \\ \mathbf{T}_b(\mathbf{I}-\mathbf{R}_a^*\mathbf{R}_b)^{-1}\mathbf{T}_a & \mathbf{R}_b + \mathbf{T}_b(\mathbf{I}-\mathbf{R}_a^*\mathbf{R}_b)^{-1}\mathbf{R}_a^*\mathbf{T}_b^* \end{vmatrix}, \quad (8.57)$$

and

$$\begin{vmatrix} \boldsymbol{\Sigma}_{a+b}^+ \\ \boldsymbol{\Sigma}_{a+b}^- \end{vmatrix} = \begin{vmatrix} \boldsymbol{\Sigma}_a^+ \\ \boldsymbol{\Sigma}_b^- \end{vmatrix} + \begin{vmatrix} \mathbf{T}_a^*(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1}\mathbf{R}_b & \mathbf{T}_a^*(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1} \\ \mathbf{T}_b(\mathbf{I}-\mathbf{R}_a^*\mathbf{R}_b)^{-1} & \mathbf{T}_b(\mathbf{I}-\mathbf{R}_a^*\mathbf{R}_b)^{-1}\mathbf{R}_a^* \end{vmatrix} \begin{vmatrix} \boldsymbol{\Sigma}_a^- \\ \boldsymbol{\Sigma}_b^+ \end{vmatrix}. \quad (8.58)$$

$\mathbf{I}$ is the identity operator. Since the interaction principle is completely general, we can apply it to the combined layer $\tau_c = \tau_a + \tau_b = \tau_3 - \tau_1$ and obtain

$$\begin{vmatrix} \mathbf{I}^+(\tau_1) \\ \mathbf{I}^-(\tau_3) \end{vmatrix} = \mathbf{S}(c) \begin{vmatrix} \mathbf{I}^-(\tau_1) \\ \mathbf{I}^+(\tau_3) \end{vmatrix} + \begin{vmatrix} \boldsymbol{\Sigma}_c^+ \\ \boldsymbol{\Sigma}_c^- \end{vmatrix}, \quad (8.59)$$

where

$$\mathbf{S}(c) = \begin{vmatrix} \mathbf{R}_c & \mathbf{T}_c^* \\ \mathbf{T}_c & \mathbf{R}_c^* \end{vmatrix}. \quad (8.60)$$

It follows that (8.56) and (8.59) must be identical, that is

$$\mathbf{S}(c) = \mathbf{S}(a+b), \quad (8.61)$$

$$\boldsymbol{\Sigma}_c^{\pm} = \boldsymbol{\Sigma}_{a+b}^{\pm}. \quad (8.62)$$

Equating the individual components of (8.61), we obtain the rule for combining the reflection and transmission operators,

$$\mathbf{R}_c = \mathbf{R}_a + \mathbf{T}_a^*(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1}\mathbf{R}_b\mathbf{T}_a, \quad (8.63)$$

$$\mathbf{T}_c = \mathbf{T}_b(\mathbf{I}-\mathbf{R}_a^*\mathbf{R}_b)^{-1}\mathbf{T}_a, \quad (8.64)$$

$$\mathbf{R}_c^* = \mathbf{R}_b + \mathbf{T}_b(\mathbf{I}-\mathbf{R}_a^*\mathbf{R}_b)^{-1}\mathbf{R}_a^*\mathbf{T}_b^*, \quad (8.65)$$

$$\mathbf{T}_c^* = \mathbf{T}_a^*(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1}\mathbf{T}_b^*. \quad (8.66)$$

The physical meaning of (8.63)–(8.66) becomes clear from Fig. 8.2. Consider the incident stream $\mathbf{I}^-(\tau_1)$. The reflected stream can be written as a superposition of streams that have undergone single and multiple

reflections. By inspection,

$$\mathbf{R}_1 = \mathbf{R}_a, \tag{8.67}$$

$$\mathbf{R}_2 = \mathbf{T}_a^* \mathbf{R}_b \mathbf{T}_a, \tag{8.68}$$

and, in general,

$$\mathbf{R}_n = \mathbf{T}_a^* (\mathbf{R}_b \mathbf{R}_a^*)^{n-1} \mathbf{R}_b \mathbf{T}_a \qquad (n = 2, 3, \ldots). \tag{8.69}$$

Therefore, the total reflected intensity is

$$\begin{aligned} \mathbf{I}^+(\tau_1) &= \left[\mathbf{R}_a + \mathbf{T}_a^* \sum_{n=1}^{\infty} (\mathbf{R}_b \mathbf{R}_a^*)^{n-1} \mathbf{R}_b \mathbf{T}_a \right] \mathbf{I}^-(\tau_1), \\ &= [\mathbf{R}_a + \mathbf{T}_a^* (\mathbf{I} - \mathbf{R}_b \mathbf{R}_a^*)^{-1} \mathbf{R}_b \mathbf{T}_a] \mathbf{I}^-(\tau_1), \\ &= \mathbf{R}_c \mathbf{I}^-(\tau_1). \end{aligned} \tag{8.70}$$

This explains the meaning of (8.63). Similarly, inspection of Fig. 8.2 yields

$$\mathbf{T}_c = \sum_{n=1}^{\infty} \mathbf{T}_n, \tag{8.71}$$

where

$$\mathbf{T}_1 = \mathbf{T}_b \mathbf{T}_a, \tag{8.72}$$

$$\mathbf{T}_2 = \mathbf{T}_b \mathbf{R}_a^* \mathbf{R}_b \mathbf{T}_a, \tag{8.73}$$

and

$$\mathbf{T}_n = \mathbf{T}_b (\mathbf{R}_a^* \mathbf{R}_b)^{n-1} \mathbf{T}_a \qquad (n = 3, 4, \ldots). \tag{8.74}$$

Substituting (8.72)–(8.74) into (8.71) leads to (8.64). A straightforward extension of the above reasoning to an upward directed stream $\mathbf{I}^+(\tau_3)$ provides a similar physical interpretation of (8.65) and (8.66).

8.3.2. *The star semigroup*

Given the operator $\mathbf{S}(a)$ for layer a, as defined by (8.53), and $\mathbf{S}(b)$ for layer b, as defined by (8.55), (8.57) states a rule for computing $\mathbf{S}(a + b)$ for the combined layer $a + b$. This simple binary composition rule may be formally defined as a star product,

$$\mathbf{S}(a + b) = \mathbf{S}(a) * \mathbf{S}(b), \tag{8.75}$$

where the order of adding the layers is important. In general $\mathbf{S}(a+b) \neq \mathbf{S}(b+a)$, and the star multiplication is noncommutative, except for the special case in which the combined layer is homogeneous. It can be shown that the star product is associative for the combination of three layers a, b, and c,

$$\mathbf{S}(a+b+c) = \mathbf{S}(a) * [\mathbf{S}(b) * \mathbf{S}(c)] = [\mathbf{S}(a) * \mathbf{S}(b)] * \mathbf{S}(c). \quad (8.76)$$

For a layer of zero optical thickness we define

$$\mathbf{S}(0) = \begin{pmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix}. \quad (8.77)$$

It is clear from (8.57) that

$$\mathbf{S}(0+a) = \mathbf{S}(0) * \mathbf{S}(a) = \mathbf{S}(a) * \mathbf{S}(0) = \mathbf{S}(a+0) = \mathbf{S}(a), \quad (8.78)$$

for any layer a. Note that $\mathbf{S}(0)$ plays the role of an identity operator for star multiplication.

Let A be an arbitrary collection of plane-parallel slabs. Then the set of operators $\{\mathbf{S}(a): a \in A\}$ forms a semigroup with respect to star multiplication. Note that the essential elements of the semigroup structure are the binary composition rule (8.75), the associative law (8.76), and the existence of an identity operator. The semigroup differs from the group in that the inverse of $\mathbf{S}(a)$ generally does not exist, i.e., for an arbitrary operator $\mathbf{S}(a)$ we cannot find an $\mathbf{S}(b)$ such that

$$\mathbf{S}(a) * \mathbf{S}(b) = \mathbf{S}(0). \quad (8.79)$$

The physical reason is obvious. In multiple scattering, the entropy of radiation always increases, except in the trivial case of scattering by a layer of zero optical thickness, which conserves entropy. Hence no combination of nontrivial layers can simulate the net effect of a trivial layer.

8.3.3. *Doubling and Adding*

The results of the previous sections, § 8.3, § 8.3.1, and § 8.3.2, suggest that there is an alternative approach to study the equation of radiative transfer. Instead of solving for the specific intensities, we can study the reflection and transmission properties of a medium using the simple combination laws derived from the interaction principle. We shall first give a precise definition of the diffuse reflection and transmission

functions for a plane-parallel layer bounded by τ_1 and τ_2 ($\tau_a = \tau_2 - \tau_1$, see Fig. 8.1),

$$I_r^+(\tau_1; \mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R(\tau_a; \mu, \phi; \mu', \phi') I^-(\tau_1; \mu', \phi') \mu' \, d\mu' \, d\phi', \tag{8.80}$$

$$I_t^-(\tau_2; \mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T(\tau_a; \mu, \phi; \mu', \phi') I^-(\tau_1; \mu', \phi') \mu' \, d\mu' \, d\phi', \tag{8.81}$$

$$I_r^-(\tau_2; \mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R^*(\tau_a; \mu, \phi; \mu', \phi') I^+(\tau_2; \mu', \phi') \mu' \, d\mu' \, d\phi', \tag{8.82}$$

$$I_t^+(\tau_1; \mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T^*(\tau_a; \mu, \phi; \mu', \phi') I^+(\tau_2, \mu', \phi') \mu' \, d\mu' \, d\phi', \tag{8.83}$$

where $I_r^+(\tau_1; \mu, \phi)$ and $I_t^+(\tau_1; \mu, \phi)$ are the reflected and transmitted streams at the upper boundary, and $I_r^-(\tau_2; \mu, \phi)$ and $I_t^-(\tau_2; \mu, \phi)$ are the corresponding streams at the lower boundary. In previous chapters and sections we have maintained a consistent vector notation for ξ, except when we discussed discrete ordinates, $\pm\xi_i$, defined for $\xi_i > 0$ only. Since intensities, reflection, and transmission functions are now defined separately for upward and downward streams we no longer need the automatic signing provided by the signed ξ, and it is convenient to define $\mu = |\xi|$ as the continuous variable associated with ξ_i.

As an illustration of the above formulas, consider sunlight incident on top of a layer with $\tau_1 = 0$ and $\tau_2 = \tau_a$. According to the discussion of the solar beam in § 2.3.4, it may be approximated by the product of Dirac δ-functions,

$$I_\odot^-(0; \mu, \phi) = \delta(\mu - \mu_0)\delta(\phi - \phi_0) f(0), \tag{8.84}$$

where $\mu_0 = |\xi_\odot|$ and $\phi_0 = \phi_\odot$. Note that $I_\odot^-$ denotes a downward traveling beam and μ_0 is a positive quantity. Using (8.80) and (8.81), we obtain

$$R(\tau_a; \mu, \phi; \mu_0, \phi_0) = \pi I_r^+(0; \mu, \phi)/\mu_0 f(0), \tag{8.85}$$

and

$$T(\tau_a; \mu, \phi; \mu_0, \phi_0) = \pi I_t^-(\tau_a; \mu, \phi)/\mu_0 f(0). \tag{8.86}$$

For an optically thin layer $\tau_a \ll 1$, it can be shown (see § 8.4.1)

$$R(\tau_a; \mu, \phi; \mu_0, \phi_0) = \frac{a}{4(\mu + \mu_0)} P(\mu, \phi; -\mu_0, \phi_0) \times \left\{1 - \exp\left[-\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)\tau_a\right]\right\} \approx \frac{a\tau_a}{4\mu\mu_0} P(\mu, \phi; -\mu_0, \phi_0), \tag{8.87}$$

$$T(\tau_a; \mu, \phi; \mu_0, \phi_0) = \frac{a}{4(\mu - \mu_0)} P(-\mu, \phi; -\mu_0, \phi_0) \times \left[\exp\left(-\frac{\tau_a}{\mu}\right) - \exp\left(-\frac{\tau_a}{\mu_0}\right)\right] \approx \frac{a\tau_a}{4\mu\mu_0} P(-\mu, \phi; -\mu_0, \phi_0) \quad (\mu \neq \mu_0) = \frac{a\tau_a}{4\mu_0^2} P(-\mu_0, \phi; -\mu_0, \phi_0) e^{-\tau_a/\mu_0} \quad (\mu = \mu_0). \tag{8.88}$$

Similar results are obtained for R^* and T^* by considering a narrow beam incident on the lower boundary,

$$R^*(\tau_a; \mu, \phi; \mu_0, \phi_0) = \frac{a}{4(\mu + \mu_0)} P(-\mu, \phi; \mu_0, \phi_0) \times \left\{1 - \exp\left[-\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)\tau_a\right]\right\}, \tag{8.89}$$

$$T^*(\tau_a; \mu, \phi; \mu_0, \phi_0) = \frac{a}{4(\mu - \mu_0)} P(\mu, \phi; \mu_0, \phi_0) \times \left[\exp\left(-\frac{\tau_a}{\mu}\right) - \exp\left(-\frac{\tau_a}{\mu_0}\right)\right] \quad (\mu \neq \mu_0) = \frac{a\tau_a}{4\mu_0^2} P(\mu, \phi; \mu_0, \phi_0) e^{-\tau_a/\mu_0} \quad (\mu = \mu_0). \tag{8.90}$$

Each specific intensity in (8.80)–(8.83) can be expanded as a Fourier series according to (8.10), and each reflection or transmission function admits a similar Fourier decomposition

$$X(\tau; \mu, \phi; \mu', \phi') = \sum_{m=0}^{N} (2 - \delta_{0m}) X^m(\tau; \mu, \mu') \cos m(\phi - \phi'), \tag{8.91}$$

where $X = R$, T, R^*, or T^*. Replacing the relevant functions in

(8.80)–(8.83) by their appropriate Fourier series, we have for each Fourier component

$$I_r^+(\tau_1, \mu) = 2\int_0^1 R(\tau_a; \mu, \mu')I^-(\tau_1, \mu')\mu'\, d\mu', \tag{8.92}$$

$$I_t^-(\tau_2, \mu) = 2\int_0^1 T(\tau_a; \mu, \mu')I^-(\tau_1, \mu')\mu'\, d\mu', \tag{8.93}$$

$$I_r^-(\tau_2, \mu) = 2\int_0^1 R^*(\tau_a; \mu, \mu')I^+(\tau_2, \mu')\mu'\, d\mu', \tag{8.94}$$

$$I_t^+(\tau_1, \mu) = 2\int_0^1 T^*(\tau_a; \mu, \mu')I^+(\tau_2, \mu')\mu'\, d\mu', \tag{8.95}$$

where the azimuthal index m (which should label all the specific intensities and the reflection and transmission functions) has been omitted for simplicity. Following the standard procedure of discretizing the angle variable μ using Gaussian divisions and replacing the integrals by Gaussian quadratures (see § 8.2.2), we obtain the following vector equations,

$$\mathbf{I}_r^+(\tau_1) = \mathbf{R}_a\mathbf{I}^-(\tau_1), \tag{8.96}$$

$$\mathbf{I}_t^-(\tau_2) = \mathbf{T}_a\mathbf{I}^-(\tau_1), \tag{8.97}$$

$$\mathbf{I}_r^-(\tau_2) = \mathbf{R}_a^*\mathbf{I}^+(\tau_2), \tag{8.98}$$

$$\mathbf{I}_t^+(\tau_1) = \mathbf{T}_a^*\mathbf{I}^+(\tau_2), \tag{8.99}$$

where the vectors $\mathbf{I}^{\pm}(\tau)$ are defined by (8.17), and the reflection and transmission operators, first defined in (8.50) and (8.51), are $n \times n$ matrices whose elements are given by

$$X_{aij} = 2X(\tau_a; \mu_i, \mu_j)a_j\mu_j, \tag{8.100}$$

for $X_a = R_a$ or R_a^*, and by

$$Y_{aij} = e^{-\tau_a/\mu_i}\delta_{ij} + 2Y(\tau_a; \mu_i, \mu_j)a_j\mu_j, \tag{8.101}$$

for $Y_a = T_a$ or T_a^*. Note that in the expression for the transmission operator (8.101), we have included the direct as well as diffuse transmission. Also, the right-hand side of (8.100) may be viewed as the multiplication of two matrices $2X(\tau_a; \mu_i, \mu_j)$ and $(\delta_{jk}a_k\mu_k)$ where the second matrix is diagonal. The same viewpoint applies to (8.101). Since a nontrivial diagonal matrix can be inverted, we may use (8.100) and (8.101) to express $X(\tau_a; \mu_i, \mu_j)$ and $Y(\tau_a; \mu_i, \mu_j)$ in terms of X_{aij} and Y_{aij}.

From the formulas for reflection and transmission functions (8.87)–(8.90), we can accurately compute the **S** operator, i.e., the reflection and transmission operators, using (8.100) and (8.101) for an extremely thin layer, e.g., $\tau_a = 2^{-20}\tau_0$, where τ_0 is of order of unity. Repeated application of (8.63)–(8.66), putting together two identical layers, allows us to generate the **S** operator for an homogeneous layer of finite optical thickness. Since

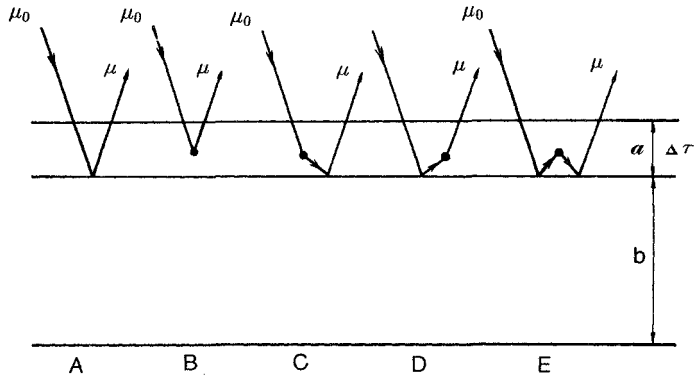
$$\mathbf{S}(2^p\tau_a) = \mathbf{S}(2^{p-1}\tau_a) * \mathbf{S}(2^{p-1}\tau_a) \qquad p = 1, 2, 3, \ldots, \tag{8.102}$$

we may generate $\mathbf{S}(\tau_0)$ in 20 steps in this case, compared with 2^{20} steps required for the direct adding method (8.61). The algorithm (8.102) is known as the *doubling method.* This algorithm works only for a homogeneous layer. For a nonhomogeneous atmosphere, the usual strategy is to divide the atmosphere into a large number of thin but finite homogeneous layers. The doubling method is applied to generate the **S** operator for each homogeneous layer from an infinitesmal layer. The distinct finite layers are then combined, using the adding method. The boundary conditions at the upper and lower boundaries are simulated by layers with special properties. Having obtained the **S** operators for the entire atmosphere, it is straightforward to apply them to calculate the source terms using (8.58).

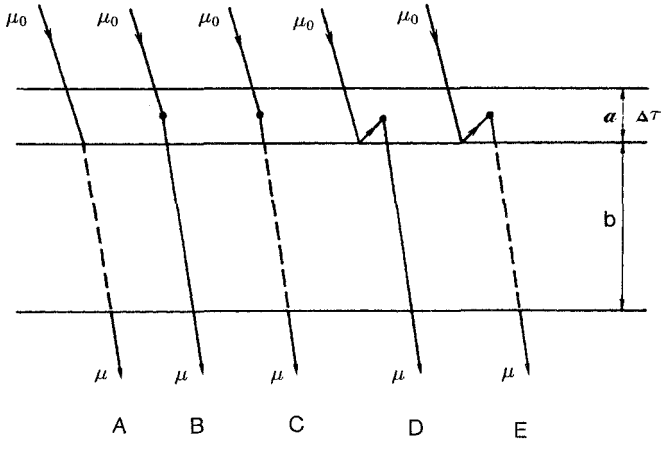
The chief advantages of the doubling and adding method are (1) simple physical interpretation, (2) simple mathematical operations such as matrix multiplications, (3) the reflection and transmission operators are obtained for all incident and emergent angles at once, (4) there is no difference in principle in the computations for isotropic and anisotropic phase functions, although in practice more Gaussian points are needed for the latter case, and (5) results are obtained for a range of optical thickness between the final value and the starting value. The main drawback of the method seems to be the somewhat artificial manner of dealing with nonhomogeneous atmospheres. This criticism is of course academic, since we can always approximate the properties of a nonhomogeneous atmosphere to arbitrary accuracy using a finite number of homogeneous layers. However, there is an alternative method, which is based on a natural and elegant extension of the present method to a nonhomogeneous atmosphere. This will be the subject of the following section.

8.3.4. *Invariant imbedding*

Consider the combination of two layers a and b, as shown in Fig. 8.3, in which the top layer is infinitesimally thin ($\tau_a = \Delta\tau$) and the bottom layer is of finite optical depth ($\tau_b = \tau$). The reflection and transmission operators for the combination can be computed from (8.63)–(8.66). Let



(a)



(b)

FIG. 8.3. Adding an infinitesimal layer to a finite layer. There is, at most, one scattering in the infinitesimal layer. A dashed line represents diffusive transmission (scattering). The five processes are described in the text.

us first consider (8.63). If we expand the operator $(\mathbf{I}-\mathbf{R}_b\mathbf{R}_a^*)^{-1}$, we obtain an infinite series in $\mathbf{R}_b\mathbf{R}_a^*$. Since $\mathbf{R}_a^*$ is proportional to $\Delta\tau$, the result is an infinite series in $\Delta\tau$. Keeping only the terms to order $\Delta\tau$, (8.63) becomes

$$\mathbf{R}_{a+b} \approx \mathbf{R}_a + \mathbf{T}_a^*\mathbf{R}_b\mathbf{T}_a + \mathbf{T}_a^*\mathbf{R}_b\mathbf{R}_a^*\mathbf{R}_b\mathbf{T}_a. \tag{8.103}$$

This equation will enable us to derive a differential equation in τ for the

reflection and transmission functions. From (8.100), (8.101), and (8.87)–(8.90), the reflection and transmission operators can be expressed in terms of the corresponding functions to order $\Delta\tau$,

$$R_{bij} = 2R(\tau_b; \mu_i, \mu_j)a_j\mu_j = 2R(\tau; \mu_i, \mu_j)a_j\mu_j, \tag{8.104}$$

$$R_{(a+b)ij} = 2R(\tau_a + \tau_b; \mu_i, \mu_j)a_j\mu_j = 2R(\tau + \Delta\tau; \mu_i, \mu_j)a_j\mu_j, \tag{8.105}$$

$$R_{aij} = 2R(\tau_a; \mu_i, \mu_j)a_j\mu_j = \frac{a\,\Delta\tau}{2\mu_i} P(\mu_i, -\mu_j)a_j, \tag{8.106}$$

$$R^*_{aij} = 2R^*(\tau_a; \mu_i, \mu_j)a_i\mu_j = \frac{a\,\Delta\tau}{2\mu_i} P(-\mu_i, \mu_j)a_j, \tag{8.107}$$

$$T_{aij} = e^{-\tau_a/\mu_i}\delta_{ij} + 2T(\tau_a; \mu_i, \mu_j)a_j\mu_j = \left(1 - \frac{\Delta\tau}{\mu_i}\right)\delta_{ij} + \frac{a\Delta\tau}{2\mu_i} P(\mu_i, \mu_j)a_j, \tag{8.108}$$

$$T^*_{aij} = e^{-\tau_a/\mu_i}\delta_{ij} + 2T^*(\tau_a; \mu_i, \mu_j)a_j\mu_j = \left(1 - \frac{\Delta\tau}{\mu_i}\right)\delta_{ij} + \frac{a\Delta\tau}{2\mu_i} P(-\mu_i, -\mu_j)a_j. \tag{8.109}$$

Substituting (8.104)–(8.109) into (8.103) and simplifying the algebra, we have, to order $\Delta\tau$,

$$\begin{aligned} R(\tau + \Delta\tau; \mu_i, \mu_j) = {} & R(\tau; \mu_i, \mu_j) - \left(\frac{1}{\mu_i} + \frac{1}{\mu_j}\right)\Delta\tau R(\tau; \mu_i, \mu_j) \\ & + \frac{a\,\Delta\tau}{4\mu_i\mu_j} P(\mu_i, -\mu_j) + \frac{a\,\Delta\tau}{2\mu_j} \sum_{k=1}^{n} R(\tau; \mu_i, \mu_k)a_k P(\mu_k, \mu_j) \\ & + \frac{a\,\Delta\tau}{2\mu_i} \sum_{k=1}^{n} P(-\mu_i, -\mu_k)a_k R(\tau; \mu_k, \mu_j) \\ & + a\,\Delta\tau \sum_{k=1}^{n} R(\tau; \mu_i, \mu_k)a_k \\ & \times \left[\sum_{l=1}^{n} P(-\mu_k, \mu_l)a_l R(\tau; \mu_l, \mu_j)\right]. \end{aligned} \tag{8.110}$$

On taking the limit $\Delta\tau \to 0$ and $n \to \infty$, we obtain a nonlinear

integrodifferential equation for $R(\tau; \mu, \mu_0)$,

$$\frac{\partial R(\tau; \mu, \mu_0)}{\partial \tau} = -\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) R(\tau; \mu, \mu_0) + \frac{a(\tau)}{4\mu\mu_0} P(\mu, -\mu_0)$$
$$+ \frac{a(\tau)}{2\mu_0} \int_0^1 R(\tau; \mu, \mu') P(\mu', \mu_0)\, d\mu'$$
$$+ \frac{a(\tau)}{2\mu} \int_0^1 P(-\mu, -\mu') R(\tau; \mu', \mu_0)\, d\mu'$$
$$+ a(\tau) \int_0^1 R(\tau; \mu, \mu') \left[\int_0^1 P(-\mu', \mu'') R(\tau; \mu'', \mu_0)\, d\mu'' \right] d\mu'. \tag{8.111}$$

This is known as the *invariant imbedding equation* for the reflection function. The physical meaning of (8.111) is clear from an inspection of Fig. 8.3a. The five terms on the right-hand side of (8.111) correspond to the five processes marked A, B, C, D, E in the same order in the figure.

Starting from (8.64), and applying the same procedure as above, we obtain an integrodifferential equation for the transmission function,

$$\frac{\partial T(\tau; \mu, \mu_0)}{\partial \tau} = -\frac{1}{\mu} T(\tau; \mu, \mu_0) + \frac{a(\tau)}{4\mu\mu_0} e^{-\tau/\mu_0} P(-\mu, -\mu_0)$$
$$+ \frac{a(\tau)}{2\mu} \int_0^1 P(-\mu, -\mu') T(\tau; \mu', \mu_0)\, d\mu'$$
$$+ \frac{a(\tau)}{2\mu_0} e^{-\tau/\mu_0} \int_0^1 R(\tau; \mu, \mu') P(\mu', -\mu_0)\, d\mu'$$
$$+ a(\tau) \int_0^1 R(\tau; \mu, \mu') \left[\int_0^1 P(\mu', -\mu'') T(\tau; \mu'', \mu_0)\, d\mu'' \right] d\mu'. \tag{8.112}$$

The physical meaning of the five terms on the right-hand side of (8.112) is illustrated by A, B, C, D, E in Fig. 8.3b. Note that (8.111) is a nonlinear integrodifferential equation in $R(\tau; \mu, \mu_0)$. We can start from the lower boundary, where the reflection function $R(\tau_1; \mu, \mu_0)$ must be prescribed, and use (8.112) to integrate upward to obtain the reflection function for all values of τ. In contrast, (8.112) is a linear integrodifferential equation in $T(\tau; \mu, \mu_0)$. However, the function $R(\tau; \mu, \mu_0)$ must be computed first.

We have derived (8.111) and (8.112) by adding an infinitesimal layer to the top of a finite layer. Obviously we can accomplish the same purpose by placing an infinitesimal layer at the bottom of the finite layer.

This gives rise to two more equations,

$$\frac{\partial R(\tau;\mu,\mu_0)}{\partial\tau}=\frac{a(\tau)}{4\mu\mu_0}\exp\left[-\tau\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)\right]P(\mu,-\mu_0)$$
$$+\frac{a(\tau)}{2\mu}e^{-\tau/\mu}\int_0^1 P(\mu,-\mu')T(\tau;\mu',\mu_0)\,d\mu'$$
$$+\frac{a(\tau)}{2\mu_0}e^{-\tau/\mu_0}\int_0^1 T(\tau;\mu,\mu')P(\mu',-\mu_0)\,d\mu'$$
$$+a(\tau)\int_0^1 T(\tau;\mu,\mu')\left[\int_0^1 P(\mu',-\mu'')T(\tau;\mu'',\mu_0)\,d\mu''\right]d\mu', \tag{8.113}$$

$$\frac{\partial T(\tau;\mu,\mu_0)}{\partial\tau}=-\frac{1}{\mu_0}T(\tau;\mu,\mu_0)+\frac{a(\tau)}{4\mu\mu_0}e^{-\tau/\mu}P(-\mu,-\mu_0)$$
$$+\frac{a(\tau)}{2\mu}e^{-\tau/\mu}\int_0^1 P(-\mu,\mu')R(\tau;\mu',\mu_0)\,d\mu'$$
$$+\frac{a(\tau)}{2\mu_0}\int_0^1 T(\tau;\mu,\mu')P(-\mu',-\mu_0)\,d\mu'$$
$$+a(\tau)\int_0^1 T(\tau;\mu,\mu')\left[\int_0^1 P(-\mu',\mu'')R(\tau;\mu'',\mu_0)\,d\mu''\right]d\mu'. \tag{8.114}$$

These two equations do not provide additional information on the reflection and transmission functions. In fact, (8.113) and (8.114) are coupled nonlinear integrodifferential equations and are more difficult to solve than (8.110) and (8.112). Except for the simple cases referred to in § 8.3.5, the invariant imbedding method has not been as extensively used as the doubling and adding method. Based on limited available information, the method is almost as efficient as the doubling method for a homogeneous atmosphere but can be N times faster than the adding method for a nonhomogeneous atmosphere approximated by N homogeneous layers.

8.3.5. *X, Y, and H functions*

The invariant imbedding equations derived in the previous section are applicable to an arbitrary nonhomogeneous atmosphere with anisotropic scattering. In this section we shall examine the particularly simple case of isotropic scattering in a homogeneous atmosphere. Setting $P(\mu,\mu_0)=1$

and $a(\tau) = a$ in (8.111)–(8.114) we have

$$\frac{\partial R(\tau; \mu, \mu_0)}{\partial \tau} + \left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) R(\tau; \mu, \mu_0) = \frac{a}{4\mu\mu_0} X(\mu) X(\mu_0), \quad (8.115)$$

$$\frac{\partial R(\tau; \mu, \mu_0)}{\partial \tau} = \frac{a}{4\mu\mu_0} Y(\mu) Y(\mu_0), \quad (8.116)$$

$$\frac{\partial T(\tau; \mu, \mu_0)}{\partial \tau} + \frac{1}{\mu} T(\tau; \mu, \mu_0) = \frac{a}{4\mu\mu_0} X(\mu) Y(\mu_0), \quad (8.117)$$

$$\frac{\partial T(\tau; \mu, \mu_0)}{\partial \tau} + \frac{1}{\mu_0} T(\tau; \mu, \mu_0) = \frac{a}{4\mu\mu_0} X(\mu_0) Y(\mu), \quad (8.118)$$

where *Chandrasekhar's X and Y functions* are defined by

$$X(\mu) = 1 + 2\mu \int_0^1 R(\tau; \mu, \mu')\, d\mu', \quad (8.119)$$

$$Y(\mu) = e^{-\tau/\mu} + 2\mu \int_0^1 T(\tau; \mu, \mu')\, d\mu'. \quad (8.120)$$

Eliminating $\partial R/\partial \tau$ from (8.115) and (8.116), we obtain

$$R(\tau; \mu, \mu_0) = \frac{a}{4(\mu + \mu_0)} [X(\mu) X(\mu_0) - Y(\mu) Y(\mu_0)], \quad (8.121)$$

and, eliminating $\partial T/\partial \tau$ from (8.117) and (8.118), we have

$$T(\tau; \mu, \mu_0) = \frac{a}{4(\mu_0 - \mu)} [X(\mu) Y(\mu_0) - X(\mu_0) Y(\mu)]. \quad (8.122)$$

The equations governing the X and Y functions can be derived by inserting (8.121) and (8.122) into (8.119) and (8.120),

$$X(\mu) = 1 + \tfrac{1}{2} a\mu \int_0^1 \frac{1}{\mu + \mu'} [X(\mu) X(\mu') - Y(\mu) Y(\mu')]\, d\mu', \quad (8.123)$$

$$Y(\mu) = e^{-\tau/\mu} + \tfrac{1}{2} a\mu \int_0^1 \frac{1}{\mu' - \mu} [X(\mu) Y(\mu') - X(\mu') Y(\mu)]\, d\mu'. \quad (8.124)$$

Thus the reflection and transmission functions can be expressed in closed form using the X and Y functions, which are governed by a pair of coupled nonlinear integral equations. The existence of a scattering or

reflecting surface at the lower boundary can be handled simply. An extra intensity can be defined that adds to the existing solutions, and that, in most cases, can be expressed in terms of the X and Y functions already computed for the standard problem.

For a semiinfinite atmosphere, $Y=0$, and the X function is equivalent to the Chandrasekhar's H function, first introduced in approximate form in (8.35). The reflection function is now given by

$$R(\mu, \mu_0) = \frac{a}{4(\mu + \mu_0)} H(\mu)H(\mu_0), \tag{8.125}$$

and the integral equation for H is

$$H(\mu) = 1 + \tfrac{1}{2}a\mu H(\mu) \int_0^1 \frac{H(\mu')}{\mu + \mu'} d\mu'. \tag{8.126}$$

The method of X, Y, and H functions can, in principle, be applied to scattering problems with arbitrary phase functions. In practice, the amount of algebraic manipulations is too great, and not much progress has been made beyond the Rayleigh phase function (§ 7.3).

8.4. Miscellaneous methods

In this section we shall briefly examine methods that are not as commonly used as those discussed in the previous sections, but are nevertheless important for their theoretical ideas. These methods are generally less efficient for solving problems of radiative transfer in a plane-parallel atmosphere than the standard methods, but they have the advantage of being adaptable to arbitrary geometry.

8.4.1. *Successive orders of scattering*

The specific intensity can be considered to be composed of a series,

$$I^{\pm}(\tau; \mu, \phi) = \sum_{n=1}^{\infty} I_n^{\pm}(\tau; \mu, \phi), \tag{8.127}$$

where $I_n^{\pm}$ is the partial intensity that has been scattered n times. For a plane-parallel atmosphere of optical depth τ_1 (see Fig. 2.7), we can use (2.86), (2.98), and (2.99) to derive expressions for each order I_n,

$$I_n^+(\tau; \mu, \phi) = \frac{1}{\mu} \int_{\tau}^{\tau_1} J_n^+(t; \mu, \phi) e^{-(t-\tau)/\mu} dt, \tag{8.128}$$

$$I_n^-(\tau; \mu, \phi) = \frac{1}{\mu} \int_0^{\tau} J_n^-(t; \mu, \phi) e^{-(\tau-t)/\mu} dt. \tag{8.129}$$

The corresponding term of the source function is given by

$$J_n^{\pm}(\tau;\mu,\phi)=\frac{a(\tau)}{4\pi}\int_0^{2\pi}\left[\int_0^1 P(\tau;\pm\mu,\phi;\mu',\phi')I_{n-1}^{+}(\tau;\mu',\phi')\,d\mu'\right.$$
$$\left.+\int_0^1 P(\tau;\pm\mu,\phi;-\mu',\phi')I_{n-1}^{-}(\tau;\mu',\phi')\,d\mu'\right]d\phi'$$
$$+\delta_{1n}\Sigma^{\pm}(\tau;\mu,\phi), \tag{8.130}$$

where $\Sigma^{\pm}(\tau;\mu,\phi)$ is the primary source term and $I_0^{\pm}(\tau;\mu,\phi)=0$.

The equations (8.128)–(8.130) can be used recursively to obtain higher order partial intensities and source functions. In principle this method is intuitive and computationally simple, and can be applied to homogeneous as well as nonhomogeneous atmospheres. In practice, the series (8.127) converges very slowly for $\tau_1 \geqslant 1$ unless $a(\tau) \ll 1$. So the most useful computation that can be performed by this method is for scattering by an optically thin layer. In this case the first-order solution is sufficient.

Consider illumination by sunlight. The primary source is due to the first scattering of incident radiation as defined by (2.116) and (8.84),

$$J_1^{\pm}(\tau;\mu,\phi)=\frac{a}{4\pi}e^{-\tau/\mu_0}P(\tau;\pm\mu,\phi;-\mu_0,\phi_0)f(0). \tag{8.131}$$

Inserting (8.131) into (8.128) and (8.129), we obtain the reflected and transmitted intensities,

$$I_r^{+}(0;\mu,\phi)=\frac{a}{4\pi(\mu+\mu_0)}\left\{1-\exp\left[-\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)\tau_1\right]\right\}$$
$$\times P(\mu,\phi;-\mu_0,\phi_0)f(0), \tag{8.132}$$

$$I_t^{-}(\tau_1;\mu,\phi)=\frac{a}{4\pi(\mu-\mu_0)}\left[\exp\left(-\frac{\tau_1}{\mu}\right)-\exp\left(-\frac{\tau_1}{\mu_0}\right)\right]$$
$$\times P(-\mu,\phi;-\mu_0,\phi_0)f(0), \tag{8.133}$$

where the phase function is averaged over the entire layer. From these expressions the reflection and transmission functions (8.84)–(8.88) can be obtained.

8.4.2. *The integral equation*

The integral equation for the specific intensity inside a medium of arbitrary shape and volume V may be derived using arguments similar to

those presented in § 2.3.1,

$$I(P, \mathbf{s}) = \int_V dV(P') \int d\omega_{s'} \delta(\mathbf{s} - \mathbf{s}_{PP'}) \frac{e^{-\bar{\tau}(P,P')}}{4\pi s^2} e_v(P') \times P(P'; \mathbf{s}, \mathbf{s}') I(P', \mathbf{s}') + I(P'', \mathbf{s}) e^{-\bar{\tau}(P,P'')}, \tag{8.134}$$

where $\mathbf{s}_{PP'}$ is a unit vector directed from P' to P, and P'' is the boundary. This equation may be regarded as a generalization of (2.86) and (2.90). The first term on the right-hand side of (8.136) arises from photons that are scattered from all other points P' in the volume into P. The second term comes from radiation incident on V from outside the system. A complete solution of (8.136) provides detailed information on the internal and external radiation fields. The boundary conditions are incorporated and the integral equation is applicable to any type of geometry. However, solving the general equation numerically is a formidable task. Consequently, the full potential of this method has not been realized. To date, only solutions involving special geometries, such as plane-parallel and spherical, have been obtained.

8.4.3. *Monte Carlo*

The absorption and scattering of photons in an atmosphere can be considered as stochastic processes. The phase function may be regarded as a probability density function for the redistribution of photons in different directions. Taking this point of view, we can simulate the solution to the equation of radiative transfer by following the history of a large number of photons on a computer. In the simplest version of the Monte Carlo method, the photons are first released from the source. The subsequent events that may happen to each photon are determined by suitable probability distributions. A set of random numbers is then used to make a particular choice of the outcome of each event. The number of photons reaching the detector is recorded. Because of its conceptual simplicity, this method is extremely flexible and can be applied to complicated problems that are virtually insolvable by any other method.

However, the Monte Carlo method has a serious drawback. The computed results are subject to statistical fluctuations, which decrease in magnitude only as the square root of the number of photons used in the calculation. Hence enormous amounts of computer time may be required to obtain results of modest accuracy. There are at least two ways to improve the performance of this method. One is using backward Monte Carlo, that is, to start with the photon at the detector and carry out a time-reversed path tracing back to the source. The other way to optimize the statistics is not to terminate a simulated photon history when it is absorbed or escapes from the atmosphere. Instead, a weighting factor is assigned to the photon and the "renormalized" photon continues its path.

Eventually a given photon history is terminated if its weighting falls below a predetermined lower limit, e.g., 10^{-5}. The use of the Monte Carlo method is currently at a developing stage. With the expected improvement in the capabilities of computers, we anticipate that this versatile method will realize its full potential.

8.4.4. *Distribution of path lengths*

An alternative approach based on the distribution of photon path lengths is widely used in the study of spectral lines in the presence of a homogeneous conservative scattering haze (scattering albedo, $a = 1$). It is convenient to introduce a probability distribution $p(\lambda; \tau; \mu, \phi)$, where $p(\lambda)\, d\lambda$ is the probability that a photon contributing to $I_{a=1}(\tau; \mu, \phi) = I_1$ has traveled an optical path between λ and $\lambda + d\lambda$ since entering the atmosphere. If $p(\lambda)$ and I_1 are known, it can be shown that the monochromatic intensity for arbitrary absorption coefficient k_ν will be given by

$$I_\nu(\tau_\nu; \mu, \phi) = I_1 \int_0^\infty p(\lambda; \tau; \mu, \phi) e^{-k_\nu \lambda}\, d\lambda \tag{8.135}$$

It is clear that (8.135) gives a simple relation between the solution for a conservative atmosphere and the solution for an atmosphere that is physically identical except for the addition of molecular absorption. The first moment,

$$\langle \lambda \rangle = \int_0^\infty p(\lambda) \lambda\, d\lambda, \tag{8.136}$$

is the mean optical path traveled by radiation in a conservative atmosphere. In practice, we must first obtain I_1 using any of the standard methods discussed in this chapter. The function $p(\lambda)$ is usually determined by Monte Carlo techniques, although in simple cases analytic solutions are available.

8.4.5. *Low-order approximations for anisotropic scattering*

We now extend the low-order approximations introduced in § 2.4.5 to the case of anisotropic phase functions. In § 2.4.5 we showed that these methods are equivalent to the use of a single discrete ordinate, a subject that was treated generally in § 8.2.2. Nevertheless, because of the great popularity of low-order approximations we shall examine them in greater detail.

The approximate equations for a stratified atmosphere have their simplest forms for isotropic source functions (§ 2.4.4, § 2.4.5). Despite

the obvious difficulties of handling a strongly anisotropic source function in a one-term approximation, there have been many attempts to do so, because the result is much easier to understand than a complete numerical solution to the equations.

When we introduce anisotropic phase functions we bring in new parameters, with the consequence that different low-order scattering approximations proliferate. As for isotropic phase functions, however, the successful treatments are essentially the same, and ours is typical.

We shall restrict attention to scattering alone, although mixed conditions of thermal emission and scattering are a straightforward extension. We start from (2.116), consider only the first Stokes' parameter (the intensity), omit the variable ν, and write a in place of s/e,

$$\xi \frac{dI(\tau; \xi, \phi)}{d\tau} = I(\tau; \xi, \phi) - a \int_{\omega'} P(\tau; \xi, \phi; \xi', \phi') I(\tau; \xi', \phi')\, d\omega'/4\pi - aP(\tau; \xi, \phi; \xi_\odot, \phi_\odot) f(\tau)/4\pi, \tag{8.137}$$

where $f(\tau)$ is given by (2.115).

In the spirit of the method of moments, as used in § 2.4.4, we shall operate on (8.137) with the integral operators $\int d\omega$ and $\int \xi\, d\omega$ and make use of the normalization (2.33). We must also make use of the fact that, for spherical particles or for a random array of nonspherical particles, the phase function is a function only of the scattering angle, θ, the angle between an incident direction designated by (ξ', ϕ') or $(\xi_\odot, \phi_\odot)$ and a scattered direction (ξ, ϕ) (see Chapter 7, Fig. 7.1 for further discussion).

In (7.69) we defined the asymmetry factor g in terms of gains (7.63). In terms of the phase function, the definition becomes

$$g = \int P(\cos\theta) \cos\theta\, d\omega/4\pi \tag{8.138}$$

$$= \begin{cases} +1 & \text{for forward scattering,} \\ 0 & \text{for isotropic scattering,} \\ -1 & \text{for backward scattering.} \end{cases}$$

If we multiply (8.137) by $d\omega$, integrate, and use the normalization (2.33), we find

$$\frac{dF}{d\tau} = (1 - a)4\pi\bar{I} - af. \tag{8.139}$$

Now multiply (8.137) by $\xi\, d\omega$ and integrate to give

$$\int \xi^2 \frac{dI}{d\tau} d\omega = F - a \int \int P(\xi, \phi; \xi', \phi')\xi I(\xi', \phi')\, d\omega\, d\omega'/4\pi$$

$$- a \int P(\xi, \phi; \xi_\odot, \phi_\odot)\xi f\, d\omega/4\pi. \tag{8.140}$$

As in § 2.4.4, we now approximate

$$\int \xi^2 \frac{dI}{d\tau} d\omega \simeq \frac{4\pi}{3} \frac{d\bar{I}}{d\tau}.$$

It is reasonable to assume that the intensity is diffuse because, in (2.149), I excludes the direct solar beam. If we wish to consider the total flux of solar radiation, scattered plus direct, we must introduce

$$\mathbf{F} = F + \xi_\odot f, \tag{8.141}$$

where the second term on the right-hand side is the vertical flux of direct solar radiation; it is directed downward ($\xi_\odot < 0$) and is therefore negative.

Equation (8.140) can be simplified with the relationship

$$\int \xi P(\xi, \phi; \xi', \phi')\, d\omega/4\pi = g\xi'. \tag{8.142}$$

This far-from-obvious result will not be derived here: it follows from the symmetry of $P(\cos\theta)$ and the geometric identity (8.2)

$$\cos\theta \equiv \xi\xi' + (1 - \xi^2)^{1/2}(1 - \xi'^2)^{1/2} \cos(\phi - \phi').$$

Substituting (8.142) in (8.140) gives

$$\frac{4\pi}{3} \frac{d\bar{I}}{d\tau} = F(1 - ag) - ag\xi_\odot f, \tag{8.143}$$

and, from (8.139),

$$\frac{d^2F}{d\tau^2} = 3(1 - a)(1 - ag)F - af[1/\xi_\odot + \xi_\odot 3g(1 - a)]. \tag{8.144}$$

For the sum of the direct plus diffuse flux the relationship is

$$\frac{d^2\mathbf{F}}{d\tau^2} = 3(1-a)(1-ag)\mathbf{F} + f(1-a)(1/\xi_\odot - 3\xi_\odot). \tag{8.145}$$

As an example of a boundary condition we may take $I^- = 0$ at $\tau = 0$. From (2.142), (2.143), and (8.139) there results

$$\frac{dF}{d\tau} = 2(1-a)F - af, \tag{8.146}$$

and

$$\frac{d\mathbf{F}}{d\tau} = 2(1-a)\mathbf{F} + f(1-a)(1-2\xi_\odot). \tag{8.147}$$

We now consider two related topics concerned with the nature of the phase function and the asymmetry factor:

1. The *similarity relations* of van de Hulst that seek to approximate solutions for complex phase functions in terms of solutions for simpler phase functions (specifically isotropic phase functions).
2. The *δ-approximations* that draw attention to the importance of sharp forward peaks in the phase function.

The question posed under (1) is as follows. If $\mathbf{F}(e, a, g)$ is a solution to a problem involving the optical parameters e, a, and g, how may we choose e', a', and g', where g' represents a more tractable phase function than g, such that

$$\mathbf{F}(e', a', g') \simeq \mathbf{F}(e, a, g)? \tag{8.148}$$

As far as the approximate equations (8.146) and (8.163) are concerned, the question is moot because they are no more difficult to solve for one value of g than for another. We may, however, use these equations to derive the similarity relations, to be applied, in practice, to a comparison between numerical solutions.

We argue that the nonhomogeneous term in (8.146) is less important than the homogeneous terms. The nonhomogeneous term changes sign as $\xi_\odot$ varies between 0 and -1 and is zero for $\xi_\odot = -1/\sqrt{3}$. For climatic averages over all solar zenith angles, therefore, it may be small. Moreover, in the interior of a cloud the direct solar flux $\xi_\odot f$ becomes rapidly smaller than the diffuse flux F as the optical depth increases. This follows because the solution to the homogeneous part of (8.144) involves the term $\exp(-\beta\tau)$ where β is the inverse scattering path for diffuse

radiation,

$$\beta = \sqrt{3(1-a)(1-ag)}. \tag{8.149}$$

The direct solar beam, on the other hand, attenuates as $\exp(\tau/\xi_\odot)$, so that when a is close to unity, as it is for all interesting cases of atmospheric scattering, the attenuation of the direct beam is more rapid (possibly very much more rapid) than that of the diffuse radiation.

Similar solutions for the homogeneous terms are ensured if $\beta\tau$ remains constant. Since τ is proportional to the extinction coefficient, e, this condition is satisfied if

$$e\beta = e'\beta'. \tag{8.150}$$

The boundary condition (8.148) scales exactly with $(1-a)\tau$, so that similarity also requires

$$e(1-a) = e'(1-a'). \tag{8.151}$$

Equations (8.150) and (8.151) can be written in the following form, that is commonly used to express the similarity relations:

$$\frac{e'}{e} = \frac{1-a}{1-a'} = \frac{1-ag}{1-a'g'}. \tag{8.152}$$

Turning now to topic (2), there is a sense in which it is an artifact of numerical methods. The scattering phase function for cloud droplets has a very sharp forward, "diffraction" peak containing as much as half of the scattered energy (see the discussion of Babinet's principle in § 7.4). Since exactly forward scattering is equivalent to no scattering at all, one might be tempted simply to ignore this diffraction peak entirely and this is, in effect, what the δ-approximations do. Before the advent of large computers experimenters did so without seriously considering the matter because their laboratory equipment generally had insufficient angular resolution to resolve the diffraction peak.

With the availability of large computers it has become standard practice to use theoretical phase functions calculated from electromagnetic theory (§ 7.6). Unlike the laboratory experimenter, the computer misses very little and the forward scattering peak is explicitly included in any calculated phase function. There was a short period, at the time when numerical methods first came into general use, when this discrepancy of a factor 2 between measured and theoretical coefficients was a matter for concern. Since some aspects of this debate continue, we shall look at the question more closely.

Let us compare two scattering media characterized by the parameters (a, e, g) and (a', e', g'), where the former has a fraction b of its scattered energy concentrated in a narrow forward peak, while the latter has no such singular behavior.

Since scattering into the forward peak is equivalent to no scattering at all, the two sets of scattering parameters must be related by

$$s' = (1 - b)s.$$

But the absorption process is independent of the scattering so that

$$k' = k.$$

These two relationships require that

$$e' = e(1 - ab), \tag{8.153}$$

and

$$a' = \frac{a(1 - b)}{1 - ab}. \tag{8.154}$$

We now require that the asymmetry factor g be unchanged and, since a forward peak makes a contribution equal to b, it follows that

$$g = (1 - b)g' + b. \tag{8.155}$$

Equations (8.153), (8.154), and (8.155) together comprise the δ-approximation. As we have already pointed out, they deal more with an illusion than with a physical phenomenon. But, if we wish to modify numerical phase functions that have energy in a forward peak, we require a number for b in order to do so. This is best assessed from the data, by inspection. But it is also possible to take a different view of the situation.

Suppose that a particular phase function under consideration scatters strongly forward but not so sharply as implied by the foregoing discussion of diffraction peaks. It may prove advantageous to replace the phase function by a more symmetric function plus a peak in the sense that we have used; the peak can then be treated exactly in the manner described, while a smoother phase function may be chosen so as to be compatible with a particular set of approximate equations.

If the approximate equations are defined by the Eddington approximation (2.147), and if the original forward phase function is the Henyey–Greenstein function (§ 7.75), it has been demonstrated that g^2 is a good choice for b. This defines the *δ-Eddington approximation.* Its significance is unclear because, if we eliminate b from (8.153), (8.154),

and (8.155), we obtain, after some awkward manipulation, van de Hulst's similarity relations, (2.147). van de Hulst's relations therefore include the information contained in the δ-approximations regardless of the presence or magnitude of the forward scattering peak. Since the two approaches are quite distinct, their agreement suggests that the range of validity is wider than either derivation would indicate.

Numerical models of stratified atmospheres usually employ two coupled, first-order differential equations in preference to second-order equations, such as (8.145). In order to form these coupled equations, integrations are performed independently over both hemispheres in the spirit of the integral equation (2.106). The resulting equations are known as *two-stream equations.*

The two-stream equations involve approximations in addition to those employed in the derivation of (8.145), but the extent of these additional approximations is no more than those employed to derive the boundary conditions using (2.142) and (2.143), and there is probably little to choose between these equations and those of § 2.4.5 and § 2.4.6. In order to demonstrate this relationship we derive the two-stream equations from (8.139) and (8.144), using the further approximations (2.142) and (2.143). If we write F^+ for πI^+ and F^- for πI^- and then add and subtract (8.139) and (8.144) we obtain

$$\begin{aligned} dF^+/d\tau &= F^+\gamma_1 - F^-\gamma_2 - fa\gamma_3, \\ dF^-/d\tau &= F^+\gamma_2 - F^-\gamma_1 + fa(1-\gamma_3), \end{aligned} \tag{8.156}$$

where

$$\begin{aligned} \gamma_1 &= \tfrac{1}{4}[7 - a(4+3g)], \\ \gamma_2 &= -\tfrac{1}{4}[1 - a(4-3g)], \\ \gamma_3 &= \tfrac{1}{4}[2 + 3g\xi_\odot]. \end{aligned} \tag{8.157}$$

The general form of (8.156) is valid for all two-stream approximations but with different values for the coefficients depending upon the approximation applied to the radiation field and the approximation applied to the phase function. The coefficients (8.157) give the "standard Eddington approximation," as it has been called. Applied to an optically thin atmosphere with a black lower boundary, the approximation leads to negative albedos for $g > 2/(3\,|\xi_\odot|)$. While this is an unsatisfactory result, it does not mean that the coefficients (8.157) give results that are worse than others, on the average. All of the stratified atmosphere approximations that we have discussed have errors of the same order of magnitude, although some manage to avoid such egregious problems as negative albedos.

8.5. Numerical results

It is not the purpose of this chapter to conduct a thorough examination of the numerical results obtained by various methods. However, we shall present representative results that demonstrate the advantages of particular methods, and provide insight into the radiative transfer problem. Also, having developed more powerful and more accurate methods in this chapter, we can now check the accuracy of the two-stream and other low-order multistream approximations, first discussed in § 2.4.

8.5.1. *The diffusion exponent*

In the method of discrete ordinates, § 8.2.2, we consider solutions of the form $\mathbf{I}^{\pm} = \mathbf{g}^{\pm} e^{k\tau}$. Different values of k represent different modes of propagation of photons in an unbounded homogeneous medium. It can be shown that the characteristic equation derived from (8.23) for isotropic scattering is

$$f(k) = 1 - a \sum_{j=1}^{n} \frac{a_j}{1 - \mu_j^2 k^2} = 0. \tag{8.158}$$

The eigenvalues correspond to the roots of (8.158). Each value is known as a *diffusion exponent* and may be physically interpreted as a diffusivity factor that, when multiplied by the normal optical depth, represents an equivalent optical path in each discrete stream. Figure 8.4 illustrates a distribution of eigenvalues for isotropic scattering with a single scattering albedo $a = 0.9$, using four ordinates ($2n = 4$).

The mode that corresponds to the smallest diffusion exponent is known as the *fundamental mode,* and is the one which is least damped. For this mode, the diffusion length ($\propto k^{-1}$) is the largest. The smallest diffusion exponent may be computed for the $n = \infty$ limit of (8.158),

$$f(k) = 1 - \frac{a}{2k} \ln \frac{1+k}{1-k} = 0. \tag{8.159}$$

The result, k versus a, is summarized by the curve $g = 0$ in Fig. 8.5. Note that as a approaches unity, the value of k becomes smaller. Physically this implies that the diffusion length k^{-1} increases and the photons can penetrate deeper into the medium via multiple scattering. Equation (8.159) can be generalized to include the case of multiple scattering with anisotropic phase functions. The results for three choices of the asymmetry factor g for a Henyey–Greenstein phase function are also summarized in Fig. 8.5. It can be seen that the net effect of forward scattering, for a given single scattering albedo, is a reduction of the value of k. This is the basis for the similarity relation discussed in § 8.4.5.

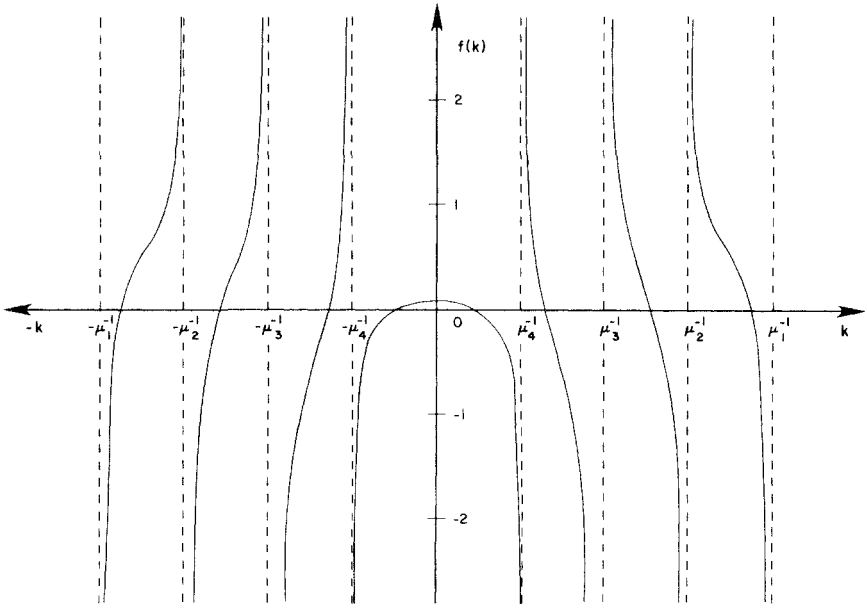


FIG. 8.4. Eigenvalues of (8.158). Four ordinates are used. The single scattering albedo is 0.95. The eigenvalues are given by the intercepts on the k axis. After Liou (1980).

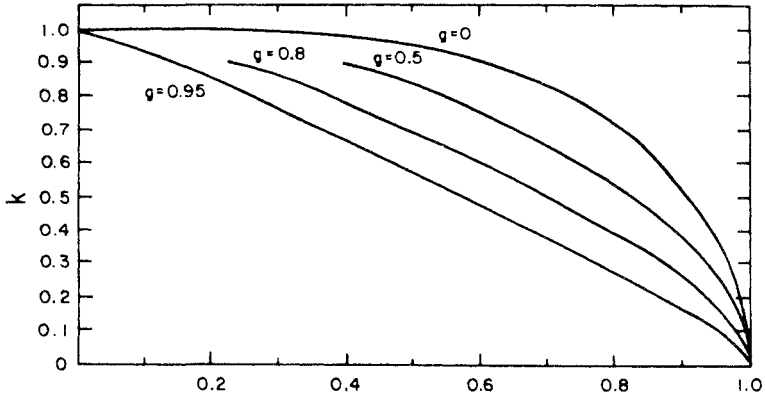


FIG. 8.5. Smallest diffusion exponent, k, for the limit $n = \infty$. The curve marked $g = 0$ is for isotropic scattering; the other curves are for a Henyey–Greenstein phase function with three choices of the asymmetry parameter, $g = 0.5$, 0.8, and 0.95. After Irvine (1975).

8.5.2. *X, Y, and H functions*

We consider the special case of a homogeneous, isotropically scattering slab of optical thickness $\tau_1 = 4$. Figure 8.6 presents values of $X(\mu)$ and $Y(\mu)$ for single scattering albedos $a = 0.6$, 0.9, and 1.0. The dashed curve gives the value for $H(\mu)$ for $a = 1.0$. From § 8.3.5, we know that $H(\mu)$ is the limit of $X(\mu)$ as $\tau_1 \to \infty$. There is a simple expression,

$$H(\mu) \approx \frac{1 + 2\mu}{1 + 2\mu\sqrt{1 - a}}, \tag{8.160}$$

which reproduces the true values of $H(\mu)$ to better than 3% everywhere. As $\tau_1 \to \infty$, $Y(\mu) \to 0$. For $\tau_1 = 4$ and $a = 0.6$, the values of $Y(\mu)$ are of the order of 10^{-3}, and are not plotted in Fig. 8.6.

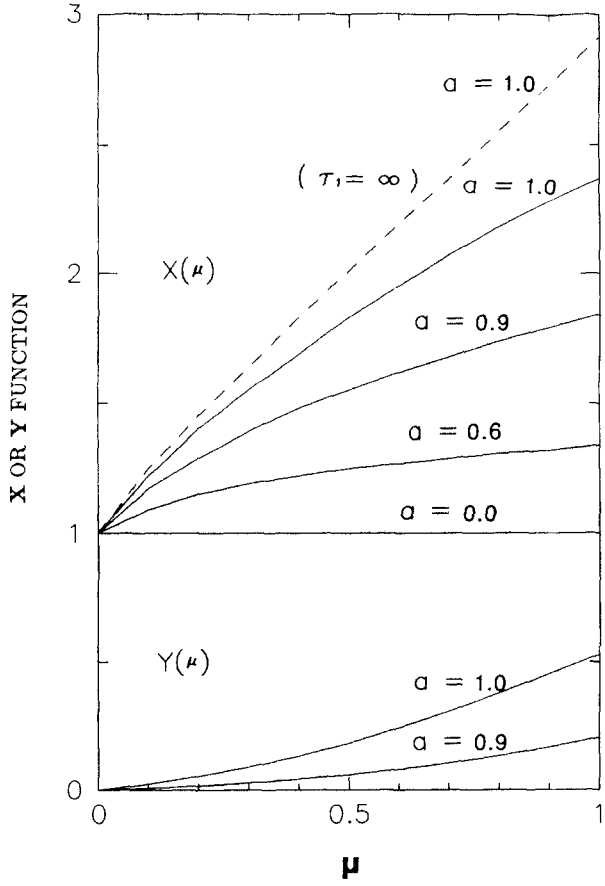


FIG. 8.6. X and Y functions for optical thickness $\tau_1 = 4$. Values of $Y(\mu)$ for $a = 0.6$ and $a = 0.0$ are close to 0 and are not shown. The dashed line, the asymptotic limit of $X(\mu)$ as $\tau_1 \to \infty$, is the H function. Numerical values are taken from Caldwell (1971).

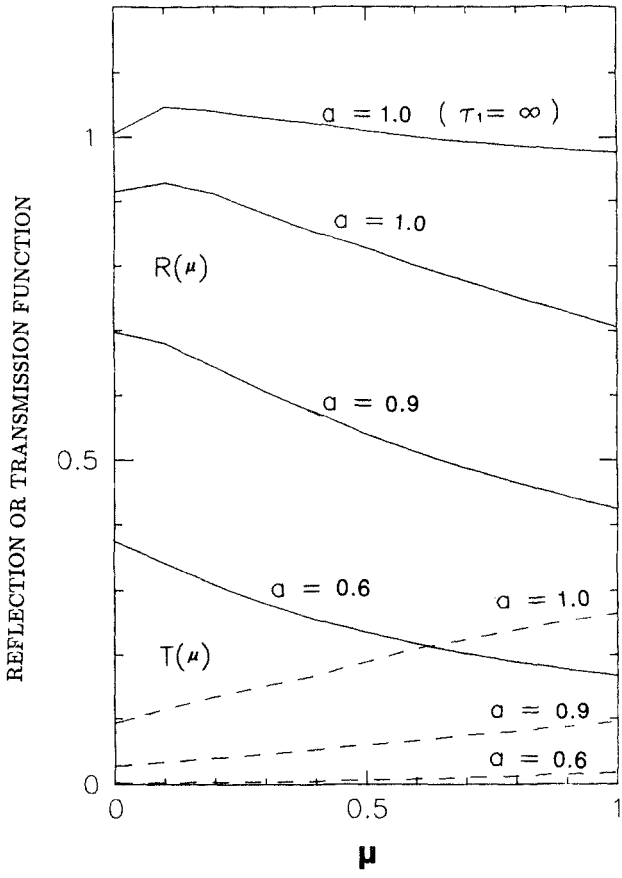


FIG. 8.7. Reflection and transmission functions. $R(\mu)$ (solid lines) and $T(\mu)$ (broken lines) are computed using values of X and Y functions given in Fig. 8.6. Unless otherwise indicated, the optical depth is $\tau_1 = 4$.

Using (8.121), (8.122), and (8.125) we can compute the reflection and transmission functions. The results are shown in Fig. 8.7.

8.5.3. *Internal radiation field*

The X, Y, and H functions provide detailed information on reflected and transmitted radiation. They do not provide information on the internal radiation field. Figure 8.8a shows the internal radiation field, $I(\tau, \xi)$, computed using the Feautrier method, for an isotropically scattering slab with optical depth $\tau_1 = 4$. For comparison, we also present in Fig. 8.8b results for the azimuthally independent component of specific intensity [$m = 0$ in (8.10)] for scattering by a Henyey–Greenstein phase function with $g = 0.5$. The isotropic and anisotropic cases are approximately

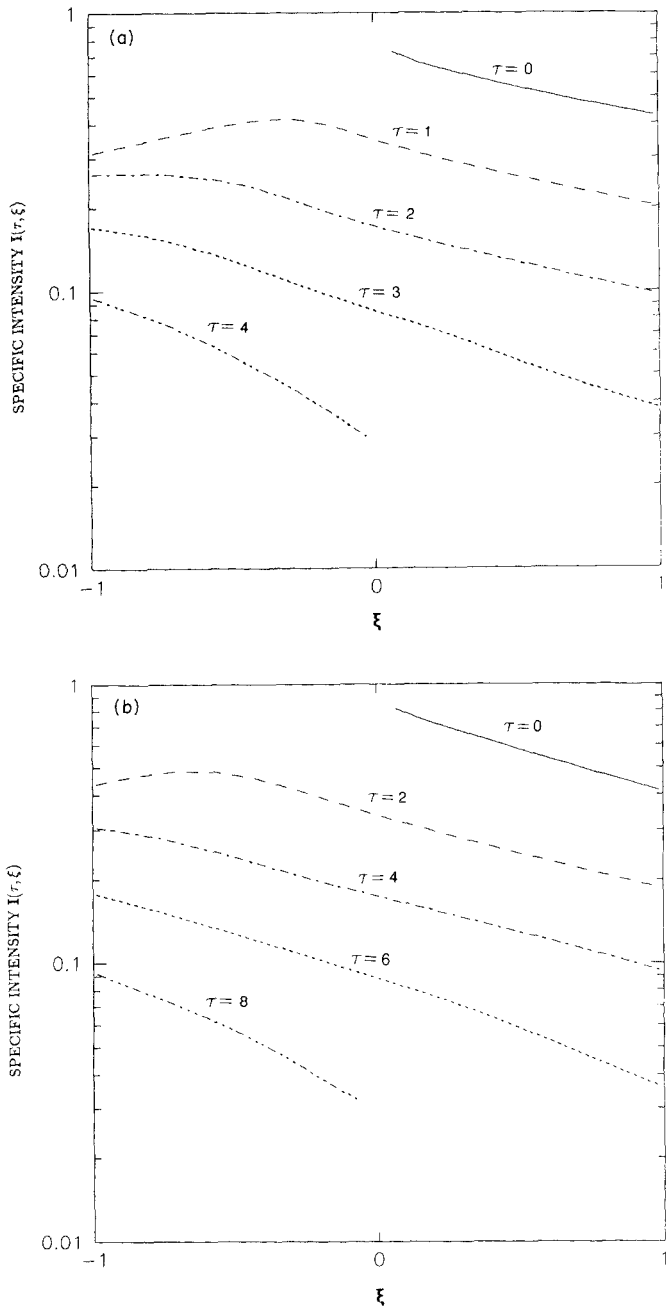


FIG. 8.8. Specific intensity as a function of angle at various depths inside a finite slab. (a) Optical depth $\tau_1 = 4$, isotropic scattering, $a = 0.9$, and $|\xi_\odot| = 1$. The incident solar flux is chosen to be $f(0) = 1$. (b) $\tau_1 = 8$, $a = 0.95$, and asymmetry factor $g = 0.5$. The values of τ, a, and g in (a) and (b) satisfy the similarity relation (8.151) exactly, and (8.152) almost exactly. Specific intensities for similar lines in the two diagrams are very close. The similarity argument developed in § 8.4.5 can readily be extended to the $m = 0$ term in (8.10).

related via the similarity transformation (§ 8.4.5). These results demonstrate how well this approximation works.

8.5.4. *Scattering by haze*

Figure 8.9 shows the reflected and transmitted intensities for a medium with the optical properties of a natural haze computed using a discrete-ordinate method and a doubling method. The phase function used is a double Henyey–Greenstein phase function with an effective asymmetry factor, $g = 0.786$. Only the azimuthally independent intensities are plotted in the figure.

8.5.5. *Convergence of successive scatterings*

From § 8.4.1 we may compute the reflection function by

$$R(\tau_1; \mu, \mu_0) = \sum_{n=1}^{\infty} a^n R_n(\tau_1; \mu, \mu_0), \tag{8.161}$$

where $R_n(\tau_1; \mu, \mu_0)$ represents the contribution from nth order multiple

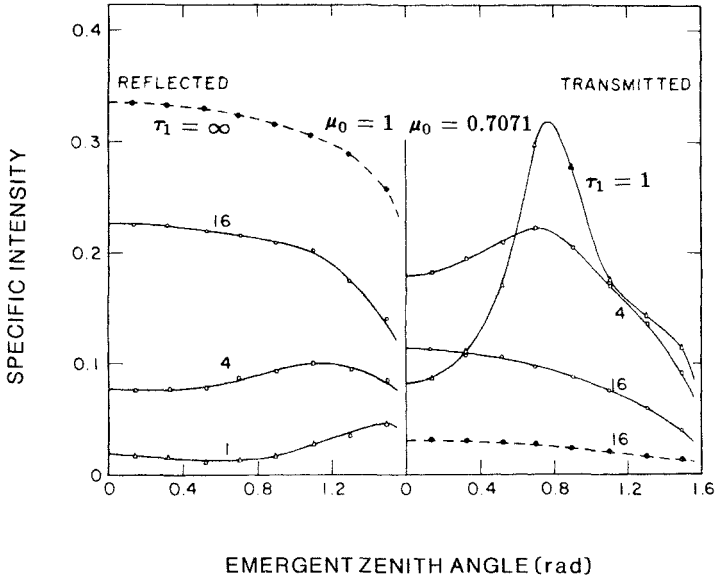


Fig. 8.9. Comparison of doubling and discrete ordinate methods. The points are calculated using 16 discrete ordinates. The solid and broken lines are for the doubling method. The solid lines are for a double Henyey–Greenstein phase function with an effective asymmetry factor $g = 0.786$. Broken lines are for an isotropic phase function. After Liou (1973).

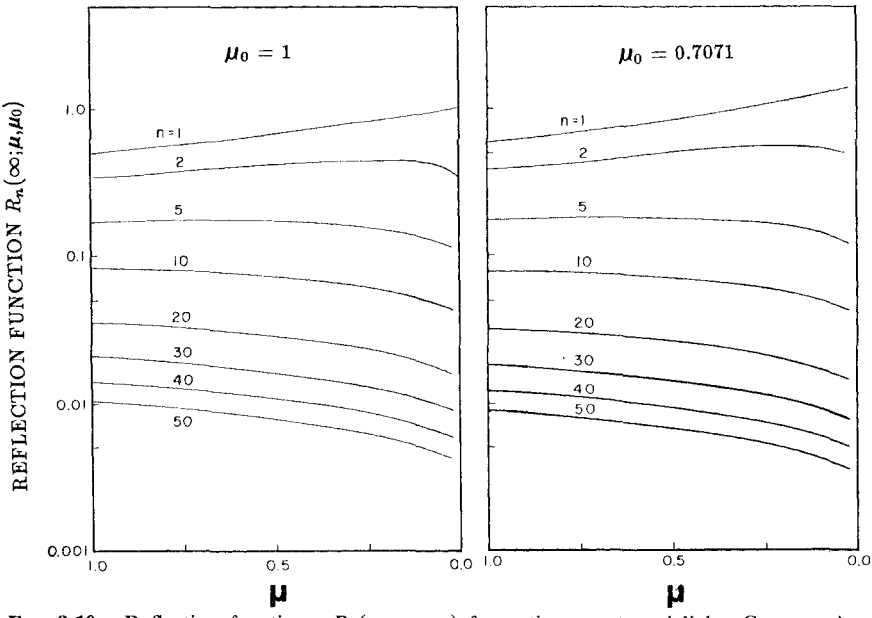


FIG. 8.10. Reflection functions, $R_n(\infty; \mu, \mu_0)$ for n times scattered light. Conservative, isotropic scattering is assumed in a semiinfinite atmosphere. After Uesugi and Irvine (1970).

scattering. Numerical experiments have been performed with conservative isotropic scattering in an infinite atmosphere. The results are summarized in Fig. 8.10 for $\mu_0 = 1$ and 0.7071. It is clear that convergence of (8.161) is extremely slow. The 50th order scattering can still contribute about 1% to the total reflection function.

8.5.6. *The accuracy of low-order approximations*

The two-stream equations (8.156) for a homogeneous atmosphere can be solved analytically. For a slab of optical thickness τ_1 illuminated by solar radiation, the reflectivity, $R = [F^+(0)]/[\mu_0 f(0)]$, and transmissivity, $T = [F^-(\tau_1)]/[\mu_0 f(0)]$, are (with $\mu_0 = -\xi_\odot$)

$$R = \frac{a}{(1-k^2\mu_0^2)[(k+\gamma_1)e^{k\tau_1} + (k-\gamma_1)e^{-k\tau_1}]}\Big[(1-k\mu_0)(\alpha_2+k\gamma_3)e^{k\tau_1} - (1+k\mu_0)(\alpha_2-k\gamma_3)e^{-k\tau_1} - 2k(\gamma_3-\alpha_2\mu_0)e^{-\tau_1/\mu_0}\Big], \tag{8.162}$$

$$T = e^{-\tau_1/\mu_0}\Big\{1 - \frac{a}{(1-k^2\mu_0^2)[(k+\gamma_1)e^{k\tau_1} + (k-\gamma_1)e^{-k\tau_1}]} \times [(1+k\mu_0)(\alpha_1+k\gamma_4)e^{k\tau_1} - (1-k\mu_0)(\alpha_1-k\gamma_4)e^{-k\tau_1} - 2k(\gamma_4+\alpha_1\mu_0)e^{\tau_1/\mu_0}]\Big\}, \tag{8.163}$$

where the parameters γ_1, γ_2, and γ_3 have been defined by (8.157), and the new parameters are

$$\gamma_4 = 1 - \gamma_3, \tag{8.164}$$

$$\alpha_1 = \gamma_1\gamma_4 + \gamma_2\gamma_3, \tag{8.165}$$

$$\alpha_2 = \gamma_1\gamma_3 + \gamma_2\gamma_4, \tag{8.166}$$

$$k = (\gamma_1^2 - \gamma_2^2)^{1/2}. \tag{8.167}$$

One use of the accurate methods described in § 8.2 and § 8.3 is to check the accuracy of simplified but useful solutions such as (8.162) and

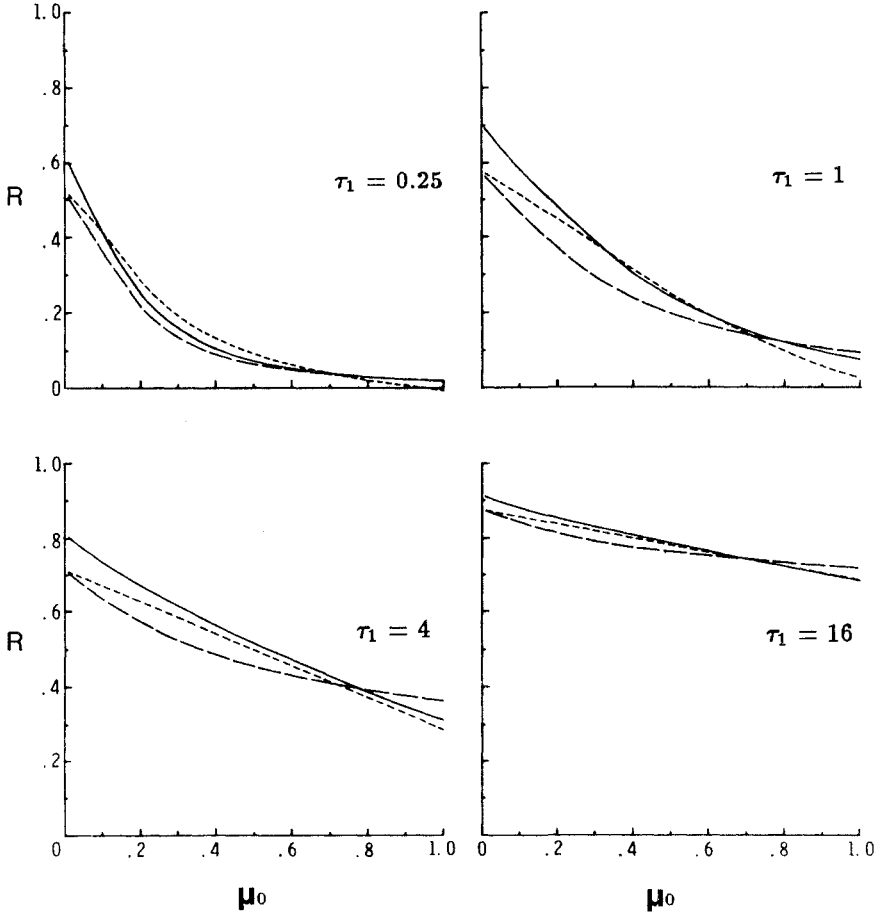


FIG. 8.11. Reflectivity R for two-stream approximations. The solid curves are exact calculations for a plane-parallel atmosphere. The standard Eddington (short dashes) and modified Eddington (long dashes) approximations are given for comparison. The scattering is conservative, and the phase function is Henyey–Greenstein with $g = 0.75$. After Meador and Weaver (1980).

(8.163). Figure 8.11 shows the comparison between exact calculations and approximate solutions given by the standard Eddington and a modified Eddington approximation [for which γ_1 and γ_2 are unchanged but $\gamma_3 = 1 - (1/2)\int_0^1 P(\xi_\odot, \xi')\, d\xi'$].

Simple formulas, analogous to (8.162) and (8.163), may also be derived for a four-stream approximation. The details will not be discussed here. The accuracy of this method may be appreciated by inspection of Fig. 8.12. It is clear that such low-order multiple-stream approximations are correct to about 10% in most representative cases. Now, inexact knowledge of the fundamental physical parameters such as extinction coefficients, phase functions, and atmospheric properties are often responsible for even greater uncertainties and in many practical computations, these low-order approximations are adequate.

8.6. Applications

The austere and formal character of scattering theory is relieved by the variety of interesting applications that can be made of it. We conclude this chapter with a brief account of a few areas for which good observations and satisfactory theory exist.

8.6.1. *Solar and thermal fluxes in stratocumulus clouds*

Figures 8.13 and 8.14 show aircraft measurements of the radiation field in stratocumulus clouds. The solar flux must be dominated by scattering because, in the absence of scattering, the upward flux component would be unmeasurable. The measured solar heating rate in Fig. 8.14 just exceeds the root-mean-square error but it also exceeds the theoretical prediction at most levels. The theory employs Mie calculations with measured drop size distributions and liquid water contents. A two-stream approximation was used with coefficients adjusted to fit certain doubling and adding calculations on model clouds. The discrepancy between theory and observation appears to be typical for this type of observation, and throws doubts on the common assumption that the optical properties of cloud drops are those of pure water.

The thermal flux observations show outgoing fluxes at both boundaries corresponding to black bodies at the cloud surface temperature confirming, in this instance, a classical assumption in the meteorological literature. The thermal heating data in Fig. 8.14 emphasize the low precision obtainable even for high-quality observations. The observed data do not show the strong cooling that is predicted near the cloud top.

The theoretical calculation of thermal flux employs an emissivity model (§ 6.4.2) with emissivities derived from "exact" calculations involving liquid water and water vapor. This implies that, to first order, only the droplet absorption matters and that scattering may be neglected. The reason for this can be seen from the approximate equation, (8.145).

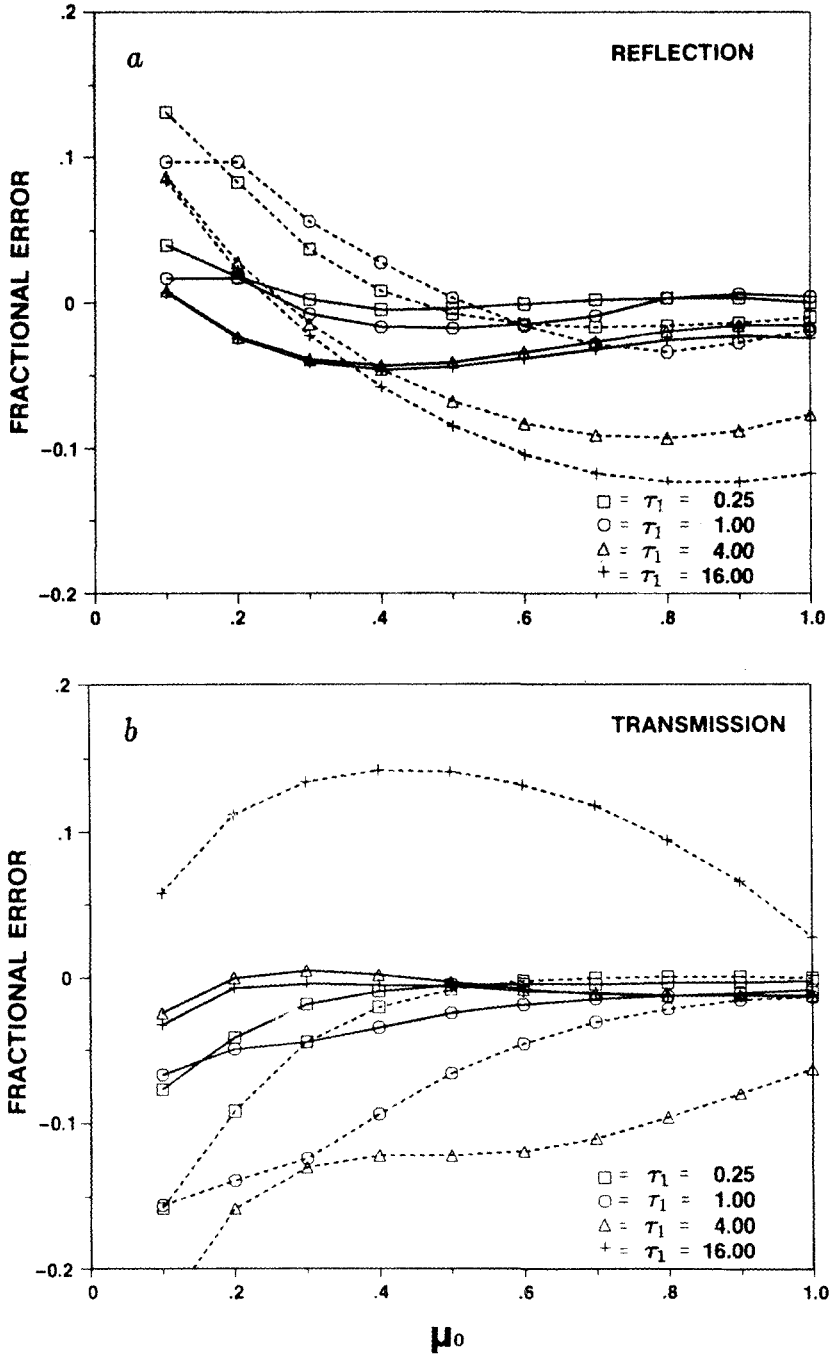


FIG. 8.12. Fractional error of the emergent intensity for a four-stream (two discrete ordinates) approximation. The calculations are for $a = 0.9$, Henyey–Greenstein phase function $g = 0.75$, and underlying surface reflectivity = 0.3. The solid curves are results from the delta four-stream method; dashed curves are results from the delta two-stream method. After Cuzzi et al. (1982).

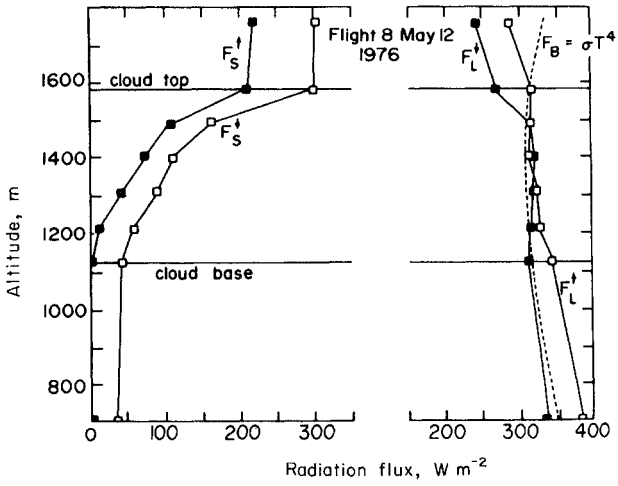


FIG. 8.13. Thermal and solar flux components measured in stratocumulus clouds. $F_S^{\uparrow}$ = upward solar component, $F_S^{\downarrow}$ = downward solar component, $F_L^{\uparrow}$ = upward thermal component, $F_L^{\downarrow}$ = downward thermal component, $F_B = \sigma\theta^4$, θ = local temperature. After Stephens et al. (1978).

If a thermal radiation term were used in place of scattering from the direct solar beam, the difference between including and not including scattering would lie in the value of the diffusion factor β. For absorption alone, an emissivity treatment is appropriate; then, $a = 0$ and $\beta = 3^{1/2}$, corresponding to the diffusivity factor, r, of § 6.1.2. For thermal

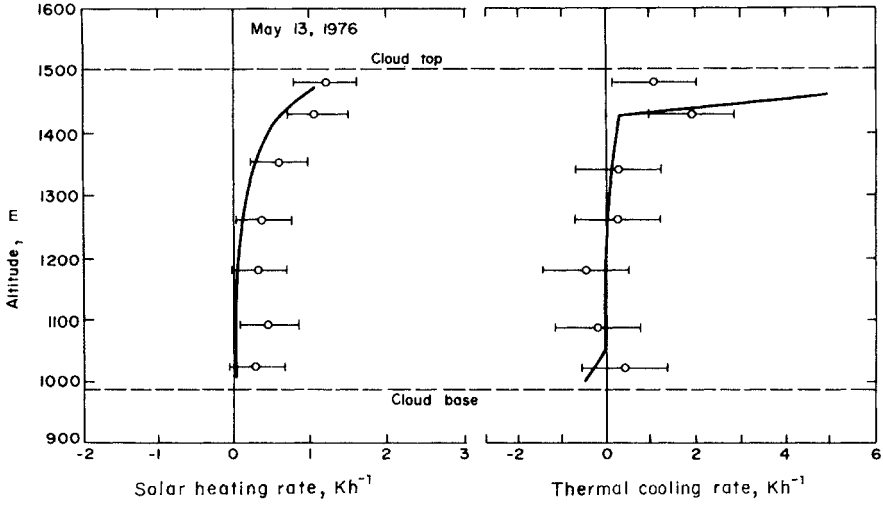


FIG. 8.14. Thermal and solar heating rates in stratocumulus clouds. The points were obtained by differencing data similar to those shown in Fig. 8.13. The horizontal lines show root mean square errors. The solid lines are based on theoretical calculations. After Stephens et al. (1978).

radiation, scattering from small particles will be close to isotropic ($g = 0$), and for absorbing droplets, $Q_s \approx Q_a \approx 1$; then, $a \sim 0.5$ and $\beta = 1.5^{1/2}$. In fact, the single-scattering albedo will be smaller than 0.5 because of the additional gaseous absorption. The difference between diffusivity factors with and without scattering is, in the light of other uncertainties, rather small and is, to first order, taken up in an empirical determination of emissivities.

This conclusion can be generalized. If, at any level, the optical depth for scattering is smaller than that for absorption, the possibility exists that scattering may be neglected. This statement applies to clear-sky absorption of solar radiation in the near infrared bands of water vapor and in the ultraviolet bands of oxygen and ozone. These gases are the major sources of solar heating in the lower and middle atmospheres; there are good physical reasons why successful atmospheric calculations could be made before the ready availability of modern scattering protocols.

8.6.2. *Polarization of light reflected from Venus*

A classical series of measurements by Lyot showed the polarization of reflected light from Venus to vary in a complicated manner with wavelength and phase angle (the sun–Venus–earth angle). The data are reproducible and, clearly, carry important information about the nature of the Venus clouds. One set of measurements, for a single wavelength, is shown in Fig. 8.15.

The data in Fig. 8.15 show vestiges of a rainbow (~15° phase angle) and a glory (~155° phase angle). The latter will be observed only if there is a very narrow distribution of drop sizes (a variance of 0.07 gives the best fit), and the polarization in the rainbow is very sensitive to the mean particle size (a radius of 1.05 μm gives the best fit).

The theoretical calculations shown in Fig. 8.15 are based on a doubling method with a matrix formulation for the Stokes parameters and Mie theory calculations for the assumed particle sizes and refractive indices. In addition to the particle radius and variance, the best fit between theory and observations gave a value for the pressure (more precisely, the volume scattering coefficient) at the cloud top; and the variation of particle refractive index with wavelength led to a tentative identification of sulfuric acid for the droplet composition.

8.6.3. *Scattered light in the stratosphere*

A calculation of photolysis rates of oxygen and ozone in the stratosphere requires a knowledge of the total density of actinic photons, both direct and scattered. Most early calculations included only the contribution from the direct beam, but scattering, particularly from the lower atmosphere, should also be included. Figure 8.16 shows the results of

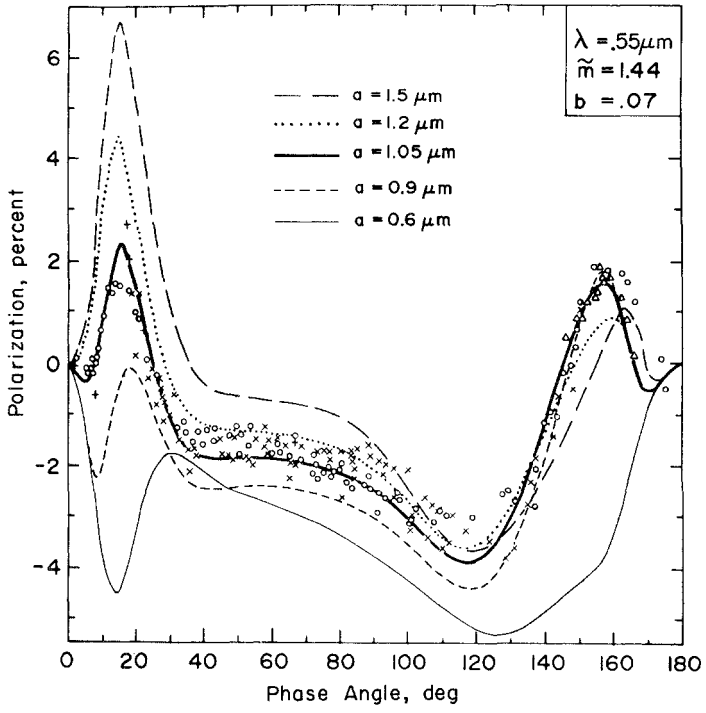


FIG. 8.15. Polarization of reflected light from Venus. The wavelength is 0.55 μm. The crosses and circles are observations by two different observers. Theoretical calculations (solid, dotted, and broken lines) are for spheres with refractive index, $\tilde{m} = 1.44$, and a size distribution with mean radius, a, and variance, b. After Hansen and Hovenier (1974).

balloon measurements of the two components for the spectral range 190–310 nm.

A theoretical calculation was made by a matrix multiplication method (§ 8.3.1), combining a number of homogeneous layers in which there was molecular scattering and absorption by oxygen and ozone. The ozone distribution was unknown a priori and could be adjusted for best fit to the observations. With this degree of freedom the agreement between theory and observation (as shown in Fig. 8.16) was exceptionally good.

In the light of the discussion of single-scattering albedos in § 8.6.1, it is instructive to calculate this quantity for the atmosphere at 40 km (Table 8.1). The calculated albedos are small, consistent with the small fraction of scattered light that is observed; the shape of the curve in Fig. 8.16 follows, more or less, the shape of the absorption coefficient in the Hartley bands of ozone and the Schumann–Runge bands of oxygen.

Although the fraction of scattered light is small, it is not negligible;

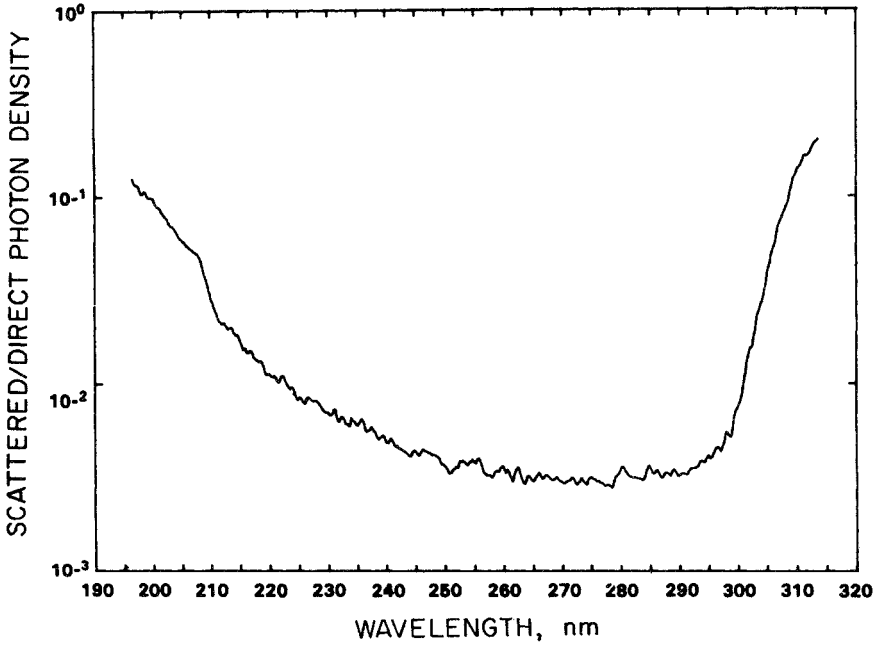


FIG. 8.16. Ratio of scattered to direct beam solar photons at 40 km. The solar zenith angle was 41.6°. After Herman and Mentall (1982).

near 300 nm it amounts to about 20% of the total and has a small but significant effect upon the calculated equilibrium ozone concentration.

8.6.4. *Scattered light in clear water*

Closely related to atmospheric scattering is the diffusion of light in the upper layers of deep waters; and we may take advantage of the relatively

Table 8.1. Single-scattering albedos at 40 km[a]

Wavelength (nm)	s_n (cm^2)	k_n(cm^2) Ozone	k_n(cm^2) Oxygen	Albedo
300	5.7×10^{-26}	3×10^{-19}	—	2.5×10^{-2}
250	1.3×10^{-25}	1.1×10^{-17}	—	1.6×10^{-3}
200	3.6×10^{-25}	2×10^{-19}	1.5×10^{-23}	8×10^{-2}
	$n(O_3)/n(M) = 7 \times 10^{-6}$			
	$n(O_2)/n(M) = 2 \times 10^{-1}$			

[a] For absorption coefficients of oxygen and ozone, see Figs. 5.4 and 5.11. For number densities, see Appendix 3, Fig. 1.7, and Table 1.1.

placid conditions in lakes and oceans to obtain better data than are available for mists and clouds. One set of observations, to a depth of 275 m in the Mediterranean, is shown in Fig. 8.17.

The cusp on the 50 m curve and the vertical bars in Fig. 8.17 indicate the intensity of the direct solar beam. At 100 m depth and below, the

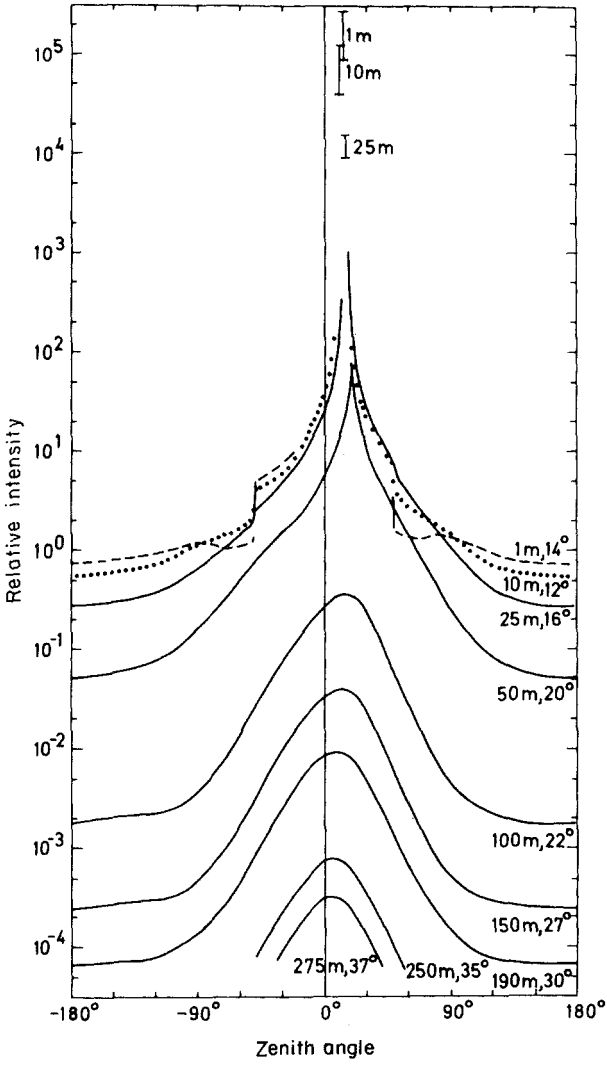


Fig. 8.17. Normalized intensities in the solar vertical in the Mediterranean. The data were observed in clear water. The zenith angle of the sun changed during the course of the observations; its value is given on the right-hand side of the figure. The angular resolution of the radiometer is 1.3°. Readings taken in the solar direction are indicated by vertical bars. After Jerlov (1976).

direct beam has vanished, leaving only the diffuse radiation field, which preserves its angular distribution at all greater depths. This asymptotic state is the fundamental mode, discussed in § 8.5.1. It corresponds to the term varying as $\exp(-\beta\tau)$ in the solution to (8.144) where, from (8.149), $\beta = [3(1-a)(1-ag)]^{1/2}$. The direct solar beam, on the other hand, attenuates more rapidly, as $\exp(-\tau/\mu_0)$. These attenuation coefficients can be determined approximately from Fig. 8.17: they are $1.4 \times 10^{-1}\,\mathrm{m}^{-1}$ and $4.6 \times 10^{-2}\,\mathrm{m}^{-1}$ for the direct and diffuse components, respectively. For a vertical sun and small particles ($g = 0$), we may calculate an approximate value for the single-scattering albedo, $a = 0.964$.

8.6.5. *CO_2 lines in the reflection spectrum of Venus*

Figure 8.18a shows an observed reflection spectrum of Venus for one of the combination bands of carbon dioxide, and Fig. 8.18b shows a synthetic spectrum that provides an acceptable fit to the data. The synthetic spectrum is for a semiinfinite, homogeneous, isotropically scattering atmosphere using a two-stream approximation and Eddington's second approximation to obtain angular dependence for the reflected radiation.

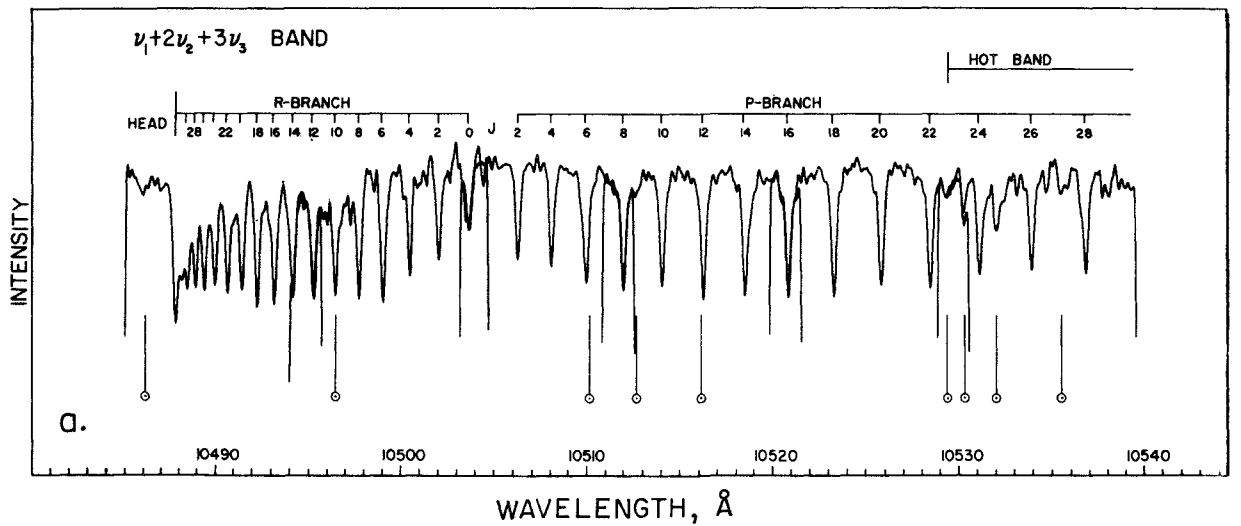
Lines (intensity I_ν) are superimposed on a continuum (intensity I_C). The variables are the single-scattering albedos for the cloud alone (a_C) and for the mixture of cloud and carbon dioxide (a_ν). The following expression may be obtained,

$$\frac{I_\nu}{I_C} = \frac{a_\nu[1+\mu\sqrt{3(1-a_C)}][1+\mu_0\sqrt{3(1-a_C)}]}{a_C[1+\mu\sqrt{3(1-a_\nu)}][1+\mu_0\sqrt{3(1-a_\nu)}]}, \tag{8.168}$$

and is used in Fig. 8.18b. μ and μ_0 are the direction cosines for the observed and solar beams, respectively. The properties of the cloud are defined by the quantity $M = n(CO_2)/s_C n_s$, where s_C is the scattering coefficient of the cloud and n_s is Loschmidt's number. From the best fit to the data, the cloud appears to be very tenuous, with a scattering mean free path of 4 km (under the assumption of isotropic scattering).

8.6.6. *The color and polarization of skylight*

The color and polarization of the daytime sky are paradigms of the scattering problem. The fundamental reason for both color and polarization was given by Rayleigh in 1871 with his theory of scattering by molecules, but a full account of the multiple scattering problem had to await Chandrasekhar's method of discrete ordinates and his recognition of the way to handle the polarization, using Stokes parameters. Numerical compilations of the Stokes parameters for scattering in a molecular atmosphere are available for all conditions encountered on earth.


$\nu_1+2\nu_2+3\nu_3$ BAND
HOT BAND
HEAD
R-BRANCH
P-BRANCH
28 22 18 16 14 12 10 8 6 4 2 0 J 2 4 6 8 10 12 14 16 18 20 22 24 26 28
INTENSITY
a.
10490
10500
10510
10520
10530
10540
WAVELENGTH, Å

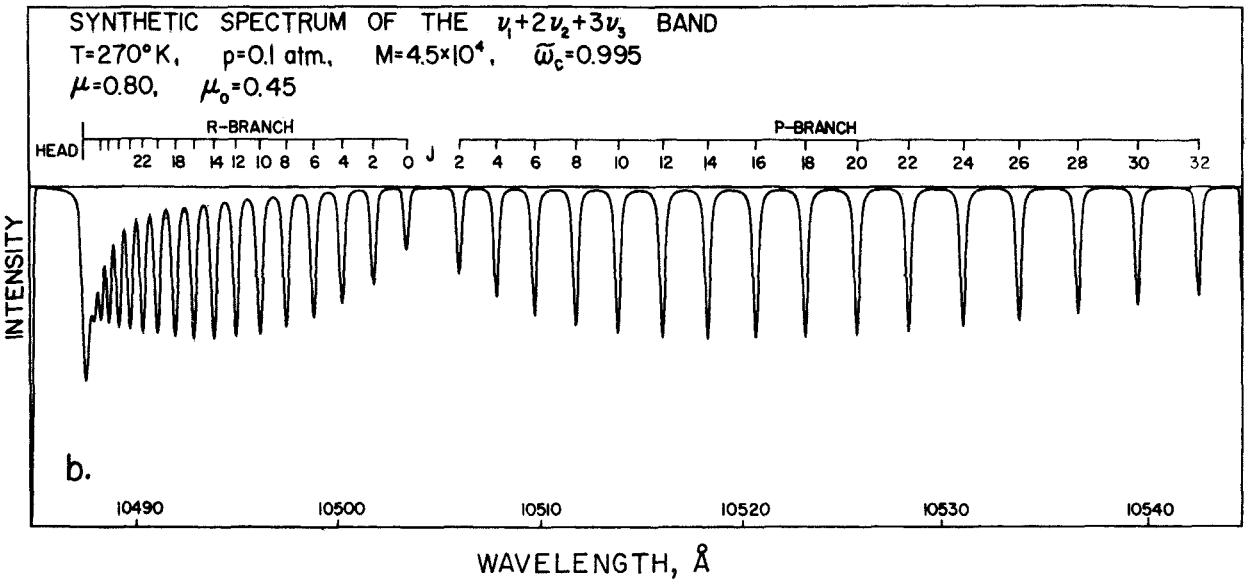


FIG. 8.18. The reflection spectrum of Venus in the $\nu_1+2\nu_2+3\nu_3$ band of carbon dioxide at 10,500 Å. (a) The observed spectrum. The vertical lines indicate overlapping spectral regions from which the entire band was put together. Fraunhofer lines are indicated by the symbol ⊙. (b) a synthetic spectrum. Based on a two-stream approximation for a semiinfinite, homogeneous atmosphere and the parameters indicated. The Fraunhofer lines and the upper state band [*hot band* in (a)] are omitted. After Belton et al. (1968).

There are a large number of ways to compare theory and observation. One is to compare the degree of polarization in the sun's vertical. Figure 8.19 shows observations at a wavelength of 3200 Å, while Fig. 8.20 shows theoretical results with which they may be compared. The agreement is fair except for the negative polarizations near to the sun. The data are for clear conditions. For turbid conditions or for longer wavelengths (for which the haze is relatively more important), there may

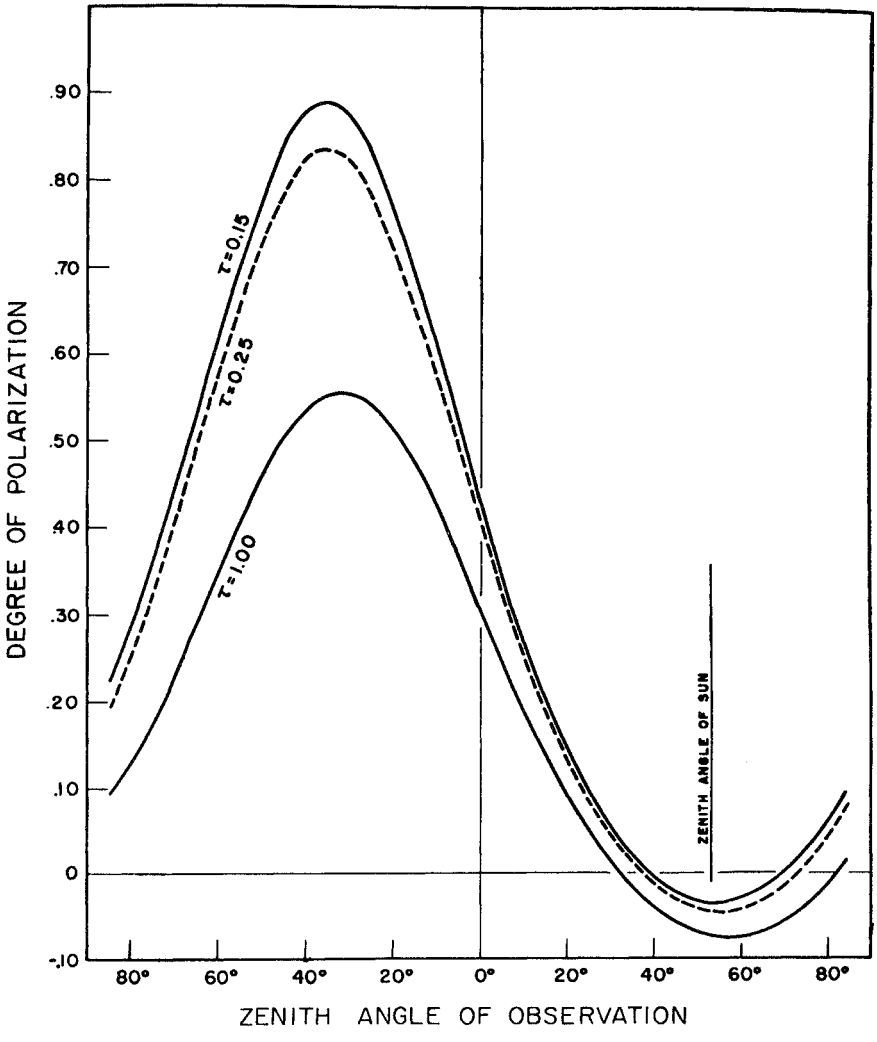


FIG. 8.19. Calculated degree of polarization in the sun's vertical. Calculations are presented for a molecular atmosphere with three optical depths. The solar zenith angle is 53.1°. The underlying surface is assumed to be nonreflecting. After Sekara (1956).

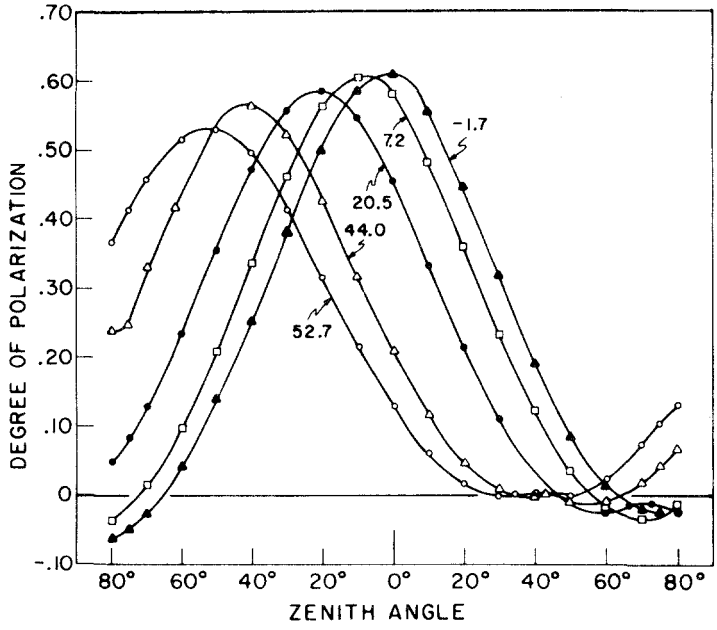


FIG. 8.20. Measured degree of polarization in the sun's vertical. The solar elevations (90° – zenith angle) are marked on the curves. The wavelength is 3200 Å and the optical depth for molecular scattering is approximately unity. Compare the left-hand curve with the lowest curve in Fig. 8.19. The agreement is fair except for the negative polarizations near to the sun angle. After Coulson (1971).

be large discrepancies between observations and theoretical results for a molecular atmosphere.

BIBLIOGRAPHY

8.1. Introduction

There are a number of treatises on multiple scattering:

Chandrasekhar, S., 1960, *Radiative transfer.* New York: Dover.

Sobolev, V. V., 1975, *Light scattering in planetary atmospheres.* Oxford: Pergamon.

van de Hulst, H. C., 1980, *Multiple light scattering, Vols. 1 and 2.* New York: Academic Press.

The computational techniques are surveyed in two review papers:

Hansen, J. E., and Travis, L. D., 1974, "Light scattering in planetary atmospheres," *Space Sci. Rev.* **16,** 527.

Irvine, W. M., 1975, "Multiple scattering in planetary atmospheres," *Icarus* **25,** 175,
and in a monograph.

Lenoble, J., 1985, *Radiative transfer in scattering and absorbing atmospheres: Standard computational procedures.* Hampton, Virginia: Deepak.

8.2. Integrodifferential equation
The Fourier series expansion was given by Chandrasekhar (1960), § 8.1. The treatment of discrete ordinates follows

Liou, K.-N., 1973, "A numerical experiment on Chandrasekhar's discrete-ordinate method for radiative transfer: Applications to cloudy and hazy atmospheres," *J. Atmos. Sci.* **30,** 1303.

Liou, K.-N., 1980, *An introduction to atmospheric radiation.* New York: Academic Press.

Stamnes, K., and Swanson, R. A., 1981, "A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmospheres," *J. Atmos. Sci.* **38,** 387.

———, and Dale, H., 1981, "A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmospheres. II: Intensity computations," *J. Atmos. Sci.* **38,** 2696.

———, and Conklin, P., 1984, "A new multi-layer discrete ordinate approach to radiative transfer in vertically inhomogeneous atmospheres," *J. Quant. Spectr. Radiat. Transfer* **31,** 273.

Feautrier's method was first given by

Feautrier, P., 1964, "Sur la résolution numérique de l'équation de transfert," *Comptes Rendus* **258**, 3189.

It has been used by

Prather, M. J., 1974, "Solution of the inhomogeneous Rayleigh scattering atmosphere," *Astrophys. J.* **192,** 787.

8.3. Interaction Principle
The basic ideas were advanced by

Redheffer, R. M., 1962, "On the relation of transmission-line theory to scattering and transfer," *J. Math. Phys.* **41,** 1.

They were developed by a number of groups

van de Hulst, H. C., and Grossman, K., 1968, "Multiple light scattering in planetary atmospheres," *The atmospheres of Venus and Mars* (J. C. Brandt, and M. B. McElroy, Eds.). New York: Gordon and Breach, p. 35.

Hunt, G. E., and Grant, I. P., 1969, "Discrete space theory of radiative transfer and its application to problems in planetary atmospheres," *J. Atmos. Sci.* **26,** 963.

Grant, I. P., and Hunt, G. E., 1969, "Discrete space theory of radiative transfer. I. Fundamentals," *Proc. Roy. Soc. London Ser. A* **313,** 189.

The last reference established and proved the star semigroup mathematical structure of radiative transfer in a plane-parallel atmosphere. Doubling and adding were proposed by

Twomey, S., Jacobowitz, H., and Howell, H. B., 1966, "Matrix method for multiple-scattering problems," *J. Atmos. Sci.* **23,** 289.

Invariant imbedding was developed by

Bellman, R., Kalaba, R., and Prestrud, M., 1963, *Invariant imbedding and radiative transfer in slabs of finite thickness.* New York: Elsevier.

An efficient algorithm for anisotropic scattering was developed by

Sato, M., Kawabata, K., and Hansen, J. E., 1977, "A fast invariant imbedding method for multiple scattering calculations and an application to equivalent widths of CO_2 lines on Venus," *Astrophys. J.* **216,** 947.

8.4. Miscellaneous methods

The integral equation method is discussed in

Shia, R.-L., and Yung, Y. L., 1986, "Radiative transfer in a sphere illuminated by a parallel beam: An integral equation approach," *Astrophys. J.* **301,** 554.

Low-order discrete ordinate methods are compared in

Meador, W. E., and Weaver, W. R., 1980, "Two-stream approximation to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement," *J. Atmos. Sci.* **37,** 630.

Cuzzi, J. N., Ackerman, T. P., and Helmle, L. C., 1982, "The delta-four-stream approximation for radiative flux transfer," *J. Atmos. Sci.* **39,** 917.

Monte Carlo simulations have been carried out by

Adams, C. N., and Kattawar, G. W., 1978, "Radiative transfer in spherical shell atmospheres I. Rayleigh scattering," *Icarus* **35,** 139.

Kattawar, G. W., and Adams, C. N., 1978, "Radiative transfer in spherical shell atmospheres II. Asymmetric phase functions," *Icarus* **35,** 436.

The path distribution technique was formulated by

Irvine, W. M., 1964, "The formation of absorption bands and the distribution of photon optical paths in a scattering atmosphere," *Bull. Astron. Inst. Neth.* **17,** 266.

It is extensively discussed in Chapter I (part II) of Lenoble's text, and Chapter 17 of van de Hulst's text (§ 8.1).

8.5. Numerical results

Figure 8.4 is from Liou (1980), § 8.2. Figure 8.5 is from Irvine (1975), § 8.1.

Numerical values of X, Y, and H functions are taken from

Carlstedt, J. L., and Mullikin, T. W., 1964, "Chandrasekhar's X- and Y-functions," *Astrophys. J. Suppl.* **113,** 449.

Caldwell, J., 1971, "The X- and Y-functions of Chandrasekhar," *Astrophys. J.* **163,** 111.

Figure 8.8a and b were computed by the Feautrier method using 90 levels, 20 streams, and 10 Fourier components.

The approximate expression for the H-function (8.160) is from

Hapke, B., 1981, "Bi-directional reflectance spectroscopy I. Theory," *J. Geophys. Res.* **86,** 3039.

Figure 8.9 is from Liou (1973), § 8.2. Figure 8.10 is from

Uesugi, A., and Irvine, W. M., 1970, "Multiple scattering in a plane-parallel atmosphere I. Successive scattering in a semi-infinite medium," *Astrophys. J.* **159,** 12.

Figure 8.11 and Table 8.1 are from Meador and Weaver (1980), § 8.4. The four stream solution was first derived by

Liou, K.-N., 1974, "Analytic two-stream and four-stream solutions for radiative transfer," *J. Atmos. Sci.* **31,** 1473.

8.6. Applications

8.6.1. *Solar and thermal fluxes in stratocumulus clouds*
This section follows

Stephens, G. L., Paltridge, G. W., and Platt, C. M. R., 1978, "Radiation profiles in extended water clouds. III: Observations," *J. Atmos. Sci.* **35,** 2133.

8.6.2. *Polarization of light reflected from Venus*
Many investigators contributed to this important work on the nature of the Venus clouds. Figure 8.15 is from

Hansen, J. E., and Hovenier, J. W., 1974, "Interpretation of the polarization of Venus," *J. Atmos. Sci.* **31,** 1137.

8.6.3. *Scattered light in the stratosphere*
The measurements described are by

Herman, J. R., and Mentall, J. E., 1982, "The direct and scattered solar flux within the stratosphere," *J. Geophys. Res.* **87,** 1319.

8.6.4. *Scattered light in clear water*
A modern text on scattering in sea water is

Jerlov, N. G., 1976, *Marine optics*. New York: Elsevier.

8.6.5. *CO_2 lines in the reflection spectrum of Venus*
This presentation follows

Belton, M. J. S., Hunten, D. M., and Goody, R. M., 1968, "Quantitative spectroscopy of Venus in the region 8,000–11,000 Å," *The atmospheres of Mars and Venus* (J. C. Brandt and M. B. McElroy, Eds.). New York: Gordon and Breach, p. 69.

8.6.6. *The color and polarization of skylight*

The outstanding contributions to this problem by Chandrasekhar (1960), § 8.1, set the stage for many of the advances in the theory of multiple scattering described in this chapter.

A compendium of Stokes parameters for a molecular atmosphere is

Coulson, K. L., Dave, J. V., and Sekara, Z., 1960, *Tables related to radiation emerging from a planetary atmosphere with Rayleigh scattering.* Berkeley and Los Angeles: University of California Press.

Figure 8.19 is from

Sekara, Z., 1956, "Recent developments in the study of polarization of skylight," *Adv. Geophys.* **3,** 43.

The observed data in Fig. 8.20 are from

Coulson, K. L., 1971, "On the solar radiation field in a polluted atmosphere," *J. Quant. Spectrosc. Radiat. Transfer* **11,** 739.

9

ATMOSPHERES IN RADIATIVE EQUILIBRIUM

9.1. Introduction

In this chapter we discuss *radiative equilibrium models* of the earth's atmosphere and the closely related *radiative–convective models,* for which small-scale convection is included in a highly parameterized form. In both cases, heat transports by planetary-scale motions are neglected.

Despite their limitations, radiative equilibrium and radiative–convective studies have provided stimuli for many of the fundamental ideas discussed in this book. Their value is principally heuristic. The radiative equilibrium state is one conceivable state of a planetary atmosphere that may be analyzed so that the implications of parameteric changes can be understood in simple terms (e.g., changes in atmospheric composition, earth orbital elements, solar emission, etc.). The same cannot yet be said of any dynamic model. While numerical solutions are available from general circulation models, their behavior is often no easier to interpret than that of the atmosphere itself.

For studies that are not based on the existence of day-to-day observations, radiative equilibrium considerations provide the irreplaceable first step in a number of fields: the atmospheres of other planets, stellar atmospheres, the earth's primitive atmosphere; and much of the progress in studies of climate change has been based on the simplest energy balance models. In addition to their value in examining general principles, there is a recurrent, although disputed theme that radiative equilibrium has direct relevance to the observed atmospheric structure. This proposition embraces a number of instructive ideas but, before examining them, we consider some of the observational evidence that motivates them.

From the earliest days following the discovery of the stratosphere, theoretical workers assumed that the stratosphere, unlike the troposphere, was in radiative equilibrium. The reasoning was that no forms of heat transport, other than radiative, could be important in a highly stable atmosphere. Since nothing was known about planetary-scale motions at that time, this conclusion was premature.

If we turn to modern data, Fig. 9.1a presents the observed climatological temperatures in the middle atmosphere, to be compared

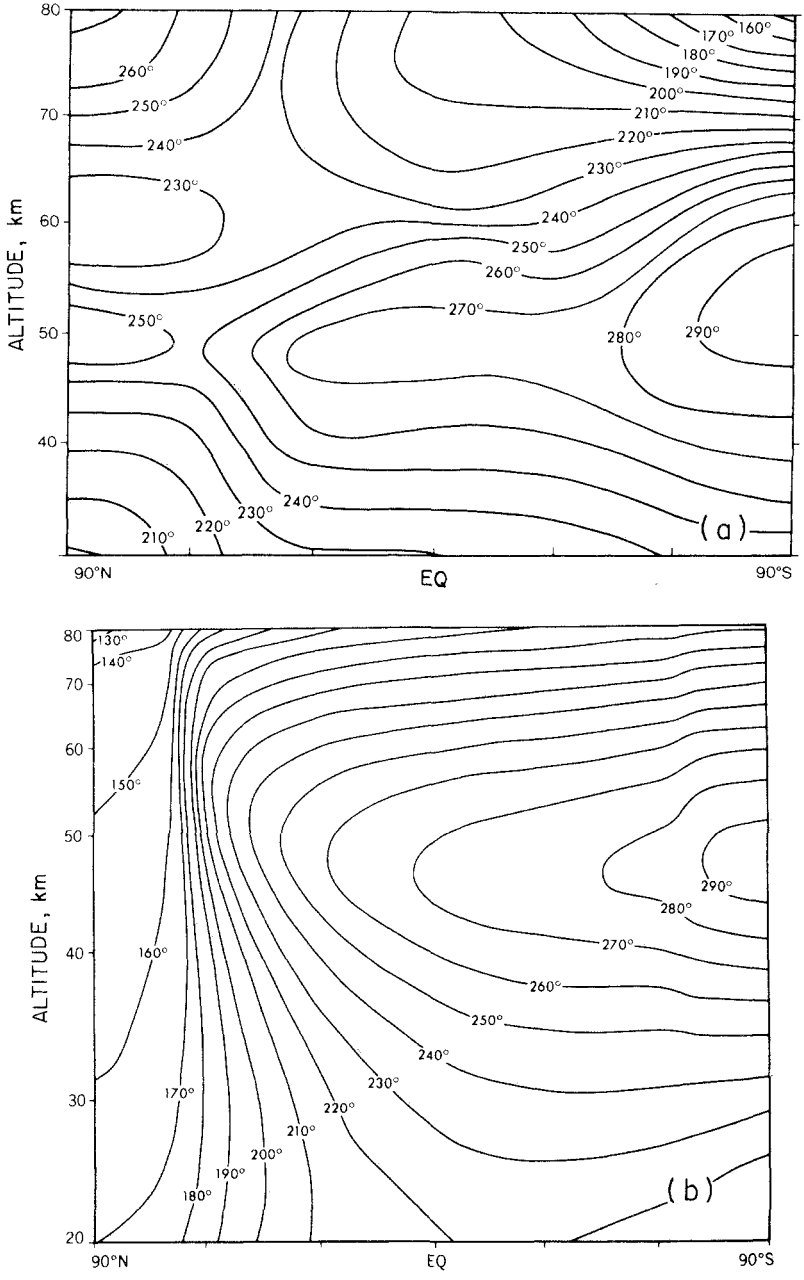


FIG. 9.1. Radiative equilibrium temperatures in the middle atmosphere. (a) Observed zonal mean temperatures for January. (b) Theoretical calculation for average conditions on January 15. Ozone concentrations are also calculated in a time-marching radiative–convective–photochemical model. Allowance must be made for the different vertical scales. The agreement is fair between 90°S and 30°N. Part of the disagreement may be attributable to the photochemical model. Discrepancies in the winter polar regions would be larger but for the fact that the calculation assumes the observed lower atmosphere temperatures. After Fels (1985).

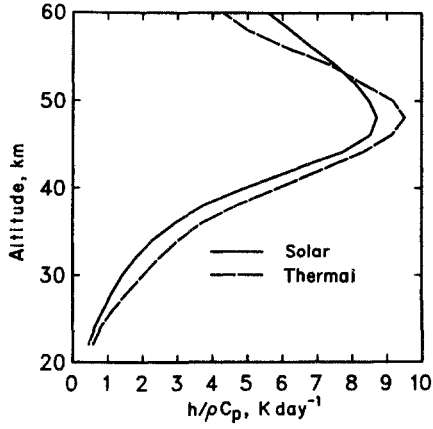


FIG. 9.2. Globally averaged heating and cooling rates in the middle atmosphere. The solid line is solar heating, while the broken line is cooling by thermal radiation. Calculations are for the month of January. Observed values are used for absorber densities. After Kiehl and Solomon (1986).

with radiative equilibrium calculations shown in Fig. 9.1b. Agreement between theory and observation is fairly good except in the region of the polar winter. Better agreement exists for horizontal averages over the entire planet. Figure 9.2 shows calculations of globally averaged solar heating rates and thermal cooling rates; the differences between the two are small compared to either individually.

Tropospheric radiative–convective calculations of atmospheric temperature also exhibit good agreement with observations. Later in this chapter, in Fig. 9.13, we show agreement for averages over the entire globe, at all altitudes between the surface and 40 km. It is, however, important to emphasize that the apparent agreement in the troposphere has little substance. The tropospheric temperature gradient for the model has been assumed to be equal to the observed global average, and most of the radiation to space originates in the troposphere. These two facts, taken together with a requirement for flux balance at the top of the atmosphere, imply that the height of the tropopause is the only undetermined parameter, and this may be calculated from a condition of continuity of temperature at the tropopause, given the stratospheric temperature. The only important question is, therefore, why is the stratosphere close to radiative equilibrium, given the existence of strong atmospheric motions throughout the middle atmosphere? A complete answer to this question has not yet been given but the following are some of the considerations.

1. We shall show in § 10.4 that a meridional, Hadley circulation driven by radiative imbalance will have temperatures close to radiative equilibrium if the atmospheric mass is small enough.

Stratospheric motions are more complicated than this, although the strong zonal winds blow orthogonally to the average temperature gradient, and transport no heat. Consequently, it is reasonable to anticipate that, above a certain level, for which the overlying mass is small enough, temperatures should be close to radiative equilibrium.

2. Tropospheric motions act to decrease the equator-to-pole temperature difference in the lower atmosphere. The equator-to-pole difference in thermal radiation passing upward into the stratosphere is thereby decreased. Strong motions in the troposphere can act to decrease the thermally driven motions in the middle atmosphere.
3. For reasons that may not be independent from those discussed in (1) and (2), isentropic surfaces in the middle atmosphere are often continuous over the globe and, for some levels, the surfaces are approximately horizontal. It can be demonstrated that, for averages over isentropic surfaces, contributions to the heating from vertical and horizontal motions cancel, to first order. Source functions, averaged over these quasihorizontal surfaces, are, therefore, principally determined by radiative balance.

The significance of these rationalizations is hard to assess but, together, they probably account for the important place occupied by radiative–convective models in the meteorological literature.

9.2. An elementary solution

The two-stream solution for a grey-absorbing atmosphere, (2.141) and (2.146), can be applied to a planetary atmosphere; we assume no incoming thermal radiation from space (at $\tau = 0$) and include heating by solar radiation at the ground and in the atmosphere.

The total flux of radiation is

$$F_R = F_S + F_T, \tag{9.1}$$

where F_S is the solar flux and F_T is the thermal flux. For radiative equilibrium in a stratified atmosphere, F_R is constant and equal to zero, because there is flux balance at the upper boundary. As a result,

$$F_S = -F_T. \tag{9.2}$$

9.2.1. *Without solar absorption*

In the absence of solar absorption the solar flux is constant (negative). The equilibrium condition is now

$$F_T = -F_S = \text{const.} \tag{9.3}$$

We use Planck's source function because we are mainly interested in the lower atmosphere. From (2.141) we have

$$B(\tau) = B(0) + \frac{3F_T\tau}{4\pi}, \tag{9.4}$$

and, from (2.146),

$$B(0) = \frac{F_T}{2\pi}.$$

At the surface ($\tau = \tau_1$),

$$B^*(\tau_1) = \frac{F_T}{2\pi}\left(2 + \frac{3\tau_1}{2}\right).$$

The solution is

$$\begin{aligned} B(\tau) &= \frac{\sigma\theta(\tau)^4}{\pi} = \frac{-F_S(1 + 3\tau/2)}{2\pi}, \\ B^*(\tau_1) &= \frac{\sigma\theta_g^4}{\pi} = \frac{-F_S(2 + 3\tau_1/2)}{2\pi}, \end{aligned} \tag{9.5}$$

where θ_g is the ground temperature.

If we wish to write (9.5) with height as the independent variable we require a relationship between τ and z. The relationship,

$$\tau = \tau_1 \exp\left(\frac{-z}{H}\right), \tag{9.6}$$

applies approximately to many gases. In particular, it is approximately true for water vapor in the troposphere with $H = 2$ km.

The solution, (9.5), although based upon many simplifications, has features that are instructive for planetary atmospheres.

1. The ground temperature θ_g is always greater than it would be if the atmosphere were transparent ($\tau_1 = 0$). This elevation of surface temperature is known as the *greenhouse effect.* There is some debate as to whether this is an appropriate term (does it account for the behavior of a domestic greenhouse?), but there can be no argument about its importance. It is, for example, the fundamental cause for the 730 K surface temperature that we find on Venus.

2. There are discontinuities,

$$\Delta B = \frac{F_S}{2\pi},$$

at both boundaries. At the upper boundary, this leads to a finite temperature as the atmosphere merges with space (the *skin temperature*),

$$\theta_S = \left(\frac{-F_S}{2\sigma}\right)^{1/4}. \tag{9.7}$$

3. The temperature lapse rate can be written

$$-\frac{\partial\theta}{\partial z} = \frac{(\theta/H)\tau}{1+3\tau/2}. \tag{9.8}$$

The lapse rate increases with τ and has its greatest value at the surface where the temperature is highest, a point of importance when we come to consider free convection.

We shall continue to use the two-stream approximation because its simplicity is a valuable aid to understanding. We should, therefore, compare its results to those for an exact solution. This is done for a semiinfinite atmosphere (i.e., no lower surface) in Fig. 9.3. Unless we are dealing with fine points that require high precision, the comparison is satisfactory. The two-stream approximation contains, in approximate form, all the physics of the complete equations; for a discussion of general principles it can be used with few reservations.

9.2.2. *Absorption of solar radiation*

We now consider an atmosphere with an absorption coefficient for solar radiation α times that for thermal radiation. Equation (2.115) can be written

$$F_S(\tau) = F_S(0)\exp\left(\frac{\alpha\tau}{\xi_\odot}\right), \tag{9.9}$$

where τ is the optical depth for thermal radiation. For radiative equilibrium,

$$F_T(\tau) = -F_S(0)\exp\left(\frac{\alpha\tau}{\xi_\odot}\right). \tag{9.10}$$

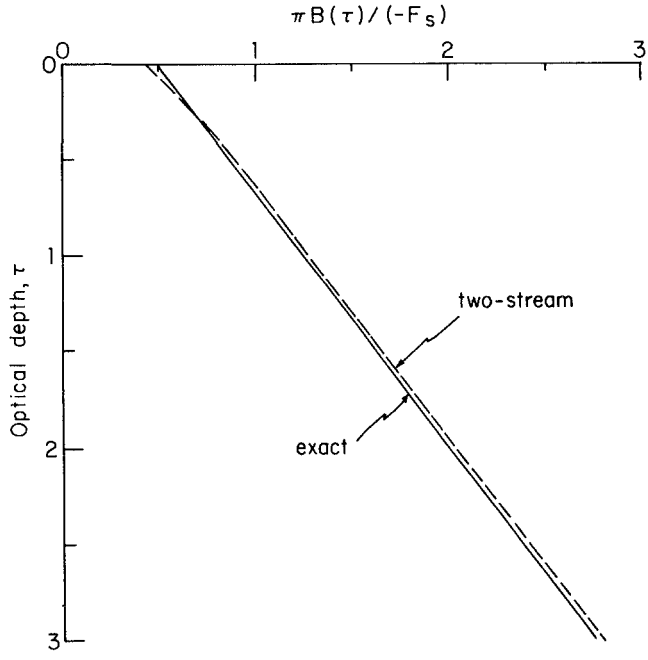


FIG. 9.3. Comparison of a two-stream solution for a semiinfinite atmosphere with the exact solution.

The two-stream equation, (2.140), now yields

$$4\pi \frac{dB(\tau)}{d\tau} = \left(3 - \frac{\alpha^2}{\xi_\odot^2}\right) F_T,$$

or

$$B(\tau) - B(0) = \frac{F_S(0)\xi_\odot}{4\pi\alpha}\left[1 - \exp\left(\frac{\alpha\tau}{\xi_\odot}\right)\right]\left(3 - \frac{\alpha^2}{\xi_\odot^2}\right). \tag{9.11}$$

The boundary condition (2.145) is, for $I^-(0) = 0$,

$$B(0) = \frac{F_S(0)}{2\pi}\left(\frac{\alpha}{2\xi_\odot} - 1\right). \tag{9.12}$$

The solution is then

$$B(\tau) = \frac{F_S(0)}{2\pi}\{(\alpha/2\xi_\odot - 1) + (3\xi_\odot/2\alpha - \alpha/2\xi_\odot)[1 - \exp(\alpha\tau/\xi_\odot)]\}. \tag{9.13}$$

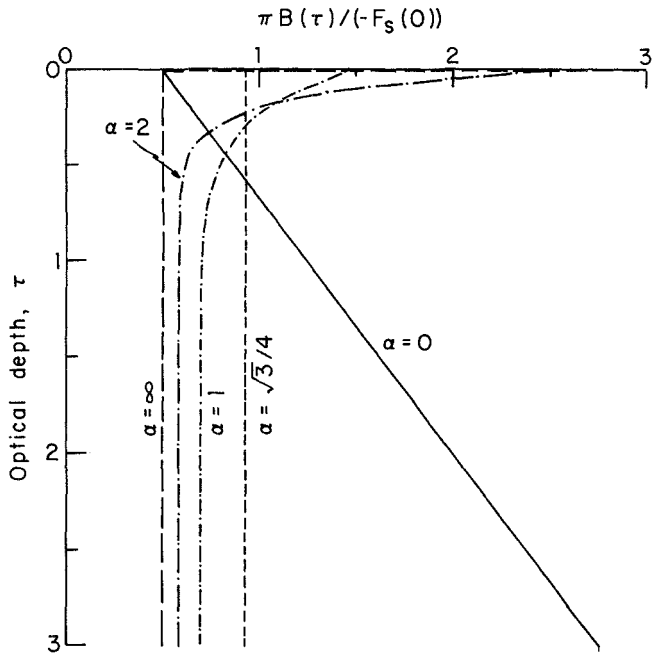


FIG. 9.4. Source functions calculated from (9.13) for $\xi_{\odot} = -1/4$.

The lower boundary condition does not influence the solution above the boundary.

The curve for $\alpha = 0$ in Fig. 9.4 is for no solar absorption and is the same as the two-stream curve in Fig. 9.3 (*a greenhouse solution*). For finite α, solar radiation is absorbed in the upper layers, which heat up, while lower layers cool. This behavior corresponds qualitatively to that of the warm layer at about 50 km in the earth's atmosphere. For this layer, the absorbing gas is ozone. The similarity to the earth's atmosphere becomes more apparent if height rather than optical depth is used as the independent variable. From the barometric law, it may be demonstrated that the temperature gradient tends to zero as $z \to \infty$ for all the profiles in Fig. 9.4, as is observed in the earth's upper atmosphere (Fig. 1.4). Solutions for $\alpha > 1$ are *thermosphere solutions.*

If solar radiation is absorbed in the upper atmosphere, the lower atmosphere is, not surprisingly, cooled. In the limit $\alpha \to \infty$, for which the skin temperature tends to infinity, the lower atmosphere temperature is the same as the skin temperature corresponding to the greenhouse solution. For earth this temperature is about 200 K.

The lower atmosphere of the grey thermosphere solution differs greatly from earth as it is now. Ozone absorbs enough solar energy to form a high-altitude, high-temperature layer, but not enough to prevent

the surface from being strongly heated by the sun. It would be different for the *nuclear winter*. Then, solar radiation of all wavelengths is absorbed by dust and the surface may be screened entirely. At high levels the atmosphere will be heated but low temperatures may develop at the surface. Such events have been observed on Mars during a global dust storm. Details of the solution depend upon the optical properties of the dust, and upon scattering, which we have omitted.

9.3. Nongrey atmospheres

Radiative equilibrium calculations for nongrey atmospheres are usually performed numerically. Before discussing such work we look at formal treatments based on simple spectral models.

9.3.1. *Models without pressure broadening*

For a grey atmosphere, the k distribution function $f(k)$ (§ 4.8) is a Dirac δ-function. For a nongrey atmosphere, the distribution function is broadened and there are characteristic differences between random and regular models (Fig. 4.20a and b). It is reasonable to anticipate a close relationship between the shape of the k distribution and the equilibrium temperature distribution.

Figure 9.6 shows radiative equilibrium solutions for the illustrative regular models displayed in Fig. 9.5. The grey solution (GR) corresponds to (9.5), but with a different presentation. None of the regular models differs radically from the grey solution. With some rescaling of the effective temperature (e.g., by scaling the incident solar radiation) and minor rescaling of the mean absorption coefficient, the curves could be made to correspond closely.

The nongrey solutions in Fig. 9.6 all have finite skin temperatures. Comparison of Figs. 9.5 and 9.6 shows the skin temperature decreasing as the k distribution broadens.

Random models have k distributions that, other things being equal, are broader than those of regular models (§ 4.8). The comparison between equilibrium curves for random and regular models shown in Fig. 9.7 bears out the general conclusion that the wider the k distribution the greater the difference in equilibrium conditions from those for grey absorption.

9.3.2. *Pressure broadening*

The effect of pressure upon line width is not included in the foregoing models. There is, however, an important difference in the skin temperature between models with and without pressure broadening; for pressure-broadened lines the skin temperature can be zero.

Figure 9.8 deals with this question in the context of the k distribution

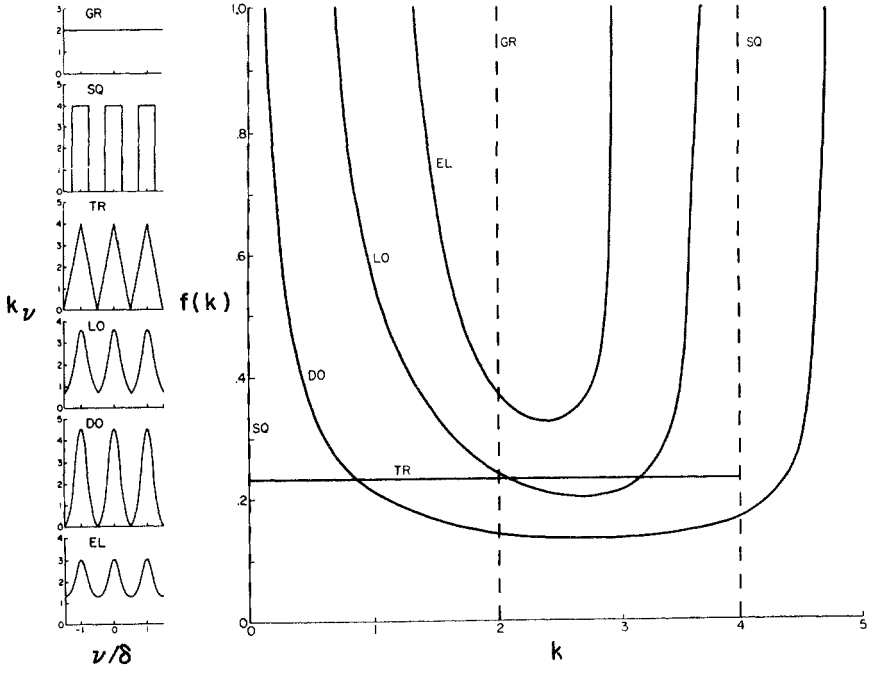


FIG. 9.5. k distributions for regular band models. The bands are shown on the left side. GR is grey absorption, SQ is square lines, TR is triangular lines, LO is the Schnaidt model with Doppler lines, and EL is the Elsasser model for Lorentz lines (§ 4.5.1). Each band has the same ratio of line width to line spacing ($y = 0.25$) and the same average absorption coefficient ($\bar{k} = 2$). For the grey model, the k distribution is a δ-function at $k = 2$. For square lines, the k distribution consists of two δ-functions at $k = 0$ and 4. After Arking and Grossman (1972).

technique for regular bands. The lines are square and their width is varied while maintaining the same mean absorption coefficient. The skin temperature decreases as the lines narrow.

The same general result is demonstrated in Fig. 9.9 for a single line of Lorentz shape. This was obtained by an ingenious analytic technique using the expression, (6.19), for the equivalent width of a single line for a gas of constant mixing ratio. A diffusivity factor $r = 1.66$, (6.6), may be included in the definition of $\bar{u}$ to allow for the diffuse character of the radiation. It can be shown that the heating rate equation (6.60), when equated to zero, has exact solutions for integral values of $\bar{u}$. These are tedious to evaluate, however, and results for only $\bar{u} = 1$, 2, and 3 are shown in Fig. 9.9. The result for $\bar{u} \gg 1$ was obtained by an approximate numerical method.

Vanishing skin temperatures for lines of all intensities are shown in Fig. 9.9. This result can be explained in the following simple terms. Consider a very thin layer above the rest of the atmosphere. If no

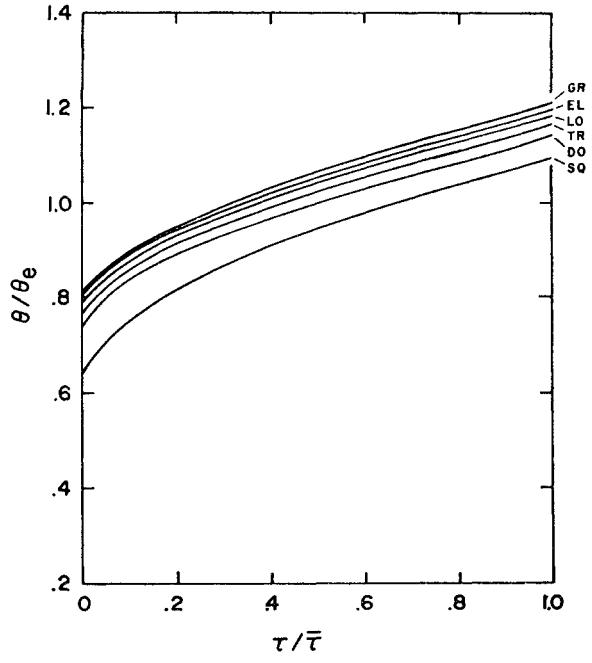


FIG. 9.6. Radiative equilibrium for the band models in Fig. 9.5. θ_e is the effective temperature, (1.2). In terms of quantities used previously in this chapter,

$$\frac{\theta}{\theta_e} = \left[\frac{\pi B}{(-F_s)}\right]^{1/4}.$$

$\bar{\tau}$ is the optical depth based on the mean absorption coefficient. The model is transparent to solar radiation and has a black surface at $\bar{\tau} = 2$. After Arking and Grossman (1972).

thermal radiation is incident from above, the layer absorbs radiation from one direction only, but emits in two. For grey absorption, this leads directly to the relationship,

$$\theta_s^4 = \tfrac{1}{2}\theta_e^4, \tag{9.14}$$

where θ_s is the skin temperature and θ_e is the effective temperature. This result is the same as (9.7) because the outward flux is, by definition, equal to $\sigma\theta_e^4$.

For line absorption, on the other hand, the only radiation that affects the skin layer is that emitted by intense line centers; but most of the flux, which determines θ_e, comes from the line wings. Emission in line centers takes place high in the atmosphere where the temperature, in the absence of solar absorption, is well below the effective temperature. The temperature used on the right-hand side of (9.14) should, under these

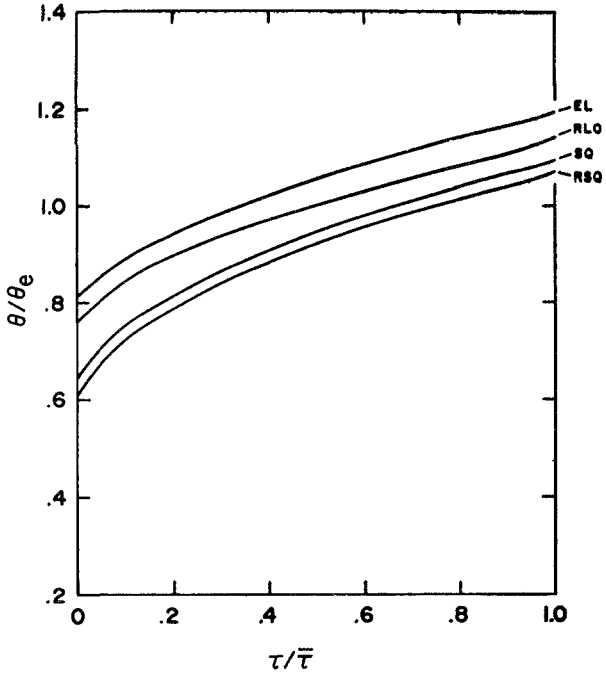


FIG. 9.7. Radiative equilibrium for random and regular bands. RLO is a random Lorentz model. RSQ is a random model of square lines. EL and SQ are the same as for Fig. 9.6. $\bar{k} = 2$ and $y = 0.25$ for all models. k distributions for random and regular models are compared in Fig. 4.20. The random model has an exponential distribution of line intensities. All calculations are without absorption of solar radiation and for a black surface at $\bar{\tau} = 2$. After Arking and Grossman (1972).

circumstances, be replaced by a temperature lower than θ_e and the calculated skin temperature will be lower. The greater the contrast between line centers and wings, the lower the skin temperature. If the Lorentz shape is used down to zero pressure (as in this calculation), the skin temperature is zero.

9.3.3. *Numerical methods*

Numerical methods for calculating heating rates for both solar and terrestrial radiation are discussed in Chapter 6. In an appropriate algorithm the temperature can be adjusted until the net heating is less than a predetermined value. Two iterative techniques have been employed. Both start from an assumed initial state. The first method then calculates the net heating at each level and uses the heating rate to determine an adjustment to the temperature at that level. For example,

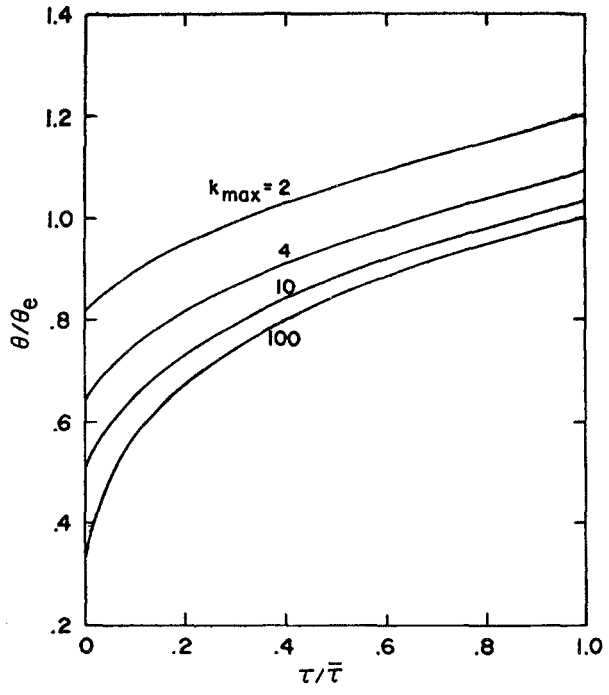


FIG. 9.8. Equilibrium temperatures for a regular array of square lines. The mean absorption coefficient is the same for each profile ($\bar{k} = 2$). The absorption coefficient between lines is fixed at zero while the coefficient at the line center (k_{max}) is varied, and is shown against each curve. The line width varies inversely with the coefficient at the line center. After Arking and Grossman (1972).

the temperature, $\theta(z)$, may be adjusted in one time step, Δt, to

$$\theta'(z) = \theta(z) + \frac{h_R(z)\,\Delta t}{\rho c_p}. \tag{9.15}$$

For this *time-marching* or *relaxation method,* the temperature has a physical meaning at each time step. But such methods are very slow, requiring 200 or more days to reach a steady state in the troposphere. With a time step every 8 hours and a detailed radiative computation at each step, this can be a lengthy process.

A more economical technique involves inversion of the Curtis matrix, (6.60). This yields directly an equilibrium source function for each level. To make matters simpler the source functions J_i^s can be expanded to first order in the temperature, about a mean state. Equation (6.60) then becomes a matrix equation involving the temperature, but with modified matrix elements.

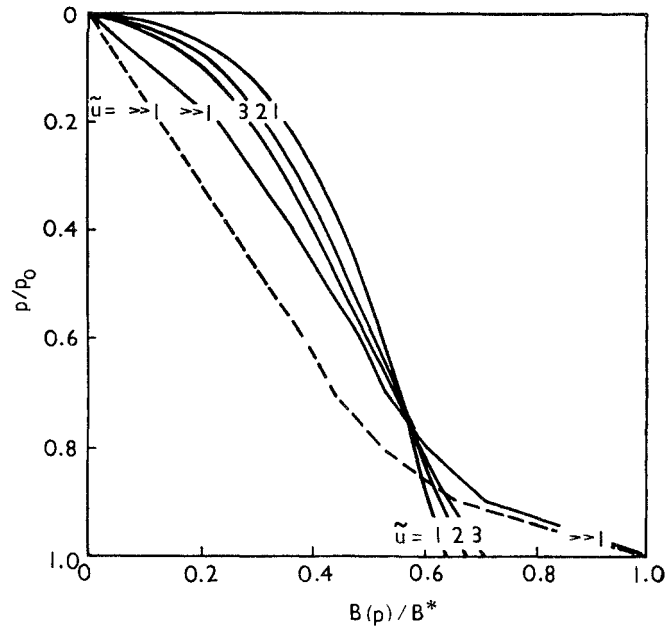


FIG. 9.9. Source functions for a single Lorentz line. p_0 is the surface pressure. B^* is the source function for the lower surface. The ratio $B(p)/B^*$ is equivalent to the quantity plotted in Figs. 9.3 and 9.4.; the outward flux has no significance for an atmosphere containing only a single absorption line. After King (1952).

Efficient algorithms for matrix inversions are available. A single inversion is not sufficient because the matrix elements are functions of temperature and must be recalculated whenever the temperatures are changed. They are not sensitive functions of temperature, however, and four iterations are normally enough for the calculation to converge. This has obvious advantages over the $1-2 \times 10^3$ iterations required for the time-marching technique.

Figure 9.10 shows the first well-known radiative equilibrium calculation for the earth's atmosphere made with a digital computer. Figure 9.11 shows a breakdown into solar and terrestrial components for each of the three gases considered. More refined calculations, published subsequently, do not change the general nature of these results.

In Fig. 9.10, the equilibrium temperature distribution below 15 km corresponds quite closely to that for a grey atmosphere, (9.5) and (9.6), with $\tau_1 = 4$. There are also differences between the grey and nongrey solutions. The grey solution has a temperature discontinuity at the surface of 11.3 K, while the numerical calculation has a discontinuity of only 0.06 K. This discrepancy is not as important as appears at first sight. The temperature gradients near to the surface also differ in the two

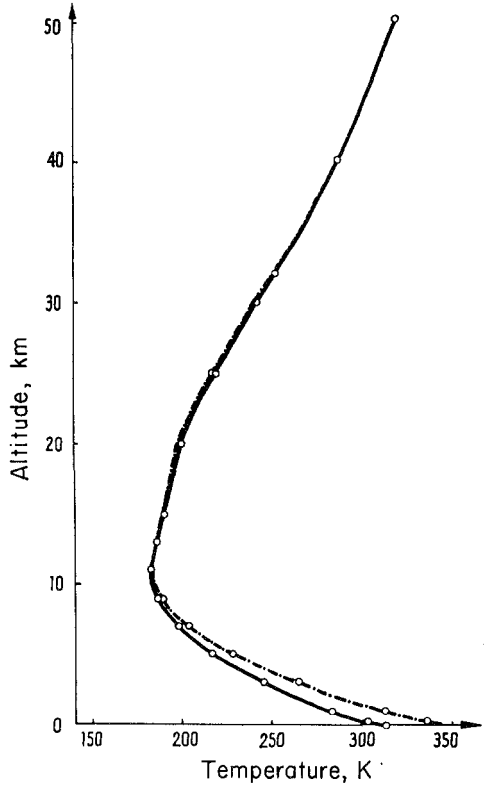


FIG. 9.10. Radiative equilibrium in the earth's atmosphere. Conditions correspond to the annual average at 40°N and a mean $\bar{\xi}_{\odot} = -0.5$. Full line (——) = clear sky conditions; dash-dot line (-·-·-) = cloudiness. After Möller and Manabe (1961).

solutions. For the numerical model, the gradient is $90\,\mathrm{K\,km^{-1}}$; for the grey model, it is $36\,\mathrm{K\,km^{-1}}$. The grey discontinuity has been exchanged for a steeper gradient in the lowest few kilometers in the numerical model. When free convection is introduced into the calculation there is relatively little difference between these two cases.

Between 15 and 30 km, Fig. 9.10 shows an increase of temperature with height. This is caused by solar absorption by ozone; the balance is cooling by water vapor and carbon dioxide in roughly equal amounts. In the troposphere, water vapor is the single most important gas, dominating both the thermal and the solar radiation fields.

9.4. The troposphere and the stratosphere

9.4.1. *Introduction*

The discovery of the stratosphere by Teisserenc de Bort in the year 1900 led to a historic series of papers on radiative equilibrium in planetary

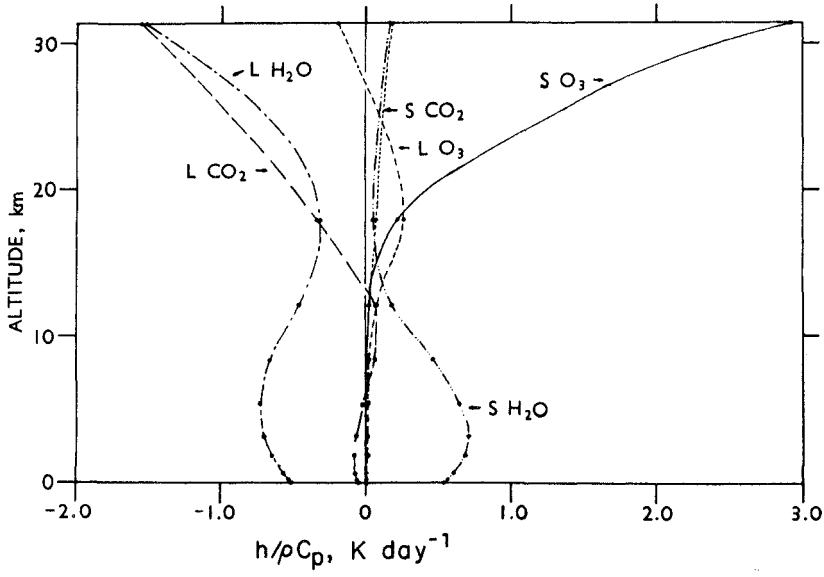


FIG. 9.11. Contributions to radiative balance. L and S refer to long-wave (terrestrial) and short-wave (solar) radiation. H_2O, CO_2, and O_3 refer to the three gaseous components. The conditions are similar to those for Fig. 9.10 but are not identical. After Manabe and Möller (1961).

atmospheres, all based on the hypothesis that the lower stratosphere is close to radiative equilibrium. In this section we shall pursue the ideas of Emden (1913), who showed that a state of radiative equilibrium in a grey atmosphere is unstable in its lower layers and that convective activity must develop near the planet's surface. This was the first attempt to account for the existence of the convective troposphere. We shall show that a remarkably good first approximation to the thermal state of the atmosphere results from models that contain a simple parametric treatment of the convective layer. In these models it is convenient to use the terms stratosphere and troposphere to refer to the radiative and convective layers even though the correspondence between atmosphere and model may be imperfect.

A historical note should draw attention to the work of Humphreys (1909) and E. Gold (1909), who approached the problem simultaneously but from two different points of view. Humphreys regarded the stratosphere as a manifestation of the finite skin temperature of an atmosphere in radiative equilibrium, (9.7).[1] This temperature is related to the

[1] To be consistent with the global average solar flux that was used in § 1.1, we must write

$$\bar{F}_S = f(1-a)\bar{\xi}_\odot, \tag{9.16}$$

where, from (1.1) and (1.2), $\bar{\xi}_\odot = -0.25$. With the numbers employed in § 1.1,

$$\bar{F}_S = -241.5 \text{ W m}^{-2}.$$

effective emission temperature by (9.14). The skin temperature corresponding to an effective temperature of 255.5 K is 214.8 K, close to the average observed temperature of the lower stratosphere.

Gold, on the other hand, made no attempt to explain the existence of an isothermal stratosphere but, assuming it to exist and to have the temperature of the tropospheric layer in contact with it, he computed where the base must lie in order to maintain an overall radiative heat balance for the stratosphere. He found a basal pressure equal to $p_0/4$, where p_0 is the surface pressure (in good agreement with the tropopause pressure in middle latitudes). Milne (1922) criticized a number of aspects of Gold's work, including the assumption of an isothermal layer in the lower stratosphere, since, for a grey-absorbing atmosphere, the temperature must decrease with height if the outgoing thermal flux is to be positive (2.141).

An interesting review of these early contributions was written by Pekeris (1932). It throws light on the progress of thought in this subject and its relationship to early astrophysical enquiries. We may note that Gold's work is not a theory but a test for self-consistency. Gold does not show why there should be an isothermal layer in equilibrium with the troposphere immediately below, but an extension of Emden's ideas leads to criteria similar to those adopted by Gold, as we shall show.

9.4.2. *The troposphere and the stratosphere*

Figure 9.12 and Table 9.1 present the radiative equilibrium solution, (9.5), in geometric coordinates by assuming an exponential distribution with height of the optical depth, (9.6). The data shown in Fig. 9.11 suggest that water vapor is the dominant thermal radiator in the troposphere, so we adopt a scale height of 2 km, similar to that observed for water vapor in the troposphere.

All of the profiles shown in Fig. 9.12 have large negative temperature discontinuities at the surface. A negative temperature discontinuity is hydrostatically unstable and convection will ensue for all models. This instability is reinforced for $\tau_1 > 0.5$ by lapse rates near to the surface that are also unstable. Following most investigators in the field, we assume that convection is so efficient that it constrains the lapse rate to a maximum value equal to the neutral, marginally stable lapse rate. While it would be reasonable to use the adiabatic lapse rate for the neutral condition, it is more common to use the observed average lapse rate in the troposphere, $6.5\ \mathrm{K\ km^{-1}}$. This has the illusory advantage of ensuring a degree of similarity between the solutions and the earth's atmosphere. The column labeled $z_{6.5}$ shows the height to which the lapse rate is unstable under this criterion.

We now seek a condition governing the height of the convective layer. The first point to note is that, for a grey atmosphere, the solution

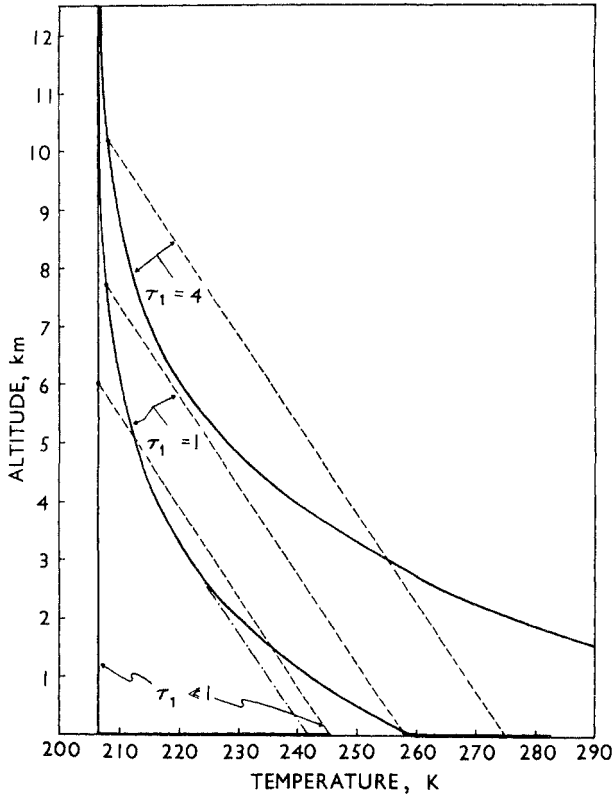


FIG. 9.12. Temperature profiles for radiative and radiative–convective states. The solid lines are radiative equilibrium profiles while the broken lines are the modified profiles after 6.5 K km^{-1} convective lapse-rate adjustment. The chain line will be discussed in the text. The heavy lines on the abscissa show the surface temperature discontinuity, $\Delta\theta(0)$.

in the radiative region does not depend at all upon the region below it. The solution, (9.5), follows from the upper boundary condition alone; the lower boundary condition is only needed to determine the surface conditions. This behavior is not followed by nongrey models but it simplifies the present discussion.

Consider a point, located in the radiative region, but separated from the convective region by an infinitesimally small distance. The thermal heat flux in the radiative region is equal to $-F_S$. The downward component of the thermal heat flux is unchanged by events in the convective region because the conditions in the radiative region are unchanged. Consequently, the convective region is required to deliver to the radiative region the same upward flux component as did the atmosphere that it replaces.

Table 9.1. Numerical data for Fig. 9.12[a,b]

τ_1	$\theta(0)$ (K)	$-\Delta\theta(0)$ (K)	$-\left(\frac{\partial\theta}{\partial z}\right)_0$ (K km^{-1})	$z_{6.5}$ (km)	$\tau(s)$	$z(s)$ (km)	$\theta(s)$ (K)	$\theta'(0)$ (K)
$\ll 1$	206.6	39.1	0.0	0.0	0	6.02	206.6	245.7
0.5	237.5	28.4	12.0	1.8	0.0159	6.90	207.2	252.0
1.0	259.7	22.8	19.5	3.2	0.0212	7.70	208.2	258.2
2.0	292.1	16.7	27.4	4.6	0.0239	8.87	208.3	265.9
4.0	335.9	11.3	36.0	6.0	0.0243	10.20	208.4	274.7

[a] Definitions:

$\theta(0)$ = atmospheric temperature at $z = 0$
$-\Delta\theta(0)$ = temperature discontinuity at $z = 0$
$-\left(\frac{\partial\theta}{\partial z}\right)_0$ = temperature lapse rate at $z = 0$
$z_{6.5}$ = height at which lapse rate = 6.5 K km^{-1}
} Under radiative equilibrium

$\tau(s)$ = optical depth of radiative layer
$z(s)$ = height of radiative–convective transition
$\theta(s)$ = temperature at transition
$\theta'(0)$ = temperature at ground level
} Under radiative–convective equilibrium

For $\tau_1 \gg 4$

$$z_s = -2 \ln \frac{0.0245}{\tau_1} \text{ km},$$

$\theta(s) = 208.5$ K,
$\theta'(0) = 208.5 + 6.5z(s)$ K.

[b] $F_S/2\sigma = 1.82 \times 10^9$ K^4; absorption scale height = 2 km. The solar flux used in this table and in Fig. 9.12 is 17% lower than the value adopted elsewhere in this book. The higher value would give $\theta(0) = 214.8$ K and $-\Delta\theta(0) = 40.7$ K in the first row of the table.

In Fig. 9.12, the profile for $\tau_1 = 1$ is unstable below 3.2 km because the lapse rate exceeds 6.5 K km^{-1}. The chain line represents a hypothetical atmosphere in which the lapse rate has been adjusted to 6.5 K km^{-1} below 3.2 km without a temperature discontinuity at the transition between the radiative and the convective regions. This profile does not satisfy the upward flux requirement. The temperature at every level in the convective region is lower than that at a corresponding level in the radiative region that it has replaced. It, therefore, delivers too small an upward flux component at the transition. The first effect of this will be cooling in the radiative region and the development of a negative temperature discontinuity at the transition.

This negative discontinuity is itself unstable. Convection will start above the transition and the top of the convective layer will rise. It must continue to rise until it can deliver the correct upward flux without an unstable, negative temperature discontinuity at the transition. This condition has been applied numerically to establish the broken curve marked $\tau_1 = 1$. This is the solution to the problem as posed. Solutions for $\tau_1 = 4$ and for $\tau_1 \ll 1$ are also given in Fig. 9.12 and Table 9.1.

The solution for $\tau_1 = 4$ is remarkably close to the average conditions

in the terrestrial atmosphere. The calculated surface temperature is 274.7 K, greatly reduced from the radiative equilibrium value of 335.9 K. Convection has carried heat up into regions where it can more easily radiate to space. More efficient convective processes would further reduce the ground temperature.

Note that a convective layer is always required in Emden's treatment. Even for a transparent atmosphere, a convective layer must extend to 6.02 km. In the limit of zero opacity, however, the convective heat flux is zero and the solution is trivial.

Nongrey radiative–convective calculations have been performed numerically on many occasions. Different numerical procedures are used to converge on a solution for the convective layer, but the essential requirements are the same as for the grey model: the convective region must deliver the required upward flux component to the radiative region, and there must be no discontinuity of temperature at the transition. The temperatures in the radiative region are now no longer independent of events in the convective region, but that creates no difficulties for an iterative numerical procedure.

Early nongrey calculations are shown in Fig. 9.13. The agreement with the U.S. Standard Atmosphere is remarkable. The surface temperature for the more realistic, cloudy model is 286.9 K and the tropopause height is 13 km. Although these two figures are close to those for a grey model with $\tau_1 = 4$, the transition region differs from the grey models in being gradual rather than sharp. This is not obvious in Fig. 9.13 because of the wide spacing of the data levels. It is much clearer in Fig. 9.14 in which three methods of convective parameterization are compared (see § 9.4.3 for discussion). Although no sharp feature occurs on these profiles there is, in fact, a level of transition between the convective and the radiative regions. For the $6.5\,\mathrm{K\,km^{-1}}$ adjustment model this level is 13.93 km; for the cumulus convection model it is 9.40 km. Above the transition level, the lapse rate starts to decrease with height, but so gradually that the change of gradient is hardly noticeable.

9.4.3. *Convective models*

The results given by radiative–convective models depend, as might be expected, on the convective model employed. The $6.5\,\mathrm{K\,km^{-1}}$ lapse-rate adjustment, as used in previous sections, may be regarded as the simplest possible convective model. There are, however, theories of convection that give the convective flux explicitly in terms of macroscopic variables; combined with the radiative transfer equation, these theories can lead to complete numerical solutions for the radiative–convective system. In this section we briefly discuss the influence of the convective model. Although such discussion is instructive, its value is uncertain because the convective models are of debatable value.

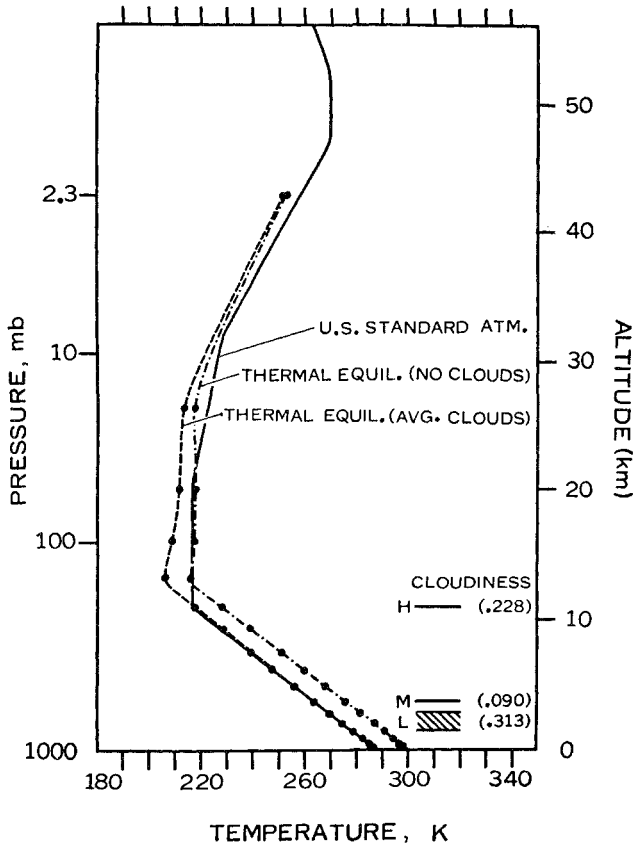


FIG. 9.13. Radiative–convective profiles in the earth's atmosphere. The lapse rate for convective adjustment is 6.5 K km^{-1}. The chain line is without clouds and a net solar flux of 299 W m^{-2}. The broken line is for the cloud layers indicated at the bottom right (cloud amounts in parentheses) and a net solar flux of 228 W m^{-2}. The solar zenith angle is taken to be 60°. The U.S. Standard Atmosphere represents a global average of observed temperatures. After Manabe and Strickler (1964).

We distinguish between convection in the free atmosphere and boundary layer convection. The latter refers to the lowest 10–100 m of the atmosphere, where the presence of the surface strongly influences the mixing processes. First, we consider the free atmosphere. This has typically been treated by means of *mixing-length theories,* for which the convective heat flux may be expressed by

$$F_C = K_H \rho c_p (\partial\theta/\partial z + \Gamma). \tag{9.17}$$

K_H is a heat diffusivity and Γ is the neutral lapse rate for marginal stability. A consistent theory would use either the dry adiabatic lapse rate

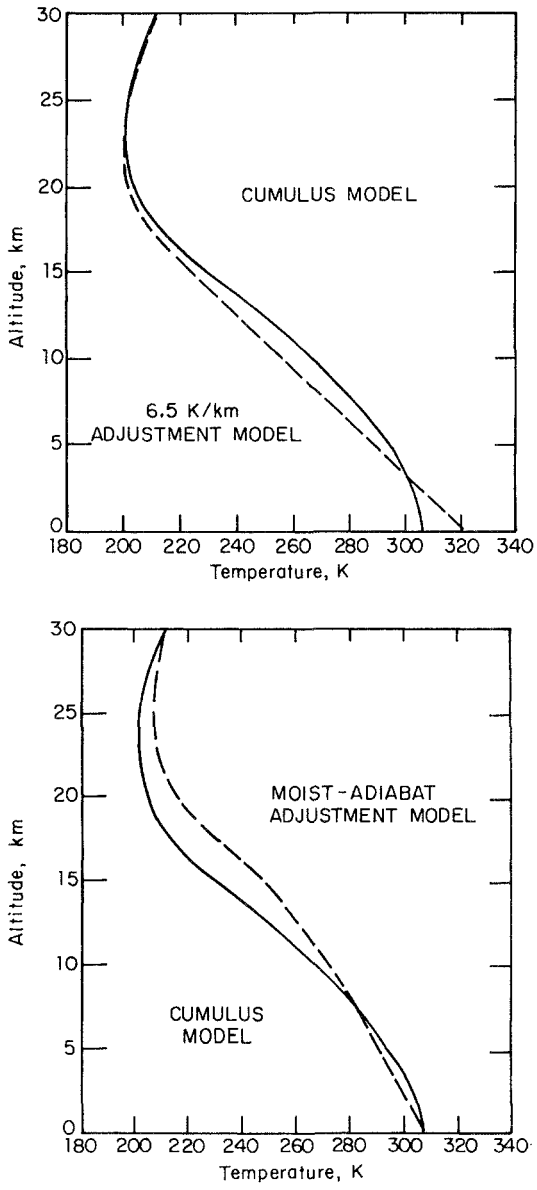


FIG. 9.14. Radiative–convective states compared for three convective models. The conditions are tropical, with a net solar flux equal to 353 W m^{-2}. The three convective models are discussed in § 9.4.3. The cumulus profile is the same in the two panels. After Lindzen et al. (1982).

($\Gamma_d = 9.8\ \mathrm{K\ km^{-1}}$) or the moist adiabatic lapse rate (Γ_m), depending upon whether condensation occurs. For low temperatures and low water vapor densities, the moist adiabatic lapse rate tends to the dry adiabatic lapse rate but, for surface conditions in the tropics, it can be as small as $3\ \mathrm{K\ km^{-1}}$.

The diffusivity K_H is central to mixing-length theories. It has been given in terms of thermal parameters alone (*free convection*), mechanical parameters alone (*shear flow turbulence*), or both together. The instructive feature of such theories is that K_H is usually very large in the sense that, for any reasonable flux of heat, (9.17) gives $\partial\theta/\partial z \sim -\Gamma$. This is the justification for the lapse-rate adjustment models.

Of greater current interest to theoretical meteorology are attempts to base theories of convection on macroscopic phenomena, most importantly on *cumulus convection.* Convection is assumed to take place in narrow buoyant columns that entrain ambient air until the warm air has lost its buoyancy. Ascent ceases at the *detrainment level.* Below that level, convective heating, according to one theory (there are several), is given by

$$h_C = c_p M_C(\partial\theta/\partial z + \Gamma_d), \tag{9.18}$$

where M_C is the vertical mass flux of air in the cumulus columns.

In the context of numerical calculations using 1 km vertical grids, the boundary layer is a discontinuity. We may approach boundary layer theory with somewhat more confidence than that of the free atmosphere because a large body of measurements exists for this accessible region of the atmosphere, and predictions can be tested. Depending upon the stability of the boundary layer, it may be treated as a phenomenon in shear flow or as free convection or as a combination of the two.

A variety of free atmosphere and boundary layer models has been used in radiative–convective theories. A comparison between a cumulus convection model, a $6.5\ \mathrm{K\ km^{-1}}$ convective adjustment model, and a moist adiabatic adjustment model is given in Fig. 9.14. The cumulus model and the moist adiabatic adjustment model are both influenced by the water vapor density and have smaller lapse rates at low levels than does the $6.5\ \mathrm{K\ km^{-1}}$ adjustment model. This has the effect of decreasing the sensitivity of radiative–convective models to externally imposed changes, such as changes in the incident solar radiation or changes in gaseous densities, for the reason illustrated in Fig. 9.15. Let us assume that absorption is strong and that radiation lost to space comes from a restricted layer or emission level [see the discussion of (6.66)]. If the solar radiation increases, the atmosphere must respond by increasing the temperature of the emission layer from θ_e to $\theta_e + \Delta\theta$. For a convective layer with a fixed lapse rate, the surface temperature must change by the

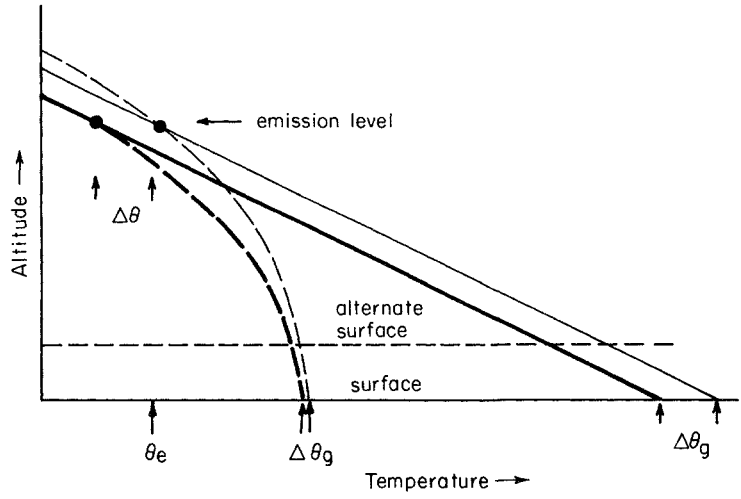


FIG. 9.15. Illustration of different model sensitivities. The "alternate surface" is a device for changing the amount of absorber while maintaining the emission to space.

same amount. For a moist adiabatic atmosphere, on the other hand, the surface temperature change may be much less than $\Delta\theta$.

A modern paradigm for climate models is the effect of doubling atmospheric carbon dioxide. Figure 9.15 illustrates a way to hold the emission constant while changing the amount of gas in the atmosphere, namely by moving the surface relative to the emission level. Again, the effect upon surface temperatures is greater for the dry adiabatic than for the moist adiabatic model.

9.4.4. *Nonlocal dissipation*

Apart from a single study of the martian atmosphere, all radiative–convective models involve the implicit assumption that free convection always carries heat upward. This statement is correct only if convective motions are generated and dissipated in the same region. Nonlocal dissipation has, however, been shown to be important for other meteorological phenomena and may have relevance to theories of the tropopause. Convective velocities may be large enough for their inertia to carry hot air from the convective region into the overlying radiative region (*convective overshoot*); the convective motions must now be destroyed by working on the stable region, where they will cause a downward convective heat flux.

An approximate treatment of this problem has been made by assuming the complete absence of viscous dissipation. The only external

heating is radiative. Motions carry enthalpy and entropy from one part of the system to another and, integrated over the whole atmosphere, their contributions vanish. The enthalpy and entropy conservation relations are

$$\frac{\partial}{\partial t}\int_0^{p_0} \rho c_p \theta \, dp = \int_0^{p_0} h_R \, dp, \tag{9.19}$$

$$\frac{\partial}{\partial t}\int_0^{p_0} \rho s \, dp = \int_0^{p_0} \frac{h_R}{\theta} \, dp, \tag{9.20}$$

where s is the specific entropy and p_0 is the surface pressure.

In the radiative region there is no transport, and both enthalpy and entropy are locally balanced; consequently the lower limits in (9.19) and (9.20) can both be replaced by the pressure at the radiative–convective boundary. If we now assume that the lapse rate in the convective region is known, (9.19) and (9.20) provide two relations from which the average temperature of the convective region and its vertical extent can be calculated.

A calculation for the martian atmosphere, based upon these ideas, is shown in Fig. 9.16. The low-dissipation case is that discussed here; the high-dissipation case is the dry adiabatic lapse-rate adjustment model discussed in previous sections. Comparison of the two cases is complicated by the large time dependence of temperature in the martian atmosphere. However, the effect of convective overshoot in the region of the radiative–convective transition is apparent.

9.4.5. *Semiconvection*

Clouds are an important factor for climate because they strongly affect both thermal and solar radiation. Here we look at one unusual aspect of cloud models, namely the nature of the transition region between convective and radiative regions when clouds are present.

The condensed phase, whether it be water or ice, appears initially as micron-sized particles. Before precipitation commences, cloud particles fall very slowly and even weak convection will bring about an approximation to a constant mixing ratio of aerosol to air. This implies a discontinuity of aerosol properties at the top of the convective region; the sharp tops observed for many cloud types support this idea. Since aerosol particles absorb solar radiation, this model has a discontinuity of solar heating at the transition, with more heating in the convective region.

When we discussed the steady states at the tropopause in Fig. 9.12, there was an implicit assumption that all heating mechanisms were continuous at the transition between the radiative and convective regions. If we introduce a discontinuity into the solar heating, the earlier

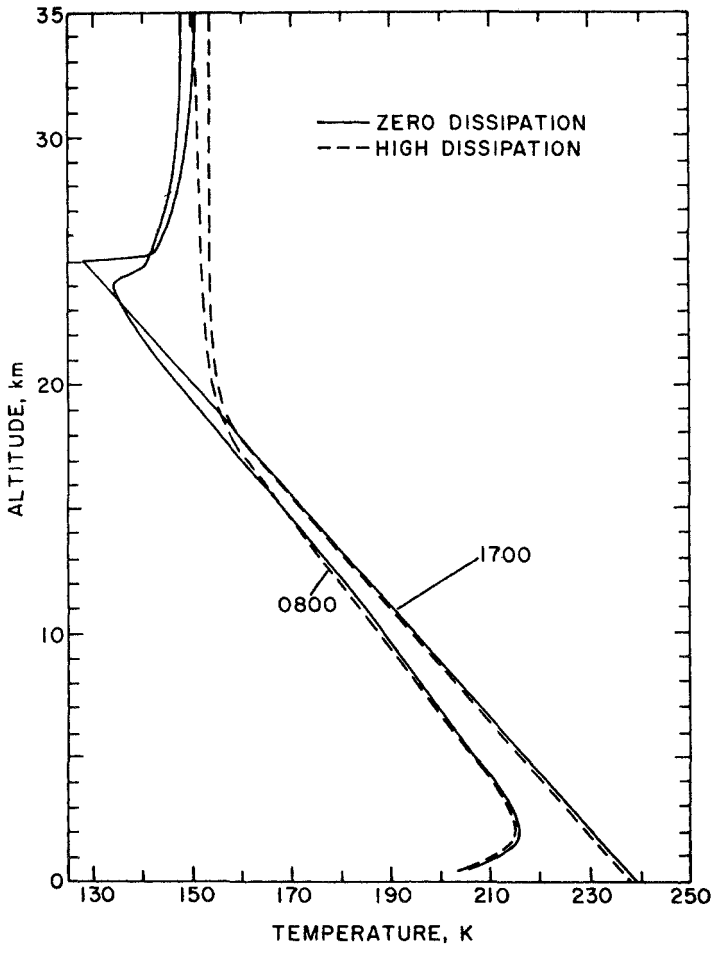


FIG. 9.16. Diurnal extreme radiative–convective profiles for Mars for both high and low dissipation. Conditions appropriate to the equatorial equinox have been chosen. The 1700 h profile has a dry adiabatic lapse rate in the convective region. The 0800 h profile is not yet unstable. After Gierasch (1971).

discussion is no longer correct. To satisfy the radiative and convective equations a temperature discontinuity is now required at the transition, with the lower temperature in the stratosphere. We have already remarked, however, that this temperature configuration is hydrostatically unstable; consequently, we are led to the conclusion that no steady-state solution exists to the problem, as posed.

A solution to this paradox is illustrated in Fig. 9.17. In place of a surface of transition between convective and radiative regions there is now an extended transition region. This region has an adiabatic lapse

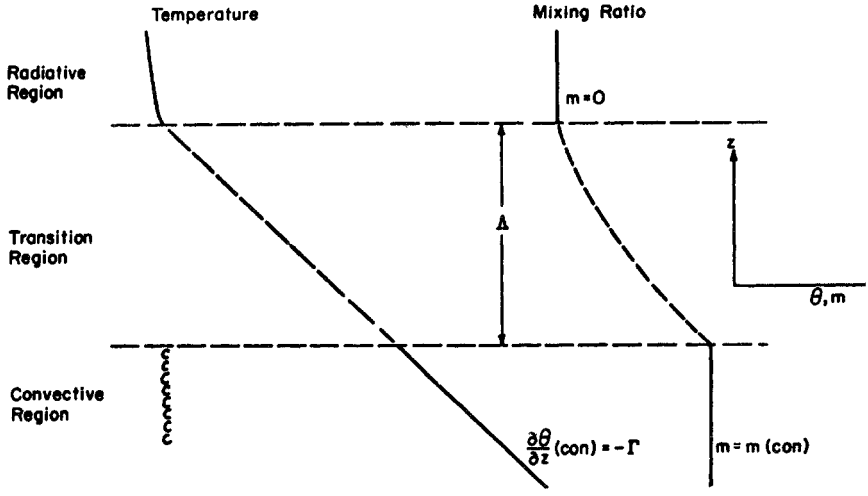


FIG. 9.17. The transition region at a cloud top. After Gierasch and Goody (1970).

rate, but is in radiative equilibrium, maintained by variable aerosol mixing ratio and, consequently, a variable solar heating. An important point is that the convection required to maintain this distribution of micron-sized particles is too small to influence the heat balance. This state of *semiconvection* can be realized in approximation.

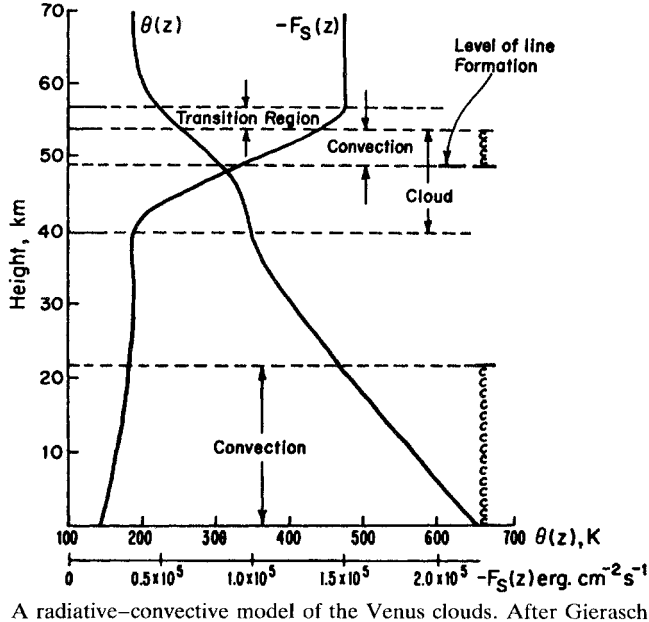


FIG. 9.18. A radiative–convective model of the Venus clouds. After Gierasch and Goody (1970).

Figure 9.18 shows a calculation for the Venus clouds. The model is time dependent and is based upon constraints imposed by spectroscopic observations. Two convective regions occur and, in this model, clouds, can exist in a nonconvective region.

9.5. The runaway greenhouse

9.5.1. *History of ideas*

The vapor pressure of water is observed to be, on the average, one-half of its saturated value throughout the troposphere (a constant relative humidity of 50%). If we accept this statement as characteristic of the earth's atmosphere, we are assuming a coupling between temperatures and vapor concentrations that can lead to an instability known as the *runaway greenhouse* effect.

A number of numerical investigations have been concerned with the change in ground temperature accompanying imposed perturbations, such as the doubling of the carbon dioxide concentration. The model sensitivity depends upon the behavior of the water vapor. Fixed relative humidity models (i.e., models in which temperatures and humidities are coupled) have been reported to be 2 to 10 times more sensitive than those with fixed vapor density. This suggests the possibility of a runaway greenhouse effect but outside the range of parameters normally encountered on earth.

The ideas behind the runaway greenhouse effect have been put forward independently on several occasions. The first was connected with an early criticism of Emden's theory of the troposphere. Hergesell (1919) pointed out that Emden's temperature profiles were supersaturated under the assumed conditions (for $\tau_1 = 1$ in Fig. 9.12 the temperature is very low between 1 and 6 km and the atmosphere could support very little water vapor). Hergesell performed an iterative calculation in which water vapor was limited by its saturated vapor pressure, and concluded that the self-consistent solution was a cold, dry atmosphere, similar to that for $\tau_1 \ll 1$ in Fig. 9.12. Hergesell's calculation contained numerical errors, but he was correct in pointing to the great importance of the feedback between water vapor density and temperature.

The next contribution to this discussion was part of a series of important ideas about atmospheric radiation put forward by G. C. Simpson. In 1927, Simpson pointed to a paradox: an atmosphere dominated by water vapor could not emit to space a flux greater than certain critical value. If the incident solar flux were to exceed this value, the atmosphere could not respond. For earth, Simpson concluded that the critical flux is exceeded by the solar flux in tropical regions. He decided that his radiative model must be in error and he developed a

more satisfactory nongrey model that did not exhibit this unacceptable behavior. The paradox remains, however, but not for conditions close to those presently occurring on earth.

For the next 30 years, the runaway greenhouse received little attention. In 1964, T. Gold raised the possibility that a runaway process has occurred on Venus and is responsible for the high surface temperatures on that planet, but he gave no details.

The physical state of the atmosphere of Venus also stimulated the enquiries of Ingersoll (1969) and Rasool and de Bergh (1970). Ingersoll saw the problem, as did Simpson, in terms of a limit to the outgoing radiation flux in the existing atmosphere, while Rasool and de Bergh employed the context of planetary evolution. We shall briefly describe the latter approach.

Theories of planetary evolution envisage the formation of the atmospheres of the inner planets by slow emission of gases from the crust by tectonic processes. The emergent gases are those found in the present-day atmosphere, or loosely bound to the planet in the oceans and in crustal carbonaceous rocks, principally water and carbon dioxide. Rasool and de Bergh studied the successive equilibrium states in an atmosphere created slowly by tectonic emission. The amounts of water vapor and carbon dioxide are limited not only by the amount emitted but also by surface interactions, the water by condensation or sublimation and the carbon dioxide by chemical reactions with the surface rocks. Both interactions are characterized by an equilibrium pressure that is a function of temperature only; if, for a given temperature, the equilibrium vapor pressure is exceeded, the gas concerned, whether it be carbon dioxide or water vapor, will be incorporated into the surface and will not pass into the atmosphere.

In the beginning, the atmosphere was cold ($\tau_1 \ll 1$ in Fig. 9.12) but the pressure was low and water vapor, at first, resided solely in the atmosphere. As τ_1 approaches unity for water vapor, the greenhouse effect starts to raise the surface temperature. The important question is whether this occurs before or after the vapor pressure has risen to the equilibrium value for a condensed phase. In the former case, gases will continue to enter the atmosphere and the temperature will continue to increase; in the latter case, the additional tectonic gases will pass into the crust and the surface temperature will remain constant. The earth is now in the latter state; almost all of the water vapor and carbon dioxide reside in the crust rather than in the atmosphere. The interesting question is what changes might lead to the alternate condition. According to Rasool and de Bergh the important difference between earth and Venus is that earth is further from the sun and had a lower surface temperature at the beginning of the evolutionary process. In the following sections we examine the theoretical basis for these ideas.

9.5.2. *Simpson's paradox*

The outgoing radiation from a stratified atmosphere is given by (6.11) with $z = \infty$ or, alternately, by the integral of the "radiation-to-space" term in the heating rate, (6.66). In connection with the derivation of (6.66), it was pointed out that, for monochromatic (or grey) absorption and for a constant source function, the integrand corresponds to the Chapman layer (Fig. 6.9). The scale height for water vapor in the troposphere is approximately 2 km and the full width of the Chapman layer is, therefore, about 4 km; for a layer of this thickness it is not unreasonable to replace the Planck function by its value, $B(z_{max})$, at the level of the maximum of the Chapman layer. The integral in (6.11) can now be performed ($0 \leq T_\nu^f \leq 1$) to give

$$\frac{F_\nu(z=\infty)}{\pi} = B_\nu(z_{max}). \tag{9.21}$$

z_{max} is the level, (6.54), for which the slant optical path is unity. For diffuse radiation, we may use the diffusivity approximation (§ 6.12)

$$\rho(z_{max}) = (k_m r H)^{-1}, \tag{9.22}$$

where k_m is the mass absorption coefficient, r is the diffusivity factor ($\simeq 1.66$), H is the scale height, and ρ is the absorber density.

The condition for fixed relative humidity can be written

$$\rho(z_{max}) = \eta \rho_{sat}[\theta(z_{max})], \tag{9.23}$$

where ρ_{sat} is the saturated vapor density and η is the relative humidity. Since the saturated vapor density varies very rapidly with temperature, it is possible to treat the product $\eta k_m r H$ as approximately constant, so that (9.22) and (9.23) approximately determine the temperature $\theta(z_{max})$ at the peak of the Chapman function. In which event, the emitted flux (9.21) is also determined.

Simpson, and others of his time, used a grey atmosphere that absorbed 90% of incident thermal radiation. Combined with other available information this led to a maximum emission temperature of about 250 K, slightly less than that obtained in (1.1) and (1.2) for the global average insolation. Within the uncertainties of this calculation, it appears to be possible to balance the incoming solar radiation averaged over a cloudy earth, but not in the tropics and not for a cloud-free earth or (even less) for a cloud-free Venus.

If we reexamine Simpson's paradox for a nongrey atmosphere it still has substance. The distribution with height of the outgoing thermal

radiation from a strong band has a layered character, similar to but broader than a Chapman layer. If atmospheric emission were entirely from strong water vapor bands Simpson's paradox would still hold. The only new factor would be the pressure broadening of spectral lines, which modifies but does not destroy the above argument. As we see from the recorded spectrum of outgoing radiation in Fig. 6.1, however, most radiation comes from the water vapor window, modified by emission in the ν_2 band of carbon dioxide and the ν_3 band of ozone. This window is translucent and most of the outgoing radiation originates at the ground (or cloud tops). This is not the condition for Simpson's paradox to occur. The numerical models discussed earlier establish that there is still some positive feedback between temperatures and water vapor densities but not with the cataclysmic character that Simpson identified (and rejected). However, the situation envisaged by Simpson should arise if the water vapor window were to become opaque. A minimum mass absorption coefficient in the window for ground-level conditions is about $1\ \mathrm{g^{-1}\ cm^2}$. For $\eta = 0.5$, $r = 1.66$, and $H_{\mathrm{air}}/H_{\mathrm{water}} = 4$, (9.22) and (9.23) lead to a saturation vapor pressure of 78 mb and a saturation temperature of 314 K. Equation (9.21) then gives $551\ \mathrm{W\ m^{-2}}$ for the outgoing radiation, about twice the global average for earth. The surface temperature for this condition would be higher than 314 K by an amount depending upon the assumed tropospheric lapse rate; clearly the surface would be very hot indeed.

9.5.3. *An evolving atmosphere*

Simpson was anxious to avoid a paradoxical result, while Ingersoll and Rasool and de Bergh were looking for such a result to account for the differences between the atmospheres of earth and Venus; both found what they looked for.

In the following discussion we use τ_1 as an independent variable. It is related to the surface temperature in two different ways. For a saturated atmosphere with a constant scale height, H,

$$\tau_1 = \eta k_{\mathrm{m}} H \rho_{\mathrm{sat}}(\theta_{\mathrm{g}}). \tag{9.24}$$

If all of the other parameters can be treated as constants, (9.24) defines a relationship between θ_{g} and τ_1. Since vapor pressure varies approximately exponentially with temperature, this relation gives close to a straight line on a plot of θ_{g} against $\ln \tau_1$ (see the broken line in Fig. 9.19). This line is a vapor pressure equilibrium curve. Above the line only the vapor can exist; below, only the condensed phase. The two phases are in equilibrium for points on the line itself.

The second relationship between τ_1 and θ_{g} is the radiative equilibrium condition, (9.5). This relationship is plotted as solid lines in Fig.

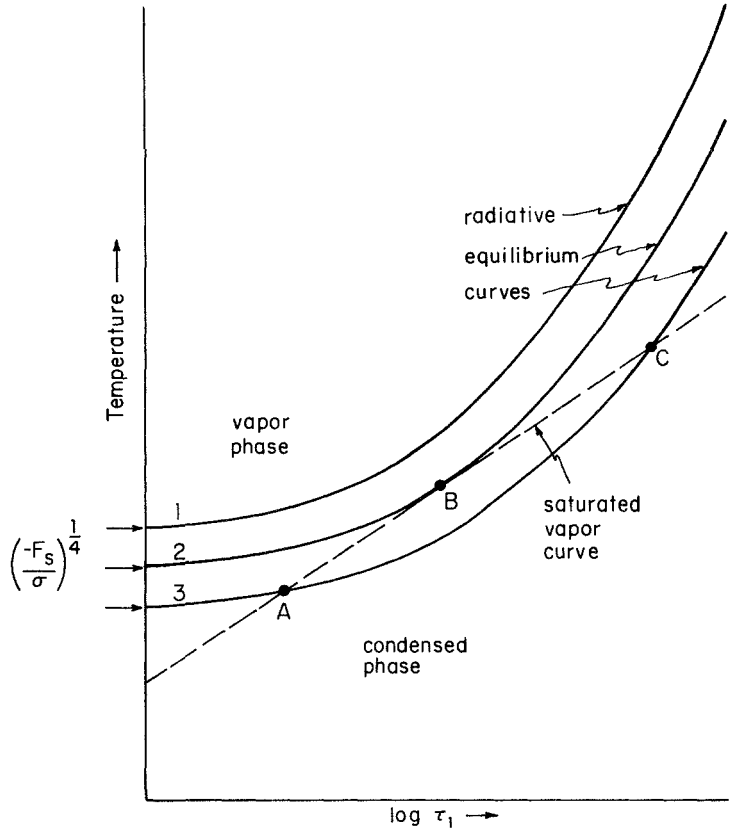


FIG. 9.19. The runaway greenhouse effect. The broken line represents (9.24). The solid lines are from (9.5) with differing amounts of incident solar radiation. Points A, B, and C are possible solutions that are discussed in the text.

9.19 for a series of values of the initial temperature,

$$\theta_g = \left(\frac{-F_S}{\sigma}\right)^{1/4} \qquad \text{for } \tau_1 = 0. \tag{9.25}$$

Where the broken and solid lines intersect conditions are defined under which a saturated atmosphere can be in radiative equilibrium; the amount stored in the surface is undetermined.

When the radiative equilibrium curve lies above the phase curve, on the other hand, all the absorbing gas is in the vapor phase. Radiative equilibrium solutions can exist at any point on the curve.

From an evolutionary standpoint, the abscissa may also be interpreted as a nonlinear time scale. In the beginning, the atmosphere was

unsaturated and evolution developed along the radiative equilibrium curves. Curve 1 represents an atmosphere for which none of the tectonic emission is ever returned to the crust. The temperature can increase indefinitely with time. On curve 3, point A gives the time at which surface incorporation commences, for that particular value of the solar constant. When A is reached, the temperature and saturation vapor pressure no longer change with time, and further gaseous emission goes directly into the surface.

Solution C is inaccessible by the evolutionary route and is, moreover, unstable. This can be seen by considering a series of solutions for different solar constants connecting the points B and C. They represent a thermodynamically unacceptable relationship between increasing solar radiation and decreasing surface temperature. The temperature corresponding to the tangent point, B, is the highest for which a stable solution, involving a condensed phase, exists.

9.5.4. *Influence of the tropospheric lapse rate*

Figure 9.20 shows the results of nongrey calculations for two convective models of the earth's atmosphere, with variable insolation. The 6.5 K km^{-1} adjustment model shows an increase of ground temperature with increase of insolation, consistent with a runaway condition. The

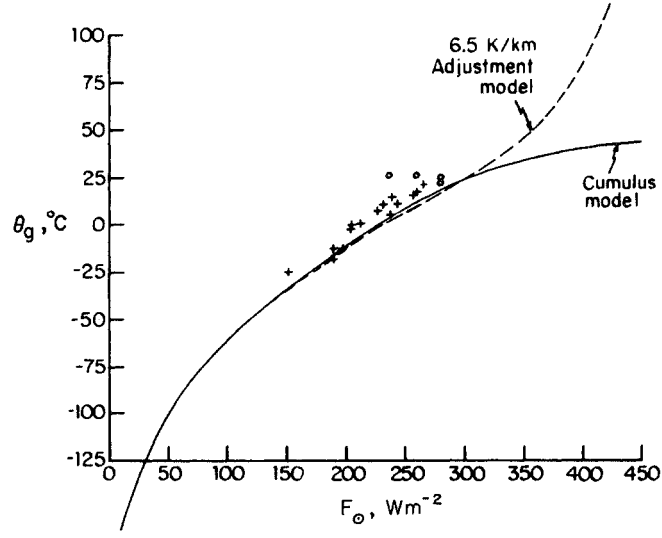


FIG. 9.20. Surface temperature as a function of solar flux. The models employed are the same as those used to derive the profiles shown in Fig. 9.14. The circles and the crosses are terrestrial climatological data, illustrating both the validity of the model and the relationship of conditions on earth to those that could lead to a runaway condition. After Lindzen et al. (1982).

cumulus convection model (discussed in § 9.4.3), on the other hand, does not appear to exhibit a runaway behavior.

Earlier discussion of cumulus convection and moist adiabatic models led to an expectation of smaller sensitivity to external change for models with smaller tropospheric lapse rates. In the limit of an isothermal convective region (unknown but imaginable) the surface temperature would equal the effective emission temperature and would increase with the fourth root of the insolation. Such an atmosphere cannot escape the Simpson paradox, however, the existence of which did not depend on the tropospheric lapse rate. If the solar flux exceeds the Simpson limit, the model breaks down because it contains no compensating sink to take up the excess radiation.

BIBLIOGRAPHY

9.1. Introduction

Radiative equilibrium and radiative–convective states were first studied as part of the field of stellar structure. A classic in this area is

Chandrasekhar, S., 1939, *An introduction to the study of stellar structure.* Chicago: University of Chicago Press.

The transfer of these ideas to atmospheric studies was reviewed by

Pekeris, C. L., 1932, *The development and the present status of the heat balance of the atmosphere.* MIT Professional Note, No. 5, Cambridge, Mass: Massachusetts Institute of Technology.

Many numerical studies of the stratosphere and mesosphere have indicated that the net radiative heating is considerably smaller than either the solar heating or the cooling by thermal radiation and, in this limited sense, is close to radiative equilibrium.

The radiative–photochemical equilibrium calculations shown in Fig. 9.1 are from

Fels, S. B., 1985, "Radiative-dynamical interactions in the middle atmosphere," *Adv. Geophys.* **28A,** 277.

This paper contains a review of general circulation model calculations.

The radiative balance of the mesosphere has recently been studied by

Gille, J. C., and Lyjak, L. V., 1986, "Radiative heating and cooling rates in the middle atmosphere," *J. Atmos. Sci.* **43,** 2215.

The stratospheric calculations in Fig. 9.2 are from

Kiehl, J. T., and Solomon, S., 1986, "On the radiative balance of the stratosphere," *J. Atmos. Sci.* **43,** 1525.

The following is a selection of earlier work on the radiative balance of the

middle atmosphere:

Ohring, G., 1958, "The radiative budget of the stratosphere," *J. Meteorol.* **15,** 440.

Leovy, C., 1964, "Radiative equilibrium of the mesosphere," *J. Atmos. Sci.* **21,** 238.

Hering, W. S., Touart, C. N., and Borden, T. R., 1967, "Ozone heating and radiative equilibrium in the lower stratosphere," *J. Atmos. Sci.* **24,** 402.

The radiative–convective state of the martian atmosphere was studied by

Gierasch, P., and Goody, R., 1968, "A study of the thermal and dynamical structure of the martian lower atmosphere," *Planet. Space Sci.* **16,** 615.

When *in situ* measurements were made on Mars, the temperatures measured were close to these predictions.

9.2. An elementary solution

The solution, (9.13), was given by

Gierasch, P., and Goody, R. M., 1970, "Models of the Venus clouds," *J. Atmos. Sci.* **27,** 224.

The nuclear winter is discussed by

Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B., and Sagan, C., 1983, "Nuclear winter: Global consequences of multiple nuclear explosions," *Science* **222,** 1283.

9.3. Nongrey atmospheres

9.3.1. *Models without pressure broadening*

The treatment of the k-distribution discussion follows,

Arking, A., and Grossman, K., 1972, "The influence of line shape and band structure on temperature in planetary atmospheres," *J. Atmos. Sci.* **29,** 937.

An earlier analytic discussion of radiative equilibrium in an atmosphere absorbing in an Elsasser band reached conclusions similar to Arking and Grossman:

King, J. I. F., 1955a, "The Gaussian quadrature formula for an Elsasser band-absorption model," *Astrophys. J.* **121,** 425.

———, 1955b, "Radiative equilibrium of a line-absorbing atmosphere. I," *Astrophys. J.* **121,** 711.

———, 1956, "Radiative equilibrium of a line-absorbing atmosphere. II," *Astrophys. J.* **124,** 272.

9.3.2. *Pressure broadening*

Figure 9.9 is from

King, J. I., 1952, "Line absorption and radiative equilibrium," *J. Meteorol.* **9,** 311.

9.3.3. *Numerical methods*
A review, containing many references is by

Ramanathan, V., and Coakley, J. A., 1978, "Climate modeling through radiative-convective models," *Rev. Geophys. Space Phys.* **16,** 465.

The early results presented in Figs. 9.10 and 9.11 are by

Manabe, S., and Möller, F., 1961, "On radiative equilibrium and heat balance of the atmosphere," *Mon. Weath. Rev.* **89,** 503.

Möller, F., and Manabe, S., 1961, "Über das Strahlungsgleichgewicht der Atmosphäre," *Z. Meteorol.* **15,** 3.

9.4. The troposphere and the stratosphere

9.4.1. *Introduction*
The early work is reviewed by Pekeris (1932), § 9.1. Specific references are

Emden, R., 1913, "Über Strahlungsgleichgewicht und atmosphärische Strahlung," *Sitz. Bayerische Akad. Wiss.*, Math.-Phys. Klasse, 55.

Humphreys, W. J., 1909, "Vertical temperature gradient of the atmosphere in the region of the upper inversion," *Astrophys. J.* **29,** 14.

Gold, E., 1909, "The isothermal layer of the atmosphere and atmospheric radiation," *Proc. Roy. Soc. London Ser. A* **82,** 43.

Milne, E. A., 1922, "Radiative equilibrium, the insolation of the atmosphere," *Phil. Mag.* **144,** 872.

9.4.2. *The troposphere and the stratosphere*
The temperature profiles shown in Fig. 9.13 are from

Manabe, S., and Strickler, R. F., 1964, "Thermal equilibrium of the atmosphere with a convective adjustment," *J. Atmos. Sci.* **21,** 361.

Other important calculations are discussed by Ramanathan and Coakley (1978), § 9.3.3. Some of these are

Manabe, S., and Wetherald, R. T., 1967, "Thermal equilibrium of the atmosphere with a given distribution of relative humidity," *J. Atmos. Sci.* **24,** 241.

Coakley, J. A., Jr., 1977, "An efficient approach to radiative-convective equilibrium," *J. Atmos. Sci.* **34,** 1402.

Ramanathan, V., 1976, "Radiative transfer within the earth's troposphere and stratosphere: A simplified radiative-convective model," *J. Atmos. Sci.* **33,** 1330.

Figure 9.14 is from

Lindzen, R. S., Hou, A. Y., and Farrell, B. F., 1982, "The role of convective model choice in calculating the climate impact of doubling CO_2," *J. Atmos. Sci.* **39,** 1189.

Some fundamental ideas concerning the thermal equilibrium at the tropopause are to be found in

Goody, R. M., 1949, "The thermal equilibrium at the tropopause and the temperature of the lower stratosphere," *Proc. Roy. Soc. London Ser. A* **197,** 487.

Dobson, G. M. B., Brewer, A. W., and Cwilong, B. M., 1946, "Meteorology of the lower stratosphere," *Proc. Roy. Soc. London Ser. A* **185,** 144.

9.4.3. *Convective models*
The effect upon climate calculations of different neutral gradients in the convective region is discussed by

Chylek, P., and Kiehl, J. T., 1981, "Sensitivities of radiative-convective models," *J. Atmos. Sci.* **38,** 1105.

A shear-flow convection model has been investigated by

Liou, K.-N., and Ou, S.-C. 1983, "Theory of equilibrium temperatures in radiative-turbulent atmospheres," *J. Atmos. Sci.* **40,** 214.

9.4.4. *Nonlocal dissipation*
An example of the treatment of a meteorological problem involving nonlocal dissipation is

Lilly, D. K., 1968, "Models of cloud-topped mixed layers under a strong inversion," *Quart. J. R. Meteorol. Soc.* **94,** 292.

The only application of these ideas to radiative-convective models is by

Gierasch, P. J. 1971, "Dissipation in atmospheres: The thermal structure of the martian lower atmosphere with and without viscous dissipation," *J. Atmos. Sci.* **28,** 315.

9.4.5. *Semiconvection*
An example of a numerical calculation for an evolving cloud is

Herman, G., and Goody, R., 1976, "Formation and persistence of summertime arctic stratus clouds," *J. Atmos. Sci.* **33,** 1537.

Figure 3 in this paper shows the cooling at the cloud top and the consequent rising of the cloud top, as discussed in the text.

The discussion of semiconvection follows Gierasch and Goody (1970), § 9.2. Semiconvection in stars is reviewed by

Spiegel, E. A., 1969, "Semiconvection," *Comm. Astrophys. Space Phys.* **1,** p. 57.

9.5. The runaway greenhouse
The following paper demonstrated the temperature–water vapor feedback in climate models:

Möller, F., 1963, "On the influence of changes in the CO_2 concentration in air on the radiation balance of the earth's surface and on the climate," *J. Geophys. Res.* **68,** 3877.

A more restricted treatment by Manabe and Wetherald (1967), see § 10.4.2., gave a smaller increase. The following paper gives sensitivity increases between 2 and 10:

Ramanathan, V., 1981, "The role of ocean-atmosphere interaction in the CO_2 climate problem," *J. Atmos. Sci.* **38,** 918.

Early work on the runaway greenhouse effect is reviewed by Pekeris (1932), § 9.1. Some original references are

Hergesell, H., 1919, "Die Strahlung der Atmosphäre unter Zugrundelegung vom Lindenbergen Temperatur- und Feuchtigheitsmessungen," *Arb. preuss. aero. Obs.* **13.**

Simpson, G. C., 1927, "Some studies in terrestrial radiation," *Mem. R. Meteorol. Soc.* **II,** #16.

———, 1928, "Further studies of terrestrial radiation," *Mem. R. Meteorol. Soc.* **III,** #21.

———, 1929, "The distribution of terrestrial radiation," *Mem. R. Meteorol. Soc.* **III,** #23.

Later papers on the runaway greenhouse effect are by

Gold, T., 1964, "Outgassing processes on the Moon and Venus," *The origin and evolution of atmospheres and oceans,* (P. J. Brancazio and A. G. W. Cameron, Eds.). New York: John Wiley, p. 249.

Ingersoll, A. P., 1969, "The runaway greenhouse: A history of water on Venus," *J. Atmos. Sci.* **26,** 1191.

Rasool, S. I., and de Bergh, C., 1970, The runaway greenhouse and the accumulation of CO_2 in the Venus atmosphere," *Nature* (*London*) **226,** 1037.

A more recent discussion is by

Matsui, T., and Abe, Y., 1986, "Impact-induced atmospheres and oceans on earth and Venus," *Nature* (*London*) **322,** 526.

Figure 9.20 is from Lindzen et al. (1982), § 9.4.1.

10

EVOLUTION OF A THERMAL DISTURBANCE

10.1. Introduction

The thermodynamic equation for an ideal gas is

$$\rho c_v \frac{d\theta}{dt} = h_R + \frac{dp}{dt}, \tag{10.1}$$

where $\rho c_v \theta$ is the internal energy per unit volume and h_R is the radiative heating rate.[1] For the sake of clarity we omit diabatic terms additional to the radiative heating.

We may expand the left-hand side of (10.1) and write it in the form of an enthalpy equation,

$$\rho c_p \frac{\partial \theta}{\partial t} = h_R + d, \tag{10.2}$$

where

$$d = \frac{\partial p}{\partial t} + \mathbf{v} \cdot \nabla p - \rho c_p \mathbf{v} \cdot \nabla \theta,$$

and c_p is the specific heat at constant pressure. The left-hand side of (10.2) includes both internal and potential energy. d is the dynamic heating. If the vertical coordinate is pressure, $\partial p / \partial t = 0$ and all terms in d tend to zero as the velocities tend to zero.

Solutions to (10.2) in conjunction with the equations of motion, the equation of continuity, and the gas law are the matter of dynamic meteorology. In this chapter, we look at a single aspect, namely the coupling between radiation and dynamics as expressed by the thermodynamic equation, (10.2).

[1] The expressions for h_R developed in Chapter 6 were static. In order to use these expressions in the time-dependent equation (10.1), we must assume the travel time of radiation to be zero and we must neglect Doppler shifts from fluid motions. The former assumption is always justified but there are occasions, e.g., when considering line shapes, when Doppler effects cannot be ignored. For fluid motions much slower than the speed of sound, however, the neglect of Doppler shifts due to atmospheric motions is also justified.

We shall use the methods of perturbation theory. Assume the existence of a basic, steady-state (suffix 0) for which

$$\rho c_p \frac{\partial \theta_0}{\partial t} = h_{R,0} + d_0 = 0. \tag{10.3}$$

This basic state could be a state of radiative equilibrium or a state dominated by dynamic transports. We assume

$$\begin{aligned} \theta &= \theta_0 + \theta', \\ h_R &= h_{R,0} + h'_R, \\ d &= d_0 + d', \end{aligned} \tag{10.4}$$

where $(\theta', h'_R, d') \ll (\theta_0, h_{R,0}, d_0)$. Equation (10.2) can then be written formally,

$$\frac{\partial \theta'}{\partial t} = -(N_R + N_D)\theta', \tag{10.5}$$

where

$$N_R = \frac{-h'_R}{\rho c_p \theta'}, \qquad \text{and} \qquad N_D = \frac{-d'}{\rho c_p \theta'}, \tag{10.6}$$

are *radiative* and *dynamic relaxation rates.*

N_R and N_D represent the stiffness of the system for radiative or dynamic processes, respectively. The larger N indicates which process is more effective in restoring the system to its steady state, after a perturbation.

The usefulness of this approach depends upon the ease and accuracy with which N_R and N_D can be calculated. As far as N_D is concerned, we shall use only order-of-magnitude arguments. For example, if the system has a horizontal extent, y, and fluid velocity, v,

$$N_D \sim \frac{|v|}{|y|}. \tag{10.7}$$

To go further with N_D requires that we solve the full Navier–Stokes equations and we shall not attempt to do so. We shall, however, look closely at N_R. Since h'_R is made up of contributions from different frequencies, it is convenient also to define a spectral relaxation rate such that

$$N_R = \int_0^\infty N_\nu \, d\nu = \int_0^\infty \frac{h'_\nu}{\rho c_p \theta'} \, d\nu. \tag{10.8}$$

Studies of the radiative relaxation rate follow two distinct protocols, the *eigenvalue problem* and the *Newtonian cooling* approach, confusing because of their similarity, but with different objectives. The former seeks a formal solution to (10.5) for $\theta'(t)$ as a function of time. In principle, the solution could and should include both radiative and dynamic terms, but we are not yet able to do so, and in §§ 10.2 and 10.3 we consider the radiation only case, $N_D = 0$.

We seek separable solutions of the form

$$\theta'(\mathbf{s}, t) = \psi(\mathbf{s}) \exp(\pm Nt), \tag{10.9}$$

where $\mathbf{s}$ is a position vector. Equations (10.9) and (10.6) have solutions only for certain specific functions $\psi^i(\mathbf{s})$, and for these eigenfunctions there are discrete eigenvalues each with a specific value of the relaxation rate N^i. Each eigenstate is associated with a different scale of disturbance. Because the solutions are separable, the eigenfunctions are form preserving. If they form a complete set, an arbitrary initial disturbance can be expressed in terms of them, and the time evolution of the disturbance can be discovered from (10.9).

A special case is that of an infinite, homogeneous medium (the *Spiegel problem*). The eigenfunctions are harmonic (wavenumber, n) and the spectrum of eigenfunctions, $\tilde{N}(n)$, is continuous. Fourier integral techniques are applicable in this case (§ 10.3).

This formal approach to radiation problems has found little direct application to practical atmospheric problems. Nevertheless, the theoretical results are important for order-of-magnitude considerations. The relaxation rate, $\tilde{N}(n)$, may be taken to be representative for any disturbance of scale, n^{-1}. An example is given in § 10.4.

The *Newtonian cooling approximation* has become popular for a specific application, the damping of gravity waves in the middle atmosphere. The implicit assumption is made that the form of the thermal disturbance is determined by events in the troposphere, and that radiation is a perturbation that does not affect the form. The gravity waves are not exact eigenfunctions for the problem and are, therefore, not form preserving. Given the form of the disturbance, (10.6) can be solved to give an empirical relaxation rate $N'(n)$, a function of position as well as the scale of the gravity wave. Given the relaxation rate, (10.6) may then be used as a heat equation,

$$h'_R = -N'(n)\rho c_p \theta'. \tag{10.10}$$

This is Newton's law of cooling, the simplest and perhaps the most important of all physical laws of heat transport.

This definition of the Newtonian approximation is more general than is used by many investigators. The term is commonly applied to the

special case $N'(0)$, for which the perturbation is not a function of position. $n=0$ is also the condition for the boundary-exchange approximation, N^{BE}(§ 6.4.1), the dominant term of which is usually the radiation-to-space term, N^{SP}. It is N^{SP} that is sometimes referred to as the *scale-independent Newtonian approximation.*

We now discuss two approximations that are widely used, although, with additional effort, both can be avoided. From (2.18), for a thermal source function, to first order in the primed quantities,

$$h'_\nu = 4\pi k'_{\nu,\nu}(\bar{I}_{\nu,0} - B_{\nu,0}) + 4\pi_{k\nu,\nu,0}(\bar{I}'_\nu - B'_\nu). \tag{10.11}$$

$\bar{I}'_\nu$ is a function of B'_ν everywhere in the medium. Everywhere, we may write

$$B'_\nu = \left(\frac{\partial B_\nu}{\partial \theta}\right)_{\theta_0} \theta'. \tag{10.12}$$

Since θ_0 varies throughout the medium, so does $(\partial B_\nu/\partial\theta)_{\theta_0}$, but in all that follows we shall treat $(\partial B_\nu/\partial\theta)$ as a constant.

The second approximation involves the neglect of the first term on the right-hand side of (10.11) in comparison to the second. If the unperturbed state is close to radiative equilibrium, the first term is close to zero when integrated over the entire spectrum (exactly so for grey absorption). In addition,

$$k'_{\nu,\nu} = \frac{\partial k_{\nu,\nu}}{\partial \theta}\theta', \tag{10.13}$$

is positive for some lines and negative for others. On both counts, the approximation is plausible. Estimates of the error involved have been made by a number of investigators, who generally reach the conclusion that the error is small but not negligible.

In § 10.6, we shall discuss the absorption of solar radiation in the middle atmosphere. The source function is no longer thermal and the volume absorption coefficient is strongly influenced by temperature-dependent photochemistry. In this case, the first term in (10.11) exceeds the second in magnitude.

Radiative relaxation is simple to treat in the transparent and opaque limits, introduced in § 2.4.2. In the *transparent limit* (with the consistent assumption $k'_{\nu,\nu}=0$), (2.120) gives

$$h'_\nu = -4\pi k_{\nu,\nu}\frac{\partial B_\nu}{\partial \theta}\theta'. \tag{10.14}$$

Hence

$$N_\nu(\infty) = \frac{4\pi k_{\nu,\nu}}{\rho c_p}\frac{\partial B_\nu}{\partial \theta} = \frac{4\pi q_a k_{m,\nu}}{c_p}\frac{\partial B_\nu}{\partial \theta}, \tag{10.15}$$

where

$$q_a = \frac{\rho_a}{\rho} \tag{10.16}$$

is the mass mixing ratio of the absorber, and ρ_a is the absorber density. The spectral integral of (10.15) is

$$N(\infty) = \int_0^\infty N_\nu(\infty)\, d\nu = \frac{4\pi q_a \bar{k}_{P,m}}{c_p}\frac{\partial B}{\partial \theta}, \tag{10.17}$$

where $\bar{k}_{P,m}$ is the Planck mean absorption coefficient. The parenthesis (∞) indicates that (10.15) and (10.17) are valid for small spatial scales, less than the mean free path of the radiation. This limit for small scales is common to all relaxation rates, whether for eigenfunctions or for empirical perturbations. It provides the fundamental rate scale for radiative processes, independently of the particular problem or circumstances.

In the *opaque limit*, (2.127), we have

$$N_\nu = -\frac{4\pi}{3\rho c_p k_{\nu,\nu}}\frac{\partial B_\nu}{\partial \theta}\frac{\nabla^2 \theta'}{\theta'}. \tag{10.18}$$

This expression cannot be evaluated without knowing the form of the perturbation. For the harmonic form,

$$\psi(\mathbf{s}) = \exp(i\mathbf{n}\cdot\mathbf{s}), \tag{10.19}$$

we find

$$\frac{\tilde{N}(n)}{N(\infty)} = \frac{n^2}{3\bar{k}_{P,\nu}\bar{k}_{R,\nu}}, \tag{10.20}$$

where $\bar{k}_{R,\nu}$ is the Rosseland mean absorption coefficient.

To conclude this section we review the different radiative relaxation rates that have been mentioned, and introduce additional rates that

appear in later sections:

$N_R = N = -h'_R/\rho c_p \theta'$, a general definition, (10.6)
N^i = relaxation rate for the ith eigenmode
$\bar{N}(n)$ = Spiegel's solution for an infinite, homogeneous medium
$N'(n)$ = empirical relaxation rate, (10.10)
$N^{BE} = N'(0)$ = boundary-exchange approximation
N^{SP} = radiation-to-space term in N^{BE}
$N(\infty)$ = transparent limit, (10.15)
N_E = any rate calculated with an equilibrium (Planck) source function, § 10.3.4
$\bar{N}$ = planetary relaxation rate, § 10.4.1
$N'_P(n)$ = photochemical–radiative rate, § 10.6.

10.2. The radiation eigenvalue problem

10.2.1. *The integral equation*

General solutions for radiative relaxation rates have been obtained only in geometries with plane-parallel symmetry. As far as the atmosphere is concerned, this restricts us to a stratified model, with disturbances that vary in the vertical direction only. This limitation is important only for intermediate scales of disturbance because the transparent and opaque limits are not so restricted. We use (2.103) and (2.18) with (10.12) to give

$$h'_\nu = 2\pi k_{v,\nu}(\partial B_\nu/\partial\theta)[-2\theta'(\tau_\nu) + \theta^{*\prime} E_2(\tau_{1,\nu} - \tau_\nu) + \int_{\tau_\nu}^{\tau_{1,\nu}} \theta'(t) E_1(t - \tau_\nu)\, dt + \int_0^{\tau_\nu} \theta'(t) E_1(\tau_\nu - t)\, dt]. \quad (10.21)$$

In (10.21), we have included ν suffices and a lower boundary perturbation $\theta^{*\prime}$. The upper boundary cannot be perturbed. In the following treatment, we shall assume the same for the lower boundary, and omit $\theta^{*\prime}$. The treatment given here can be extended to include a lower boundary perturbation, but without knowledge of conditions inside the lower boundary we cannot predict either its amplitude or its phase.

With $N_D = 0$ in (10.5) (radiation only),

$$\frac{\partial\theta'(m,t)}{\partial t} = \int_0^\infty \frac{h_\nu}{\rho c_p}\, d\nu = \int_0^\infty \frac{4\pi k_{m,\nu}(m) q_a(m)}{c_p} \frac{\partial B_\nu}{\partial\theta} \left\{ \frac{1}{2} \int_0^{m_1} \theta'(m', t) k_{m,\nu}(m') \times E_1[|\tau_\nu(m') - \tau_\nu(m)|]\, dm' - \theta'(m,t) \right\} d\nu. \quad (10.22)$$

The independent variable has been changed to the absorber amount,

$$m(z) = \int_z^\infty \rho_a(z')\, dz',$$
$$d\tau_\nu(z) = k_{m,\nu}(z)\, dm(z). \tag{10.23}$$

We seek solutions that are separable in the variables m and t. For solutions that satisfy (10.9), we cannot accept the indefinite increase with time indicated by the positive sign in the exponential. We accept only the decaying solutions,

$$\theta'(m, t) = \psi(m)e^{-Nt}. \tag{10.24}$$

If we substitute (10.24) into (10.22), we obtain the integral equation,

$$\psi(m) = \frac{N(\infty)}{2[N(\infty) - N]} \int_0^{m_1} K(m, m')\psi(m')\, dm', \tag{10.25}$$

where the kernel function is

$$K(m, m') = \left[\bar{k}_{m,P}(m)\frac{dB}{d\theta}\right]^{-1}$$
$$\times \int_0^\infty k_{m,\nu}(m)k_{m,\nu}(m')(\partial B_\nu/\partial\theta)E_1[|\tau_\nu(m) - \tau_\nu(m')|]\, d\nu, \tag{10.26}$$

and $N(\infty)$ is given by (10.17).

If $K(m, m')$ is symmetric, (10.25) is a Fredholm equation, for which standard methods of solution are available. Some asymmetries can be accommodated by redefining $\psi(m)$ but no asymmetries exist for a homogeneous atmosphere. The required conditions for a homogeneous atmosphere (besides constancy of $\partial B_\nu/\partial\theta$) are a constant mixing ratio and a constant Planck mean mass absorption coefficient. These conditions are restrictive and limit the value of the method of eigenvalues for practical problems.

Given these conditions, (10.25) can be evaluated by standard methods. For finite values of m_1, the solutions are discrete, with eigenvalues $N(\infty)/2[N(\infty) - N^i]$. The eigenvalues of the Fredholm equation are real, positive, and constitute a denumerably infinite set, while the eigenfunctions are orthogonal. These conditions allow us to express an arbitrary disturbance as a weighted sum of the eigenfunctions.

$m_1 = 0$ corresponds to the transparent limit. From (10.25), the only solution is $N = N(\infty)$. In the transparent limit the relaxation rate does not depend upon the form of the disturbance.

10.2.2. *Spiegel's solution*

Spiegel's solution for an infinite medium is central to most discussions of radiative relaxation. The integral in (10.25) can be split into two ranges, $0 \leqslant |m - m'| \leqslant m$ and $0 \leqslant |m - m'| \leqslant m_1 - m$, corresponding to the levels above and below the level of interest, respectively. For an infinite atmosphere, both integrals are the same, with the limits 0 and ∞.

The eigenfunctions of (10.25) depend upon the spatial variation of $\tau_\nu(m)$. This variation differs from frequency to frequency requiring, in principle, a different solution for each frequency. This difficulty can be avoided if we can assume that the mass absorption coefficient is independent of position (we have already assumed that the Planck mean is independent of position, so that this is not a major new restriction, although it does further limit the use of the formal theory),

$$\tau_\nu(m) - \tau_\nu(m') = k_{\mathrm{m},\nu}(m - m'). \tag{10.27}$$

m is now the only independent variable. The continuous spectrum of eigenfunctions can now be written

$$\psi(m) = \exp(in^* m), \tag{10.28}$$

where n^* is related to the geometric wavenumber by

$$n^* = \frac{n}{\rho_\mathrm{a}}. \tag{10.29}$$

The purpose of introducing n^* is to be able to accommodate vertical variations of density, provided that the mixing ratio remains constant.

If we expand the exponential integrals and make use of the substitution $\tilde{\tau} = [\tau_\nu(m) - \tau_\nu(m')]/\xi$, where ξ is the vertical direction cosine, (10.25) and (10.28) become

$$N(\infty) - \tilde{N}(n^*) = \frac{2\pi q_\mathrm{a}}{c_\mathrm{p}} \int_0^\infty d\nu \frac{\partial B_\nu}{\partial \theta} k_{\mathrm{m},\nu} \int_{-1}^{+1} d\xi \int_0^\infty d\tilde{\tau} \exp\left[-\tilde{\tau}\left(1 + \frac{in^*\xi}{k_{\mathrm{m},\nu}}\right)\right], \tag{10.30}$$

$$= \frac{4\pi q_\mathrm{a}}{c_\mathrm{p}} \int_0^\infty d\nu \frac{\partial B_\nu}{\partial \theta} k_{\mathrm{m},\nu} \left(\frac{k_{\mathrm{m},\nu}}{n^*} \tan^{-1} \frac{n^*}{k_{\mathrm{m},\nu}}\right). \tag{10.31}$$

Note the use of the tilde to denote the infinite domain.

Spiegel's solution is, for grey absorption,

$$\frac{\tilde{N}(n^*)}{N(\infty)} = 1 - \frac{k_\mathrm{m}}{n^*} \tan^{-1} \frac{n^*}{k_\mathrm{m}}. \tag{10.32}$$

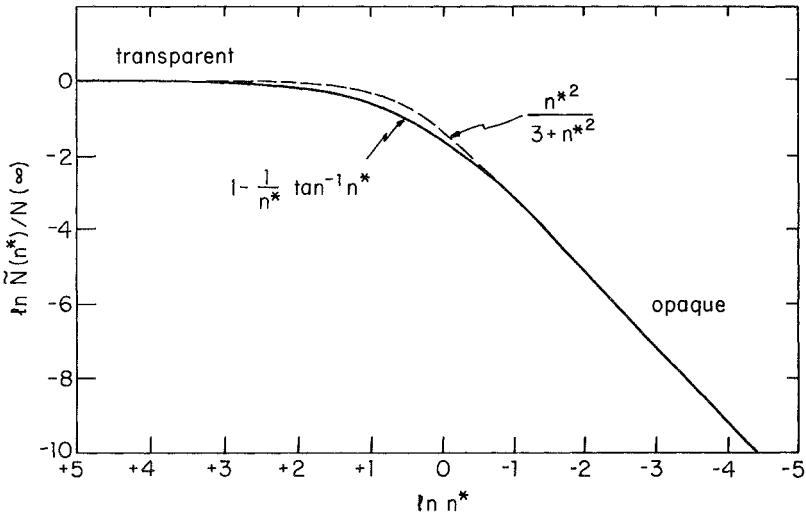


Fig. 10.1. Radiative relaxation rates for grey absorption. The broken line is obtained from a two-stream approximation (see § 10.2.3).

Equation (10.32) is plotted in Fig. 10.1. It has the asymptotic forms,

$$\frac{\tilde{N}(n^*)}{N(\infty)} \to 1, \qquad \text{as } \frac{n^*}{k_m} \to \infty, \tag{10.33}$$

$$\frac{\tilde{N}(n^*)}{N(\infty)} \to \frac{n^{*2}}{3k_m^2}, \qquad \text{as } \frac{n^*}{k_m} \to 0. \tag{10.34}$$

Equations (10.33) and (10.34) correspond to the transparent, (10.17), and the opaque, (10.20), limits, respectively. The important independent variable n^*/k_m is the optical path of a wavelength divided by 2π.

Because there is a continuum of solutions, the time evolution of an arbitrary disturbance can be studied by means of Fourier transforms. Figure 10.2 shows a calculation of $\theta'(\tau', t)/\theta'(\tau', 0)$ for an initial symmetric disturbance,

$$\theta'(\tau, 0) = \exp - |\tau - \tau'| . \tag{10.35}$$

The Fourier decomposition of (10.35) has some power at transparent frequencies ($n^*/k_m > 1$). The initial decay rate is about one-half $N(\infty)$; at later times, as the disturbance spreads, the power moves to opaque frequencies and the decay rate becomes progressively slower.

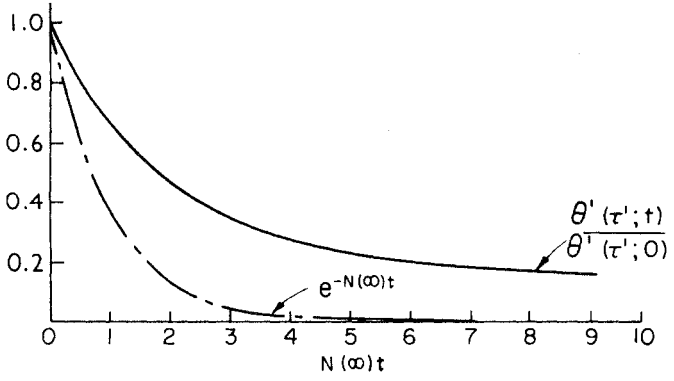


FIG. 10.2. Decay of the central temperature for a symmetric disturbance. The initial disturbance is

$$\theta'(\tau, 0) = \exp - |\tau' - \tau|.$$

After Spiegel (1957).

10.2.3. *Two-stream solution for a scattering atmosphere*

A two-stream solution for the radiative relaxation rate is compared with an exact solution in Fig. 10.1. The asymptotic regimes agree, and the correspondence is good for all scales. It bears repeating that two-stream approximations contain all of the essential physics of radiative transfer and that they are exact in transparent and opaque limits. We now derive the two-stream result for a mixed scattering and absorbing atmosphere, having a single-scattering albedo a_ν.

From (2.23) and (2.32), the source function for mixed scattering and absorption is

$$J_\nu = (1 - a_\nu)B_\nu + a_\nu \int PI_\nu \frac{d\omega}{4\pi}. \tag{10.36}$$

If we use this source function in (8.137), and perform the same operations as those used to obtain (8.144), we obtain an equation for the flux,

$$\frac{d^2F_\nu}{d\tau_\nu^2} = (1 - a_\nu)\left[3F_\nu(1 - a_\nu g_\nu) - 4\pi \frac{dB_\nu}{d\tau_\nu}\right] - a_\nu f_\nu \left[\frac{1}{\xi_\odot} + 3(1 - a_\nu)g_\nu \xi_\odot\right]. \tag{10.37}$$

Differentiating (10.37) with respect to z and equating perturbation terms

(f_ν is not perturbed),

$$\frac{d^2h'_\nu}{d\tau_\nu^2} = (1-a_\nu)\left[3h'_\nu(1-a_\nu g_\nu) - 4\pi \frac{\partial B_\nu}{\partial \theta} e_{\nu,\nu} \frac{d^2\theta'}{dz^2}\right]. \tag{10.38}$$

If we define N_ν from (10.10) and use the spatial perturbation (10.28), we find

$$\frac{\tilde{N}_\nu(n^*)}{N_\nu(\infty)} = \frac{n^{*2}}{n^{*2} + 3k^2_{m,\nu}(1-a_\nu g_\nu)/(1-a_\nu)}. \tag{10.39}$$

$N_\nu(\infty)$ is defined from (10.15) with $k_{\nu,\nu} = (1-a_\nu)e_{\nu,\nu}$.

For $a_\nu = 0$ (no scattering), (10.39) gives the broken line in Fig. 10.1. In a later section, we shall obtain a similar result using a diffusivity approximation to the integral equations. As discussed in § 2.4.5, the two-stream approximation and the diffusivity approximation give similar results. In this particular case, the two approaches agree for the diffusivity factor $r = 3^{1/2}$. The same result is obtained for $g_\nu = 1$ (forward scattering) as for $a_\nu = 0$. There is no formal distinction between strict forward scattering and no scattering at all.

10.2.4. *Effect of boundaries*

In a finite domain, eigenvalues and eigenfunctions are discrete. The eigenfunctions form a complete basis set and can be used to describe the time evolution of an arbitrary disturbance in a finite atmosphere. The

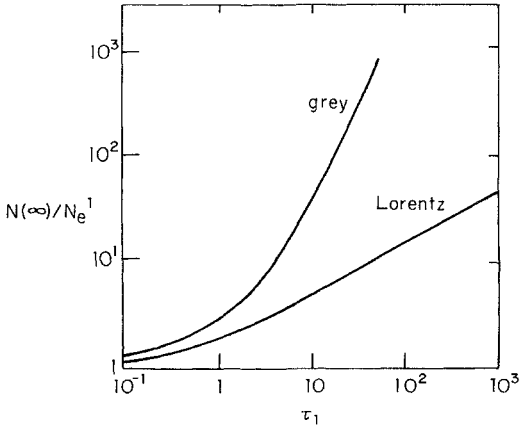


FIG. 10.3. The first even eigenvalues, N_c^1, for a perturbation between two boundaries. τ_1 is the optical depth for a medium contained between two boundaries (the optical depth at the line center for a single Lorentz line). $N(\infty)$ is the transparent rate (10.17). After Gay and Thomas (1981).

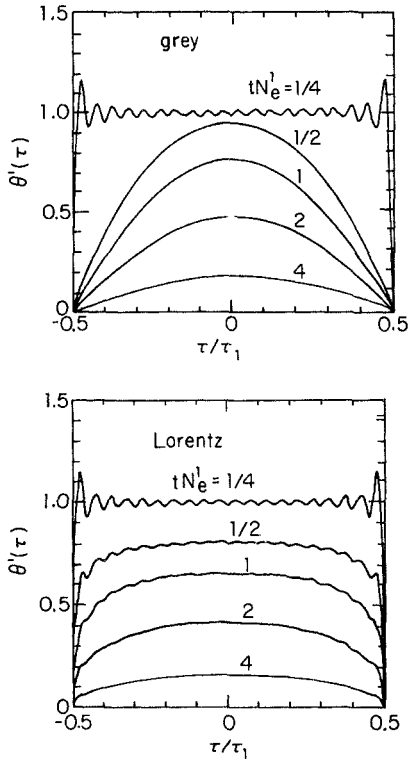


FIG. 10.4. Decay of a temperature perturbation between two boundaries. The initial perturbation is $\theta'(\tau, 0) = 1$. The optical depth τ_1 is 50 in both cases. The elapsed times for the successive curves are, from top to bottom, $tN_e^1 = 1/4$, 1/2, 1, 2, 4. From Fig. 10.3, $N_e^1/N(\infty)$ is 1/800 for grey absorption and 1/14 for the Lorentz line. After Gay and Thomas (1981).

decay of an initially uniform disturbance between two thermal boundaries is illustrated in Figs. 10.3 and 10.4.

Results are shown for grey absorption and for a single Lorentz line. In the latter case, τ_1 refers to the optical depth at the line center. The solutions are obtained by successive approximations. An analytic fit is made to the kernel function, and a first approximation to the eigenfunctions is based on an analytic solution to a related problem.

The calculation is performed separately for odd and even parts of the eigenfunction; for this problem only even functions are involved. The lowest eigenvalues for the even eigenfunctions are shown in Fig. 10.3. These correspond to the smallest possible rates and are related (inexactly) to boundary-exchange terms. The highest eigenvalues are for the smallest disturbances and correspond to the lower horizontal axis in Fig. 10.3. Other eigenvalues lie on intermediary curves.

For an infinite medium, the chosen perturbation would not decay. All of the behavior of the solutions shown in Fig. 10.4 is attributable to the presence of the boundaries. The ripple is caused by truncating the series used to represent the eigenfunctions.

10.3. Numerical results

10.3.1. $N_i(\infty)$ *for atmospheric bands*

Partial relaxation rates may be calculated for individual atmospheric bands. Since relaxation rates are proportional to heating rates, they are additive and, for several bands,

$$N = \sum_i N_i. \tag{10.40}$$

The appropriate Planck mean is (2.130)

$$\bar{k}_{\mathrm{P,m},i} = S_{\mathrm{m},i} \frac{dB_i/d\theta}{dB/d\theta}, \tag{10.41}$$

where $S_{\mathrm{m},i}$ is the band intensity per unit mass.

Values for $S_{\mathrm{m},i}$ and $N_i(\infty)$ for six atmospheric bands are shown in Table 10.1. For typical surface mixing ratios, the maximum total relaxation rate is about $3 \times 10^{-3}\,\mathrm{s}^{-1}$ and is dominated by the water vapor rotation band. At higher temperatures, the 6.3 μm band would be more important. In the middle atmosphere, the 15 μm carbon dioxide band is the most important, and the maximum radiative rate is $\sim 10^{-4}\,\mathrm{s}^{-1}$ at all levels.

Table 10.1. Radiative rates for six atmospheric bands, at 250 K[a]

Band	v/c (cm^{-1})	$S_{n,i}$ (cm)	q_n	$N_i(\infty)$ (s^{-1})
H_2O, rot.	~150	5.27×10^{-17}	6×10^{-3}	2×10^{-2} to 8×10^{-6}
H_2O, 6.3 μm	1595	1.04×10^{-17}	to	5×10^{-4} to 2×10^{-6}
H_2O, 2.7 μm	3756	7.5×10^{-18}	2×10^{-5}	4×10^{-6} to 1×10^{-10}
CO_2, 15 μm	677	9.41×10^{-18}	5×10^{-4}	1×10^{-4}
CO_2, 4.3 μm	2349	1.05×10^{-16}		1×10^{-5}
O_3, 9.6 μm	1042	1.52×10^{-17}	1.3×10^{-5}	3×10^{-6}

[a] Data in the first three columns are from Chapter 5. The two mixing ratios given for water vapor correspond to ground level and mesospheric conditions. The ozone mixing ratio is the maximum, which occurs at about 36 km. For bands with frequencies greater than $1000\,\mathrm{cm}^{-1}$, $N_i(\infty)$ is very sensitive to temperature. For $x = h\nu/k\theta \gg 1$, the Planck function varies as $\exp(-x)$. For the 6.3 μm band at 250 K, $x = 9.2$.

10.3.2. *Special absorption laws*

A useful form for (10.30) is obtained by performing the frequency integral before the path integral. After one partial integration we find

$$\tilde{N}(n^*) = N(\infty)\int_0^\infty \frac{1}{m}\frac{\partial \epsilon_c^*(m)/\partial m}{[\partial \epsilon_c^*(m)/\partial m]_{m=0}}\left[\frac{\sin(n^*m)}{n^*m} - \cos(n^*m)\right] dm, \quad (10.42)$$

where

$$\epsilon_c^*(m) = \int_0^\infty \frac{(\partial B_\nu/\partial\theta)[1-\exp(-k_{m,\nu}m)]\,d\nu}{dB/d\theta} = \sum_i \frac{(\partial B_i/\partial\theta)[1-\bar{T}_i(m)]}{dB/d\theta} \quad (10.43)$$

is related to the modified emissivity, introduced in (6.73), and can be calculated with equal confidence ($\bar{T}$ replaces $\bar{T}^f$). Given the necessary data, the quadrature in (10.42) is very simple to perform.

For a power law,

$$\epsilon_c^*(m) \propto m^a, \quad (10.44)$$

$$\tilde{N}(n^*)/N(\infty) \propto (n^*)^{1-a}. \quad (10.45)$$

For a transparent path, absorption is linear in the amount of absorber ($a = 1$) and $\tilde{N}(n^*)$ is constant. For independent, saturated, Lorentz lines (4.15), $a = 0.5$ and $\tilde{N}(n^*) \propto (n^*)^{0.5}$.

Equation (10.42) has been integrated for an Elsasser band using a diffusivity approximation. For a diffusivity factor, r, the equivalent expression to (10.39), with $a_\nu = 1$, is

$$\frac{\tilde{N}_\nu(n^*)}{N_\nu(\infty)} = \frac{n^{*2}}{n^{*2} + r^2 k_{m,\nu}^2}. \quad (10.46)$$

This provides a comparison for the Elsasser band, for which

$$\frac{\tilde{N}(n^*)}{N(\infty)} = \frac{\frac{r}{2}\{[(z^2-1)^2 + 4z^2\coth 2\pi y]^{1/2} + 1 - z^2\}^{1/2}}{[2(z^2-1)^2 + 8z^2\coth 2\pi y]^{1/2}}, \quad (10.47)$$

where

$$z = \frac{r\bar{k}_{P,m}}{n^*}, \qquad y = \frac{\alpha_L}{\delta}.$$

In the limit of strongly overlapping lines ($y \to \infty$), (10.47) and (10.46) are the same.

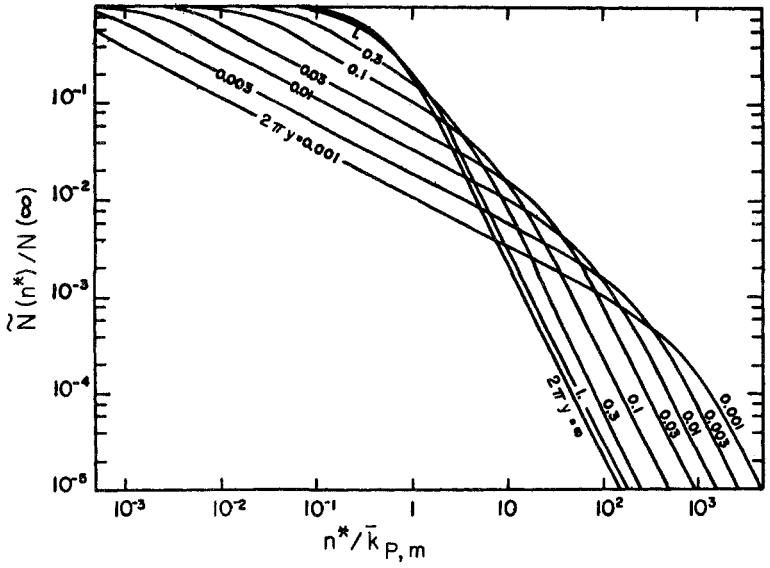


FIG. 10.5. Radiative relaxation rates for an Elsasser band. A diffusivity approximation has been used, with the diffusivity factor $r = 2$. After Sasamori and London (1966).

The complete expression (10.47), with $r = 2$, is plotted in Fig. 10.5. The curve for $y = \infty$ differs slightly from the broken line in Fig. 10.1 because of a different diffusivity factor. The difference is within the uncertainties inherent in the two-stream and diffusivity approximations. In addition to the transparent and opaque limits, Fig. 10.5 shows, for $y \ll 1$, an intermediate region with $\tilde{N}(n^*) \propto (n^*)^{1/2}$ characteristic, as we have shown, of strongly absorbed Lorentz lines.

Strong-line limits for absorption by complete bands were discussed in § 4.9. Theory and measurement suggest a logarithmic dependence on path length for path lengths beyond a threshold (see Table 4.3)

$$\epsilon_c^*(m) = a(p, \theta) + b \ln(m). \tag{10.48}$$

With this emissivity, (10.42) gives

$$\frac{\tilde{N}(n^*)}{N(\infty)} = \frac{\pi b}{4} \frac{n^*}{\bar{k}_{P,m}}. \tag{10.49}$$

In this logarithmic regime, $\tilde{N}(n)$ [but not $\tilde{N}(n^*)$] is independent of the absorber density ρ_a, because both $N(\infty)$ and $(n^*)^{-1}$ are proportional to it.

10.3.3. *Radiative relaxation for earth and Mars*

As examples of relaxation rates, we may look at data for water vapor in the earth's lower atmosphere, for water drops, and at a comparison between the lower atmospheres of earth and Mars.

Figure 10.6 shows that for water vapor the entire range of scales may be divided into transparent, logarithmic, and opaque regimes. The transition between transparent and logarithmic regimes lies at $n^*/\bar{k}_{P,m} = 20$ and that between logarithmic and opaque regimes lies at $n^*/\bar{k}_{P,m} = 7 \times 10^{-4}$. Based on a water vapor partial pressure of 3 mb, these data correspond to length scales (n^{-1}) of 3 m and 100 km, respectively. Since the vertical scale of water vapor is about 2 km in the troposphere, we do not expect ever to encounter the opaque limit. The transparent regime

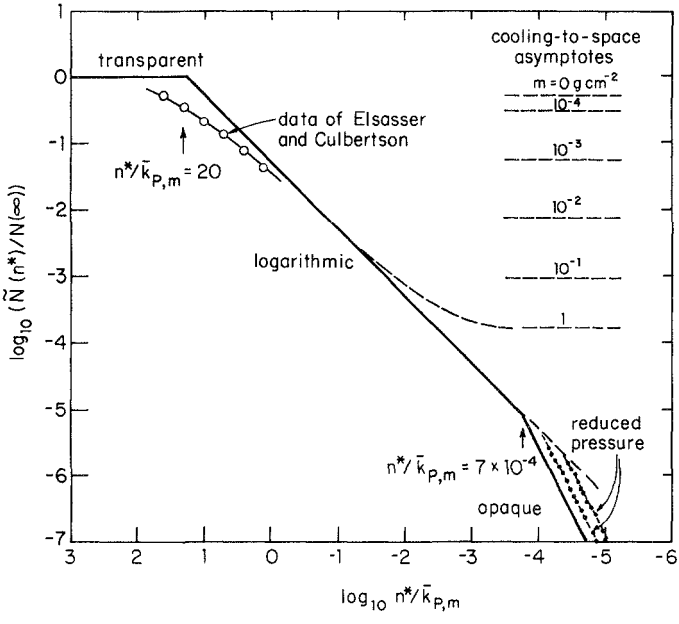


FIG. 10.6. Radiative relaxation rates for water vapor. Calculations are for 1 bar and 273 K. The curve for the *opaque limit* follows (10.20); we used $\bar{k}_P/\bar{k}_R = 10^3$. The *logarithmic region* follows (10.49), with $b = 0.067$, from data given by Rodgers (1967, § 6.4.2). The points were calculated by numerical quadrature from (10.42), using the data of Elsasser and Culbertson (1960, § 6.4.2). The air pressure has no influence on the *transparent region* and only a second-order influence on the logarithmic region. The opaque region is dominated by the water vapor continuum that depends upon both the air pressure and the partial pressure of the vapor (§ 5.4.3). The dotted lines represent schematically the effect of reducing either pressure below typical ground-level values. $n^*/\bar{k}_{m,P}$ is dimensionless. To obtain a dimensional wavenumber, use $n = (n^*/\bar{k}_{m,P})\rho_a \bar{k}_{m,P}$, where $\bar{k}_{m,P} = 200\,\mathrm{g}^{-1}\,\mathrm{cm}^2$ and $\rho_a = 8 \times 10^{-7}\,\mathrm{g\,cm}^{-3}$ for 1 mb partial pressure at 273 K. The broken lines on the right-hand side of the figure are cooling-to-space asymptotes, calculated from Elsasser and Culbertson's data (see § 10.4.2 for details).

will also rarely be encountered except in the boundary layer; the lower atmosphere is dominated by the logarithmic regime.

In the middle atmosphere, carbon dioxide is the dominant gas. Numerical calculations by a number of authors suggest that the transparent limit is valid for n^{-1} less than 100 m at an altitude of 70 km, and for n^{-1} less than 500 m at 80 km.

The Planck mean mass absorption coefficient is much larger for water in the liquid phase than in the vapor phase. If we wish to compare the relaxation rates for the same amount of water in the liquid and vapor phases, however, this is not the relevant comparison. We need to know both the absorption and scattering coefficients per unit volume and, unlike the vapor, the liquid is aggregated into droplets that occupy only a small fraction of the volume. To compare the effect of equal amounts of water per unit volume we must know the size of the droplets and calculate their optical properties by the methods of Chapter 7. Figure 10.7 shows radiative rates for the same amount of water, in the form of droplets of radius 3 or 10 μm, or in the form of vapor. For scales less than about 50 m, the relaxation rate for the droplets exceeds the relaxation rate for the vapor by a factor up to 10.

In Fig. 10.8 we compare the relaxation times (N^{-1}) for earth and Mars in dimensional terms. Other things being equal relaxation times are inversely proportional to the absorber mixing ratio. For Mars, the mixing ratio of carbon dioxide is unity and this is one reason why the relaxation

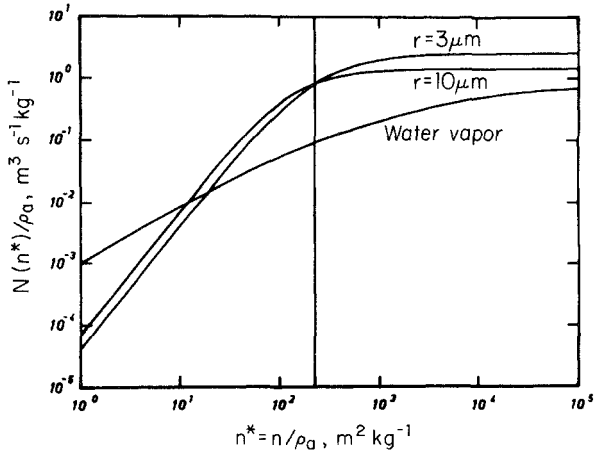


FIG. 10.7. Relaxation rates for water as droplets or as vapor. The absorber density, ρ_a, is the mass of water per unit volume of air whether in the vapor or in the liquid phase. The droplet data are taken from Irvine and Pollack (1968). The presentation is equivalent to that in Fig. 10.4, with a reversal of the horizontal axis and shifts along both axes, equal to $\ln[\bar{k}_{m,P}]$. For $\rho_a = 8 \times 10^{-7}$ g cm^{-3}, the vertical line corresponds to $n = 0.18$ m^{-1}. After Coantic and Simonin (1984).

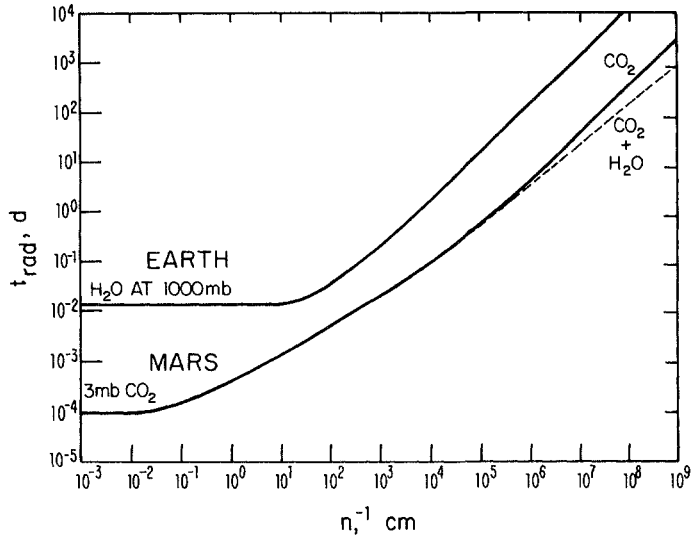


FIG. 10.8. Radiative relaxation times for earth and Mars. Note the use of relaxation times rather than rates ($t = N^{-1}$). The martian data are for 3 mb (the center of mass of the atmosphere) and a water vapor mixing ratio of 10^{-3}. The terrestrial data are for 3 mb of water vapor at s.t.p. After Goody and Belton (1967, § 4.9).

times are, on the average, 30 times less for Mars than for earth. This difference has profound effects on the diurnal response of the atmosphere. For earth, the radiative response time of the entire atmosphere (see § 10.4) is about 100 days. The sol (the solar day) is approximately equal on the two planets. For earth, the day is too short for the atmosphere to respond to diurnal forcing with significant amplitude, but for Mars the meteorology has a strongly tidal character.

10.3.4. *Nonequilibrium source functions*

For a two-level model, the source function involves two time constants, $\phi(l, u)$ and $\eta(l, u)$, which are given in terms of molecular constants by (2.72) and (2.73). For a time-dependent problem, the rate equations should be solved, in place of the steady-state treatment given in § 2.2. This can be complicated enough with only two time constants; if more are involved the problem can become intractable. Fortunately, we can show that, for radiative relaxation, the two-level problem is quasistatic and that the source function developed in § 2.2 is valid. This statement does not apply to all time-dependent problems. The effect of vibrational relaxation on the transmission of ultrasound is not quasistatic, and leads to an established method of measuring vibrational relaxation times.

We may establish the quasistatic condition by demonstrating that the natural lifetime of the excited state is always much less than the least radiative relaxation time. Both $N(\infty)$, (10.17), and ϕ^{-1}, (2.72), are proportional to the band intensity. The relationship between the two can be made explicit by introducing the vibrational contribution to the specific heat,

$$c_{\text{vib}} = n\mathbf{k}\frac{\partial}{\partial\theta}\left[\theta^2\frac{\partial \ln f(\theta)}{\partial\theta}\right], \tag{10.50}$$

where n is the number of molecules per unit mass and $f(\theta)$ is the vibrational partition function for a single vibrational mode,

$$f(\theta) = [\exp(\mathbf{h}\nu_0/\mathbf{k}\theta) - 1]^{-g}. \tag{10.51}$$

g is the upper-state degeneracy ($= g_u/g_l$ in the symbols of § 2.2).

With this definition of c_{vib}, it can be shown that

$$N(\infty)\phi(u, l) = \frac{q_a c_{\text{vib}}}{c_p}. \tag{10.52}$$

The right-hand side of (10.52) is the ratio of the volume heat capacity of the vibrational modes of the absorbers alone to that for all modes of energy (including gravitational potential energy) of all molecules, neutral or absorbers. Even for a pure absorber, this ratio is less than unity and, for low temperatures, it may be very much less than unity. For all atmospheres, but particularly for earth's with its small absorber mixing ratios,

$$\phi^{-1} \gg N(\infty) > \tilde{N}(n^*). \tag{10.53}$$

Vibrational and rotational state populations adjust much more rapidly than the radiative relaxation process, and they may be treated quasistatically.

The 15 μm band of carbon dioxide is the most important band involving nonequilibrium source functions. This band is narrow, and the thermal source function may be treated as constant over the band. From (2.74) and (2.21), we may write a perturbation equation (omitting all i suffices),

$$J' = B' + \frac{\eta}{2\phi}\frac{h'}{2\pi S_v}. \tag{10.54}$$

From (10.6), (10.12), (10.15), and (10.41), we may write

$$\frac{\partial J}{\partial \theta} = \frac{\partial B}{\partial \theta} - \frac{\eta}{\phi}\frac{\rho c_p N}{4\pi S_\nu}$$

$$= \frac{\partial B}{\partial \theta}\left[1 - \frac{\eta}{\phi}\frac{N}{N_E(\infty)}\right]. \tag{10.55}$$

The subscript (E) is used to indicate the use of the equilibrium (Planck) source function.

In both the transparent and the opaque limits, (10.15) and (10.18), N is proportional to $(\partial B/\partial\theta)$. For a more general source function, it is only necessary to replace this term by $\partial J/\partial\theta$. Hence,

$$\frac{N}{N_E} = \frac{\partial J/\partial\theta}{\partial B/\partial\theta} = 1 - \frac{\eta}{\phi}\frac{N}{N_E(\infty)}. \tag{10.56}$$

There is a question about intermediary scales. To the extent that $\partial B/\partial\theta$ is a constant, it can be taken outside the integral in (10.30), and (10.56) is still valid. However, η varies inversely with the pressure and $\partial J/\partial\theta$ is not necessarily constant, even if $\partial B/\partial\theta$ is. We may, nevertheless, expect that (10.56) is a fairly good approximation at all times. It may be written

$$\frac{N(n)}{N_E(n)} = \left[1 + \frac{\eta}{\phi}\frac{N_E(n)}{N_E(\infty)}\right]^{-1}, \tag{10.57}$$

with the special case

$$\frac{N(\infty)}{N_E(\infty)} = \left(1 + \frac{\eta}{\phi}\right)^{-1}. \tag{10.58}$$

We may recall that, well below 70 km, $\eta/\phi \ll 1$ and thermodynamic equilibrium prevails. Well above 70 km, where $\eta/\phi \gg 1$, $N(\infty)$ has an upper limit $(\phi/\eta)N_E(\infty) \ll N_E(\infty)$.

Numerical calculations for an infinite medium, having the properties of the earth's atmosphere at altitudes of 70 and 80 km, are shown in Fig. 10.9. A very large decrease in the radiative rate accompanying breakdown of thermodynamic equilibrium is shown at 80 km. The values of η/ϕ were not given in the referenced paper, but they should be close to 2.5 at 80 km and 0.5 at 70 km.

Although this discussion has been presented in the context of an infinite medium, the result, (10.57), is quite general.

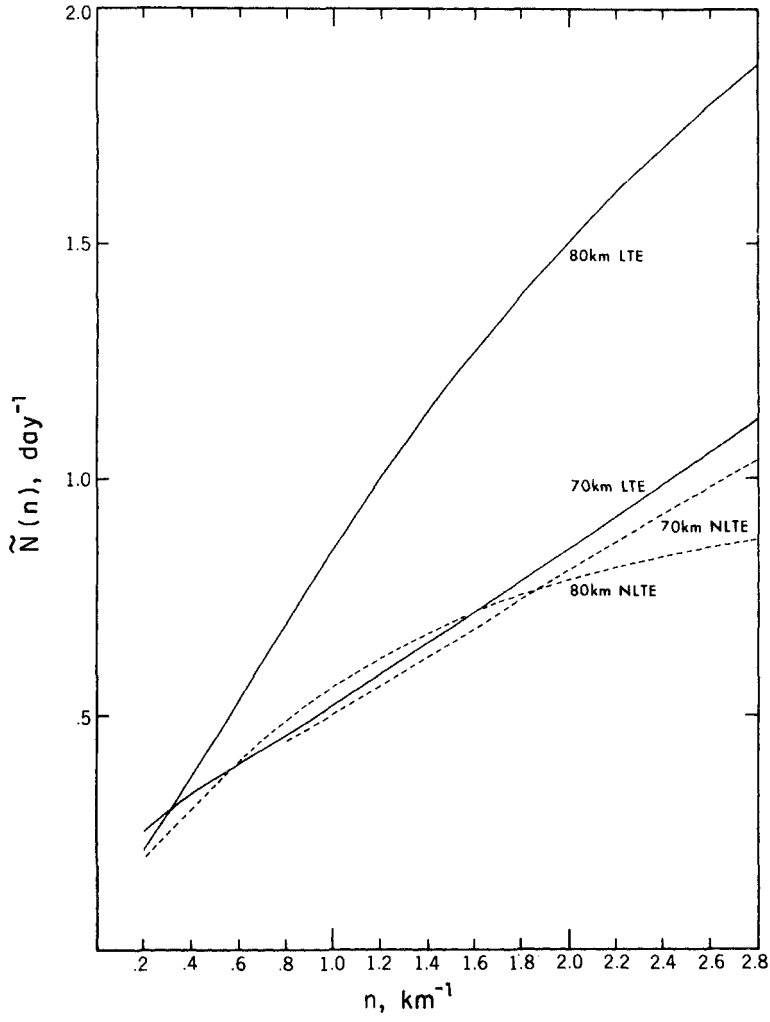


FIG. 10.9. Radiative relaxation times for carbon dioxide in the middle atmosphere. LTE indicates calculations using the Planck function (incorrect at these altitudes). NLTE uses the correct source function. After Fels (1984).

10.4. Planetary-scale relaxation

The formal results of §§ 10.2 and 10.3 are valuable for order-of-magnitude arguments. One interesting application is to the scale analysis of planetary-scale, steady-state circulations. Radiation is the fundamental atmospheric drive and can never be neglected, but we may still anticipate that circumstances will occur when the atmosphere is closer to an adiabatic state (strong dynamic influence) rather than to radiative

equilibrium. We may develop criteria for whether temperatures are closer to radiative equilibrium or to an adiabatic state in terms of the radiative and the dynamic relaxation rate.

10.4.1. *The planetary relaxation rate*

We first require an estimate of the globally averaged radiative relaxation rate. The average solar flux $[F_S = -f(1-a)/4]$ is distributed throughout the atmosphere by radiative and dynamic processes and balanced by emission to space equal to $\sigma\theta_e^4$, where θ_e is the effective emission temperature (we also use θ_e as a measure of the general level of temperature over the planet). The *planetary mean relaxation rate* ($\bar{N}$) is the incoming solar flux divided by the enthalpy of the entire atmosphere ($p_0 c_p \theta_e / g$ per unit area),

$$\bar{N} = \frac{F_S g}{c_p \theta_e p_0} = \frac{\sigma \theta_e^3 g}{c_p p_0}. \tag{10.59}$$

p_0/g is the mass per unit area of the atmosphere. The surface pressure for planets varies from that of Mercury (10^{-15} bar) to that of Jupiter (10^8 bar). Even among the terrestrial planets it varies from 10^2 bar for Venus to about 10^{-2} bar for Mars. The surface pressure, through its effect on the radiative relaxation rate, is a dominant parameter for assessing differences in the average thermal state of planetary atmospheres.

The planetary relaxation rate is interestingly related to the relaxation rate, (10.32), for an infinite medium. The right-hand side of (10.32) has a flat maximum at $k_{m,v}/n^* = 0.65$. The upper limit to $N(n^*)$ for a given value of $n(= n^* \rho_a)$ is

$$\begin{aligned} \bar{N}_{max}(n) &= \frac{0.9\pi n (dB/d\theta)}{\rho c_p} \\ &= \frac{3.6 n \sigma \theta_e^3}{\rho c_p}. \end{aligned} \tag{10.60}$$

As far as global-scale motions are concerned, the relevant length scale is the shortest distance between branches of the circulation, which we shall take to be the atmospheric scale height, H. With $n = H^{-1}$, (10.60) becomes

$$N_{max}(H^{-1}) = 3.6\bar{N}. \tag{10.61}$$

In order to achieve the maximum radiative relaxation rate, the absorption coefficient must be such as to make the optical depth

($\rho_a k_{m,\nu} H = \rho_a k_{m,\nu}/n$) approximately unity. If a significant fraction of the atmospheric spectrum has optical depths between, say, 0.1 and 10, we may expect the relaxation rate to approach $\bar{N}$. This is the case for earth and also for some other planets, and it is the reason why the planetary relaxation rate is a valuable concept. Note that provided that the atmosphere is sufficiently nongrey, the planetary relaxation rate depends neither upon its absorption spectrum nor upon its composition, but only on its mass. In this respect a nongrey absorber differs fundamentally from a grey absorber.

10.4.2. *The temperature of a nonrotating atmosphere*

Two extremes for the temperature distribution for a nonrotating atmosphere are illustrated in Fig. 10.10. Very strong mixing, as a result of intense dynamic activity, will cause the temperature to be horizontally uniform, at θ_e. On the other hand, the radiative equilibrium temperature, θ_{re}, is high in the tropics and zero at the poles (at an equinox). The observed state should lie somewhere between these extremes.

For a nonrotating planet, the global dynamic relaxation rate has been shown to be

$$N_D \approx (r\theta_e)^{1/2} R_0, \tag{10.62}$$

where r is the specific gas constant and R_0 is the radius of the planet. We may expect that the overall thermal state of the planet should be governed by the ratio $N_D/\bar{N}$. The following result has been obtained:

for *strong radiative control*, $N_D/\bar{N} \ll 1$, $\theta \to \theta_{re}$, and

$$\frac{\langle|\theta - \theta_{re}|\rangle}{\theta_e} \approx \frac{N_D}{\bar{N}} \left(\frac{\langle|\theta_{re} - \theta_e|\rangle}{\theta_e} \right)^{3/2}; \tag{10.63}$$

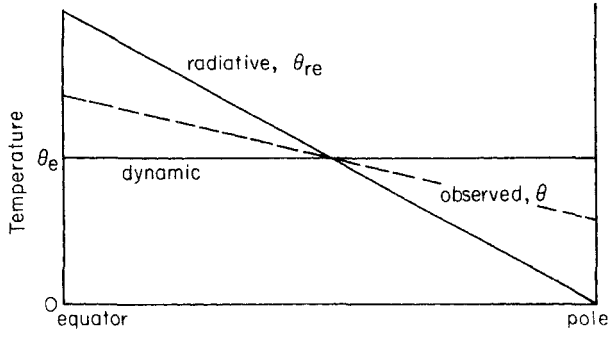


FIG. 10.10. Schematic global distribution of atmospheric temperature.

for *weak radiative control*, $N_D/\bar{N} \gg 1$, $\theta \to \theta_e$, and

$$\frac{\langle|\theta - \theta_e|\rangle}{\theta_e} \approx \left(\frac{\bar{N}}{N_D}\frac{\langle|\theta_e - \theta_{re}|\rangle}{\theta_e}\right)^{2/3}. \tag{10.64}$$

The angle brackets indicate global averages.

For Venus, $N_D \sim 3 \times 10^{-5}\,s^{-1}$, $\bar{N} \sim 7 \times 10^{-5}\,s^{-1}$, and the planet is under weak radiative control with a nearly adiabatic atmosphere. Substituting in (10.64) gives $\langle|\theta - \theta_e|\rangle \sim 0.1$ K. Observations confirm that there are only small latitudinal variations of temperature on Venus.

10.5. The Newtonian cooling approximation

10.5.1. *Transparent and boundary-exchange approximations*

We have made references to Newtonian cooling in two previous chapters: in § 2.4.2 and (2.120), in connection with the transparent approximation, and in § 6.4.1 and (6.69), in connection with boundary-exchange and cooling-to-space approximations. The transparent limit, $N(\infty)$, has already been derived, (10.17). This limit is valid under all circumstances for all disturbances, whether or not they are eigenmodes of the problem. The derivation is consistent with the integral formulation (10.25); since $\psi(m)$ is finite, by definition, the limit $m_1 \to 0$ can be reached only if $N \to N(\infty)$.

All other disturbances that we consider in this section are not form preserving but are imposed by extraneous circumstances. Consistency requires that, in each instance, radiative damping must be small enough that the form is not disturbed. An example is the boundary-exchange approximation, for which we assume a constant perturbation, $\psi(m) = \psi(m') = \text{constant}$. If this is substituted in (10.25), $\psi(m)$ can be taken outside the integral and we find

$$\frac{N^{BE}}{N(\infty)} = 1 - \frac{1}{2}\int_0^{m_1} K(m, m')\,dm'. \tag{10.65}$$

After some manipulation, and making use of the recurrence relations for exponential integrals (Appendix 6), (10.65) can be put into the form

$$\frac{N^{BE}}{N(\infty)} = \left(4\bar{k}_{\nu,P}\frac{dB}{d\theta}\right)^{-1}\int_0^{\infty}\frac{\partial B_\nu}{\partial\theta}\left[\frac{\partial T_\nu^f(z,\infty)}{\partial z} - \frac{\partial T_\nu^f(0,z)}{\partial z}\right]d\nu, \tag{10.66}$$

where $T_\nu^f(x) = 2E_3[\tau_\nu(x)]$.

Equation (10.66) can also be obtained from the boundary-exchange equations (6.65) and (6.66) if they are put into perturbation form. The radiation-to-space component, N^{SP}, is given in terms of transmission

functions in (6.70); it corresponds to the first term in square brackets in (10.66).

For (10.66) to be a useful approximation, we require either that $n^* \to 0$ (the condition under which it was obtained), or else that one of the terms in the square brackets be so large that internal exchange terms can be neglected in comparison; this latter condition requires that the transmission to one of the boundaries not be small.

10.5.2. *Internal gravity waves*

The damping of internal gravity waves in the middle atmosphere is a central concern for studies of the dynamic state of that region. Vertically propagating gravity waves, generated in the troposphere, carry fluxes of heat and momentum into the stratosphere. If these waves are damped, the fluxes diverge, and sources of heat and momentum are created.

Gravity waves have frequencies less than approximately $2 \times 10^{-2}\ \mathrm{s}^{-1}$ and, as the name implies, involve gravitational restoring forces. Radiative damping has been included in the thermodynamic equation by many investigators, but only in the Newtonian approximation, and only when this damping is small. Wavelengths are measured in hundreds or thousands of meters and are not negligible compared to a scale height. As a result, radiation exchange with the boundaries must be considered. A debate has centered around the question of whether it is sufficient to use a cooling-to-space term, (10.66), or whether scale-dependent terms are also involved. Numerical computations have resolved this debate in favor of the importance of scale-dependent terms.

We assume the form of the wave to be

$$\theta'(m, t) = \theta' \exp(i\omega t + in^* m), \tag{10.67}$$

where ω is real, but $n^*\ (= n_r^* + in_i^*)$ may be imaginary. The real and imaginary parts of the refractive index must be calculated in terms of the frequency, ω, and the atmospheric structure.

Numerical calculations of relaxation rates start from (10.21), or from an equivalent expression. The rate, $N'(n^*)$, is defined as a Newtonian coefficient by substituting (10.21) into (10.10), and using the harmonic disturbance (10.67). The result may be written

$$N(\infty) - N'(n^*, m) = \frac{2\pi q_a}{c_p}\left(\int_0^m dm' + \int_m^{m_1} dm'\right)\exp[in^*(m' - m)]$$
$$\times \int_0^\infty d\nu \frac{\partial B_\nu}{\partial \theta} E_1(|\tau_\nu(m') - \tau_\nu(m)|) k_{m,\nu}(m) k_{m,\nu}(m'). \tag{10.68}$$

The two integrals with respect to m' are over the atmosphere above and

below the level of interest, respectively. We write $N'(n^*, m)$ for the empirical rate to indicate that it is now a function of m and m_1, and not of n^* alone.

The numerical evaluation of (10.68) is straightforward, even including vertical variations of q_a and $\partial B_\nu/\partial\theta$ or, for that matter, a nonharmonic form for the disturbance. The equation can be written in many convenient forms; one that is suitable for our discussion uses the modified emissivity, ϵ^*; from (6.73) and (6.7)

$$N(\infty) - N'(n^*, m) = \frac{\pi q_a}{c_p}\left(\int_0^m dm' + \int_m^{m_1} dm'\right)\exp[in^*(m' - m)] \times \frac{dB}{d\theta}\frac{\partial^2\epsilon^*(m, m')}{\partial m\,\partial m'}. \tag{10.69}$$

Integration of (10.69) gives complex values for $N'(n^*, m)$; the perturbed heating and the perturbed temperature are not in phase. This difficulty is commonly avoided by taking the real part of (10.69), equivalent to using the perturbation $\cos n^*(m' - m)$.

The results of two calculations are shown in Fig. 10.11. One is for radiation-to-space alone while the other makes use of numerical integrations of equations similar to (10.69). The latter results, for an infinite wavelength, correspond well with the radiation-to-space calculation (the differences are probably accounted for by differences in the data used). However, even for a wavelength as long as 12.6 km, the radiation-to-space term is not dominant and the structure of the disturbance affects the relaxation rate. We examine these results more closely in terms of results obtained in previous sections.

For the infinite-domain result, (10.31), the two terms in (10.68) and (10.69) are equal. We now take the lower atmosphere to be semiinfinite and evaluate the term for the upper levels. We find

$$\frac{N(\infty)}{2} + \frac{\tilde{N}(n^*)}{2} - N'(n^*, m) = \frac{\pi q_a}{c_p}\int_0^m \exp[in^*(m' - m)]\frac{dB}{d\theta}\frac{\partial^2\epsilon^*(m, m')}{\partial m\,\partial m'}dm'. \tag{10.70}$$

Very high in the atmosphere, we may set $m = 0$ and

$$N'(n^*, m) = \frac{N(\infty) + \tilde{N}(n^*)}{2}. \tag{10.71}$$

The radiative rate is the average of the rates for the transparent upper layers and that for the semiinfinite lower atmosphere.

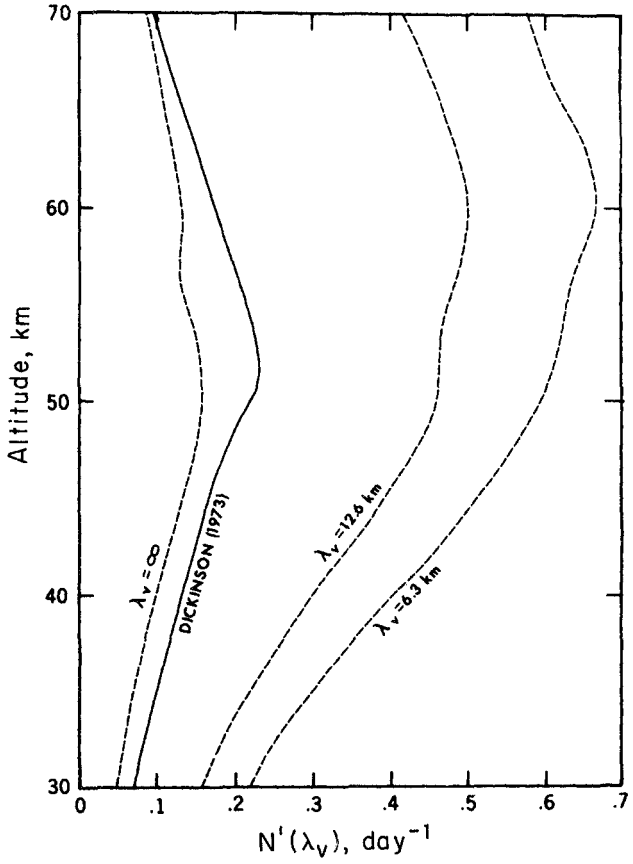


FIG. 10.11. Newtonian relaxation rates in the middle atmosphere. λ_v $(=2\pi/n)$ is the vertical wavelength. The broken lines are from a numerical calculation by Fels (1982). The solid line is a cooling-to-space calculation by Dickinson (1973).

The radiation-to-space limit is found by allowing n^* and $\tilde{N}(n^*)$ both to tend to zero. Using the result $\text{Lim}_{m\to 0}\, \partial\epsilon^*(m)/\partial m = 2\bar{k}_{\text{m,P}}$, (10.70) becomes

$$N'(0, m) = \frac{\pi q_a}{c_p}\frac{dB}{d\theta}\frac{\partial\epsilon^*(m, 0)}{\partial m}. \tag{10.72}$$

$N'(0, m)$ can be shown to equal N^{SP}, (10.66).

Approximations to (10.70) have been proposed. The following was used to draw the broken asymptotes in Fig. 10.6:

$$N'(n^*, m) = \tilde{N}(n^*) + N^{\text{SP}}(m). \tag{10.73}$$

Equation (10.73) is correct in the limits $m \to \infty$ and $n^* \to 0$. For $m \to 0$, $N^{SP} \to N(\infty)/2$ and (10.73) differs from (10.71), and is incorrect.

The harmonic disturbance, (10.67), is not a good representation of an upward-propagating gravity wave. Such waves are closer to being harmonic in a space coordinate rather than in the amount of absorber and, to first order, their amplitude varies as $\exp(z/2H)$, where H is the scale height of the neutral atmosphere. These features can be included in a numerical integration of (10.70). If they are not, certain limitations must be placed on the use of the results. The correct limits are complicated but they are expressed qualitatively by $\rho_a H \gg (n^*)^{-1}$ and $\rho_a H \gg \tilde{m}$. The first condition requires that the wavelength be less than the scale height; the second requires that the absorber amount above the level of interest exceed an amount, $\tilde{m}$, defined to be equal to the range of $(m - m')$ that includes most of the integrands in (10.68). The second condition appears to be satisfied for carbon dioxide in the middle atmosphere. The first is rendered less important by the dominance of the radiation-to-space term for long wavelengths.

10.6. Solar radiation in the middle atmosphere

Temperature dependence of chemical coefficients involved in the formation of ozone causes its steady-state density to vary with air temperature. This, in turn, affects the absorption of solar radiation and couples heating and temperature in ways different from that discussed in previous sections.

The volume absorption coefficient is proportional to the absorber density. If it is perturbed by the amount $k'_{v,\nu}$, the solar heating perturbation is, from (6.50),

$$\frac{h'_\nu(z)}{h_\nu(z)} = \frac{k'_{v,\nu}(z)}{k_{v,\nu}(z)} - \int_z^\infty \frac{k'_{v,\nu}(z')}{|\xi_\odot|} dz'. \tag{10.74}$$

$h_\nu(z)$ has the form of a Chapman layer, as shown in Fig. 6.9, for wavelengths that do not penetrate to the surface. The most important ozone absorption is between 2430 and 2770 Å in the Hartley bands and it gives rise to a Chapman layer with a maximum heating rate (for vertical incidence) close to 40 km.

The effect of temperature upon either the molecular or mass absorption coefficient is relatively small compared to the effect of temperature upon absorber density. We therefore attribute all variations of the volume absorption coefficient to variations in ozone density,

$$\frac{k'_{v,\nu}}{k_{v,\nu}} = \frac{\partial \ln \rho_a}{\partial \theta} \theta'. \tag{10.75}$$

If we use the harmonic perturbation (10.28), the photochemical relaxation rate becomes

$$N'_{\mathrm{P}}(n^*, m) = \frac{-h'_\nu(m)}{\rho c_{\mathrm{p}} \theta'(m)}$$

$$= -\frac{h'_\nu(m)}{\rho c_{\mathrm{p}}} \frac{\partial \ln \rho_{\mathrm{a}}}{\partial \theta}$$

$$\times \left\{1 - \frac{k_{\mathrm{m},\nu}(m)}{in^* |\xi_\odot|} [\exp(in^* m) - 1]\right\}. \tag{10.76}$$

The relaxation rate is scale dependent and imaginary, indicating phase differences between temperature and heating rate perturbations.

The in-phase component has the following transparent and opaque limits. As $n^* \to \infty$,

$$\mathcal{R}N'_{\mathrm{P}}(\infty) = -\frac{h_\nu}{\rho c_{\mathrm{p}}} \frac{\partial \ln \rho_{\mathrm{a}}}{\partial \theta}. \tag{10.77}$$

As $n^* \to 0$,

$$\mathcal{R}N'_{\mathrm{P}}(0) = -\frac{h_\nu}{\rho c_{\mathrm{p}}} \frac{\partial \ln \rho_{\mathrm{a}}}{\partial \theta} \left[1 - \frac{m k_{\mathrm{m},\nu}(m)}{|\xi_\odot|}\right]. \tag{10.78}$$

The relaxation rate, (10.78), changes sign when $mk_{\mathrm{m},\nu}(m)/|\xi_\odot| = 1$, the same level at which the Chapman function has its maximum. This sign change is associated with the atmospheric absorption above the level under consideration; an increase in absorption coefficient decreases the solar radiation available for heating, which may offset the increase of heating caused by a *local* increase of absorption coefficient. The second term in square brackets in (10.78) has been called the *extinction term.*

The nature of the ozone perturbation may be discussed in terms of a simplified chemical scheme involving only the oxygen allotropes, as shown in Table 10.2.

For a steady state the ozone density is given by

$$[\mathrm{O}_3] = [\mathrm{O}_2] \left(\frac{J_1 \kappa_{2,1} [\mathrm{M}]}{J_3 \kappa_{3,1}}\right)^{1/2}. \tag{10.79}$$

The rate constants are very sensitive to temperature,

$$\frac{\kappa_{2,1}}{\kappa_{3,1}} \propto \theta^{-2.3} \exp\left(\frac{2060}{\theta}\right). \tag{10.80}$$

Table 10.2. Ozone photochemistry[a]

	Reaction	Rate of reaction
1.	$O_2 + h\nu_2 \rightarrow 2O$	$J_1[O_2]$
2.	$O_2 + O + M \rightarrow O_3 + M$	$\kappa_{2,1}[O][O_2][M]$
3.	$O_3 + h\nu_3 \rightarrow O_2 + O$	$J_3[O_3]$
4.	$O_3 + O \rightarrow 2O_2$	$\kappa_{3,1}[O][O_3]$

[a] The square brackets indicate number densities. J_1 and J_3 are photodissociation rates per molecule. $\kappa_{2,1}$ and $\kappa_{3,1}$ are reaction rate coefficients.

If the photochemical reactions take place quickly, so that equilibrium is maintained, (10.79) and (10.80) lead to a negative value for $\partial \ln \rho_a / \partial \theta$ and a positive photochemical relaxation rate (if the extinction is small). Thermal relaxation by carbon dioxide occurs simultaneously and the two rates must be added to give the net relaxation rate. Most calculations suggest that the photochemical relaxation rate is of similar magnitude to the radiation-to-space rate in the neighborhood of 50 km. Figure 10.12 shows the radiation-to-space rate alone and combined with

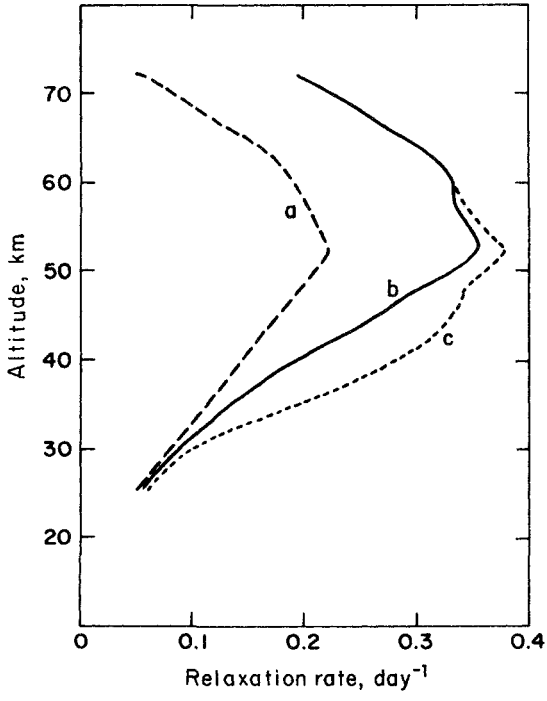


FIG. 10.12. Relaxation rates for $n^* = 0$. (a) Thermal relaxation alone (cf. Fig. 10.11). (b) Thermal plus photochemical relaxation. (c) As for (b), but without the extinction term. After Hartmann (1978).

the photochemical rate, with and without the extinction effect. This figure gives a good impression of the magnitudes of the different rates under discussion.

In the foregoing discussion, we assumed photochemical reactions to be instantaneous, but the reaction rates depend strongly upon the intensity of solar radiation and rates may be slow at low altitudes. Above 45 km, a quasistatic treatment is a good approximation; below 35 km ozone behaves more like an inert gas. The simplest treatment assumes that there is a single rate-limiting step in the chemistry. Then we may write

$$\frac{\partial \rho_a'}{\partial t} = -\alpha_{\rho,\theta}\theta' - N_\rho \rho_a' \tag{10.81}$$

$$\frac{\partial \theta'}{\partial t} = \beta_{\rho,\theta}\rho_a' - N'(n^*)\theta'. \tag{10.82}$$

In these equations, N_ρ is the limiting chemical rate and N' is the thermal relaxation rate, discussed in previous sections. $\alpha_{\rho,\theta}$ and $\beta_{\rho,\theta}$ are density–temperature interaction coefficients, representing the effect of temperature upon density and the effect of density on temperature (through the volume absorption coefficient), respectively.

The quasistatic case is represented by the limit N_ρ large. Equations (10.81) and (10.82) becomes

$$\rho_a' = -\frac{\alpha_{\rho,\theta}}{N_\rho}\theta', \tag{10.83}$$

$$\frac{\partial \theta'}{\partial t} = -[N_P'(n^*) + N'(n^*)]\theta', \tag{10.84}$$

where

$$N_P' = \frac{\alpha_{\rho,\theta}\beta_{\rho,\theta}}{N_\rho}. \tag{10.85}$$

In the general case, however, (10.81) and (10.82) must be solved as simultaneous, second-order differential equations with complex solutions for the temperature and density perturbations.

In addition to these photochemical relaxation effects, ozone is involved in another coupling between temperature perturbations and heating perturbations through the motions that occur in gravity waves, and the effect of those motions on the ozone density. If the ozone mixing ratio varies with height, vertical motions convect ozone and change the solar heating. The coupling can be made explicit only in a dynamic model, a description of which would take us beyond the scope of this chapter; what follows is a summary of one approach.

The linear equations governing propagating gravity waves can be solved in terms of an attenuation coefficient (amplitude varies as $\exp[-\alpha z]$). This coefficient is proportional to the combined thermal and photochemical rates,

$$\alpha = \frac{n_h N_B}{2\hat{\omega}(\hat{\omega}^2 - f^2)^{1/2}} [N'(n^*) + N'_{P,1} + N'_{P,2}]. \tag{10.86}$$

In (10.86), n_h is the horizontal wave number, N_B is the Brunt–Väisälä frequency, $\hat{\omega}$ is the Doppler-shifted wave frequency, and f is the Coriolis parameter. There are now two photochemical rates,

$$N'_{P,1} = \frac{\beta_{\rho,\theta}(\partial q_a/\partial z)\rho}{(\partial\theta/\partial z + \Gamma)[1 + (N_\rho/\hat{\omega})^2]}, \tag{10.87}$$

$$N'_{P,2} = \frac{\alpha_{\rho,\theta}\beta_{\rho,\theta}}{N_\rho[1 + (\hat{\omega}/N_\rho)^2]}. \tag{10.88}$$

q_a is the absorber mixing ratio and Γ is the adiabatic lapse rate. Both of these rates involve the interaction coefficients introduced in (10.81) and (10.82). The second term, $N'_{P,2}$ is, apart from a frequency-dependent attenuation, the same as the photochemical relaxation rate in (10.85). $N'_{P,1}$ is a dynamic rate. It is proportional to the vertical gradient of the mixing ratio and would be zero in a well-mixed atmosphere.

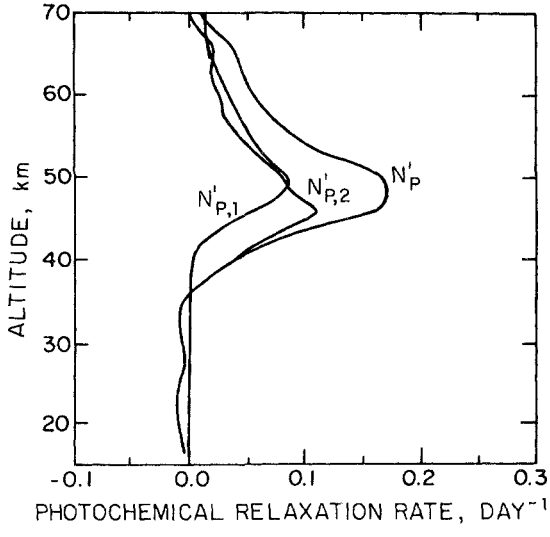


FIG. 10.13. Photochemical and dynamic relaxation rates. Midlatitude, summer conditions with $\hat{\omega} = 10^{-4}\,s^{-1}$. After Zhu and Holton (1986).

Results of a calculation of $N'_{P,1}$ and $N'_{P,2}$ are shown in Fig. 10.13 for a wave frequency of $10^{-4}\,s^{-1}$. The extinction term was omitted from these calculations but, where N'_P is significant, it is of little importance. For comparison, see the term N' in Fig. 10.12a. All of the effects discussed in this section appear to have some importance at this wave frequency.

BIBLIOGRAPHY

10.1. Introduction

A recent book on the astrophysical aspects of this subject is,

Mihalas, D., and Mihalas, B. W., 1984, *Foundations of radiation hydrodynamics.* New York: Oxford University Press.

The technique of linearizing the radiative transfer equations was first proposed by

Spiegel, E. A., 1957, "The smoothing of temporal fluctuations by radiative transfer," *Astrophys. J.* **126,** 202.

10.2. The radiation eigenvalue problem

10.2.1. *The integral equation*
The treatment given here follows

Gay, C., and Thomas, G. E., 1981, "Radiative temperature dissipation in a finite atmosphere—I. The homogeneous case," *J. Quant. Spectrosc. Radiat. Transfer* **25,** 351.

This paper contains references to the extensive literature dealing with neutron diffusion, for which subject these methods are in common use.

10.2.2. *Spiegel's solution*
See Spiegel (1957, § 10.1).

10.2.3. *Two-stream solution for a scattering atmosphere*
The theory is slightly modified from that given by

Coantic, M., and Simonin, O., 1984, "Radiative effects on turbulent temperature spectra and budgets in the planetary boundary layer," *J. Atmos. Sci.* **41,** 2629.

10.2.4. *Effect of boundaries*
Calculations of the effect of boundaries upon the decay of a given thermal perturbation are referenced in § 10.5. The discussion here concerns the eigenvalue problem and the decay of an arbitrary thermal disturbance. The treatment follows Gay and Thomas (1981, § 10.2.1).

10.3. Numerical results

10.3.1. $N_i(\infty)$ *for atmospheric bands*

10.3.2. *Special absorption laws*
The calculation for an Elsasser band, shown in Fig. 10.3, was obtained using a

diffusivity approximation by

Sasamori, T., and London, J., 1966, "The decay of small temperature perturbations by thermal radiation in the atmosphere," *J. Atmos. Sci.* **23,** 543.

An exponential band model that, for strong absorption, leads to a logarithmic absorption law has been treated by

Nelson, D. A., and Edwards, D. K., 1973, "Radiative decay times for thermal perturbations in a nongrey medium", *J. Quant. Spectrosc. Radiat. Transfer* **13,** 747.

10.3.3. *Radiative relaxation for earth and Mars*
Figure 10.7 is from Coantic and Simonin (1984, § 10.2.3). The data on droplets that are used in this figure are from

Irvine, W. M., and Pollack, J. B., 1968, "Infrared optical properties of water and ice spheres," *Icarus* **8,** 324.

Figure 10.8 is from Goody and Belton (1967, § 4.9).

10.3.4. *Nonequilibrium source functions*
The first treatment of this problem was by

Hodges, R. R., 1969, "Radiative decay in carbon dioxide," *J. Geophys. Res.* **74,** 862.

This paper demonstrated that the source function is quasistatic for the thermal decay problem. Similar conclusions were reached by

Shved, G. M., and Utyakovskiy, D. P., 1983, "Radiative damping of temperature perturbations in the earth's upper atmosphere, taking into account the breaking-up of local thermodynamic equilibrium," *Atmos. Oc. Phys.* **19,** 353.

A quasistatic treatment was employed by

Fels, S. B., 1984, "The radiative damping of short scale vertical waves in the mesosphere," *J. Atmos. Sci.* **41,** 1755.

All of the time dependencies of the problem are displayed explicitly by

Shved, G. M., 1978, "Influence of radiative transfer on certain types of motions in planetary atmospheres," *Adv. Heat Transfer* **14,** 249.

10.4. Planetary-scale relaxation

10.4.1. *The planetary relaxation rate*

10.4.2. *The temperature of a nonrotating atmosphere*
These sections follow

Gierasch, P., Goody, R., and Stone, P. 1970, "The energy balance of planetary atmospheres," *Geophys. Fluid Dyn.* **1,** 1.

A similar treatment of temperature gradients in a rotating atmosphere is

Stone, P. H., 1972, "A simplified radiative-dynamical model for the static stability of rotating atmospheres," *J. Atmos. Sci.* **29,** 405.

10.5. The Newtonian cooling approximation

10.5.1. *Transparent and boundary-exchange approximations*

The Newtonian cooling approximation has been rediscovered many times; its simplicity has made it almost irresistible for analytical investigations of internal gravity waves. One of the first uses of the boundary-exchange approximation was described by

Lindzen, R., and Goody, R., 1965, "Radiative and photochemical processes in mesospheric dynamics, Part I. Models for radiative and photochemical processes," *J. Atmos. Sci.* **22,** 341.

The cooling-to-space approximation is one aspect of the influence of boundaries upon the decay of a thermal perturbation. The following are some more formal studies of boundary effects:

Prinn, R. G., 1977, "On the radiative damping of atmospheric waves," *J. Atmos. Sci.* **34,** 1386.

Nelson, D. A., 1979, "Thermal radiation relaxation times near the boundary of a semi-infinite nongrey medium," *J. Quant. Spectrosc. Radiat. Transfer* **21,** 99.

Traugott, S. C., 1977, "Infrared cooling rates for two-dimensional thermal perturbations in a nonuniform atmosphere," *J. Atmos. Sci.* **34,** 863.

10.5.2. *Internal gravity waves*

Shved (1978, § 10.1.5) attempts to treat the problem of radiative decay of acoustic-gravity waves in general terms.

One question not treated here is the radiative damping of acoustic waves, see

Stein, R. F., and Spiegel, E. A., 1967, "Radiative damping of sound waves," *J. Acoust. Soc. Am.* **42,** 866.

Gille, J. C., 1968, "Acoustic wave propagation in a non-grey radiating atmosphere," *J. Atmos. Sci.* **25,** 808.

The following paper also uses the acoustic equations including radiative damping, but applies the theory to the propagation of a step-function in fluid properties:

Olfe, D. B., and DePlomb, E. P., 1970, "The decay of perturbations in a radiating gas," *J. Fluid Mech.* **40,** 127.

The first integrations of equations of the type (10.68) were by Sasamori and London (1966, § 10.3.2). The most complete study is

Fels, S. B., 1982, "A parameterization of scale-dependent radiative damping rates in the middle atmosphere," *J. Atmos. Sci.* **39,** 1141.

The cooling-to-space curve in Fig. 10.11 is from

Dickinson, R. E., 1973, "Method of parameterization for infrared cooling between altitudes of 30 and 70 km," *J. Geophys. Res.* **78,** 4451.

10.6. Solar Radiation in the middle atmosphere

The feedback caused by temperature-dependent solar absorption coefficients was first identified by

Craig, R. A., and Ohring, G., 1958, "The temperature dependence of ozone radiational heating rates in the vicinity of the mesopeak," *J. Meteorol.* **15,** 59.

The extinction effect was first discussed by

Hartmann, D. L., 1978, "A note concerning the effect of varying extinction on radiative-photochemical relaxation," *J. Atmos. Sci.* **35,** 1125.

Strobel, D. F., 1978, "Photo-chemical-radiative damping and instability in the stratosphere," *Geophys. Res. Lett.* **4,** 424.

The linear equations (10.81) and (10.82) were first used by Lindzen and Goody (1965, § 10.5.1). A more detailed scheme has been given by

Blake, D., and Lindzen, R. S., 1973, "Effects of photochemical models on calculated equilibria and cooling rates in the stratosphere," *Mon. Wea. Rev.* **101,** 783.

The photochemical–dynamic relaxation associated with gravity waves is discussed by

Zhu, X., and Holton, J. R., 1986, "Photochemical damping of inertio-gravity waves," *J. Atmos. Sci.* **43,** 2578.

Observational support for some of these ideas is given by

Ghazi, A., Wang, P.-H., and McCormick, M. P., 1985, "A study on radiative damping of planetary waves utilizing stratospheric observations," *J. Atmos. Sci.* **42,** 2032.

APPENDIX 1

PHYSICAL CONSTANTS

General constants

Stefan–Boltzmann constant, $\boldsymbol{\sigma} = 5.67032 \times 10^{-8}\ \mathrm{W\ m^{-2}\ K^{-4}}$
Speed of light *in vacuo*, $\mathbf{c} = 2.997924580 \times 10^{8}\ \mathrm{m\ s^{-1}}$
Boltzmann constant, $\mathbf{k} = 1.380662 \times 10^{-23}\ \mathrm{J\ K^{-1}}$
Planck constant, $\mathbf{h} = 6.626176 \times 10^{-34}\ \mathrm{J\ s}$
Radiation constants:

$$C_1 = 2\pi\mathbf{hc}^2 = 3.741832 \times 10^{-16}\ \mathrm{W\ m^2}$$

$$C_2 = \mathbf{hc/k} = 1.438786 \times 10^{-2}\ \mathrm{m\ K}$$

Atomic mass unit, amu $= 1.6605654 \times 10^{-27}$ kg
Avogadro number, $\mathbf{L} = 6.022045 \times 10^{23}\ \mathrm{g\text{-}mol^{-1}}$
Molar gas constant, $\mathbf{R} = 8.31441\ \mathrm{J\ K^{-1}\ g\text{-}mol^{-1}}$
Triple-point temperature of water, $\theta_s = 273.16$ K

Sun

Equatorial radius $= 6.599 \times 10^{8}$ m
Effective emission temperature $= 5783$ K
Mean angle subtended by photosphere at earth $= 31.988$ arc min

Earth

Equatorial radius $= 6.378388 \times 10^{6}$ m
Mean earth–sun distance, AU $= 1.4960 \times 10^{11}$ m
Eccentricity of orbit $= 0.016750$
Inclination of rotation axis $= 23.45$ deg
Standard surface gravity, $g = 9.80665\ \mathrm{m\ s^{-2}}$

Earth's atmosphere

Standard surface pressure, $p_s = 1.01325$ bar
Loschmidt number, $n_s = 2.686754 \times 10^{25}\ \mathrm{m^{-3}}$ (at s.t.p.)
Density at s.t.p. $\rho_s = 1.2925\ \mathrm{kg\ m^{-3}}$
Specific gas constant, $R_a = 2.8700 \times 10^{2}\ \mathrm{J\ K^{-1}\ kg^{-1}}$

Mean molecular mass, $M = 28.964$ amu
Specific heats at s.t.p.:

$$c_p = 1.006 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$$
$$c_v = 7.18 \times 10^2 \text{ J kg}^{-1} \text{ K}^{-1}$$
$$\gamma = c_p/c_v = 1.401$$

The above data are for dry air. For data on water vapor, see Appendix 4.

BIBLIOGRAPHY

Kaye, G. W. C., and Laby, T. H., 1973, *Tables of physical and chemical constants and some mathematical functions, 14th Edition.* London: Longman.

Allen, C. W., 1973, *Astrophysical quantities.* London: The Athlone Press.

APPENDIX 2

SPECTROSCOPIC UNITS

Wavelength

The use of wavelength, Table A.2.1, to specify the quality of electromagnetic radiation suffers from the defect that the wavelength is inversely proportional to the refractive index of the medium in which it is measured. Wavelengths are commonly corrected to vacuum, but the correction is negligible for most atmospheric studies.

Table A.2.1. Wavelength units

Unit	Symbol	Size	Spectral region
Ångström unit	Å	10^{-10} m	Soft X-ray, ultraviolet, visible
Nanometer	nm	10^{-9} m	
Micron	μm	10^{-6} m	Visible, infrared
Millimeter	mm	10^{-3} m	Radio waves
Centimeter	cm	10^{-2} m	
Meter	m	1 m	

Wavenumber and frequency

Since the energy of a photon is proportional to its frequency, it is convenient to perform spectral analysis in frequency rather than wavelength units. Frequency is not a function of the medium in which it is measured and there is no confusion as to whether vacuum corrections have or have not been applied. The symbol, ν, is universal and the unit is the *hertz* (Hz or s^{-1}).

The hertz, kilohertz, megahertz, etc. are used by radio engineers. Spectroscopists, however, have shown a strange aversion to the use of this simple and unambiguous unit, preferring the *wavenumber* or reciprocal wavelength, ν/c'. This practice is too widely adopted to attempt to go against it; however, two difficulties should be noted. First, the refractive index of the medium has reappeared through its influence on c', the local speed of light. Second, there is no general agreement as to the name or unit for this quantity. Some authors use the symbol ν for both frequency and wavenumber and refer to both as "frequencies." But

quantities with different physical dimensions deserve distinct symbols and units, and there have been attempts to introduce the symbol, k, and the unit *kayser* for wavenumber. This proposal has not been widely adopted, however, and the symbol, k, is generally preempted in quantitative spectroscopy to denote the absorption coefficient. We shall take the term wavenumber to imply its vacuum value, and we shall use the symbol $\nu/\mathbf{c}$ and the unit cm^{-1}.

Energy

One of the great achievements of physical science has been the recognition of the unity of different forms of energy. For all, one unit (the joule) should suffice. But the trend of physics has been to adopt a different unit of energy for each field of study; nowhere is this profusion of units richer than in spectroscopy. The four units in Table A.2.2 are not the only units that may be encountered; of the four, only the joule is defined in a straightforward way. The energy unit, cm^{-1}, denotes the energy (**hc**) of a photon of unit wavenumber; the electron volt (eV) is the electrical potential energy of an electron with respect to a potential drop of 1 volt; **k** kelvin is the thermal energy ($\mathbf{k}\theta$) associated with two degrees of freedom of a molecule at a temperature of 1 K.

Table A.2.2. Numerical conversion factors for energy units

Unit	cm^{-1}	joule	eV	k kelvin
1 cm^{-1}	1	1.986478×10^{-23}	1.239852×10^{-4}	1.438786
1 joule	5.034036×10^{22}	1	6.241460×10^{18}	7.242902×10^{22}
1 eV	8.065478×10^{3}	1.602189×10^{-19}	1	1.160450×10^{4}
1 **k** kelvin	6.950302×10^{-1}	1.380662×10^{-23}	8.617346×10^{-5}	1

Extinction coefficient

In Chapter 2, the *extinction coefficient* is defined in four different ways according to the definition of the amount of matter in the optical path. Table A.2.3 shows the dimensions of these quantities and the relationships between them (cgs units are used because they are more commonly used in the literature than are MKS).

Table A.2.3. Dimensions and conversion factors for extinction coefficients[a]

Symbol	e_v	e_m	e_n	e_s
Name	Volume e.c.	Mass e.c.	Molecular e.c.	e.c. per cm s.t.p.
Dimensions	cm^{-1}	$g^{-1}\,cm^2$	cm^2	cm^{-1}
$e_v = 1$	1	ρ^{-1}	n^{-1}	n_s/n
$e_m = 1$	ρ	1	m	ρ_s
$e_n = 1$	n	m^{-1}	1	n_s
$e_s = 1$	n/n_s	ρ_s^{-1}	n_s^{-1}	1

[a] ρ = density of absorbing gas ($g\,cm^{-3}$).

ρ_s = density of absorbing gas at s.t.p. ($g\,cm^{-3}$).

n = molecular number density (cm^{-3}).

n_s = molecular number density at s.t.p. (Loschmidt number, cm^{-3}).

m = molecular mass (g).

The volume extinction coefficient, e_v, is the quantity entering atmospheric calculations when distance is the independent variable. Data cannot be presented in this unit, however, without specifying the density of absorbing material and tabulations must use one of the other three units. Of these, e_s is the popular choice of spectroscopists. It is an unfortunate unit because, through Loschmidt's number, it contains a standard temperature in its definition, even when the measurements may be made at an entirely different temperature. Thus, it is possible to encounter such ambiguous statements as "the extinction coefficient per centimeter at s.t.p., at a temperature of 400 K, is . . . ." This same coefficient is also given in units of (cm-Amagat)$^{-1}$. The Amagat is a dimensionless measure of density in terms of the density at s.t.p. (ρ/ρ_s). For the highest precision, the conversion takes account of the small variation in Loschmidt's number from one gas to another, but the distinction is not important for atmospheric applications.

The most satisfactory unit is e_n (unit, cm^2), and its alternative name, *extinction cross section,* conveys a valuable picture of the extinction process in terms of the collision cross section between photon and absorbing molecule. In Chapter 7, we also use e_n to mean the extinction coefficient per particle, for the case when the absorbing matter is aggregated into particles. The relationships in Table A.2.3 remain the same except that n is now the particle rather than the molecular density.

Some authors use decadic coefficients,

$$e' = e/2.3026. \tag{A.2.1}$$

Related to the decadic coefficient is the power loss in db km^{-1},

$$e\ (\text{db km}^{-1}) = 10^6 e'_v\ (\text{cm}^{-1}). \qquad \text{(A.2.2)}$$

Band and line intensities

The integrated absorption coefficient over a line or band has the dimensions of an extinction coefficient times a wavelength, a wavenumber, or a frequency. e_m, e_n, and e_s are equally suitable for the extinction coefficient and, as a consequence, we have nine different definitions of intensity; fortunately, these may be readily distinguished because they all have different dimensions (Table A.2.4).

Table A.2.4. Physical dimensions of intensities

	e_m (g^{-1} cm^2)	e_n (cm^2)	e_s (cm^{-1})
Wavelength, cm	$g^{-1}\,cm^3$	cm^3	1
Wavenumber, cm^{-1}	$g^{-1}\,cm$	cm	cm^{-2}
Frequency, Hz	$g^{-1}\,cm^2\,s^{-1}$	$cm^2\,s^{-1}$	$cm^{-1}\,s^{-1}$

APPENDIX 3

A MODEL ATMOSPHERE

Table A.3.1. U.S. Standard Atmosphere, 0 to 30 km

Altitude (km)	Temperature (K)	Pressure (mb)	Density ($g\ cm^{-3}$)	Pressure scale height (km)	Number density (cm^{-3})	Molecular weight (amu)
0	288.150	1.01325×10^{3}	1.2250×10^{-3}	8.4345	2.5470×10^{19}	28.964
2	275.154	7.9501×10^{2}	1.0066×10^{-3}	8.0592	2.0928×10^{19}	28.964
4	262.166	6.1660×10^{2}	8.1935×10^{-4}	7.6836	1.7036×10^{19}	28.964
6	249.187	4.7217×10^{2}	6.6011×10^{-4}	7.3078	1.3725×10^{19}	28.964
8	236.215	3.5651×10^{2}	5.2579×10^{-4}	6.9317	1.0932×10^{19}	28.964
10	223.252	2.6499×10^{2}	4.1351×10^{-4}	6.5554	8.5976×10^{18}	28.964
12	216.650	1.9399×10^{2}	3.1194×10^{-4}	6.3656	6.4857×10^{18}	28.964
14	216.650	1.4170×10^{2}	2.2786×10^{-4}	6.3696	4.7375×10^{18}	28.964
16	216.650	1.0352×10^{2}	1.6647×10^{-4}	6.3736	3.4612×10^{18}	28.964
18	216.650	7.5652×10^{1}	1.2165×10^{-4}	6.3776	2.5292×10^{18}	28.964
20	216.650	5.5293×10^{1}	8.8910×10^{-5}	6.3816	1.8486×10^{18}	28.964
22	218.574	4.0475×10^{1}	6.4510×10^{-5}	6.4423	1.3413×10^{18}	28.964
24	220.560	2.9717×10^{1}	4.6938×10^{-5}	6.5049	9.7591×10^{17}	28.964
26	222.544	2.1883×10^{1}	3.4257×10^{-5}	6.5675	7.1225×10^{17}	28.964
28	224.527	1.6161×10^{1}	2.5076×10^{-5}	6.6302	5.2138×10^{17}	28.964
30	226.509	1.1970×10^{1}	1.8410×10^{-5}	6.6929	3.8278×10^{17}	28.964

Table A.3.2. U.S. Standard Atmosphere, 30 to 100 km

Altitude (km)	Temperature (K)	Pressure (mb)	Density ($g\ cm^{-3}$)	Pressure scale height (km)	Number density (cm^{-3})	Molecular weight (amu)
30	226.509	1.1970×10^{1}	1.8410×10^{-5}	6.6929	3.8278×10^{17}	28.964
35	236.513	5.7459	8.4634×10^{-6}	6.9995	1.7597×10^{17}	28.964
40	250.350	2.8714	3.9957×10^{-6}	7.4206	8.3077×10^{16}	28.964
45	264.164	1.4910	1.9663×10^{-6}	7.8423	4.0882×10^{16}	28.964
50	270.650	7.9779×10^{-1}	1.0269×10^{-6}	8.0474	2.1351×10^{16}	28.964
55	260.771	4.2525×10^{-1}	5.6810×10^{-7}	7.7658	1.1812×10^{16}	28.964
60	247.021	2.1958×10^{-1}	3.0968×10^{-7}	7.3678	6.4387×10^{15}	28.964
65	233.292	1.0929×10^{-1}	1.6321×10^{-7}	6.9691	3.3934×10^{15}	28.964
70	219.585	5.2209×10^{-2}	8.2829×10^{-8}	6.5699	1.7222×10^{15}	28.964
75	208.399	2.3881×10^{-2}	3.9921×10^{-8}	6.2449	8.3003×10^{14}	28.964
80	198.639	1.0524×10^{-2}	1.8458×10^{-8}	5.9617	3.8378×10^{14}	28.964
85	188.893	4.4568×10^{-3}	8.2196×10^{-9}	5.6780	1.7090×10^{14}	28.964
90	186.870	1.8359×10^{-3}	3.4160×10^{-9}	5.6360	7.1160×10^{13}	28.910
95	188.420	7.5966×10^{-4}	1.3930×10^{-9}	5.7270	2.9200×10^{13}	28.730
100	195.080	3.2011×10^{-4}	5.6040×10^{-10}	6.0090	1.1890×10^{13}	28.400

Table A.3.3. U.S. Standard Atmosphere, 100 to 1000 km

Altitude (km)	Temperature (K)	Pressure (mb)	Density ($g\ cm^{-3}$)	Pressure scale height (km)	Number density (cm^{-3})	Molecular weight (amu)
100	195.080	3.2011×10^{-4}	5.6040×10^{-10}	6.0090	1.1890×10^{13}	28.400
120	360.000	2.5382×10^{-5}	2.2220×10^{-11}	12.0910	5.1070×10^{11}	26.200
140	559.630	7.2028×10^{-6}	3.8310×10^{-12}	20.0250	9.3220×10^{10}	24.750
160	696.290	3.0395×10^{-6}	1.2330×10^{-12}	26.4140	3.1620×10^{10}	23.490
180	790.070	1.5271×10^{-6}	5.1940×10^{-13}	31.7030	1.4000×10^{10}	22.340
200	854.560	8.4736×10^{-7}	2.5410×10^{-13}	36.1830	7.1820×10^{9}	21.300
300	976.010	8.7704×10^{-8}	1.9160×10^{-14}	51.1930	6.5090×10^{8}	17.730
400	995.830	1.4518×10^{-8}	2.8030×10^{-15}	59.6780	1.0560×10^{8}	15.980
500	999.240	3.0236×10^{-9}	5.2150×10^{-16}	68.7850	2.1920×10^{7}	14.330
600	999.850	8.2130×10^{-10}	1.1370×10^{-16}	88.2440	5.9500×10^{6}	11.510
800	999.990	1.7036×10^{-10}	1.1360×10^{-17}	193.8620	1.2340×10^{6}	5.540
1000	1000.000	7.5138×10^{-11}	3.5610×10^{-18}	288.2030	5.4420×10^{5}	3.940

BIBLIOGRAPHY

U.S. Standard Atmosphere, 1976, *Publication NOAA-S/T*76-1562. Washington D.C.: U.S. Government Printing Office.

APPENDIX 4

PROPERTIES OF WATER VAPOR

Molecular mass = 18 amu
Density at s.t.p., $\rho_w = 8.031 \times 10^{-1}$ kg m^{-3}
Specific gas constant, $R_w = 4.618 \times 10^2$ J K^{-1} kg^{-1}

Table A.4.1. Saturated vapor pressure (mb) over pure liquid water

°C	−30	−20	−10	0	+10	+20	+30	+40
0	0.5088	1.2540	2.8627	6.1078	12.272	23.373	42.430	73.777
+1	0.5589	1.3664	3.0971	6.5662	13.119	24.861	44.927	77.802
+2	0.6134	1.4877	3.3484	7.0547	14.017	26.430	47.554	82.015
+3	0.6727	1.6186	3.6177	7.5753	14.969	28.086	50.307	86.423
+4	0.7371	1.7597	3.9061	8.1294	15.977	29.831	53.200	91.034
+5	0.8070	1.9118	4.2148	8.7192	17.044	31.671	56.236	95.855
+6	0.8827	2.0755	4.5451	9.3465	18.173	33.608	59.422	100.89
+7	0.9649	2.2515	4.8981	10.013	19.367	35.649	62.762	106.16
+8	1.0538	2.4409	5.2753	10.722	20.630	37.796	66.264	111.66
+9	1.1500	2.6443	5.6780	11.474	21.964	40.055	69.934	117.40

BIBLIOGRAPHY

Publication No. 79 of the International Meteorological Organization, 1961, *Definitions and specifications of water vapour in the atmosphere.*

Table A.4.2. Saturated vapor pressure (mb) over pure ice

°C	−100	−90	−80	−70	−60	−50	−40	−30	−20	−10	0
0	1.403×10^{-5}	9.672×10^{-5}	5.472×10^{-4}	2.618×10^{-3}	1.080×10^{-2}	3.935×10^{-2}	1.283×10^{-1}	3.798×10^{-1}	1.032	2.597	6.107
+1	1.719×10^{-5}	1.160×10^{-4}	6.444×10^{-4}	3.032×10^{-3}	1.236×10^{-2}	4.449×10^{-2}	1.436×10^{-1}	4.213×10^{-1}	1.135	2.837	
+2	2.101×10^{-5}	1.388×10^{-4}	7.577×10^{-4}	3.511×10^{-3}	1.413×10^{-2}	5.026×10^{-2}	1.606×10^{-1}	4.669×10^{-1}	1.248	3.097	
+3	2.561×10^{-5}	1.658×10^{-4}	8.894×10^{-4}	4.060×10^{-3}	1.612×10^{-2}	5.671×10^{-2}	1.794×10^{-1}	5.170×10^{-1}	1.371	3.379	
+4	3.117×10^{-5}	1.977×10^{-4}	1.042×10^{-3}	4.688×10^{-3}	1.838×10^{-2}	6.393×10^{-2}	2.002×10^{-1}	5.720×10^{-1}	1.506	3.685	
+5	3.784×10^{-5}	2.353×10^{-4}	1.220×10^{-3}	5.406×10^{-3}	2.092×10^{-3}	7.198×10^{-2}	2.233×10^{-1}	6.323×10^{-1}	1.652	4.015	
+6	4.584×10^{-5}	2.796×10^{-4}	1.425×10^{-3}	6.225×10^{-3}	2.380×10^{-2}	8.097×10^{-2}	2.488×10^{-1}	6.985×10^{-1}	1.811	4.372	
+7	5.542×10^{-5}	3.316×10^{-4}	1.662×10^{-3}	7.159×10^{-3}	2.703×10^{-2}	9.098×10^{-2}	2.769×10^{-1}	7.709×10^{-1}	1.984	4.757	
+8	6.685×10^{-5}	3.925×10^{-4}	1.936×10^{-3}	8.228×10^{-3}	3.067×10^{-2}	1.021×10^{-1}	3.079×10^{-1}	8.502×10^{-1}	2.172	5.173	
+9	8.049×10^{-5}	4.638×10^{-4}	2.252×10^{-3}	9.432×10^{-3}	3.476×10^{-2}	1.145×10^{-1}	3.421×10^{-1}	9.370×10^{-1}	2.376	5.623	

APPENDIX 5

THE PLANCK FUNCTION

Table A.5.1. B_λ, λB_λ, and B_ν[a]

$\lambda\theta$ (mK)	$x = C_2/\lambda\theta$	$\int_0^\lambda B_\lambda\, d\lambda/B$	$B_\lambda/(B_\lambda)_{max}$	$\lambda B_\lambda/(\lambda B_\lambda)_{max}$	$B_\nu/(B_\nu)_{max}$
	Large x	$x^3e^{-x}/6.49394$	$x^5e^{-x}/21.20144 \times 10$	$x^4e^{-x}/4.779841$	$x^3e^{-x}/1.421436$
1.0×10^{-4}	1.438786×10^{2}	1.53064×10^{-57}	9.504319×10^{-54}	2.930061×10^{-55}	6.848045×10^{-57}
5.0×10^{-4}	2.877573×10^{1}	1.29849×10^{-9}	2.962286×10^{-7}	4.566176×10^{-8}	5.335962×10^{-9}
1.0×10^{-3}	1.438786×10^{1}	3.20745×10^{-4}	1.640767×10^{-2}	5.058275×10^{-3}	1.182204×10^{-3}
2.0×10^{-3}	7.193932	6.67279×10^{-2}	6.831367×10^{-1}	4.212047×10^{-1}	1.968853×10^{-1}
3.0×10^{-3}	4.795954	2.73225×10^{-2}	9.971416×10^{-1}	9.222183×10^{-1}	6.466139×10^{-1}
5.0×10^{-3}	2.877573	6.33722×10^{-1}	5.548877×10^{-1}	8.553239×10^{-1}	9.995179×10^{-1}
1.0×10^{-3}	1.438786	9.14156×10^{-1}	9.043926×10^{-2}	2.788128×10^{-1}	6.516326×10^{-1}
2.0×10^{-2}	7.19393×10^{-1}	9.85554×10^{-1}	8.628998×10^{-3}	5.320422×10^{-2}	2.486945×10^{-1}
1.0×10^{-1}	1.43879×10^{-1}	9.998552×10^{-1}	1.880270×10^{-5}	5.793770×10^{-4}	1.354098×10^{-2}
1.0	1.43879×10^{-2}	9.999998×10^{-1}	2.006819×10^{-9}	6.186710×10^{-7}	1.445925×10^{-4}
	Small x	$1.0 - 5.132991 \times 10^{-2} \cdot x^3$	$4.7167 \times 10^{-2} \cdot x^4$	$2.09212 \times 10^{-1} \cdot x^3$	$7.03514 \times 10^{-1} \cdot x^2$

[a] Radiation constants:

$C_1 = 3.741832 \times 10^{-16}\ \mathrm{W\,m^{-2}}$

$C_2 = 1.4387863 \times 10^{-2}\ \mathrm{m\,K}.$

$\pi B = 5.6703162\theta^4 \times 10^{-8}\ \mathrm{W\,m^{-2}}.$
$\pi(B_\lambda)_{max} = 1.286783\theta^5 \times 10^{-5}\ \mathrm{W\,m^{-2}\,m^{-1}}.$
$\pi(B_\nu)_{max} = 5.956646\theta^3 \times 10^{-19}\ \mathrm{W\,m^{-2}\,Hz^{-1}}.$
$\pi(\lambda B_\lambda)_{max} = 4.173617\theta^4 \times 10^{-8}\ \mathrm{W\,m^{-2}}.$

Table A.5.2. $dB_\nu/d\theta$, $\mathrm{W\,m^{-2}\,K^{-1}\,Hz^{-1}} \times 10^{14}$[a]

	Temperature (°C)			
$\nu(\mathrm{Hz} \times 10^{-12})$	**−80**	**−40**	**0**	**+40**
$\nu \to 0$		$\sim 2\mathbf{k}\nu^2/\mathbf{c}^2$		
1.0	3.056630×10^{-2}	3.061563×10^{-2}	3.064496×10^{-2}	3.066381×10^{-2}
5.0	6.764917×10^{-1}	7.037438×10^{-1}	7.205500×10^{-1}	7.315918×10^{-1}
10.0	1.881657×10^{0}	2.183814×10^{0}	2.390559×10^{0}	2.535594×10^{0}
15.0	2.426266×10^{0}	3.300150×10^{0}	3.994135×10^{0}	4.530919×10^{0}
20.0	2.138275×10^{0}	3.507859×10^{0}	4.800749×10^{0}	5.926128×10^{0}
30.0	8.908868×10^{-1}	2.202970×10^{0}	3.988937×10^{0}	6.009889×10^{0}
40.0	2.344571×10^{-1}	8.856262×10^{-1}	2.157125×10^{0}	4.037826×10^{0}
50.0	4.771029×10^{-2}	2.759017×10^{-1}	9.074976×10^{-1}	2.121944×10^{0}
60.0	8.246716×10^{-3}	7.303295×10^{-2}	3.246597×10^{-1}	9.496615×10^{-1}
70.0	1.273547×10^{-3}	1.727300×10^{-2}	1.037921×10^{-1}	3.799418×10^{-1}
$\nu \to \infty$		$\sim (2\mathbf{k}\nu^2/\mathbf{c}^2)x^2e^{-x}$		

[a] $dB_\nu/d\theta = (2\mathbf{k}\nu^2/\mathbf{c}^2)x^2e^x/(e^x-1)^2$, $x = C_2/\lambda\theta = \mathbf{h}\nu/\mathbf{k}\theta$.

BIBLIOGRAPHY

Weast, R. C. (Ed.), 1983, *Handbook of chemistry and physics*, 63*rd Edition*. Boca Raton, Florida: CRC Press Inc.

APPENDIX 6

THE EXPONENTIAL INTEGRALS

$$E_n(x) = \int_1^\infty e^{-\omega x}/\omega^n \, d\omega$$

$$= \int_0^1 e^{-x/\eta}\eta^{n-2} \, d\eta, \tag{A.6.1}$$

$$E_1(0) = \infty,$$

$$E_n(0) = 1/(n-1).$$

Two recurrence formulas:

$$E_n'(x) = -E_{n-1}(x), \tag{A.6.2}$$

$$nE_{n+1}(x) = e^{-x} - xE_n(x). \tag{A.6.3}$$

The following expansion always converges, but is useful only for small values of the argument:

$$E_1(x) = -\gamma - \ln|x| + \sum_{n=1}^{\infty} (-1)^{n-1}x^n/nn!, \tag{A.6.4}$$

where $\gamma = 0.5772156$ is Euler's constant.

Table A.6.1. Numerical values of the exponential integrals

x	$E_1(x)$	$E_2(x)$	$E_3(x)$
0	∞	1.000000	0.500000
0.01	4.037929	0.949671	0.490277
0.05	2.467898	0.827835	0.454919
0.10	1.822924	0.722545	0.416292
0.50	0.559774	0.326644	0.221604
1.00	0.219384	0.148496	0.109692
1.50	0.100020	0.073101	0.056739
2.00	0.048901	0.037534	0.030133
2.50	0.024915	0.019798	0.016295
3.00	0.013048	0.010642	0.008931
3.50	0.006970	0.005802	0.004945

From the recurrence relations,

$$(n-1)!\,E_n(x) = (-x)^{n-1}E_1(x) + e^{-x}\sum_{s=0}^{n-2}(n-s-2)!\,(-x)^s. \quad \text{(A.6.5)}$$

An asymptotic expansion for large x is

$$E_n(x) \sim \frac{e^{-x}}{x}[1 - n/x + n(n+1)/x^2 - \cdots]. \quad \text{(A.6.6)}$$

BIBLIOGRAPHY

Kourganoff, V., 1952, *Basic methods in transfer problems.* Oxford: Clarendon Press.

Abramowitz, M., and Stegun, I. A., 1972, *Handbook of mathematical functions.* National Bureau of Standards Applied Mathematics Series **55,** Tenth Printing. Washington, D.C.: The U.S. Government Printing Office, p. 228.

APPENDIX 7

THE LADENBURG AND REICHE FUNCTION

Table A.7.1. $L(u) = u \exp(-u)[I_0(u) + I_1(u)]$

u	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0099	0.0198	0.0295	0.0392	0.0488	0.0583	0.0676	0.0769	0.0861
0.1	0.0952	0.1042	0.1132	0.1220	0.1308	0.1395	0.1482	0.1567	0.1652	0.1735
0.2	0.1818	0.1900	0.1982	0.2063	0.2143	0.2223	0.2302	0.2380	0.2457	0.2534
0.3	0.2610	0.2685	0.2760	0.2834	0.2908	0.2981	0.3053	0.3125	0.3196	0.3267
0.4	0.3337	0.3406	0.3475	0.3543	0.3611	0.3678	0.3745	0.3811	0.3877	0.3942
0.5	0.4007	0.4071	0.4135	0.4198	0.4261	0.4324	0.4386	0.4447	0.4508	0.4569
0.6	0.4629	0.4689	0.4748	0.4807	0.4865	0.4923	0.4981	0.5038	0.5095	0.5152
0.7	0.5208	0.5264	0.5319	0.5374	0.5429	0.5483	0.5537	0.5591	0.5644	0.5697
0.8	0.5749	0.5801	0.5853	0.5905	0.5956	0.6007	0.6058	0.6108	0.6158	0.6208
0.9	0.6258	0.6307	0.6356	0.6404	0.6452	0.6500	0.6548	0.6596	0.6643	0.6690
1.0	0.6737	0.6783	0.6829	0.6875	0.6921	0.6966	0.7012	0.7057	0.7101	0.7146
1.1	0.7190	0.7234	0.7278	0.7322	0.7365	0.7408	0.7451	0.7494	0.7536	0.7578
1.2	0.7620	0.7662	0.7704	0.7746	0.7787	0.7828	0.7869	0.7910	0.7950	0.7990
1.3	0.8030	0.8070	0.8110	0.8150	0.8189	0.8228	0.8267	0.8306	0.8345	0.8364
1.4	0.8422	0.8460	0.8498	0.8536	0.8574	0.8612	0.8649	0.8686	0.8723	0.8760
1.5	0.8797	0.8834	0.8870	0.8907	0.8943	0.8979	0.9015	0.9051	0.9086	0.9122
1.6	0.9157	0.9193	0.9228	0.9263	0.9298	0.9332	0.9367	0.9402	0.9435	0.9470
1.7	0.9504	0.9538	0.9572	0.9606	0.9639	0.9673	0.9706	0.9740	0.9773	0.9806
1.8	0.9839	0.9872	0.9904	0.9937	0.9969	1.0002	1.0034	1.0066	1.0098	1.0130
1.9	1.0162	1.0194	1.0226	1.0257	1.0289	1.0320	1.0351	1.0383	1.0414	1.0445
2.0	1.0476	1.0506	1.0537	1.0568	1.0598	1.0629	1.0659	1.0689	1.0719	1.0750
2.1	1.0780	1.0809	1.0839	1.0869	1.0899	1.0928	1.0958	1.0987	1.1016	1.1046
2.2	1.1075	1.1104	1.1133	1.1162	1.1191	1.1220	1.1248	1.1277	1.1305	1.1334
2.3	1.1362	1.1391	1.1419	1.1447	1.1475	1.1503	1.1531	1.1559	1.1587	1.1615
2.4	1.1642	1.1670	1.1698	1.1725	1.1753	1.1780	1.1807	1.1835	1.1862	1.1889
2.5	1.1916	1.1943	1.1970	1.1997	1.2023	1.2050	1.2077	1.2103	1.2130	1.2156
2.6	1.2183	1.2209	1.2235	1.2262	1.2288	1.2314	1.2340	1.2366	1.2392	1.2418
2.7	1.2444	1.2470	1.2495	1.2521	1.2547	1.2572	1.2598	1.2623	1.2649	1.2674
2.8	1.2699	1.2725	1.2750	1.2775	1.2800	1.2825	1.2850	1.2875	1.2900	1.2925
2.9	1.2949	1.2974	1.2999	1.3024	1.3048	1.3073	1.3097	1.3122	1.3146	1.3171
3.0	1.3195	1.3219	1.3243	1.3268	1.3292	1.3316	1.3340	1.3364	1.3388	1.3412
3.1	1.3436	1.3459	1.3483	1.3507	1.3531	1.3554	1.3578	1.3601	1.3625	1.3649

Table A.7.1. $L(u) = u \exp(-u)[I_0(u) + I_1(u)]$ *(Continued)*

u	0	1	2	3	4	5	6	7	8	9
3.2	1.3672	1.3695	1.3719	1.3742	1.3765	1.3789	1.3812	1.3835	1.3858	1.3881
3.3	1.3904	1.3927	1.3950	1.3973	1.3996	1.4019	1.4042	1.4064	1.4087	1.4110
3.4	1.4132	1.4155	1.4178	1.4200	1.4223	1.4245	1.4268	1.4290	1.4312	1.4335
3.5	1.4357	1.4379	1.4402	1.4424	1.4446	1.4468	1.4490	1.4512	1.4534	1.4556
3.6	1.4578	1.4600	1.4622	1.4644	1.4666	1.4687	1.4709	1.4731	1.4753	1.4774
3.7	1.4796	1.4817	1.4839	1.4861	1.4882	1.4903	1.4925	1.4946	1.4968	1.4989
3.8	1.5010	1.5032	1.5053	1.5074	1.5095	1.5116	1.5137	1.5159	1.5180	1.5201
3.9	1.5222	1.5243	1.5264	1.5285	1.5305	1.5326	1.5347	1.5368	1.5389	1.5409
4.0	1.5430	1.5451	1.5471	1.5492	1.5513	1.5533	1.5554	1.5574	1.5595	1.5615
4.1	1.5636	1.5656	1.5677	1.5697	1.5717	1.5738	1.5758	1.5778	1.5798	1.5818
4.2	1.5839	1.5859	1.5879	1.5899	1.5919	1.5939	1.5959	1.5979	1.5999	1.6019
4.3	1.6039	1.6059	1.6079	1.6099	1.6118	1.6138	1.6158	1.6178	1.6197	1.6217
4.4	1.6237	1.6256	1.6276	1.6296	1.6315	1.6335	1.6354	1.6374	1.6393	1.6413
4.5	1.6432	1.6452	1.6471	1.6490	1.6510	1.6529	1.6548	1.6567	1.6587	1.6606
4.6	1.6625	1.6644	1.6663	1.6683	1.6702	1.6721	1.6740	1.6759	1.6778	1.6797
4.7	1.6816	1.6835	1.6853	1.6872	1.6891	1.6910	1.6929	1.6948	1.6967	1.6986
4.8	1.7005	1.7023	1.7042	1.7061	1.7079	1.7098	1.7117	1.7135	1.7154	1.7173
4.9	1.7191	1.7210	1.7228	1.7247	1.7265	1.7284	1.7302	1.7321	1.7339	1.7357
5	1.7376	1.7558	1.7739	1.7918	1.8095	1.8270	1.8444	1.8616	1.8786	1.8955
6	1.9123	1.9288	1.9453	1.9616	1.9778	1.9938	2.0097	2.0255	2.0412	2.0568
7	2.0722	2.0875	2.1027	2.1178	2.1328	2.1477	2.1625	2.1771	2.1917	2.2062
8	2.2206	2.2349	2.2491	2.2632	2.2772	2.2912	2.3050	2.3188	2.3325	2.3461
9	2.3597	2.3731	2.3865	2.3998	2.4130	2.4262	2.4393	2.4523	2.4653	2.4781
10	2.4910	2.5037	2.5164	2.5290	2.5416	2.5541	2.5665	2.5789	2.5912	2.6035
11	2.6157	2.6278	2.6399	2.6519	2.6639	2.6758	2.6877	2.6995	2.7113	2.7230
12	2.7347	2.7463	2.7579	2.7694	2.7809	2.7923	2.8037	2.8150	2.8263	2.8375
13	2.8487	2.8599	2.8710	2.8821	2.8931	2.9041	2.9150	2.9259	2.9368	2.9476
14	2.9584	2.9691	2.9798	2.9905	3.0011	3.0117	3.0223	3.0328	3.0433	3.0537
15	3.0641	3.0745	3.0848	3.0951	3.1054	3.1156	3.1258	3.1360	3.1461	3.1562
16	3.1663	3.1763	3.1863	3.1963	3.2063	3.2162	3.2261	3.2359	3.2457	3.2555
17	3.2653	3.2750	3.2847	3.2944	3.3041	3.3137	3.3233	3.3328	3.3424	3.3519
18	3.3614	3.3708	3.3803	3.3897	3.3990	3.4084	3.4177	3.4270	3.4363	3.4456
19	3.4548	3.4640	3.4732	3.4823	3.4914	3.5006	3.5096	3.5187	3.5277	3.5367
20	3.5457	3.6344	3.7210	3.8055	3.8883	3.9693	4.0487	4.1266	4.2030	4.2781
30	4.3519	4.4244	4.4958	4.5660	4.6352	4.7034	4.7706	4.8369	4.9022	4.9667
40	5.0304	5.0933	5.1554	5.2168	5.2775	5.3374	5.3968	5.4554	5.5135	5.5709
50	5.6277									

BIBLIOGRAPHY

These data are from

Kaplan, L. D., and Eggers, D. F., 1956, "Intensity and line width of the 15-micron CO_2 band, determined by the curve-of-growth method," *J. Chem. Phys.* **25,** 876.

Useful expansions of the equivalent width for the Lorentz profile and also for the Doppler profile are given by

Rodgers, C. D., and Williams, A. P., 1974, "Integrated absorption of a spectral line with the Voigt profile," *J. Quant. Spectrosc. Radiat. Transfer* **14,** 319.

Another tabulation for the Voigt profile is by

Jansson, P. A., and Korb, C. L., 1968, "A table of the equivalent widths of isolated lines with combined Doppler and collision broadened profiles," *J. Quant. Spectrosc. Radiat. Transfer* **8,** 1399.

APPENDIX 8

THE ELSASSER FUNCTION

The intervals in Table A.8.1 are close enough to permit interpolation to four-figure accuracy. The table may be extended by means of the asymptotic forms,

$y \to \infty$ equation (4.74),

$u \to \infty$ equation (4.78),

$u \to 0$ equation (4.14),

$y \to 0$ equation (4.83).

BIBLIOGRAPHY

Table A.8.1 has been taken from an unpublished report (1952).

Other tabulations have been given by

Kaplan, L. D., 1953, "Regions of validity of various absorption-coefficient approximations," *J. Meteorol.* **10,** 100.

Wark, D. Q., and Wolk, M., 1960, "An extension of a table of absorption for Elsasser bands," *Mon. Weather Rev.* **88,** 249.

Zachor, A. S., 1967, "Absorptance and radiative transfer by a regular band," *J. Quant. Spectrosc. Radiat. Transfer* **7,** 857.

Table A.8.1. $E(y, u) = \int_{-1/2}^{+1/2} \exp[-2\pi y u \sinh 2\pi y / (\cosh 2\pi y - \cos 2\pi x)]\, dx$

	$\log_{10} y$												
$\log_{10} u$	**0**	**−0.2**	**−0.4**	**−0.6**	**−0.8**	**−1.0**	**−1.2**	**−1.4**	**−1.6**	**−1.8**	**−2.0**	**−2.2**	**−2.4**
−1.5	0.819803	0.882179	0.923986	0.951422	0.969151	0.980480	0.987671	0.992216	0.995088	0.996901	0.998046	0.998768	0.999222
−1.25	0.702345	0.800180	0.868912	0.915393	0.946005	0.965756	0.978348	0.986328	0.991369	0.994557	0.996564	0.997849	0.998648
−1.0	0.533489	0.672748	0.779021	0.854938	0.906613	0.940540	0.962342	0.976204	0.984976	0.990522	0.994017	0.996225	0.997617
−0.75	0.327154	0.494204	0.641798	0.757900	0.841751	0.898540	0.935548	0.959223	0.974242	0.983751	0.989737	0.993526	0.995912
−0.5	0.137119	0.285620	0.455299	0.613677	0.740791	0.831771	0.892574	0.931887	0.956942	0.972824	0.982839	0.989174	0.993164
−0.25	0.029209	0.107790	0.248226	0.425665	0.597505	0.733211	0.828071	0.890574	0.930716	0.956223	0.972364	0.982558	0.988996
0	0.001867	0.019086	0.085420	0.228257	0.421162	0.602472	0.739704	0.833225	0.894115	0.933025	0.957700	0.973300	0.983149
+0.25		0.000883	0.013262	0.080820	0.243257	0.449996	0.629960	0.760129	0.846978	0.903021	0.938706	0.961130	0.975573
0.5			0.000529	0.014694	0.103825	0.292971	0.502665	0.671077	0.788405	0.865454	0.914854	0.946211	0.966040
0.75				0.000876	0.026680	0.152201	0.361635	0.563239	0.714958	0.817674	0.884390	0.927064	0.953903
1.0					0.002818	0.053621	0.218853	0.436054	0.622565	0.755982	0.844510	0.901565	0.937750
1.25						0.009639	0.099128	0.296179	0.509169	0.676843	0.792512	0.868168	0.916591
1.5						0.000532	0.027379	0.162295	0.377192	0.577164	0.724935	0.824284	0.888580
1.75							0.003209	0.061940	0.238134	0.456400	0.638321	0.766765	0.851541
2.0								0.012711	0.115335	0.320145	0.530375	0.692162	0.802726
2.25								0.000866	0.035641	0.184674	0.402508	0.597320	0.738889
2.5									0.005068	0.076825	0.264119	0.481019	0.656585
2.75									0.000186	0.018276	0.136381	0.347295	0.553160
3.0										0.001647	0.046995	0.210050	0.429006
3.25											0.008074	0.094610	0.291546
3.5											0.000411	0.025806	0.159558
3.75												0.002952	0.060695
4.0													0.012373

APPENDIX 9

THE PHYSICAL STATE OF THE SUN

A.9.1 The quiet sun

The sun is a typical main-sequence dwarf star of apparent visual magnitude −26.74 and spectral type G2V. The visible disk, or *photosphere,* has a radius $R_\odot = 6.599 \times 10^5$ km and the solar mass is 1.989×10^{30} kg. The chemical composition of the outer layers is (by mass) 71% hydrogen, 26.5% helium, and 2.5% heavier elements. The sun derives its energy from thermonuclear reactions between hydrogen atoms in the core (0 to $0.25R_\odot$) at a temperature of about 10^7 K. The energy is transferred outward by the radiation field to about $0.7R_\odot$ where the temperature is about 10^6 K. The final $0.3R_\odot$ (an uncertain figure) is accomplished by convective mixing processes maintaining the outer layers of the sun in a state of violent small-scale motion that exhibits itself, in high quality images, as surface *granulation,* rapidly varying structures of diameters approximately 1000 km and lifetimes of approximately 8 min.

At the mean sun–earth distance (1.495979×10^8 km), the spectrally integrated irradiance is a slightly variable quantity, corresponding to a black-body emission temperature of 5783 ± 20 K. The average angular diameter of the photosphere as measured at earth is 31.988 arc min, varying by ±1.7% over the year, as the earth follows its elliptic orbit.

The sun rotates at a rate that is variable with depth and with latitude. As measured by the motion of sunspots (other features give slightly different values), the synodic period (as seen from earth) is $[26.90 + 5.2 \sin^2 (\text{latitude})]$ days. The quiet sun has a general magnetic field with a value of 1–2 G at the surface near the poles.

The sun is entirely gaseous and the term *solar atmosphere* is no more than a convenient term for that part of the sun from which we receive measurable electromagnetic radiation, of any wavelength from γ rays to radio waves. The variation of temperature with height in the *photosphere and chromosphere,* measured from a zero level defined in terms of the opacity at 5000 Å, is shown in Fig. A.9.1. After falling to 4200 K in the *reversing layer,* the temperature rises to 20–50,000 K in the transition region (TR) between chromosphere and *corona.* In the corona, the temperature rises to about 1,800,000 K at a height equal to $R_\odot$.

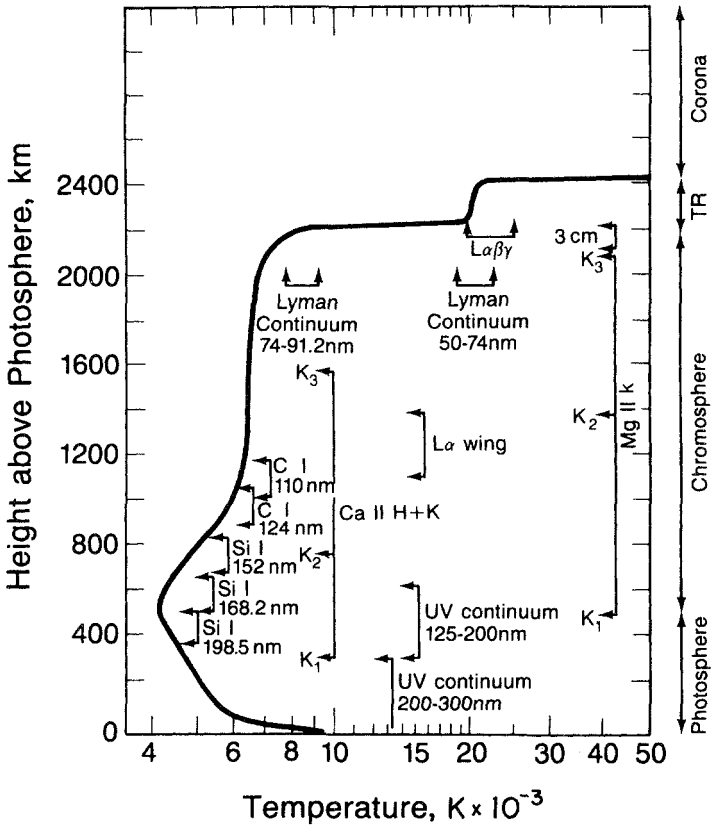


FIG. A.9.1. Temperature in the quiet photosphere and chromosphere. Compare these data to those of Allen (1973), Table 77. According to Allen, heights are usually measured from the level at which the optical depth at 5000 Å is unity and the temperature is about 6430 K. On this basis, the zero altitude in this figure corresponds to −90 km, and would have a 5000 Å optical depth equal to 20. According to Allen, the gas pressures are about 2×10^{-1} bar at the base of the photosphere and 10^{-2} bar at the base of the chromosphere. At the base of the corona the electron pressure is about 10^{-1} dyne cm^{-2}. The levels indicated by arrows correspond to unit optical depth for the named lines or continua. Note the 3-cm radio level at 2000 km. After Lean (1987).

A.9.2. The solar spectrum

The solar spectrum consists of a continuous emission with a superimposed line structure that, relative to the continuum, can appear in either absorption or in emission. The visible and infrared spectrum of the photosphere shows absorption lines only (the *Fraunhofer spectrum*), the strongest of which are caused by H, Na, Mg, Fe, Ca, Si, and singly ionized Ca and Mg. At 0.185 μm, the photospheric spectrum changes and most of the lines at shorter wavelengths are exhibited in emission. Light

from the chromosphere and the corona has emission lines at all observable wavelengths.

Most of the solar energy is carried by the continuum. This owes its existence to nonquantized electronic transitions, both bound-free (ionizing) and free–free. The most important single contributor is hydrogen, both in its neutral state and as negative ions. Neutral hydrogen has ionization continua starting at the limits of the Lyman, Balmer, Paschen, Brackett, and Rydberg series; the Balmer discontinuity at 0.3646 μm is prominent in observational data (e.g., Fig. A.9.4). Metallic atoms contribute to the continuum in the ultraviolet spectrum and are responsible for the maximum absorption coefficient at 0.15 μm in Fig. A.9.2. Through the visible and infrared spectrum, however, the continuum absorption is by negative hydrogen ions. Between 0.4 and 1.7 μm, the main contribution is from the bound–free transition; at longer wavelengths the absorption coefficient varies as (wavelength)^2, and is caused by a free–free transition.

The effect of the smoothed-out absorption by Fraunhofer lines is indicated by the broken line in Fig. A.9.2. At wavelengths less than

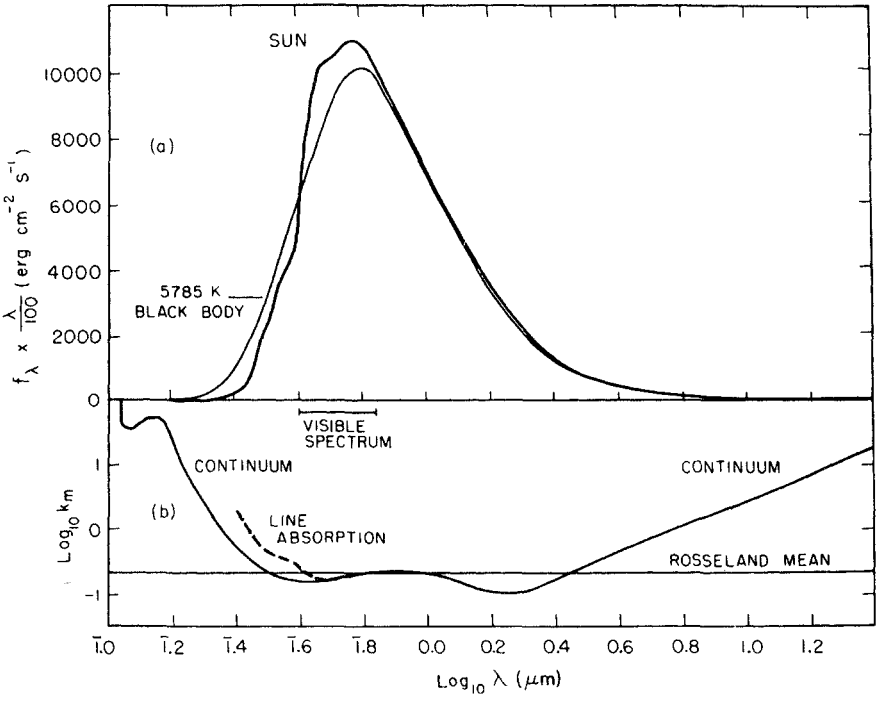


FIG. A.9.2. Spectrum of solar emission and photospheric absorption. (a) Solar spectrum compared to that of a 5785 K black body. The method of plotting gives areas ($f_\lambda \lambda d \log_{10} \lambda/100$) proportional to energy flow ($f_\lambda\, d\lambda$). (b) Mass absorption coefficient for the photosphere at a temperature of 5785 K. After Allen (1958).

0.3 μm more than half of the continuum is reabsorbed by lines. Under these circumstances, empirical separation of lines and continua in the observational data is no longer possible. At infrared wavelengths, line absorption can be neglected for practical purposes.

In terms of the discussion of § 6.5.1, emission by each feature in the solar spectrum has a kernel function, similar to the Chapman function, with a maximum emission taking place from the level of unit optical depth; the emission is approximately equal to the source function at that level. Emission levels are indicated in Fig. A.9.1 for ultraviolet continua and certain important ultraviolet lines. For the 168.2 nm (0.1682 μm) Si I line and for longer wavelengths, the continuum level has a temperature higher than the overlying line, and Si I appears in absorption. For the 110 nm C I line, on the other hand, the overlying line emits at a temperature higher than the continuum, and the line appears in emission. It is consistent with this interpretation that, when chromospheric and photospheric light are briefly separated during an eclipse, the former shows the Fraunhofer lines in emission, at wavelengths longer than 185 nm.

The increase of absorption toward both long and short wavelengths shown in Fig. A.9.2 causes continuum emission in both the X ray and radio wave regions to originate in the rarified corona. Line emission from the corona, where the temperature exceeds 10^6 K, is principally from the 0.5303 μm line of iron, 13 times ionized.

A.9.3. The intensity of solar radiation

The irradiance of the quiet sun at mean solar distance for the wavelength range 0.3–5 μm can be read from Fig. A.9.2. An extended range is shown in Fig. A.9.3. The emissions at wavelengths shorter than 0.1 μm and longer than 1 cm are coronal, and are highly variable (see below).

Table A.9.1 gives tabulated data from 0.2 to 5 μm. Shorter wavelengths are important for their effect on the temperature and composition of the upper earth's atmosphere and will be discussed further in § A.9.4. The irradiance for wavelengths between 5 μm and 1 cm can be calculated from the effective emission temperatures given in Table A.9.2.

The foregoing data refer to emission from the entire solar disk. For most wavelengths, the sun shows wavelength-dependent limb darkening. From the curves in Fig. A.9.4 it is possible to calculate the intensity at each point on the disk. Following the discussion in § 6.5.1(iii), limb darkening corresponds to emission from a region in which temperature decreases with height (the photosphere, Fig. A.9.1); for chromospheric emissions ($\lambda < 150$ nm), limb brightening occurs. Limb darkening for visible wavelengths can be seen in the photograph in Fig. A.9.5.

The integrated or total irradiance is the quantity that is most

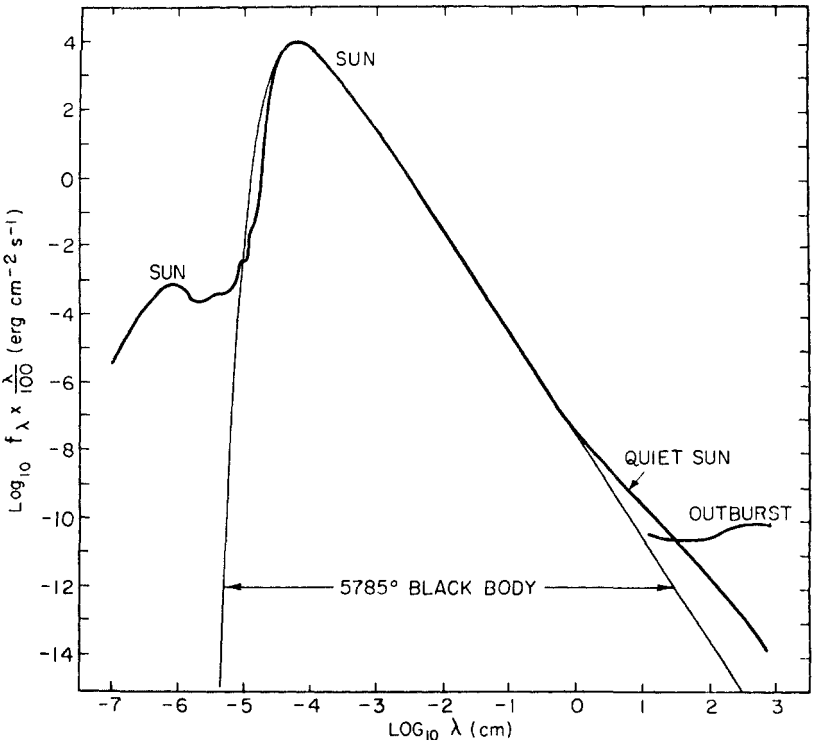


FIG. A.9.3. The solar irradiance from 10^{-7} to 10^{3} cm. Spectral features are smoothed out by averaging over intervals containing many Fraunhofer lines. Areas are not proportional to energy, as was the case in Fig. A.9.2. "Outburst" indicates a radio disturbance (see § A.9.4). After Allen (1958).

important for meteorological and climatological studies. Values quoted in the literature vary from time to time, partly perhaps because the quantity is intrinsically variable but partly also because of the difficulties of absolute measurements. The most widely accepted figure, ca 1980, is

$$f = 1373 \pm 20 \text{ W m}^{-2}, \tag{A.9.1}$$

corresponding to an emission temperature

$$\theta_e = 5783 \pm 20 \text{ K}. \tag{A.9.2}$$

A.9.4. Solar variability

The sun is a variable star. At times there have been claims that all weather and climate changes are caused by changes in the solar irradiance, countered at other times by the assertion that solar variations

Table A.9.1. Solar irradiance, 0.2–5.0 μm[a]

Wavelength (μm)	Irradiance (erg cm^{-2} Å^{-1} s^{-1})	Wavelength (μm)	Irradiance (erg cm^{-2} Å^{-1} s^{-1})
0.20	0.1	0.50	210
0.22	0.7	0.55	196
0.24	3	0.60	185
0.26	14	0.65	164
0.28	24	0.70	146
0.30	55	0.80	114
0.32	74	0.90	89
0.34	89	1.00	71
0.36	104	1.10	58
0.38	118	1.20	48
0.39	125		
		1.4	32
0.40	144	1.6	21
0.42	182	1.8	15
0.44	202	2.0	10.8
0.46	216		
0.48	217	2.5	4.9
		3.0	2.6
		4.0	0.9
		5.0	0.4

[a] The spectrum is smoothed by averaging over spectral intervals wide compared to Fraunhofer lines. Absolute spectral measurements are difficult to make and there is some disagreement between different observers. For details, consult the exhaustive reviews in White (1977). *Source*: After Allen (1973).

Table A.9.2. Effective temperatures for infrared and radio emissions

Wavelength	Emission temperature (K)
5 μm	5500
10 μm	5050
20 μm	4740
50 μm	4500
100 μm	4370
1 mm	5500
1 cm	8200

Source: After Allen (1973).

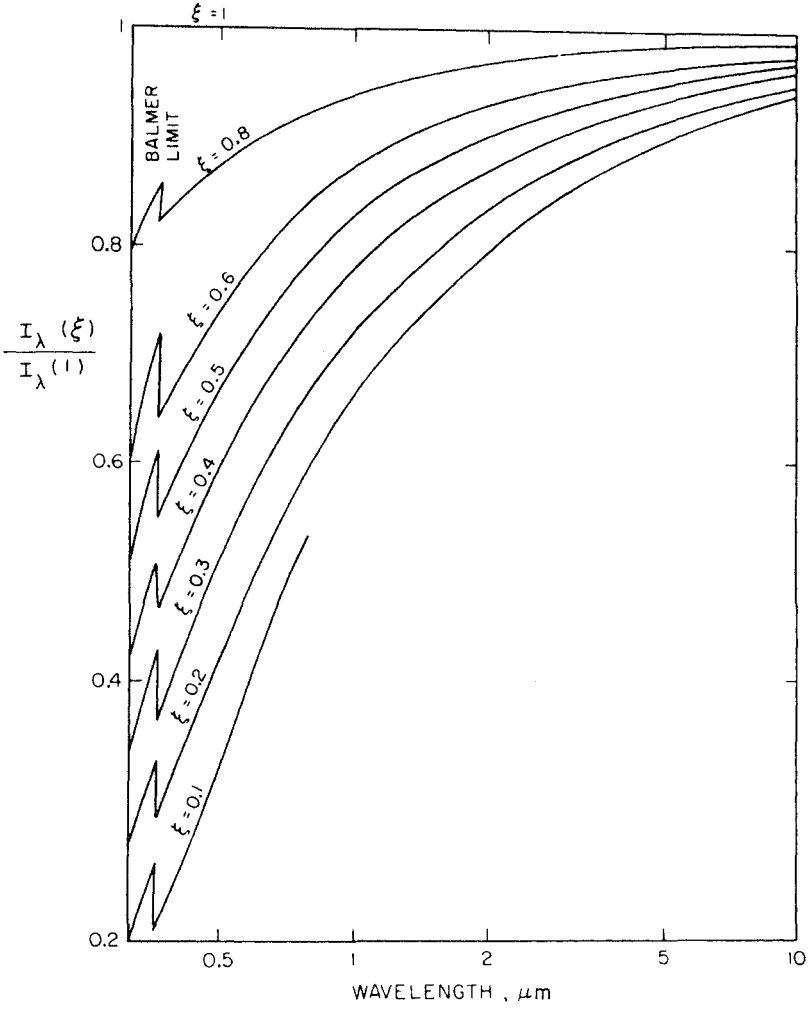


FIG. A.9.4. Solar limb darkening. ξ is the cosine of the zenith angle at the sun's surface. $\xi = 1$ corresponds to the center of the disk. After Allen (1958).

are irrelevant as far as the lower atmosphere is concerned. To put the problem into perspective, we note that large changes in the corona are well established and that coronal ultraviolet emission is the heat source for levels in the upper atmosphere where the density is very low. The thermosphere, above 150 km, is greatly influenced by variable conditions on the sun.

The greater part of the solar irradiance comes from the photosphere, however, where pressures are high and variability is low. Such variations as are observed are small compared to other familiar effects, e.g., diurnal

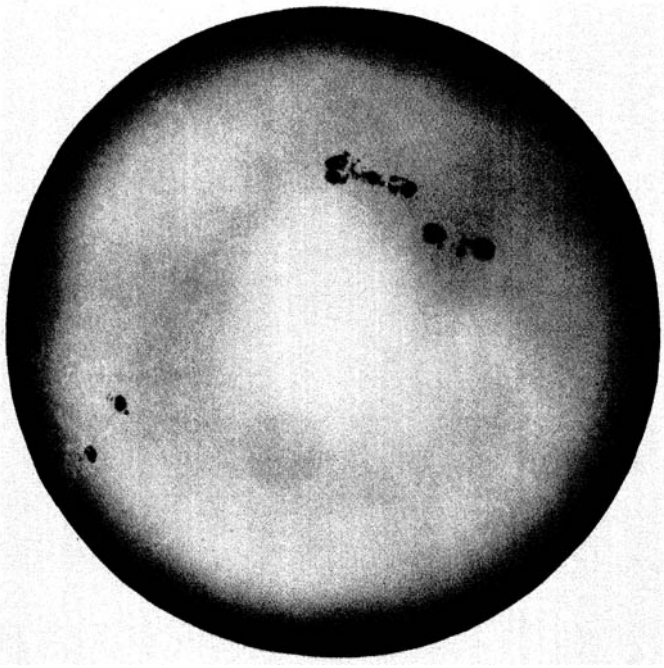


FIG. A.9.5. The photosphere in white light showing two large sunspot groups observed on April 8, 1980. Courtesy of the Big Bear Solar Observatory, California Institute of Technology.

and seasonal variations in the geometry of the insolation and variations in cloudiness, surface albedo, and aerosols. Variations in the earth's orbital elements can give rise to variations of insolation of a few percent with periods between 21,000 and 90,000 years (the *Milankovitch effect*). For the climatologist, all variations in insolation are equally important and total irradiance changes on the order of a few tenths of a percent do not command much attention.

The most striking visual disturbances on the photosphere are the *sunspots,* patches varying in diameter from a few thousand to a hundred thousand kilometers, with an emission temperature in the center about 1500 K lower than that of the undisturbed photosphere. An unusually large spot group, that could be seen with the unaided eye, is shown in Fig. A.9.5. The fraction of the photosphere covered by spots is never more than 0.2%, and their average persistence is about a week.

The Wolf *sunspot number* is an index of solar activity based on the numbers of sunspots and sunspot groups. The sunspot number for the period 1610 to 1976 AD (estimates only for the first 100 years) is shown in Fig. A.9.6. For most of this period, there was a *sunspot cycle* with an

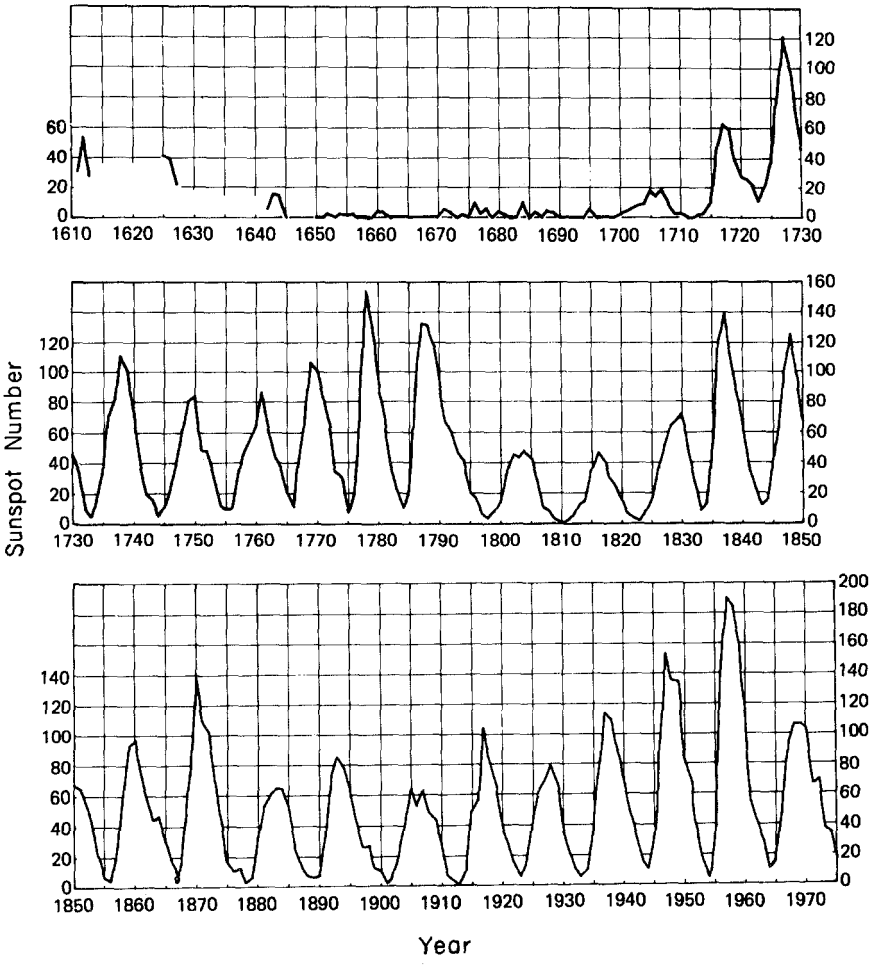


FIG. A.9.6. The sunspot cycle. After Eddy (1977).

average period of 11.04 years. During the late seventeenth century, however, evidence suggests that there were no sunspots to be seen (the *Maunder minimum*); there may have been earlier events of a similar nature. The number of spots is only one characteristic feature of the sun that changes in this rhythmic manner. Just after the minimum, spots first appear near latitude 27° in both hemispheres. As the cycle proceeds, they drift equatorward and disappear close to latitude 8°. They are rarely observed at latitudes higher than 30° or lower than 5°.

Other types of disturbance are associated with sunspots. When a spot is near to the limb it can be seen to be surrounded by a network of enhanced photospheric emission, patches of which are called *faculae*.

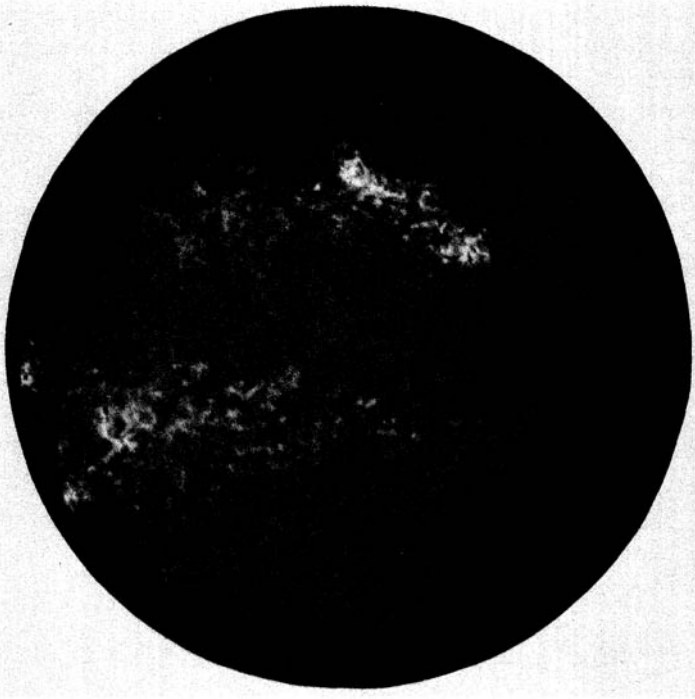


FIG. A.9.7. The photosphere in Hα light. This photograph was taken at the same time as that shown in Fig. A.9.5 and shows the same features. It was taken through a narrow filter centered on the Hα line of hydrogen. Courtesy of the Big Bear Solar Observatory, California Institute of Technology.

These have longer lifetimes than the associated spot group, appearing before and disappearing after the spots themselves.

When the sun is viewed in monochromatic light from a single element, other disturbances can be seen. Figure A.9.7 illustrates typical features in Hα light. The spots are visible on the monochromatic images but surrounding bright areas, known as *flocculi* or *plages,* are now the most prominent features, and they also occur at high latitudes, where spots do not. Occasionally, a hydrogen flocculus near a spot will brighten up, in extreme cases to the extent that the brightening is visible to the eye. These brightenings are known as *solar flares,* and they are associated with great increases of Lyman α and other ultraviolet radiations that influence the upper atmosphere.

Prominences are photospheric eruptions, extending into the chromosphere, that can be seen on the limb of the sun if precautions are taken to cut out the intense photospheric light. Many different forms occur, but a typical prominence might be 30,000 km high and 200,000 km long, with a temperature of 5000 K. Since this temperature is less than that of the

photosphere, prominences may be seen in Hα light as *dark filaments* on a bright background. More prominences occur at solar maximum than at solar minimum.

Coronal disturbances are closely related to the sunspot cycle. In visible light the corona appears more jagged at sunspot maximum than at minimum. Solar radio emission from the corona shows a marked variation with the sunspot cycle and is also correlated with shorter period changes in sunspot number. The intensity of radio emission from occasional *outbursts* suggests local condensations with temperatures of 10^7 K; outbursts are accompanied by large increases in the far-ultraviolet and X-ray emission from the sun.

All of these variations are associated with magnetic activity, resulting

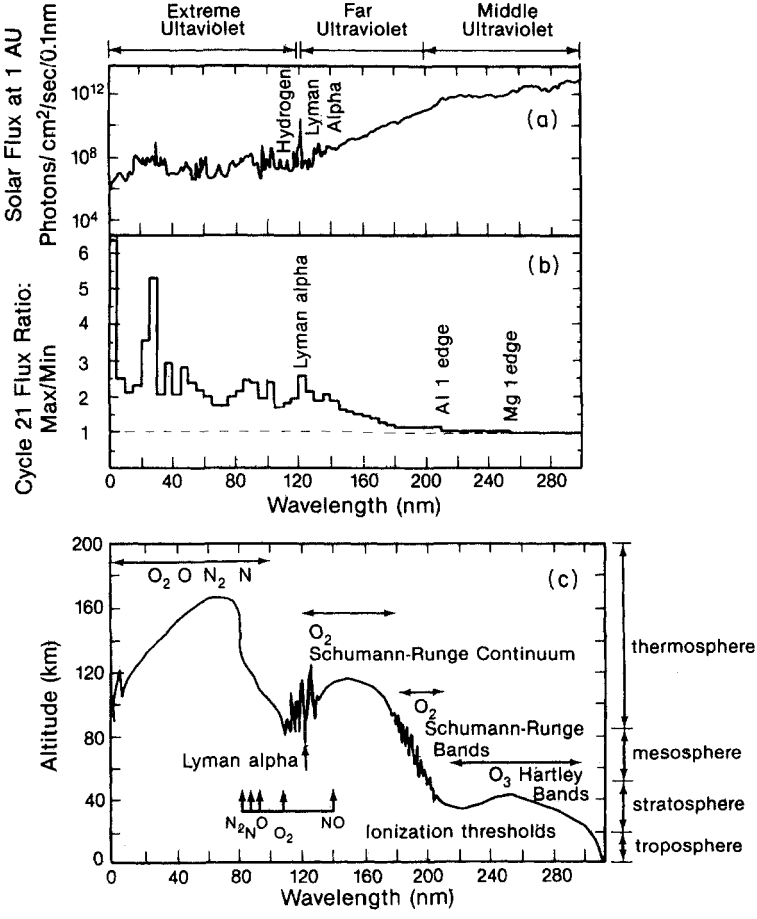


FIG. A.9.8. The solar ultraviolet spectrum and its variability. (a) The average spectral irradiance. (b) Ratio of average maximum to average minimum in a solar cycle. (c) This is the same as Fig. 5.1. It shows the level at which each spectral component is principally absorbed and it indicates ionization thresholds. After Lean (1987).

from interactions between convective motions, the solar rotation, and the general magnetic field of the sun. Sunspots contain magnetic fields up to 4000 G. The polarity in spot pairs reverses in successive sunspot cycles so that the complete cycle is sometimes taken to be 22.08 years. Magnetic fields and electric currents penetrate the chromosphere and corona, where magnetic variations have far greater influence because of the low densities.

Solar variations are now routinely monitored in a certain chromospheric lines (Ca II H and K at 3875–3933 Å and He I at 10,830 Å) and at 10.7 cm in the radio region, in addition to observations of sunspots, faculae, and plages. These indices all show a general correlation with the sunspot cycle, although they usually differ from each other on shorter time scales. This information is now supplemented, less systematically, by rocket and satellite measurements of the total irradiance and also of ultraviolet and X-ray emissions, not detectable at the earth's surface.

Solar cycle variability in the ultraviolet spectrum is illustrated in Fig. A.9.8b. The greatest changes occur at wavelengths that are absorbed in the thermosphere, where they give rise to large changes of temperature (see Fig. 1.4) and of ionization. Larger changes may be associated with solar flares and other short-period disturbances, especially at wavelengths less than 20 nm (Fig. A.9.9).

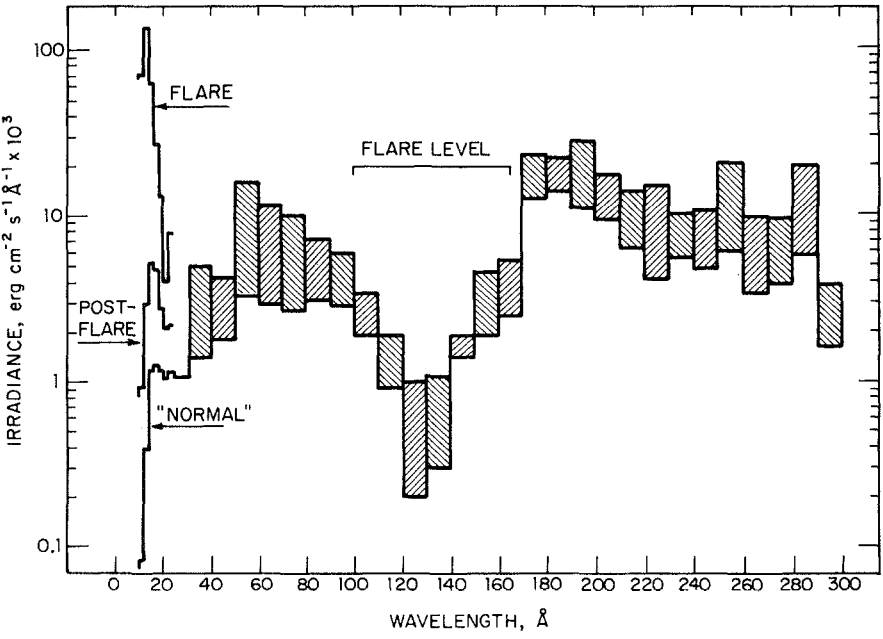


FIG. A.9.9. Spectral irradiance between 10 and 300 Å. Three states of solar activity are shown for the region 10 to 31 Å. The effect of a 2B flare is indicated for the spectral region 100–170 Å. The shaded areas include all known data in the absence of flares and give some indication of the extent of the solar-cycle changes. After Manson (1977).

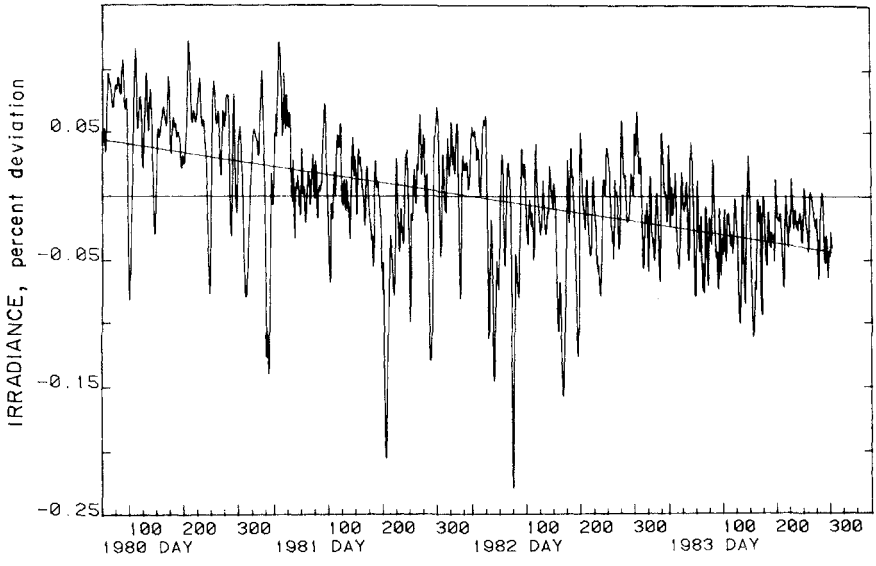


FIG. A.9.10. Percentage variation of solar irradiance corrected to mean solar distance. The measurements were made with a cavity radiometer on the Solar Maximum Mission. The mean irradiance for this period was 1367.6 W m^{-2}. After Willson (1984).

Figure A.9.10 shows variations in the total irradiance measured from space, over a period of 3 years. Short-period changes up to 0.3% occurred, together with a long-term drift of 0.09%. The long-term drift may be part of a solar cycle change; there is some evidence for a slow increase of irradiance between 1967 and 1980, and recent, unpublished results for 1985 and 1986 show a leveling out with variations around 1366.8 W m^{-2}. Short-period changes are accounted for, in part, by the geometric shadowing by sunspots, but there are additional correlations with the extent of faculae.

Theories of climatic changes on geological time scales draw attention to inevitable changes that must have taken place during the lifetime of the sun. The sun is now about 5×10^9 years old. According to widely accepted theories, it was 6% smaller, 300 K cooler, and its irradiance was 40% lower than present-day values when it formed. For the next few billion years variation may be approximated by

$$f = f_0[1 + 0.4(1 - t/t_0)]^{-1}, \tag{A.9.3}$$

where f and t are the irradiance and the age of the sun and zero suffices indicate present-day values.

BIBLIOGRAPHY

The best compilation of numerical data is

Allen, C. W., 1973, *Astrophysical quantities*. London: Athlone Press.

An earlier paper by the same author presents a valuable overview for the atmospheric scientist

Allen, C. W., 1958, "Solar radiation," *Quart. J. Roy. Meteorol. Soc.* **84,** 307.

A popular account of all aspects of solar physics is

Noyes, R. W., 1982, *The sun, our star*. Cambridge, Mass.: Harvard University Press.

The most complete modern source on solar radiation and its variations is

White, O. R. (Ed.), 1977, *The solar output and its variations*. Boulder: Colorado Associated Universities Press. This contains articles by

Eddy, J. A., 1977, "Historical evidence for the existence of the solar cycle," p. 51.

Vernekar, A. D., 1977, "Variations in insolation caused by changes in the orbital elements of the earth," p. 117.

Shimabukuro, F. I., 1977, "The solar spectrum above 1 mm," p. 133.

Mankin, W. G., 1977, "The solar spectrum between 10 and 1000 μm," p. 151.

Pierce, A. K., and Allen, R. G., 1977, "The solar spectrum between .3 and 10 μm," p. 169.

Heath, D. F., and Thekaekara, M. P., 1977, "The solar spectrum between 1200 and 3000 Å," p. 193.

Vidal-Madjar, A., 1977, "The solar spectrum at Lyman-alpha 1216 Å," p. 213.

Timothy, J. G., 1977, "The solar spectrum between 300 and 1200 Å," p. 237.

Manson, J. E., 1977, "The solar spectrum between 10 and 300 Å," p. 261.

Gough, D. O., 1977, "Theoretical predictions of variations in the solar output," p. 451.

A group of papers in *J. Geophys. Res.* **92,** Number D1, 1987, pp. 775–914 is entitled "Solar variability and its stratospheric, mesospheric and thermospheric effects." It contains several informative papers, particularly

Lean, J. 1987, "Solar ultraviolet irradiance variations: A review," *J. Geophys. Res.* **92,** 839.

Total irradiance measurements made from space are reviewed by

Willson, R. C., 1984, "Measurements of solar total irradiance and its variability," *Space Sci. Rev.* **38,** 203.

This page intentionally left blank

AUTHOR INDEX

This page intentionally left blank

SUBJECT INDEX