Principles of Planetary Climate

R. T. Pierrehumbert

February 20, 2007

Preface

When it comes to understanding the why's and wherefores of climate, there is an infinite amount one needs to know, but life affords only a finite time in which to learn it; the time available before one's fellowship runs out and a PhD thesis must be produced affords still less. Inevitably, the student who wishes to get launched on significant interdisciplinary problems must begin with a somewhat hazy sketch of the relevant physics, and fill in the gaps as time goes on. It is a lifelong process. This book is an attempt to provide the student with a sturdy scaffolding upon which a deeper understanding may be hung later.

The climate system is made up of building blocks which in themselves are based on elementary physical principles, but which have surprising and profound collective behavior when allowed to interact on the planetary scale. In this sense, the "climate game" is rather like the game of Go, where interesting structure emerges from the interaction of simple rules on a big playing field, rather than complexity in the rules themselves. This book is intended to provide a rapid entrée into this fascinating universe of problems for the student who is already somewhat literate in physics and mathematics, but who has not had any previous experience with climate problems. The subject matter of each individual chapter could easily fill a textbook many times over, but even the abbreviated treatment given here provides enough core material for the student to begin treating original questions in the physics of climate.

This is a somewhat Earth-centric book, in that the Earth provides our best-observed example of a planetary climate. Nonetheless, the central organizing principle is the manner in which the interplay of the same basic set of physical building-blocks gives rise to the diverse climates of present, past and future Earth, of the other planets in the Solar system, and of hypothetical planets yet to be discovered.

In this book I have chosen to deal only with aspects of climate that can be treated without consideration of the fluid dynamics of the Atmosphere or Ocean. Many successful scientists have spent their entire careers productively in this sphere. A sequel will treat the additional phenomena that emerge when fluid dynamics is introduced, culminating in a do-it-yourself General Circulation Model.

The short exercises embedded in the text are meant to be done "on the spot," as an immediate check of comprehension. More involved and thought-provoking problems may be found in the accompanying Workbook section at the end of each chapter. The Workbook provides an integral part of the course. Using the techniques and tools developed in the Workbook sections, the student will be able to reproduce every single computational and data analysis result included in the text. The Workbook also offers considerable opportunities for independent inquiry launching off from the results shown in the text. After having completed the course, the diligent student will be in possession of a tool kit that will be immediately useful in original research. In a modest way, the Workbook is intended to do for climate modelling what *Numerical Recipes* did for numerical

analysis.

Contents

Chapter 1

The Big Questions

This chapter will provide a qualitative overview of some of the major problems of Earth and planetary climate. Some have been solved to one extent or another, but most are largely unresolved. All involve physics that will be developed during the course

1.1 Faint Young Sun and habitability of the Earth

General introduction to the role of stellar evolution in climate, which provides necessary background also for the Early Mars problem. Co-evolution of atmosphere and solar forcing. Question of how Earth maintains its habitability as the Sun changes.

Additional basic facts about Early Earth history. The early appearance of prokaryotic life. The timing of the first glaciations, which appear surprisingly late in the game.

1.2 Earth,Mars and Venus:The Goldilocks problem

Summary of present climates and atmospheres of Venus and Mars vs. Earth

What happened to Venus? How did it keep its CO2 (compare to carbonate rocks on Earth)? Where did its water go? Did Venus go through a transient habitable phase when the Sun was fainter?

Evidence for warm, wet Early Mars What happened to Mars? Where did its atmosphere go?

How much would Earth's conditions have to change (size of planet, position of orbit) before it fell into the fate of Venus or Mars?

1.3 Extrasolar planets

Large Earths. Roasters.

1.4 The Proterozoic climate:Snowball Earth

What is the Proterozoic. Paleoproterozoic and Neoproterozoic. History of oxygenation. Oxygen and the methane catastrophe. Evolution of eukaryotes. Evolution of metazoans.

Paleoproterozoic Snowball Earth.

Neoproterozoic Snowball Earth

1.5 Hothouse climates

Characteristics of hothouse climates, like Eocene and Cretaceous. Lack of permanent polar ice. Low meridional gradients. What accounts for such climates?

Note occurrence of other climate periods with permanent polar ice, similar to the present one. How to account for transition between these and hothouse climates

1.6 Pleistocene ice ages

Onset of the Pleistocene ice ages. Variation of dominant frequency over time. Mystery of the onset of the 100Kyr cycle. The origin of the glacial-interglacial CO2 cycle. Apparent relation to orbital forcing.

1.7 Holocene climate variation and abrupt climate change

Abrupt change. Younger Dryas. Dansgaard-Oeschger events. Why did such things cease at the onset of the Holocene?

The "Climatic Optimum," and Sahara wet/dry cycles. The time of initiation of tropical mountain glaciers (based on Lonnie Thompson's work).

The Little Ice Age

1.8 The Fate of the Earth

Global warming in the context of past climate change, and CO2 history. How much CO2 would it take to cancel the next glaciation?

Lifetime of the biosphere, as the Sun continues to warm. Will Earth turn into Venus? Will Mars bloom? What will happen to Titan as the solar system warms?

Chapter 2

Thermodynamics in a Nutshell

The atmospheres which are our principal objects of study are made of compressible gases. The compressibility has a profound effect on the vertical profile of temperature in these atmospheres. As things progress it will become clear that the vertical temperature variation in turn strongly influences the planet's climate. To deal with these effects it will be necessary to know some thermodynamics though just a little. This chapter does not purport to be a complete course in thermodynamics. It can only provide a summary of the key thermodynamic concepts and formulae needed to treat the basic problems of planetary climate. It is assumed that the student has obtained (or will obtain) a more fundamental understanding of the general subject of thermodynamics elsewhere.

2.1 A few observations

The temperature profile in Figure 2.1, measured in the Earth's tropics introduces most of the features that are of interest in the study of general planetary atmospheres. It was obtained by releasing an instrumented balloon (radiosonde) which floats upwards from the ground, and sends back data on temperature and pressure as it rises. Pressure goes down monotonically with height, so the lower pressures represent greater altitudes.

Pressure is a very natural vertical coordinate to use. Many devices for measuring atmospheric profiles directly report pressure rather than altitude, since the former is generally easier to measure. More importantly, most problems in the physics of climate require knowledge only of the variation of temperature and other quantities with pressure; there are relatively few cases for which it is necessary to know the actual height corresponding to a given pressure. Pressure is also important because it is one of the fundamental thermodynamic variables determining the state of the gas making up the atmosphere. Atmospheres in essence present us with a thermodynamic diagram conveniently unfolded in height. Throughout, we will use pressure (or its logarithm) as our fundamental vertical coordinate.

However, for various reasons one might nevertheless want to know at what altitude a given pressure level lies. By altitude tracking of the balloon, or using the methods to be described in Section 2.2, the height of the measurement can be obtained in terms of the pressure. The right panel of Figure 2.1 shows the relation between altitude and pressure for the sounding shown in Figure 2.1. One can see that the height is very nearly linearly related to the log of the pressure.

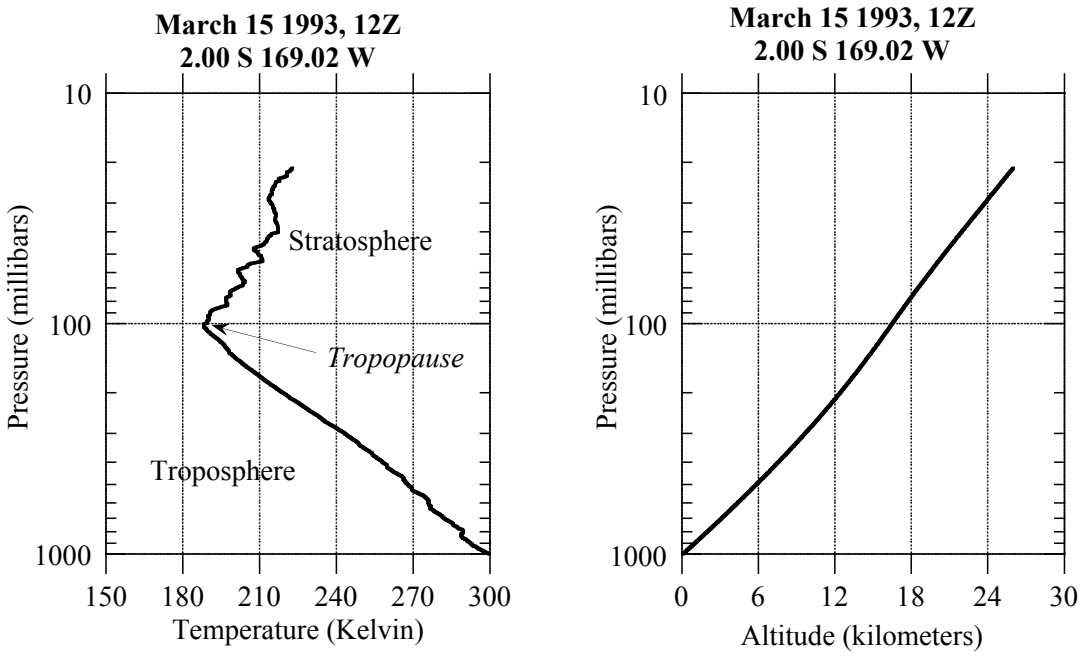

Figure 2.1: Left panel: Temperature profile measured at a point in the midlatitude Pacific. Right panel:The corresponding altitude. The measurements were obtained from a radiosonde ("weather balloon") launched at 12Z (an abbreviaton for Greenwich Mean Time) on March 15, 1993.

This is the reason it is often convenient to plot quantities vs. pressure on a log plot. If p_o is representative of the largest pressure of interest, then $-\ln(p/p_o)$ is a nice height-like coordinate, since it is positive and increases with height.

We can now return to our discussion of the critical aspects of the temperature profile. The most striking feature of the temperature sounding is that the temperature goes down with altitude. This is a phenomenon familiar to those who have experienced weather in high mountains, but the sounding shows that the temperature drop continues to altitudes much higher than sampled at any mountain peak. This sounding was taken over the Pacific Ocean, so it also shows that the temperature drop has nothing to do with the presence of a mountain surface. The temperature drop continues until a critical height, known as the tropopause, and above that height (100mb, or 16 km in this sounding) begins to increase with height. The portion of the atmosphere below the tropopause is known as the troposphere, whereas the portion immediately above is the stratosphere. "Tropo" comes from the Greek root for "turning" (as in "turning over"), while "Strato.." refers to stratification. The reasons for this terminology will become clear shortly. The stratosphere was discovered in 1900 by Lèon Phillipe Teisserenc de Bort, the French pioneer of instrumented balloon flights.

The sounding we have shown is typical. In fact, a similar pattern is encountered in the atmospheres of many other planets, as indicated in Figure 2.2 for Venus, Mars, Jupiter and Titan. In common with the Earth case, the lower portions of these atmospheres exhibits a sharp decrease of temperature with height, which gives way to a region of more gently decreasing, or even increasing, temperature at higher altitudes. The temperature decrease with height in the Earth's atmosphere has long been known from experience of mountain weather. It became a target of quantitative investigation not long after the invention of the thermometer, and was early recognized as a challenge to those seeking an understanding of the atmosphere. It was one of the central pre-occupations of the mountaineer and scientist Horace Bénédict de Saussure (1740-1799). In the quest for an explanation, many false steps were taken, even by greats such as Fourier, before the correct answer was unveiled. As will be shown in the remainder of this chapter, some simple ideas based on thermodynamics and vertical mixing provide at least the core of an explanation for the temperature decrease with height. In Chapter 7 we will present a theory of tropopause height that similarly captures the essence of the problem. Nonetheless, some serious gaps remain in the state of understanding of the rate of decrease of temperature with height, and of the geographical distribution of tropopause height. In Chapters 3 and 4 we will see that the energy budget of a planet is crucially affected by the vertical structure of temperature; therefore, a thorough understanding of this feature is central to any theory of planetary climate.

2.2 Dry thermodynamics of an ideal gas

2.2.1 The equation of state for an ideal gas

The three thermodynamic variables with which we will mainly be concerned are: temperature (denoted by T), pressure (denoted by p) and density (denoted by ρ). Temperature is a measure of the amount of kinetic energy per molecule in the molecules making up the gas. We will always measure temperature in degrees Kelvin, which are the same as degrees Celsius (or Centigrade), except offset so that absolute zero – the temperature at which molecular motion ceases – occurs at zero Kelvin. In Celsius degrees, absolute zero occurs at about -273.15C, which is then zero degrees Kelvin by definition. Pressure is a measure of the flux of momentum per unit time carried by the molecules of the gas passing through an imaginary surface of unit area; equivalently, it is a measure

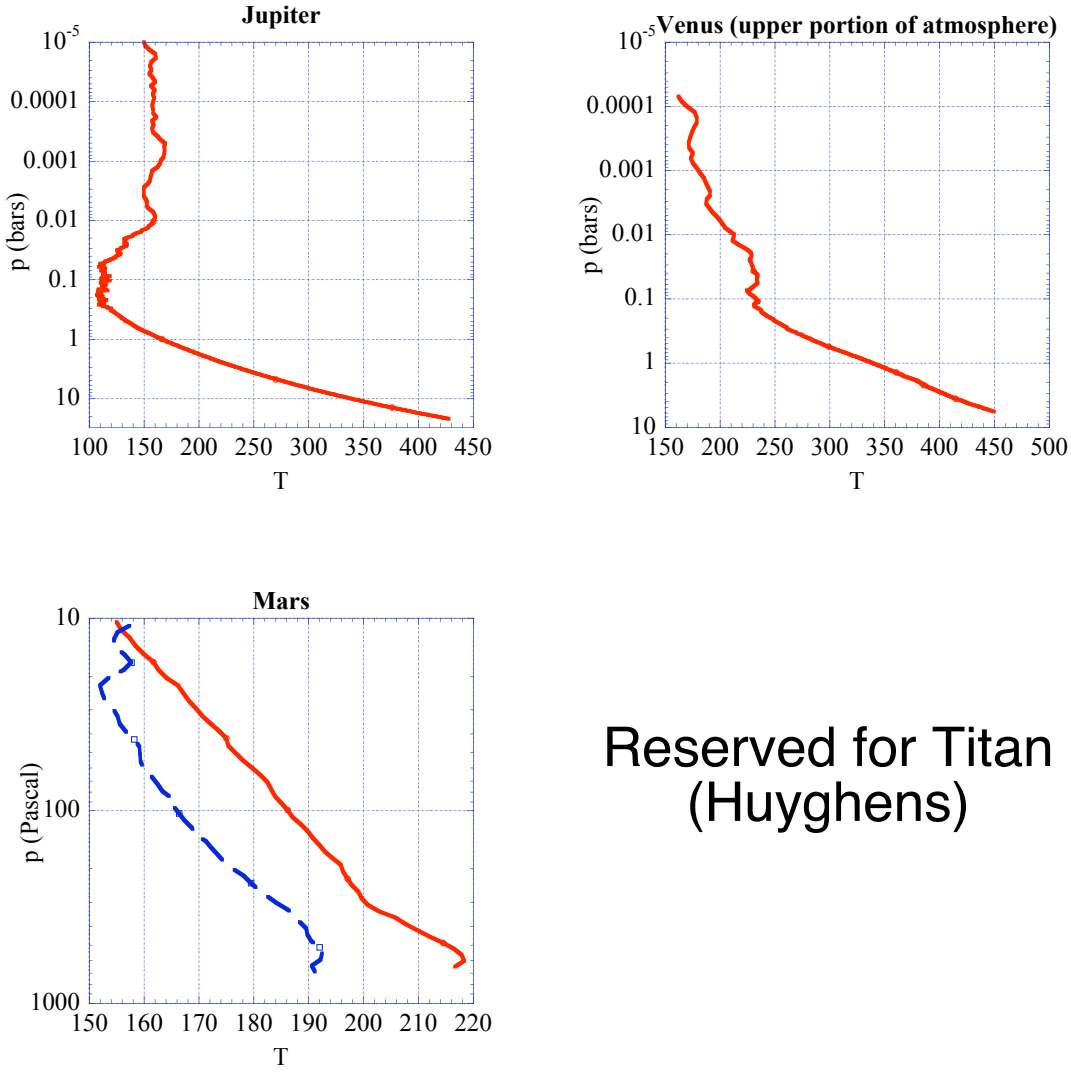

Figure 2.2: The vertical profile of temperature for a portion of the atmosphere of Jupiter (upper left panel), for the upper part of the atmosphere of Venus (upper right panel) , and for the full depth of the atmosphere of Mars (lower left panel) and Titan (lower right panel). The Venus and Mars data derive from observations of radio transmission through the atmosphere, taken by the Magellan and Global Surveyor orbiters, respectively. The Jupiter data derives from *in situ* deceleration measurements of the Galileo probe. The full Mars profile dataset reveals considerable seasonal and geographical variation. The two profiles shown here were taken at latitudes near 30N, near dawn during the Northern Hemisphere Summer.The colder sounding was taken near dawn, and the warmer in the afternoon.

of the force per unit area exerted on a surface in contact with the gas, in the direction perpendicular to the surface. In the mks units we employ throughout this book , pressure is measured in Pascals (Pa); 1 Pascal is 1 Newton of force per square meter of area, or equivalently 1 $kg/(ms^2)$. For historical reasons, atmospheric pressures are often measured in "bars" or "millibars." One bar, or equivalently 1000 millibars (mb) is approximately the mean sea-level pressure of the Earth's current atmosphere. We will often lapse into using mb as units of pressure, because the unit sounds comfortable to atmospheric scientists. For calculations, though, it is important to convert millibars to Pascals. This is easy, because 1 mb = 100 Pa. Hence, we should all learn to say "Hectopascal" in place of "millibar." It may take some time. When pressures are quoted in millibars or bars, one must make sure to convert them to Pascals before using the values in any thermodynamic calculations.

Density is simply the mass of the gas contained in a unit of volume. In mks units, it is measured in kg/m^3.

For a perfect gas, the three thermodynamic variables are related by the perfect gas equation of state, which can be written

$$p = knT \tag{2.1}$$

where p is the pressure, n is the number of molecules per unit volume (which is proportional to density) and T is the temperature. k is the *Boltzmann Thermodynamic Constant*, a universal constant having dimensions of energy per unit temperature. Its value depends only on the units in which the thermodynamic quantities are measured. To relate n to density, we divide it by the mass of a single molecule of the gas. Almost all of this mass comes from the protons and neutrons in the molecule, since electrons weigh next to nothing in comparison. Moreover, the mass of a neutron differs very little from the mass of a proton, so for our purposes the mass of the molecule can be taken to be $M \cdot \mu$ where μ is the mass of a proton and M is the *molecular weight* – an integer giving the count of neutrons and protons in the molecule. (The equivalent count for an individual atom of an element is the *atomic weight*). The density is thus $\rho = n \cdot M \cdot \mu$). If we define the *Universal Gas Constant* as $R^* \equiv k/\mu$ the perfect gas equation of state can be rewritten

$$p = \frac{R^*}{M}\rho T \tag{2.2}$$

In mks units, $R^* = 8314.5 Pa \cdot m^3/kg \cdot K$ We can also define a gas constant $R = R^*/M$ particular to the gas in question. For example, dry Earth air has a mean molecular weight of 28.97, so R_{dryair} = 287 $m^2/(s^2K)$, in mks units.

If μ is measured in kilograms, then $1/\mu$ is the number of protons needed to make up a kilogram. This large number is known as a *Mole*, and is commonly used as a unit of measurement of numbers of molecules, just as one commonly counts eggs by the dozen. For any substance, a quantity of that substance whose mass in kilograms is equal to the molecular weight of the substance will contain one Mole of molecules. For example, 2 kg of H_2 is a Mole of Hydrogen molecules, while 32 kg of the most common form of O_2 is a Mole of molecular Oxygen. If n were measured in $Moles/m^3$ instead of molecules per m^3, then density would be $\rho = n \cdot M$. One can also define the *gram-mole* (or *mole* for short), which is the number of protons needed to make a gram; this number is known as *Avogadro's number*, and is approximately $6.022 \cdot 10^{23}$.

Generally speaking, a gas obeys the perfect gas law when it is tenuous enough that the energy stored in forces between the molecules making up the gas is negligible. Deviations from the perfect gas law can be very important for the dense atmosphere of Venus, but for the purposes of the current atmosphere of Earth or Mars, or the upper part of the Jovian or Venusian atmosphere, the perfect gas law can be regarded as an accurate model of the thermodynamics– in fact, "perfect," one might say.

An extension of the concept of a perfect gas is the *law of partial pressures*. This states that, in a mixture of gases in a given volume, each component gas behaves just as it would if it occupied the volume alone. The pressure due to one component gas is called the *partial pressure* of that gas. Consider a gas which is a mixture of substance A (with molecular weight M_A) and substance B (with molecular weight M_B). The partial pressures of the two gases are

$$p_A = kn_AT, p_B = kn_BT \tag{2.3}$$

or equivalently,

$$p_A = R_A\rho_AT, p_B = R_B\rho_BT \tag{2.4}$$

where $R_A = R^*/M_A$ and $R_B = R^*/M_B$. The same temperature appears in both equations, since thermodynamic equilibrium dictates that all components of the system have the same temperature. The ratio of partial pressures of any two components of a gas is a convenient way to describe the composition of the gas. From Eq. 2.3, $p_A/p_B = n_A/n_B$, so the ratio of partial pressure of A to that of B is also the ratio of number of molecules of A to the number of molecules of B. This ratio is called the *molar mixing ratio*. When we refer to a mixing ratio without qualification, we will generally mean the molar mixing ratio. Alternately, one can describe the composition in terms of the ratio of partial pressure of one component to total presssure of the gas ($p_A/(p_A + p_B)$ in the two-component example). Summing the two partial pressure equations in Eq. 2.3, we see that this is also the ratio of number of molecules of A to total number of molecules; hence we will use the term *molar concentration* for this ratio [1]. If η_A is the molar mixing ratio of A to B, then the molar concentration is $\eta_A/(1+\eta_A)$, from which we see that for the molar concentration and molar mixing ratio are nearly the same for substances which are very dilute (i.e. $\eta_A \ll 1$).

Exercise 2.2.1 Show that a mixture of gases with molar concentrations $\eta_A = n_A/(n_A + n_B)$ and $\eta_B = n_B/(n_A + n_B)$ behaves like a perfect gas with mean molecular weight $M = \eta_AM_A + \eta_BM_B$. (i.e. derive the expression relating total pressure $p_A + p_B$ to total density $\rho_A + \rho_B$ and identify the effective gas constant). Compute the mean molecular weight of dry Earth air. (Dry Earth air consists primarily of 78.084% N_2, 20.947% O_2, and .934% Ar, by count of molecules.)

The *mass mixing ratio* is the ratio of the mass of substance A to that of substance B in a given parcel of gas, i.e. ρ_A/ρ_B. From Eq. 2.4 it is related to the molar mixing ratio by

$$\frac{\rho_A}{\rho_B} = \frac{M_A}{M_B}\frac{p_A}{p_B} \tag{2.5}$$

Throughout this book,we will use the symbol r to denote mass mixing ratios and η for molar mixing ratios, with subscripts added as necessary to distinguish the species involved. Yet another measure of composition is *specific concentration*, defined as the ratio of the mass of a given substance to the total mass of the parcel (e.g. $\rho_A/(\rho_A + \rho_B)$ in the two-component case). We'll use the symbol q, with subscripts as necessary, to denote the specific concentration of a substance. Using the law

[1] The term *volumetric* mixing ratio or concentration is often used interchangeably with the term *molar*, as in "ppmv" for "parts per million by volume." The reason for this nomenclature is that the volume occupied by a given quantity of gas at a fixed temperature and pressure is proportional to the number of molecules of the gas contained in that quantity. To see this, write $n = N/V$, where N is the number of molecules and V is the volume they occupy. Then, the ideal gas law can be written in the alternate form $V = (kT/p)N$. Hence the ratio of standardized volumes is equal to the molar mixing ratio, and so forth. Abbreviations like "ppmv" for molar mixing ratios are common and convenient, because the "v" can unambiguously remind us that we are talking about a volumetric (i.e. molar) mixing ratio or concentration, whereas in an abbreviation like "ppmm" one is left wondering whether the second "m" means "mass" or "molar."

of partial pressures, the specific concentration of substance A in a mixture is related to the molar concentration by

$$\frac{\rho_A}{\rho_{tot}} = \frac{M_A}{\bar{M}} \frac{p_A}{p_{tot}} \tag{2.6}$$

where $\bar{M}$ is the mean molecular weight of the mixture, with the mean being computed using weighting according to molar concentrations of the species, as in Exercise 2.2.1.

All of the ratios we have just defined are convenient to use because, unlike densities, they remain unchanged as a parcel of air expands or contracts, provided the constituents under consideration do not undergo condensation, chemical reaction or other forms of internal sources or sinks. Hence, for a compressible gas, two components A and B are well-mixed relative to each other if the mixing ratio between them is independant of position.

Constituents will tend to become well mixed over a great depth of the atmosphere if they are created or destroyed slowly, if at all, relative to the characteristic time required for mixing. In the Earth's atmosphere, the mixing ratio of oxygen to nitrogen is virtually constant up to about 80km above the surface. The mixing ratio of carbon dioxide in air can vary considerably in the vicinity of sources at the surface, such as urban areas where much fuel is burned, or under forest canopies when photosynthesis is active. Away from the surface, however, the carbon dioxide mixing ratio varies little. Variations of a few parts per million can be detected in the relatively slowly mixed stratosphere, associated with the industrial-era upward trend in fossil fuel carbon dioxide emissions. Small seasonal and interhemispheric fluctuations in the tropospheric mixing ratio, associated with variations in the surface sources, can also be detected. For most purposes, though, carbon dioxide can be regarded as well mixed throughout the atmosphere. In contrast, water vapor has a strong internal sink in Earth's atmosphere, because it is condensible there; hence its mixing ratio shows considerable vertical and horizontal variations. Carbon dioxide, methane and ammonia are not condensible on Earth at present, but their condensation can become significant in colder planetary atmospheres.

Exercise 2.2.2 (a) In the year 2000, the concentration of CO_2 in the atmosphere was about 370 parts per million molar. What is the ratio p_{CO_2}/p_{tot}? Estimate p_{CO_2} in *mb*. Does the molar concentration differ significantly from the molar mixing ratio? What is the mass mixing ratio of CO_2 in air? What is the mass mixing ratio of *carbon* (in the form of CO_2) in air – i.e. how many kilograms of carbon would have to be burned into CO_2 in order to produce the CO_2 in 1 kg of air? Note: The mean molecular weight of air is about 29. (b) The molar concentration of O_2 in Earth air is about 20%. How many grams of O_2 does a 1 liter breath of air contain at sea level (1000mb)? At the top of Qomolangma (a.k.a. "Mt. Everest," about 300mb)? Does the temperature of the air (within reasonable limits) affect your answer much?

2.2.2 Specific heat and conservation of energy

Conservation of energy is one of the three great pillars upon which the edifice of thermodynamics rests. When expressed in terms of changes in the state of matter, it is known as the First Law of Thermodynamics. When a gas expands or contracts, it does work by pushing against the environment as its boundaries move. Since pressure is force per unit area, and work is force times distance, the work done in the course of an expansion of volume dV is pdV. This is the amount of energy that must be added to the parcel of gas to allow the increase in volume to take place. For atmospheric purposes, it is more convenient to do write all thermodynamic relations on a per unit mass basis. Dividing V by the mass contained in the volume yields ρ^{-1}, whence the work per unit mass is $pd\rho^{-1}$. This is not the end of energy accounting. Changing the temperature of a unit

mass of the substance while holding volume fixed changes the energy stored in the various motions of the molecules by an amount $c_v dT$, where c_v is a proportionality factor known as the *specific heat at constant volume*. For example it takes about 720 Joules of energy to raise the temperature of 1kg of air by 1K while holding the volume fixed. For ideal gases, the specific heat can depend on temperature, though the dependence is typically weak. For non-ideal gases, specific heat can depend on pressure as well.

Exercise 2.2.3 There are 20 students and one professor in a well-insulated classroom measuring 20 meters by 20 meters by 3 meters. Each person in the classroom puts out energy at a rate of 100 Watts (1 Watt = 1 Joule/second). The classroom is dark, except for a computer and LCD projector which together consume power at a rate of 200 Watts. The classroom is filled with air at a pressure of 1000mb (no extra charge). How much does the temperature of the classroom rise during the course of a 1 hour lecture?

Combining the two contributions to energy change we find the expression for the amount of energy that must be added per unit mass in order to accomplish a change of both temperature and volume:

$$\delta Q = c_v dT + p d\rho^{-1} \tag{2.7}$$

Using the perfect gas law, the heat balance can be re-written in the form

$$\delta Q = c_v dT + d(p\rho^{-1}) - \rho^{-1} dp = (c_v + R)dT - \rho^{-1} dp \tag{2.8}$$

From this relation, we can identify the *specific heat at constant pressure*, $c_p \equiv c_v + R$, which is the amount of energy needed to warm a unit mass by 1K while allowing it to expand enough to keep pressure constant.

The units in which we measure temperature are an artifact of the marks one researcher or other once decided to put on some device that responded to heat and cold. Since temperature is proportional to the energy per molecule of a substance, it would make sense to set the proportionality constant to unity and simply use energy as the measure of temperature. This not being common practice, one has occasion to make use of the *Boltzmann thermodynamic constant*, k, which expresses the proportionality between temperature and energy. More precisely, each degree of freedom in a system with temperature T has a mean energy $\frac{1}{2}kT$. For example, a gas made of rigid spherical atoms has three degrees of freedom per atom (one for each direction it can move), and therefore each atom has energy $\frac{3}{2}kT$ on average; a molecule which could store energy in the form of rotation or vibration would have more degrees of freedom, and therefore each molecule would have more energy at any given temperature. The energy-temperature relation is made possible by an important thermodynamic principle, the *equipartion principle*, which states that in equilibrium, each degree of freedom accessible to a system gets an equal share of the total energy of the system. In constrast to physical constants like the speed of light, the Boltzmann constant should not be considered a fundamental constant of the Universe. It is just a unit conversion factor.

2.2.3 Entropy, reversibility and Potential temperature; The Second Law

One cannot use Eqn 2.8 to define a "heat content" Q of a state (p, T) relative to a reference state (p_o, T_o), because the amount of heat needed to go from one state to another depends on the path in pressure-temperature space taken to get there; the right hand side of Eqn 2.8 is not an exact

differential. However, it can be made into an exact differential by dividing the equation by T and using the perfect gas law as follows:

$$ds \equiv \frac{\delta Q}{T} = c_p \frac{dT}{T} - R\frac{dp}{p} = c_p d\ln(Tp^{-R/c_p}) \tag{2.9}$$

assuming c_p to be constant. This equation defines the *entropy*, $s \equiv c_p \ln(Tp^{-R/c_p})$. (Entropy can still be defined if c_p is not constant, but the expression is somewhat more complicated and need not concern us for the moment.) Entropy is a nice quantity to work with because it is a *state variable* – its change between two states is independent of the path taken to get from one to the other. A process affecting a parcel of matter is said to be *adiabatic* if it occurs without addition or loss of heat from the parcel. By definition, $\delta Q = 0$ for adiabatic processes. In consequence, adiabatic processes leave entropy unchanged. They are *reversible*. Entropy can also be defined for gases whose specific heat depends on temperature and pressure, and for non-ideal gases, but the expression is more complicated.

The *Second Law of Thermodynamics* states that entropy never decreases for energetically closed systems – systems to which energy is neither added nor subtracted in the course of their evolution. The formal derivation of the law from the microscopic properties of molecular interactions is in many ways an unfinished work of science, but the tendency towards an increase in entropy – an increase in disorder – seems to be a nearly universal property of systems consisting of a great many interacting components. The Second Law is perhaps more intuitive when restated in the following way: *In an energetically closed system, heat flows from a hotter part of the system to a colder part of the system, causing the system to evolve toward a state of uniform temperature.* To see that this statement is equivalent to the entropy-increase principle, consider a thermally insulated box of gas having uniform pressure, but within which the left half of the mass is at temperature T_1 and the right half of the mass is at temperature $T_2 < T_1$. Now suppose that we transfer an amount of heat δQ from the left half of the box to the right half. This transfer leaves the net energy unchanged, but it changes the entropy. Specifically, according to Eq. 2.9, the entropy change summed over the two halves of the gas is $ds = (\frac{1}{T_2} - \frac{1}{T_1})\delta Q$. Since $T_2 < T_1$, this change is positive only if $\delta Q > 0$, representing a transfer from the hotter to the colder portion of the gas. Entropy can be increased by further heat transfers until $T_1 = T_2$, at which point the maximum entropy state has been attained.

The Second Law endows the Universe with an arrow of time. If one watches a movie of a closed system and sees that the system starts with large fluctuations of temperature (low entropy) and proceeds to a state of uniform temperature (high entropy), one knows that time is running forward. If one sees a thermally homogeneous object spontaneously generate large temperature inhomogeneities, then one knows that the movie is being run backwards. Note that the Second Law applies only to closed systems. The entropy of a subcomponent can decrease, if it exchanges energy with the outside world and increases the entropy of the rest of the Universe. This is how a refrigerator works.

Entropy can also be used to determine how the temperature of an air parcel changes when it is compressed or expanded adiabatically. This is important because it tells us what happens to temperature is a bit of the atmosphere is lifted from low altitudes (where the pressure is high) to higher altitudes (where the pressure is lower), provided the lifting occurs so fast that the air parcel has little time to exchange heat with its surroundings. If the initial temperature and pressure are (T, p), then conservation of entropy tells us that the temperature T_o found upon adiabatically compressing or expanding to pressure p_o is given by $Tp^{-R/c_p} = T_o p_o^{-R/c_p}$ This leads us to define the *potential temperature*

$$\theta = T(\frac{p}{p_o})^{-R/c_p} \tag{2.10}$$

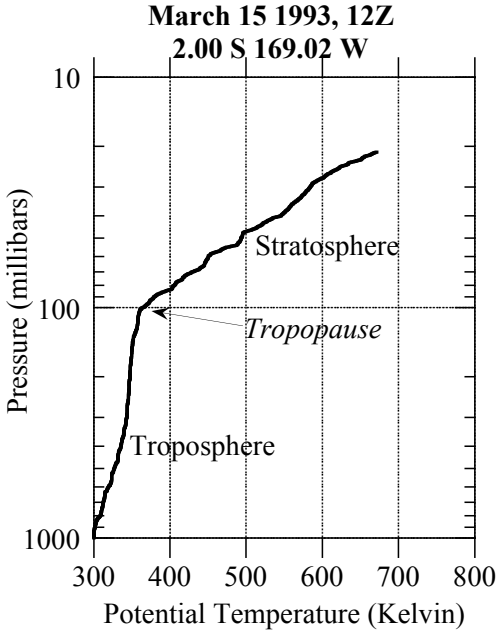

Figure 2.3: The dry potential temperature profile for the sounding in Figure 2.1

which is simply the temperature an air parcel would have if reduced adiabatically to a reference pressure p_o. Like entropy, potential temperature is conserved for adiabatic processes.

To understand why the presence of cold air above warm air in the sounding of Figure 2.1 does not succumb immediately to instability, we need only look at the corresponding profile of potential temperature, shown in Figure 2.3. This figure shows that potential temperature increases monotonically with height. This profile tells us that the air aloft is cold, but that if it were pushed down to lower altitudes, compression would warm it to the point that it is warmer than the surrounding air, and thus being positively buoyant, will tend to float back up to its original level rather than continuing its descent. We see also where the stratosphere gets its name: potential temperature increases very strongly with height there, so air parcels are very resistant to vertical displacement. This part of the atmosphere is therefore strongly stratified.

The troposphere is stable, but has much weaker gradients of θ. In a compressible atmosphere, a well-stirred layer would have constant θ rather than constant T, since it is the former that is conserved for adiabatic processes such as would be caused by rapid vertical displacements. This is the essence of the explanation for why temperature decreases with height: turbulent stirring relaxes the troposphere towards constant θ, yielding the *dry adiabat*

$$T(p) = \theta \cdot (\frac{p}{p_o})^{R/c_p} \tag{2.11}$$

In this formula, θ has the constant value $T(p_o)$. If we introduce the new vertical coordinate $\zeta = -\ln(p/p_o)$, then Equation 2.11 can be re-written $T(p) = T(p_o)exp(-(R/c_p)\zeta)$, from which we see that a dry adiabat shows up as a straight line on a plot of the logarithm of temperature vs. the logarithm of pressure.

It is evident from Figure 2.3 that something prevents θ from becoming completely well mixed. An equivalent way of seeing this is to compare the observed temperature profile with the dry adiabat. For example, if the air at 1000mb in Figure 2.3, having temperature 298K, were lifted dry-adiabatically to the tropopause, where the pressure is 100mb, then the temperature would be $298.(\frac{100}{1000})^{2/7}$, i.e. 154.3K (using the value $R/c_p = 2/7$ for Earth air). This is much colder than the observed temperature, which is 188K. We will see shortly that in the Earth's atmosphere, condensation of water vapor is one of the factors in play, though it is not the only one affecting the tropospheric temperature profile. The question of what determines the tropospheric θ gradient is at present still largely unsettled, particularly outside the Tropics.

It is no accident that the value of R/c_p for air lies close to the ratio of two small integers. It is a consequence of the equipartition principle. Using methods of statistical thermodynamics, it can be shown that a gas made up of molecules with n degrees of freedom has $R/c_p = 2/(n+2)$. Using the expression for the gas constant in terms of the specific heats, the adiabatic coefficient can also be written as $R/c_p = 1 - 1/\gamma$, where $\gamma = c_p/c_v$; for exact equipartion, $\gamma = 1 + 2/n$. The measured values of γ for a few common atmospheric gases are shown in Table 2.1. Helium comes close to the theoretical value for a molecule with no internal degrees of freedom, underscoring that excitation of electron motions plays little role in heat storage for typical planetary temperatures. The diatomic molecules have values closest to the theoretical value for $n = 5$, one short of what one would expect from adding two rotational and one vibrational internal degrees of freedom. Among the triatomic molecules, water acts roughly as if it had $n = 6$ while carbon dioxide is closer to $n = 7$. The two most complex molecules, methane and ammonia, are also characterized by $n = 7$. The failure of thermodynamics to access all the degrees of freedom classically available to a molecule is a consequence of quantum theory. Since the energy stored in states of motion of a molecule in fact comes in discrete-sized chunks, or "quanta," one can have a situation where a molecule hardly ever gets enough energy from a collision to excite even a single vibrational degree of freedom, for example, leading to the phenomenon of partial excitation or even non-excitation of certain classical degrees of freedom. This is one of many ways that the quantum theory, operating on exceedingly tiny spatial scales, exerts a crucial control over macroscopic properties of matter that can effect the very habitability of the Universe. Generally speaking, the higher the temperature gets, the more easy it is to excite internal degrees of freedom, leading to a decrease in γ. This quantum effect is the chief reason that specific heats vary somewhat with temperature.

Exercise 2.2.4 (a) A commercial jet airliner cruises at an altitude of 300mb. The air outside has a temperature of 240K. To enable the passengers to breathe, the ambient air is compressed to a cabin pressure of 1000mb. What would the cabin temperature be if the air were compressed adiabatically? How do you think airlines deal with this problem? (b) Discuss whether the lower portion of the Venus temperature profile shown in Figure 2.2 is on the dry CO_2 adiabat. (c) Assume that the Jupiter sounding is on a dry adiabat, and estimate the value of R/c_p for the atmosphere. Based on your result, what is the dominant constituent of the Jovian atmosphere likely to be? What other gas might be mixed with the dominant one?

2.3 Static stability of inhomogeneous mixtures

An atmosphere is *statically unstable* if an air parcel displaced from its original position tends to continue rising or sinking instead of returning to its original position. Such a state will tend to mix itself until it becomes stable. For a well-mixed atmosphere, the potential temperature profile tells the whole story about static stability, since, according to the ideal gas law, the density of an air parcel with potential temperature θ_0 will be $\rho_0 = p_1/(R\theta_0 \cdot (p_1/p_0)^{R/c_p})$ upon being elevated

	H_2O	CH_4	CO_2	N_2	O_2	H_2	He	NH_3
Crit. point T	647.1	190.44	304.2	126.2	154.54	33.2	5.1	405.5
Crit. point p	221.e5	45.96e5	73.825e5	34.0e5	50.43e5	12.98e5	2.28e5	112.8
Triple point T	273.15	90.67	216.54	63.14	54.3	13.95	2.17	195.4
Triple point p	611.	.117e5	5.185e5	.1253e5	.0015e5	.072e5	.0507e5	.061e5
L vap(b.p.)	22.55e5	5.1e5	–	1.98e5	2.13e5	4.54e5	.203e5	13.71e5
L vap(t.p.)	24.93e5	5.36e5	3.97e5	2.18e5	2.42e5	??	??	16.58e5
L fusion	3.34e5	.5868e5	1.96e5	.2573e5	.139e5	.582e5	??	3.314e5
L sublimation	28.4e5	5.95e5	5.93e5	2.437e5	2.56e5	??	??	19.89e5
ρ liq(b.p.)	958.4	450.2	1032.	808.6	1141.	70.97	124.96	682.
ρ liq(t.p.)	999.87	??	1110.	??	1307.	??	??	734.2
ρ solid	917.	509.3	1562.	1026.	1351.	88.	200.	822.6
c_p(0C/1bar)	1847.	2195.	820.	1037.	916.	14230.	5196.	2060.
$\gamma(c_p/c_v)$	1.331	1.305	1.294	1.403	1.393	1.384	1.664	1.309

Table 2.1: Thermodynamic properties of selected gases. Latent heats of vaporization are given at both the boiling point (the point where saturation vapor pressure reaches $1bar$) and the triple point. Liquid densities are given at the boiling point and the triple point. For CO_2 the 'boiling point' is undefined, so the liquid density is given at $253K/20bar$ instead. Note that the maximum density of liquid water is $1000.00kg/m^3$ and occurs at $-4C$. Densities of solids are given at or near the triple point. All units are mks, so pressures are quoted as Pa with the appropriate exponent. Thus, $1bar$ is written as $1e5$ in the table.

to an altitude with pressure $p_1 < p_0$. The ambient density there is $\rho_1 = p_1/(R\theta_1 \cdot (p_1/p_0)^{R/c_p})$. The displaced parcel will be negatively buoyant and return toward its original position if $\rho_0 > \rho_1$, which is true if and only if $\theta_0 < \theta_1$, i.e. if the potential temperature increases with height. For an inhomogeneous atmosphere, this is no longer the case, since the gas constant R depends on the mean molecular weight of the mixture, which varies from place to place. As an example, we may consider an atmosphere which has uniform θ, but which consists of pure N_2 for $p > p_0$ and pure CO_2 for $p < p_0$. In this case, the difference in density between a lifted N_2 parcel and that of the surrounding CO_2 is

$$\rho_0 - \rho_1 = \frac{1}{\theta}\left(\frac{1}{R_{N_2}(p_1/p_0)^{-(R/c_p)_{N_2}}} - \frac{1}{R_{CO_2}(p_1/p_0)^{-(R/c_p)_{CO_2}}}\right) \tag{2.12}$$

The value of R/c_p differs somewhat between N_2 and CO_2 but the main effect in this equation comes from the differing values of the gas constant. Since N_2 has lower molecular weight (28) than CO_2 (44), the gas constant for N_2 is considerably greater than the gas constant for CO_2. In consequence, $\rho_0 < \rho_1$, the lifted nitrogen parcel is positively buoyant, and the nitrogen layer will tend to spontaneously mix itself with the CO_2 layer despite the fact that both have the same potential temperature.

The phenomenon is very familiar: it is why helium balloons rise in air, even when they are at the same temperature as their surroundings. The low molecular weight of helium makes it lighter (i.e. lower density) than air having the same temperature and pressure.

Exercise 2.3.1 Make sense of the following statement: "For the Earth's atmosphere, moist air is lighter than dry air." Would this still be true for a planet whose atmosphere is mainly H_2?

There are a number of ways to deal with the effect of composition on static stability. For the case of moisture on the Earth, it is common to define a *virtual temperature*, which is the

temperature at which the gas law for dry air would yield the same density as the true gas law taking into account the lightening effect of water vapor. This approach has its virtues, but we find it less confusing to deal instead with *potential density*, which is the density an air parcel would have if reduced adiabatically to a standard reference pressure. Using the gas law, and the fact that mixing ratios are conserved (whence R/c_p is conserved on adiabatic compression of the parcel), the potential density is

$$\rho_p = \frac{p_o}{R\theta} = \frac{p_o}{RT}(\frac{p}{p_o})^{R/c_p} \tag{2.13}$$

From this equation it is evident that for a well-mixed system, R is independent of p, so that the system is stable precisely when θ increases with height. For an inhomogeneous mixture, the variations in R associated with varying composition can stabilize or destabilize the system. The variations in c_p can have a similar, though generally less pronounced, effect.

2.4 The hydrostatic relation

The hydrostatic relation relates pressure to altitude and the mass distribution of the atmosphere, and provides the chief reason that pressure is the most natural vertical coordinate to use in most atmospheric problems. Consider a column of any substance at rest, and suppose that the density of the substance as a function of height z is given by $\rho(z)$. Suppose further that the range of altitudes being considered is small enough that the acceleration of gravity is essentially constant; The magnitude of this acceleration will be called g, and the force of gravity is taken to point along the direction of decreasing z. Now, consider a slice of the column with vertical thickness dz, having cross sectional area A in the horizontal direction. Since pressure is simply force per unit area, then the change in pressure from the base of this slice to the top of this slice is just the force exerted by the mass. By Newton's law, then, we have

$$Adp = -Agdm = -Ag\rho dz \tag{2.14}$$

where dm is the increment of mass in the column per unit area. An immediate consequence of this relation is that

$$dm = -\frac{dp}{g} \tag{2.15}$$

which states that the amount of mass in a slab of atmosphere is proportional to the thickness of that slab, measured in pressure coordinates. A further consequence, upon dividing by dz is the relation

$$\frac{dp}{dz} = -\rho g \tag{2.16}$$

This differential equation expresses the *hydrostatic relation*. It is exact if the substance is at rest (hence the "static"), but if the material of the column is in motion, the relation is still approximately satisfied provided the acceleration is sufficiently small, compared to the acceleration of gravity. In practice, the hydrostatic relation is very accurate for most problems involving large scale motions in planetary atmospheres. It would not be a good approximation within small scale intense updrafts or downdrafts where the acceleration of the fluid may be large. Derivation of the precise conditions under which the hydrostatic approximation holds requires consideration of the equations of fluid motion, which will be taken up in a sequel to the present book.

An important consequence of the hydrostatic relation is that it enables us to determine the total mass of an atmosphere through measurements of pressure taken at the surface alone.

Integrating Eqn 2.15 from the ground ($p = p_s$) to space ($p = 0$) yields the relation

$$m = \frac{p_s}{g} \tag{2.17}$$

where m is the total mass of the atmosphere located over a unit area of the planet's surface. Note that this relation presumes that the depth of the layer containing almost all the mass of the atmosphere is sufficiently shallow that gravity can be considered constant throughout the layer. Given that gravity decays inversely with the square of distance from the planet's center, this is equivalent to saying that the atmosphere must be shallow compared to the radius of the planet. For a well mixed substance A with mass-specific concentration κ_A relative to the whole atmosphere, the mass of substance A per square meter of the planet's surface is just $m\kappa_A$

Using the perfect gas law to eliminate ρ from Eqn 2.16 yields

$$\frac{dp}{dz} = -\frac{g}{RT}p \tag{2.18}$$

where R is the gas constant for the mixture making up the atmosphere. This has the solution

$$p(z) = p_s \exp(-\frac{g}{R\bar{T}}z), \bar{T}(z) = (\frac{1}{z}\int_0^z T^{-1}dz)^{-1} \tag{2.19}$$

Here, $\bar{T}(z)$ is the harmonic mean of temperature in the layer between the ground and altitude z. If temperature is constant, then pressure decays exponentially with scale height RT/g. Because temperature is measured relative to absolute zero, the mean temperature $\bar{T}(z)$ can be relatively constant despite fairly large variations of temperature within the layer. In consequence, pressure typically decays roughly exponentially with height even when temperature is altitude-dependent.

Exercise 2.4.1 Compute the mass of the Earth's atmosphere, assuming a mean surface pressure of 1000mb. (The Earth's radius is 6378km, and the acceleration of its gravity is $9.8m/s^2$). Compute the mass of the Martian atmosphere, assuming a mean surface pressure of 6mb. (Mars' radius is 3390km, and the acceleration of its gravity is $3.7m/s^2$.)

Note that the hydrostatic relation applies only to the total pressure of all constituents; it does not apply to partial pressures individually. However, in the special case in which the gases are well mixed, the total mass of each well-mixed component can still be determined from surface data alone. One simply multiplies the total mass obtained from surface pressure, by the appropriate (constant) mass-specific concentration.

In the study of atmospheric dynamics, the hydrostatic equation is used to compute the pressure gradients which drive the great atmospheric circulations. Outside of dynamics, there are rather few problems in physics of climate that require one to know the altitude corresponding to a given pressure level. Our main use of the hydrostatic relation in this book will be in the form of Eqn 2.15, which tells us the mass between two pressure surfaces.

The hydrostatic relation also allows us to derive a useful alternate form of the heat budget, by re-writing the heat balance equation as follows:

$$\delta Q = c_p dT - \rho^{-1}dp = c_p dT - \rho^{-1}\frac{dp}{dz}dz = d(c_p T + gz) \tag{2.20}$$

assuming c_p to be constant. The quantity $c_p T + gz$ is known as the *dry static energy*. Dry static energy provides a more convenient basis for atmospheric energy budgets than entropy, since changes in dry static energy following an air parcel are equal to the net energy added to or removed from the parcel by heat sources such as solar radiation.

2.5 Thermodynamics of phase change

When a substance changes from one form to another (e.g. water vapor condensing into liquid water or gaseous carbon dioxide condensing into dry ice) energy is released or absorbed even if the temperature of the mass is unchanged after the transformation has taken place. This happens because the amount of energy stored in the form of intermolecular interactions is generally different from one form, or *phase* to another. The amount of energy released when a unit of mass of a substance changes from one phase to another, holding temperature constant, is known as the *latent heat* associated with that phase change. By convention, latent heats are stated as positive numbers, with the phase change going in the direction that releases energy. Phase changes are *reversible*. If one kilogram of matter releases L joules of energy in going from phase A to phase B, it will take the same L joules of energy to turn the mass back into phase A. The units of latent heat are energy per unit mass (Joules per kilogram in mks units).

Condensible substances play a central role in the atmospheres of many planets and satellites. On Earth, it is water that condenses, both into liquid water and ice. On Mars, CO_2 condenses into dry ice in clouds and in the form of frost at the surface. On Jupiter and Saturn, not only water but ammonia (NH_3) and a number of other substances condense. The thick clouds of Venus are composed of condensed sulfuric acid. On Titan it is methane, and on Neptune's moon Triton nitrogen itself condenses. Table 2.1 lists the latent heats for the liquid-vapor (evaporation), liquid-solid (fusion) and solid-vapor (sublimation) phase transitions are given for a number of common constituents of planetary atmospheres. Water has an unusually large latent heat; the condensation of 1 kg of water vapor into ice releases nearly five times as much energy as the condensation of 1kg of carbon dioxide gas into dry ice. This is why the relatively small amount of water vapor in Earth's present atmosphere can nonetheless have a great effect on atmospheric structure and dynamics. Ammonia also has an unusually large latent heat, though not so much so as water. In both cases, the anomalous latent heat arises from the considerable energy needed to break hydrogen bonds in the condensed phase.

Like most thermodynamic properties, latent heat varies somewhat with temperature. For example, the latent heat of vaporization of water is $2.5 \cdot 10^6 J/kg$ at 0C, but only $2.26 \cdot 10^6 J/kg$ at 100C. For precise calculations, the variation of latent heat must be taken into account, but nonetheless for many purposes it will be sufficient to assume latent heat to be constant over fairly broad temperature ranges.

The three main phases of interest are solid, liquid and gas (also called vapor), though other phases can be important in exotic circumstances. There is generally a *triple point* in temperature-pressure space where all three phases can co-exist. Above the triple point temperature, the substance undergoes a vapor-liquid phase transition as temperature is decreased or pressure is increased; below the triple point temperature vapor condenses directly into solid, once thermodynamic equilibrium has been attained. For water, the triple point occurs at a temperature of 273.15K and pressure of 6.11mb (see Table 2.1 for other gases). Generally, the triple point temperature can also be taken as an approximation to the "freezing point" – the temperature at which a liquid becomes solid – because the freezing temperature varies only weakly with pressure until very large pressures are reached. Though we will generally take the freezing point to be identical to the triple point in our discussions, the effect of pressure on freezing of liquid can nonetheless be of great importance at the base of glaciers and in the interior of icy planets or moons, and perhaps also in very dense,cold atmospheres.

Typically, the solid phase is more dense than the liquid phase, but water again is exceptional. Water ice floats on liquid water, whereas carbon dioxide ice would sink in an ocean of liquid carbon

dioxide, and methane ice would sink in a methane lake on Titan. This has profound consequences for the climates of planets with a water ocean such as Earth has, since ice formed in winter remains near the surface where it can be more readily melted when summer arrives.

Exercise 2.5.1 Per square meter, how many Joules of energy would be required to evaporate a puddle of Methane on Titan, having a depth of 20m?

Atmospheres can transport energy from one place to another by heating an air parcel by an amount δT, moving the parcel vertically or horizontally, and then cooling it down to its original temperature. This process moves an amount of heat $c_p \delta T$ per unit mass of the parcel. Latent heat provides an alternate way to transport energy, since energy can be used to evaporate liquid into an air parcel until its mixing ratio increases by δr, moving it and then condensing the substance until the mixing ratio returns to its original value. This process transports an amount of heat $L\delta r$ per unit mass of the planet's uncondensible air, and can be much more effective at transporting heat than inducing temperature fluctuations, especially when the latent heat is large. "Ordinary" heat – the kind that feels hot when you touch it, and which is stored in the form of the temperature increase of a substance – is known in atmospheric circles as "sensible" heat.

All gases are condensible at low enough temperatures or high enough pressures. On Earth (in the present climate) CO_2 is not a condensible substance, but on Mars it is. The ability of a gas to condense is characterized by the *saturation vapor pressure*, p_{sat} of that gas, which may be a function of any number of thermodynamic variables. When the partial pressure p_A of gas A is below $p_{sat,A}$, more of the gas can be added, raising the partial pressure, without causing condensation. However, once the partial pressure reaches $p_{sat,A}$, any further addition of A will condense out. The state $p_A = p_{sat,A}$ is referred to as "saturated" with regard to substance A. Each condensed state (e.g. liquid or solid) will have its own distinct saturation vapor pressure. Rather remarkably, for a mixture of perfect gases, the saturation vapor pressure of each component is independent of the presence of the other gases. Water vapor mixed with 1000 mb worth of dry air at a temperature of 300K will condense when it reaches a partial pressure of 38mb; a box of pure water vapor at 300K condenses at precisely the same 38mb. If a substance "A" has partial pressure p_A that is below the saturation vapor pressure, it is said to be "subsaturated," or "unsaturated." The degree of subsaturation is measured by the *saturation ratio* $p_a/p_{sat,A}$, which is often stated as a percent. Applied to water vapor, this ratio is called the *relative humidity*, and one often speaks of the relative humidity of other substances, e.g. "methane relative humidity" instead of saturation ratios. Note that the relative humidity is also equal to the *mixing ratio* of the substance A in a given mixture to the *mixing ratio* the air would have if the substance were saturated. This is different from the ratio of *specific humidity* to *saturation specific humidity*, or the ratio of *molar concentration* to *saturation molar concentration* except when the mixing ratio is small.

It is intuitively plausible that the saturation vapor pressure should increase with increasing temperature, as molecules move faster at higher temperatures, making it harder for them to stick together to form condensate. The temperature dependence of saturation vapor pressure is expressed by a remarkable thermodynamic relation known as the *Clausius-Clapeyron equation*. It is derived from very general thermodynamic principles, via a detailed accounting of the work done in an reversible expansion-contraction cycle crossing the condensation threshold, and requires neither approximation nor detailed knowledge of the nature of the substance condensing. The relation reads

$$\frac{dp_{sat}}{dT} = \frac{1}{T} \frac{L}{\rho_v^{-1} - \rho_c^{-1}} \tag{2.21}$$

where ρ_v is the density of the less condensed phase, ρ_c is the density of the more condensed phase, and L is the latent heat associated with the transformation to the more condensed phase. For vapor

to liquid or solid transitions, $\rho_c \gg \rho_v$, enabling one to ignore the second term in the denominator of Eqn 2.21. Further, upon substituting for density from the perfect gas law, one obtains the simplified form

$$\frac{dp_{sat}}{dT} = \frac{L}{R_A T^2} p_{sat} \tag{2.22}$$

where R_A is the gas constant for the substance which is condensing. If we make the approximation that L is constant, then Eqn 2.21 can be integrated analytically, resulting in

$$p_{sat}(T) = p_{sat}(T_o) e^{-\frac{L}{R_A}(\frac{1}{T} - \frac{1}{T_o})} \tag{2.23}$$

where T_o is some reference temperature. This equation shows that saturation water vapor content is very sensitive to temperature, decaying rapidly to zero as temperature is reduced and increasing rapidly as temperature is increased. The rate at which the change occurs is determined by the characteristic temperature $\frac{L}{R_A}$ appearing in the exponential. For the transition of water vapor to liquid, it has the value 5420K at temperatures near 300K. For CO_2 gas to dry ice, it is $3138K$, and for methane gas to liquid methane it is $1031K$. Equation 2.22 seems to imply that the p_{sat} asymptotes to a constant value when $T \gg L/R_A$. This is a spurious limit, though, since the assumption of constant L invariably breaks down over such large temperature ranges. In fact, L typically approaches zero at some *critical temperature*, where the distinction between the two phases disappears. For water vapor, this *critical point* occurs at a temperature and pressure of 647.1K and 221bars. For carbon dioxide, the critical point occurs for the vapor-liquid transition, at 304.2K and 73.825 bars. Critical points for other atmospheric gases are shown in Table 2.1. At high pressures, the solid/liquid phase boundary does not typically terminate in a critical point, but instead gives way to a bewildering variety of distinct solid phases distinguished primarily by crystal structure.

Figure 2.4 summarizes the features of a typical phase diagram. Over ranges of a few bars of pressure, the solid-liquid boundary can be considered nearly vertical. In fact the exact form of the Clausius-Clapeyron relation (Eq. 2.21) tells us why the boundary is nearly vertical and how it deviates from verticality. Because the difference in density between solid and liquid is typically quite small while the latent heat of fusion is comparatively large, Eq. 2.21 implies that the slope dp/dT is very large (i.e. nearly vertical. The equation also tells us that in the "normal" case where ice is denser than liquid, the phase boundary tilts to the right, and so the freezing temperature increases with pressure; at fixed pressure, one can cause a cold liquid to freeze by squeezing it. The unusual lightness of water ice relative to the liquid phase implies that instead the phase boundary tilts to the left; one can melt solid ice by squeezing it. Substituting the difference in density between water ice and liquid water, and the latent heat of fusion, into Eq. 2.21, we estimate that $100 bars$ of pressure decreases the freezing point temperature by about $.74K$. This is roughly the pressure caused by about a kilometer of ice on Earth. The effect is small, but can nonetheless be significant at the base of thick glaciers.

Below the triple point temperature, the favored transition is gas/solid,and so the appropriate latent heat to use in the Clausius-Clapeyron relation is the latent heat of sublimation. Above the triple point, the favored transition is gas/liquid, whence one should use the latent heat of vaporization. The triple point (T, p) provides a convenient base for use with the simplified Clausius-Clapeyron solution in Eqn. 2.23, or indeed for a numerical integration of the relation with variable L. Results for water vapor are shown in Figure 2.5. These results were computed using the constant L approximation for sublimation and vaporization, but in fact a plot of the empirical results on a logarithmic plot of this type would not be distinguishable from the curves shown. The more exact result does differ from the constant L idealization by a few percent, which can be important in some applications. Be that as it may, the figure reveals the extreme sensitivity of vapor pressure

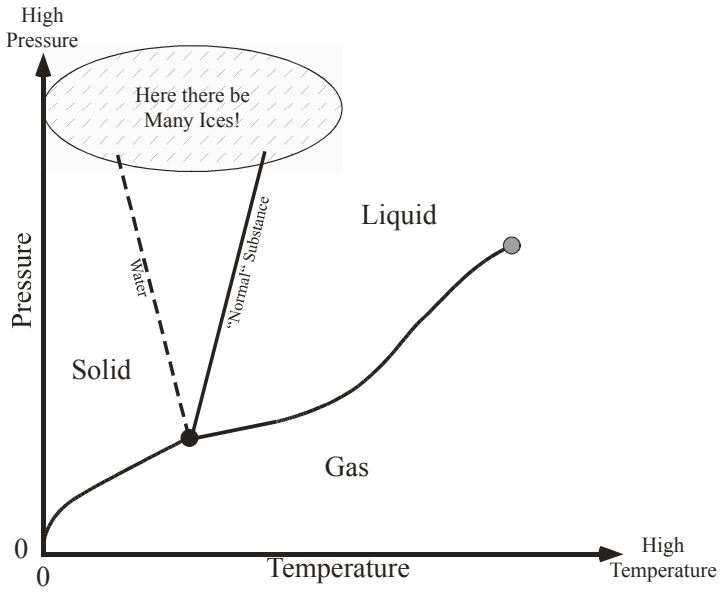

Figure 2.4: The general form of a phase diagram showing the regions of temperature-pressure space where a substance exists in solid, liquid or gaseous forms. The triple point is marked with a black circle while the critical point is marked with a grey circle. The solid-liquid phase boundary for a "normal" substance (whose solid phase is denser than its liquid phase) is shown as a solid curve, whereas the phase boundary for water (ice less dense than liquid) is shown as a dashed curve. The critical point pressure is typically several orders of magnitude above the triple point pressure, while the critical point temperature is generally only a factor of two or three above the triple point temperature. Therefore, the pressure axis on this diagram should be thought of as logarithmic, while the temperature axis should be thought of as linear. This choice of axes also reflects the fact that the pressure must typically be changed by an order of magnitude or more to cause a significant change in the temperature of the solid/liquid phase transition.

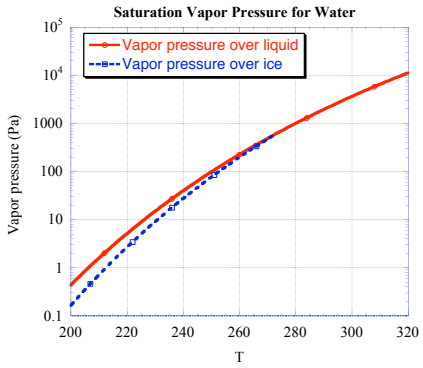

Figure 2.5: Saturation vapor pressure for water, based on the constant-L form of the Clausius-Clapeyron relation. Curves are shown for vapor pressure based on the latent heat of vaporization, and (below freezing) for latent heat of sublimation. The latter is the appropriate curve for sub-freezing temperatures.

to temperature. The vapor pressure ranges from about .1 Pascals at 200K (the tropical tropause temperature) to 35mb at a typical tropical surface temperature of 300K, rising further to 100mb at 320K. Over this span of temperatures, water ranges from a trace gas to a major constituent; at temperatures much above 320K, it rapidly becomes the dominant constituent of the atmosphere. Note also that the distinction between the ice and liquid phase transitions has a marked effect on the vapor pressure. Because the latent heat of sublimation is larger than the latent heat of vaporization, the vapor pressure over ice is lower than the vapor pressure over liquid would be, at subfreezing temperatures. At 200K, the ratio is nearly a factor of three.

Exercise 2.5.2 Let's consider once more the case of the airliner cruising at an altitude of 300mb, discussed in an earlier Exercise. Suppose that the ambient air at flight level has 100% relative humidity. What is the relative humidity once the air has been brought into the cabin, compressed to 1000mb, and chilled to a room temperature of 290K?

Once the saturation vapor pressure is known, one can compute the molar or mass mixing ratios with respect to the background non-condensible gas, if any, just as for any other pair of gases. The saturation vapor pressure is used in this calculation just like any other partial pressure. For example, the molar mixing ratio is just p_{sat}/p_a, if p_a is the partial pressure of the noncondensible background. Note that, while the saturation vapor pressure is independent of the pressure of the gas with which the condensible substance is mixed, the saturation mixing ratio is not.

Exercise 2.5.3 What is the saturation molar mixing ratio of water vapor in air at the ground in tropical conditions (1000mb and 300K)? What is the mass mixing ratio? What is the mass-specific humidity? What is the molar mixing ratio (in ppm) of water vapor in air at the tropical tropopause (100mb and 200K)?

2.6 The moist adiabat

When air is lifted, it cools by adiabatic expansion, and if it gets cold enough that one of the components of the atmosphere begins to condense, latent heat is released. This makes the lifted air parcel warmer than the dry adiabat would predict. The resulting temperature profile will be referred to as the *moist adiabat*, regardless of whether the condensing substance is water vapor (as on Earth) or something else (CO_2 on Mars or methane on Titan). We now proceed to make this quantitative.

The simplest case to consider is that of a single component atmosphere, which can attain cold enough temperatures to reach saturation and condense. This case is relevant to present Mars, which has an almost pure CO_2 atmosphere that can condense in the cold Winter hemisphere and at upper levels at any time of year. A pure CO_2 atmosphere with a surface pressure on the order of two or three bars is a commonly used model of the atmosphere of Early Mars, though the true atmospheric composition in that instance is largely a matter of speculation. Another important application of a single component condensible atmosphere is the pure steam (water vapor) atmosphere, which occurs when a planet with an ocean gets warm enough that the mass of water which evaporates into the atmosphere dominates the other gases that may be present. This case figures prominenently in the *runaway greenhouse* effect that will be studied in Chapter 4.

For a single component atmosphere, the partial pressure of the condensible substance is in fact the total atmospheric pressure. Therefore, at saturation, the pressure is related to the temperature by the Clausius-Clapeyron relation. To find the saturated moist adiabat, we simply solve for T in terms of p_{sat} in the Clausius-Clapeyron relation, and recall that $p = p_{sat}$ because we are assuming the atmosphere to be saturated – that is, any reduction in temperature or increase in pressure leads to condensation. Using the simplified form of Clausius-Clapeyron given in Eqn 2.23, the saturated moist adiabat would be

$$T(p) = \frac{T_o}{1 - \frac{RT_o}{L} \ln \frac{p}{p_{sat}(T_o)}} \tag{2.24}$$

where R is the gas constant for the substance making up the atmosphere. Without loss of generality, we may suppose that T_o is taken to be the surface temperature, so that $p_{sat}(T_o)$ is the surface pressure p_s. Since the logarithm is negative, the temperature decreases with altitude (recalling that lower pressure corresponds to higher altitude). Further, the factor multiplying the logarithm is the ratio of the surface temperature to the characteristic temperature L/R. Since the characteristic temperature is large, the prefactor is small, and as a result the temperature of saturated adiabat for a one-component atmosphere varies very little over a great range of pressures. For example, in the case of the CO_2 vapor-ice transition, an atmospheric surface pressure of $7mb$ (similar to that of present Mars) would be in equilibrium with a surface dry-ice glacier at a temperature of $149K$; at $.07mb$ – one one-hundredth of the surface pressure – the temperature on the saturated adiabat would only fall to $122K$.

Exercise 2.6.1 In the above example, what would the temperature aloft have been if there were no condensation and the parcel were lifted along the dry adiabat?

Unless there is a reservoir of condensate at the surface to maintain saturation, it would be rare for an atmosphere to be saturated all the way to the ground. Suppose now that a one-component atmosphere has warm enough surface temperature that the surface pressure is lower than the saturation vapor pressure computed at the surface temperature. In this case, when a parcel is lifted by convection, its temperature will follow the *dry* or *noncondensing* adiabat, until the

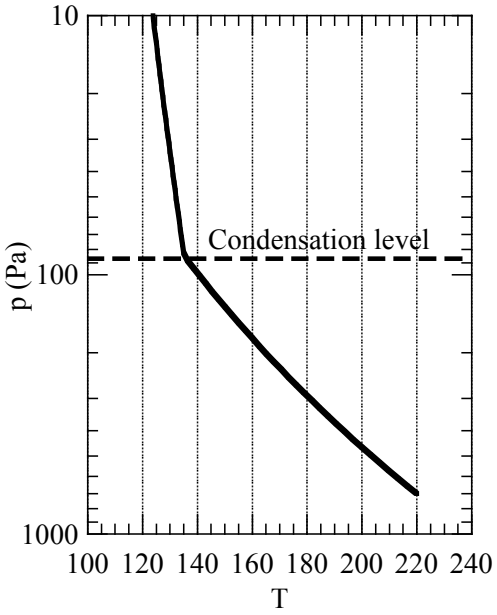

Figure 2.6: The adiabatic profile for a pure CO_2 atmosphere with a surface pressure of 700 Pa (7mb) and a surface temperature of 220K. The conditions are similar to those encountered on present-day Mars.

temperature falls so much that the gas becomes saturated. The level at which this occurs is called the *lifted condensation level*. Above the lifted condensation level, ascent causes condensation and the parcel follows the saturated adiabat. Since the temperature curve along the saturated adiabat falls with altitude so much less steeply than the dry adiabat, it is very easy for the two curves to intersect provided the surface temperature is not exceedingly large. An example for present Martian conditions is shown in Figure 2.6. A comparison with the Martian profiles in Figure 2.2 indicates that something interesting is going on in the Martian atmosphere. For the warm sounding, whose surface temperature is close to $220K$, the entire atmosphere aloft is considerably warmer than the adiabat, and the temperature nowhere comes close to the condensation threshold. Clearly, something we haven't taken into account is warming up the atmosphere. A likely candidate for the missing piece is the absorption of solar energy by dust.

Although results like Figure 2.6 show a region of weak temperature dependence aloft which bears a superficial resemblance to the stratosphere seen in Earth soundings (and also at the top of the Venus, Jupiter and Titan soundings), one should not jump to the conclusion that the stratosphere is caused by condensation. This is not generally the case, and there are other reasons for the upper atmospheric temperature structure, which will be taken up in the next few chapters.

As a final step up on the ladder of generality, let's consider a mixture of a condensible substance with a substance that doesn't condense under the range of temperatures encountered in the atmosphere under consideration. This might be a mixture of condensible methane on Titan with non-condensible nitrogen, or condensible carbon dioxide on Early Mars with non-condensible nitrogen, or water vapor on Earth with a non-condensible mixture of oxygen and

nitrogen. Whatever the substance, we distinguish the properties of the condensible substance with the subscript "c," and those of the non-condensible substance by the subscript "a" (for "air"). We now need to do the energy budget for a parcel of the mixture, assuming that it has been cooled down enough for the condensible substance to reach saturation, so that any further cooling results in formation of enough condensate (with concomitant release of latent heat) to keep the system from becoming supersaturated. We further introduce the assumption that essentially all of the condensate is immediately removed from the system, so that the heat storage in whatever mass of condensate is left in suspension may be neglected. This is a reasonable approximation for water or ice clouds on Earth, but even in that case the slight effect of the mass of retained condensate on buoyancy can be significant in some circumstances. In other planetary atmospheres the effect of retained condensate could be of greater importance. The temperature profile obtained by assuming condensate is removed from the system is called a *pseudoadiabat*, because the process is not truly reversible. One cannot return to the original saturated state, because the condensate is lost. At the opposite extreme, if all condensate is retained, it can be re-evaporated when the parcel is compressed, allowing for true reversibility.

Let the partial pressure, density, molecular weight, gas constant and specific heat of the noncondensible substance be p_a, ρ_a,M_a, R_a, and c_{pa}, and similarly for the condensible substance. Further, let L be the latent heat of the phase transition between the vapor and condensed phase of the condensible substance, and let $p_{c,sat}(T)$ be the saturation vapor pressure of this substance, as determined by the Clausius-Clapeyron relation. The assumption of saturation amounts to saying that $p_c = p_{c,sat}(T)$; if the parcel weren't at saturation, there would be no condensation and we could simply use the dry adiabat based on a noncondensing mixture of substance "a" and "c."

Now consider a parcel consisting of a mass m_a of noncondensible gas with an initial mass m_c of condensible gas. If the temperature is changed by an amount dT and the partial pressure of noncondensible gas is changed by an amount dp_a then the total heat budget of the parcel is

$$(m_a + m_c)\delta Q = m_a c_{pa} dT - \frac{m_a}{\rho_a} dp_a + m_c c_{pc} dT - \frac{m_c}{\rho_c} dp_c + L dm_c \tag{2.25}$$

where dm_c is the amount of mass lost to condensation. There is no term in this budget corresponding to heat storage in the condensed phase, since it is assumed that all condensate disappears from the parcel by precipitation. Technically, the temperature profile we will compute as a result is the *pseudoadiabat*, rather than the *adiabat*, since the removal of condensate makes the process irreversible. The usual way to change dp_a would be by lifting, causing expansion and reduction of pressure. Now, we divide by $m_c T$, make use of the perfect gas law to substitute for ρ_a and ρ_c, and make use of the fact that $m_c/m_a = (M_c/M_a)(p_c/p_a)$, since m_c/m_a is just the mass mixing ratio, denoted henceforth by r_c. This yields

$$(1 + r_c)\frac{\delta Q}{T} = c_{pa}\frac{dT}{T} - R_a\frac{dp_a}{p_a} + c_{pc} r_c \frac{dT}{T} - r_c R_c \frac{dp_c}{p_c} + \frac{L}{T} dr_c \tag{2.26}$$

The first two terms can be recognized as the contribution of the two substances to the dry entropy of the mixture, weighted according to relative abundance of each species. If there is no condensation, the mixing ratio is conserved as the parcel is displaced to a new pressure, $dr_c = 0$, and the expression reduces to the equivalent of Eqn. 2.9, leading to the dry adiabat for a mixture. At this point, we introduce the saturation assumption, which actually consists of two parts: First, we assume that the air parcel is initially saturated, so that before being displaced, $p_c = p_{c,sat}(T)$ and $r_c = r_{sat} = \epsilon p_{c,sat}(T)/p_a$, where ϵ is the ratio of molecular weights M_c/M_a and $p_{c,sat}(T)$ is determined by the Clausius-Clapeyron relation. Second, we assume that a displacement conserving r_c would cause supersaturation, so that condensation would occur and bring the partial pressure p_c back to the saturation vapor pressure corresponding to the new value of T. Usually, this

would occur as a result of ascent and cooling, since cooling strongly decreases the saturation vapor pressure. In rare circumstances, it can be compression that leads to condensation. More typically, though, the effect of compressional warming on saturation vapor pressure dominates the effect of increasing partial pressure, so that subsidence of initially saturated air follows the dry adiabat.

Assuming that the displacement causes condensation, we may replace p_c by $p_{c,sat}(T)$ and r_c by r_{sat} everywhere in Eqn. 2.26. Next, we use Clausius-Clapeyron to re-write $dp_{c,sat}$, observing that

$$\frac{dp_{c,sat}}{p_{c,sat}} = d\ln p_{c,sat} = \frac{d\ln p_{c,sat}}{dT}dT \tag{2.27}$$

and

$$dr_{sat} = \epsilon d\frac{p_{c,sat}}{p_a} = \epsilon\frac{p_c}{p_a}d\ln\frac{p_c}{p_a} = r_{sat}\cdot(d\ln p_{c,sat} - d\ln p_a) \tag{2.28}$$

Upon substituting into Equation 2.26 and collecting terms in $d\ln T$ and $d\ln p_a$ we find

$$(1+r_{sat})\frac{\delta Q}{T} = (c_{pa} + (c_{pc} + (\frac{L}{R_cT} - 1)\frac{L}{T})r_{sat})d\ln T - (1 + \frac{L}{R_aT}r_{sat})R_a d\ln p_a \tag{2.29}$$

To obtain the adiabat, we set $\delta Q = 0$, which leads to the following diferential equation defining $\ln T$ as a function of $\ln p_a$:

$$\frac{d\ln T}{d\ln p_a} = \frac{R_a}{c_{pa}}\frac{1 + \frac{L}{R_aT}r_{sat}}{1 + (\frac{c_{pc}}{c_{pa}} + (\frac{L}{R_cT} - 1)\frac{L}{c_{pa}T})r_{sat}} \tag{2.30}$$

Note that this expression reduces to the dry adiabat, as it should, when $r_{sat} \to 0$.

Exercise 2.6.2 What would the dry adiabat be for a noncondensing mixture of the two gases? Why doesn't the expression reduce to the dry adiabat for the mixture as $L \to 0$? (Hint: Think about the way Clausius-Clapeyron has been used in deriving the moist adiabat).

An examination of the typical properties of gases indicates that the c_{pc}/c_{pa} is typically of order unity, whereas $L/(R_cT)$ is typically very large, so long as the temperature is not exceedingly great. If one drops the smaller terms from the denominator of Eqn 2.30, one finds that the temperature gradient along the moist adiabiabat is weaker than that along the dry adiabat provided $\epsilon\frac{L}{c_{pa}T} > 1$, which is typically the case. It is expected that condensation should weaken the gradient, since it adds heat to the system and warms the air to greater temperatures than it would have had without condensation. This property can fail when the latent heat is weak or the noncondensible specific heat is very large, whereon the heat added by condensation has little effect on temperature. It is in this regime that there is also the possibility that condensation happens on *descent* rather than ascent; it is a very uncommon regime.

Everything on the right hand side of Eqn 2.30 is either a thermodynamic constant, or can be computed in terms of $\ln T$ and $\ln p_a$. Therefore, the equation defines a first order ordinary differential equation which can be integrated (usually numerically) to obtain T as a function of p_a. Usually one wants the temperature as a function of total pressure, rather than partial pressure of the noncondensible substance. This is no problem. Once $T(p_a)$ is known, the corresponding total pressure at the same point is obtained by computing $p = p_a + p_{c,sat}(T(p_a))$. To make a plot, or a table, one treats the problem parametrically: computing both T and p as functions of p_a. When the condensible substance is dilute, then $p_{c,sat} << p_a$, and $p \approx p_a$, so Eqn 2.30 gives the desired result directly.

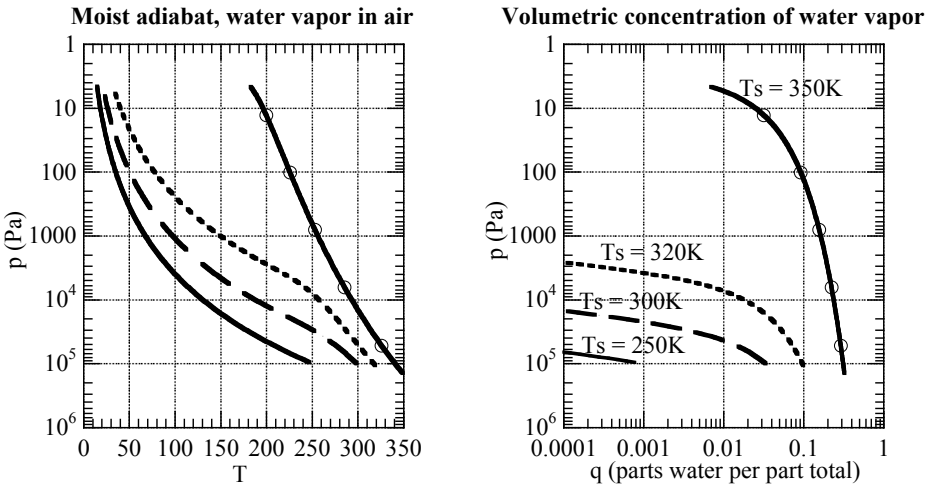

Figure 2.7: The moist adiabat for saturated water vapor mixed with Earth air having a partial pressure of 1 bar at the surface. Results are shown for various values of surface temperature, ranging from 250K to 350K. The left panel shows the temperature profile, while the right shows the profile of molar concentration of water vapor. A concentration value of .1 would mean that one molecule in 10 of the atmosphere is water vapor.

Figure 2.7 shows a family of solutions to Eqn 2.30, for the case of water vapor in Earth air. When the surface temperature is 250K, there is so little moisture in the atmosphere that the profile looks like the dry adiabat right to the ground. As temperature is increased, a region of weak gradients appears near the ground, representing the effect of latent heat on temperature. This layer gets progressively deeper as temperature increases and the moisture content of the atmosphere increases. When the surface temperature is 350K, so much moisture has entered the atmosphere that the surface pressure has actually increased to over 1300mb. Moreover, the moisture-dominated region extends all the way to 10 Pa (.1mb) , and even at 100 Pa (1mb) the atmosphere is 10% water by volume. Thus, for moderate surface temperatures, there is little water high up in the atmosphere. When the surface temperature approaches or exceeds 350K, though, the "cold trap" is lost, and a great deal of water is found aloft, where it is exposed to the destructive ultraviolet light of the sun and the possibility of thermal escape to space. In subsequent chapters, it will be seen that this phenomenon plays a major role in the life cycle of planets, and probably accounts for the present hot, dry state of Venus.

2.7 Rayleigh Fractionation

Section under development. I am considering providing an introduction to Rayleigh distillation processes in this section, since it is a natural extension of moist thermodynamics and is crucial to the understanding of many paleoclimate proxies. This would only cover the most basic equilibrium

fractionation models, and not the more intricate material associated with kinetic fractionation, retained condensate, or boundary layer kinetic processes.

Chapter 3

Elementary models of radiation balance

3.1 Energy balance and temperature

Our objective is to understand the factors governing the climate of a planet. Certainly, there is more to climate than temperature, but equally certainly temperature is a major part of what is meant by "climate," and greatly affects most of the other processes which come under that heading.

From the preceding chapter, we know that the temperature of a chunk of matter provides a measure of its energy content. Suppose that the planet receives energy at a certain rate. If uncompensated by loss, energy will accumulate and the temperature of some part of the planet will increase without bound. Now suppose that the planet loses energy at a rate that increases with temperature. Then, the temperature will increase until the rate of energy loss equals the rate of gain. It is this principle of energy balance that determines a planet's temperature. To quantify the functional dependence of the two rates, one must know the nature of both energy loss and energy gain.

The most familiar source of energy warming a planet is the absorption of light from the planet's star. This is the dominant mechanism for rocky planets like Venus, Earth and Mars. It is also possible for energy to be supplied to the surface by heat transport from the deep interior, fed by radioactive decay, tidal dissipation, or high temperature material left over from the formation of the planet. Heat flux from the interior is a major player in the climates of some gas giant planets, notably Jupiter and Saturn, because fluid motions can easily transport heat from the deep interior to the outer envelope of the planet. The sluggish motion of molten rock, and even more sluggish diffusion of heat through solid rock, prevent internal heating from being a significant part of the energy balance of rocky planets. Early in the history of a planet, when collisions are more common, the kinetic energy brought to the planet in the course of impacts with asteroids and planetesimals can be a significant part of the planet's energy budget.

There are many ways a planet can gain energy, but essentially only one way a planet can lose energy. Since a planet sits in the hard vaccuum of outer space,and its atmosphere is rather tightly bound by gravity, not much energy can be lost through heated matter streaming away from the planet. The only significant energy loss occurs through emission of electromagnetic radiation, most typically in the infrared spectrum. The quantification of this rate, and the way it is affected

by a planet's atmosphere, leads us to the subject of *blackbody radiation*.

3.2 Blackbody radiation

It is a matter of familiar experience that a sufficiently hot body emits light – hence terms like "red hot" or "white hot." Once it is recognized that light is just one form of electromagnetic radiation, it becomes a natural inference that a body with any temperature at all should emit some form of electromagnetic radiation, though not necessarily visible light. Thermodynamics provides the proper tool for addressing this question.

Imagine a gas consisting of two kinds of molecules, labeled A and B. Suppose that the two species interact strongly with each other, so that they come into thermodynamic equilibrium and their statistical properties are characterized by the same temperature T. Now suppose that the molecules A are ordinary matter, but that the "molecules" B are particles of electromagnetic radiation ("photons") or, equivalently, electromagnetic waves. If they interact strongly with the A molecules, whose energy distribution is characterized by their temperature T in accord with classical thermodynamics, the energy distribution of the electromagnetic radiation should also be characterized by the same temperature T. In particular, for any T there should be a unique distribution of energy amongst the various frequencies of the waves. This spectrum can be observed by examining the electromagnetic radiation leaving a body whose temperature is uniform. The radiation in question is known as *blackbody radiation* because of the assumption that radiation interacts strongly with the matter; any radiation impinging on the body will not travel far before it is absorbed, and in this sense the body is called "black" even though, like the Sun, it may be emitting light. Nineteenth century physicists found it natural to seek a theoretical explanation of the observed properties of blackbody radiation by applying well-established thermodynamical principles to electromagnetic radiation as described by Maxwell's classical equations. The attempt to solve this seemingly innocuous problem led to the discovery of quantum theory, and a revolution in the fundamental conception of reality.

Radiation is characterized by direction of propagation and frequency (and also polarization, which will not concern us). For electromagnetic radiation, the frequency ν and wavelength λ are related by the *dispersion relation* $\nu\lambda = c$, where c is a constant with the dimensions of velocity. Because visible light is a familiar form of electromagnetic radiation, c is usually called "the speed of light." The wavenumber, defined by $n = \lambda^{-1} = \nu/c$ is often used in preference to frequency or wavelength. Figure 3.1 gives the approximate regions of the electromagnetic spectrum corresponding to common names such as "Radio Waves" and so forth.

If a field of radiation consists of a mixture of different frequencies and directions, the mixture is characterized by a *spectrum*, which is a function describing the proportions of each type of radiation making up the blend. A spectrum is a *density* describing the amount of electromagnetic energy contained in a unit volume of the space (3D position, frequency, direction) needed to characterize the radiation. Suppose we wish to characterize the energy in the vicinity of a point $\vec{r}$ in three dimensional space, with frequency near ν and direction near that given by a unit vector $\hat{n}$. Then if $\Sigma(\vec{r}, \nu, \hat{n})$ is the energy spectrum at this point, the energy contained in a finite but small sized neighborhood of the point $(\vec{r}, \nu, \hat{n})$ is $\Sigma dV d\nu d\Omega$, where dV is a small volume of space, $d\nu$ is the width of the frequency band we wish to consider, and $d\Omega$ is a measure of the range of directions we wish to consider. A collection of directions in three-dimensional space is called a *solid angle*, and is measured in *steradians*. The measure in steriadians of a solid angle made by a collection of rays emanating from a point P is defined as the area of the patch of the unit sphere centered on P which the rays intersect. For example, a set of directions tracing out a hemisphere has measure 2π

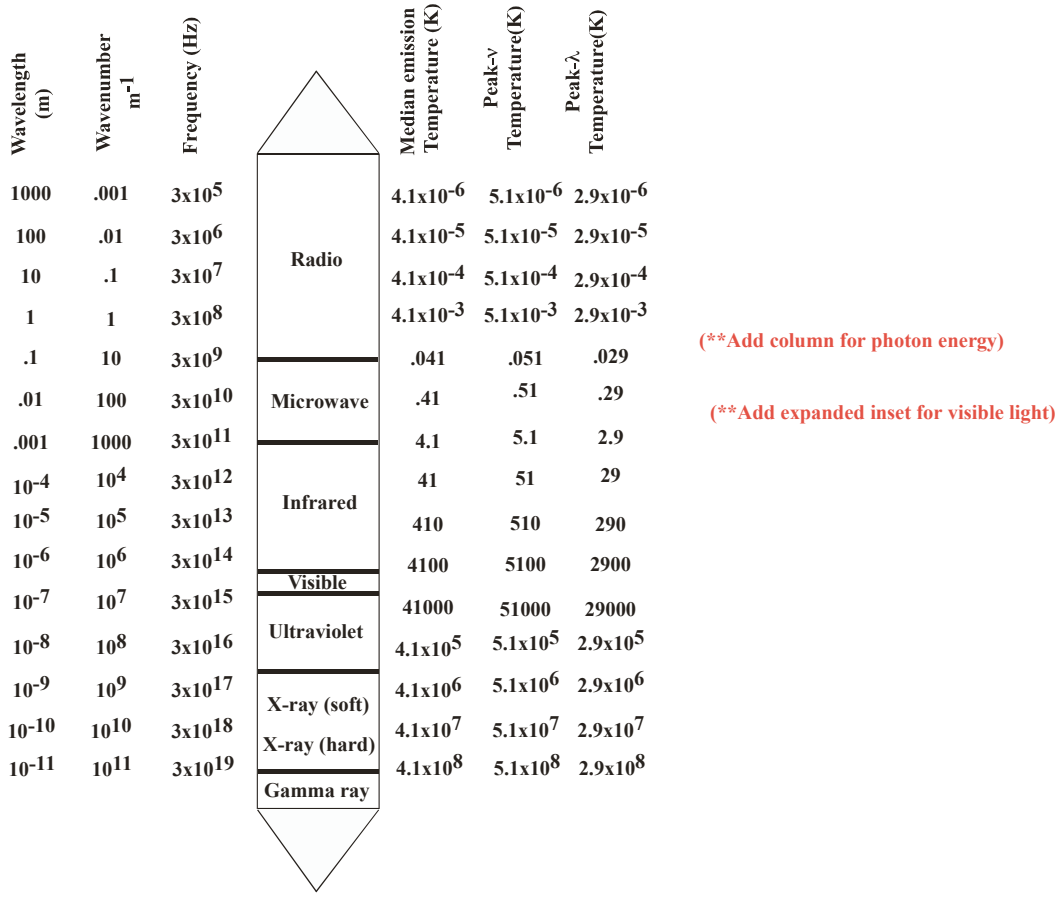

Figure 3.1: The electromagnetic spectrum. The Median Emission Temperature is the temperature of a blackbody for which half of the emitted power is below the given frequency (or equivalently, wavelength or wavenumber). The Peak-ν Temperature is the temperature of a blackbody for which the peak of the Planck density in frequency space is at the stated frequency. The Peak-λ Temperature is the temperature of a blackbody for which the peak of the Planck density in wavelength space is at the stated wavelength.

steradians, while a set of directions tracing out the entire sphere (i.e. all possible directions) has measure 4π. The set of directions contained within a cone with vertex angle θ measured relative to the altitude of the cone has measure $2\pi \sin\theta$ steradians.

Since electromagnetic waves in a vacuum move with constant speed c, the energy *flux* through a flat patch perpendicular to $\hat{n}$ with area dA is simply $c\Sigma dAd\nu d\Omega$, which defines the flux spectrum $c\Sigma$. In mks units, the flux spectrum has units of $(Watts/m^2)/(Hz \cdot steradian)$, where the Hertz (Hz) is the unit of frequency, equal to one cycle per second.

Exercise 3.2.1 The mks unit of energy is the *Joule*, J, which is 1 $Newton \cdot meter/sec$. A Watt (W) is $1J/sec$. A typical resting human in not-too-cold weather requires about $2000Calories/day$. (A *Calorie* is the amount of energy needed to increase the temperature of $1Kg$ of pure water by $1K$.) Convert this to a power consumption in W, using the fact that $1Calorie = 4184J$.

On the average, the flux of Solar energy reaching the Earth's surface is about $240W/m^2$. Assuming that food plants can convert Solar energy to usable food calories with an efficiency of 1%, what is the maximum population the Earth could support? (The radius of the Earth is about $6371km$)

The bold assumption introduced by Planck is that electromagnetic energy is exchanged only in amounts that are multiples of discrete *quanta*, whose size depends on the frequency of the radiation, in much the same sense that a penny is the quantum of US currency. Specifically, the quantum of energy for electromagnetic radiation having frequency ν is $\Delta E = h\nu$, where h is now known as *Planck's constant*. It is (so far as currently known) a constant of the universe, which determines the granuarity of reality. h is an exceedingly small number $(6.626 \cdot 10^{-34} Joule - seconds)$, so quantization of energy is not directly manifest as discreteness in the energy changes of everyday objects. A 1 watt blue nightlight (wavelength .48 microns, or frequency $6.24 \cdot 10^{14} Hz$ emits $2.4 \cdot 10^{18}$ photons each second, so it is no surprise that the light appears to be a continuous stream. If a bicycle were hooked to an electrical brake that dissipated energy by driving a blue light, emitting photons, the bike would indeed slow down in discontinuous increments, but the velocity increment, assuming the bike and rider to have a mass of $80kg$, would be only $10^{-10} m/s$; if one divides a $1m/s$ decrease of speed into 10^{10} equal parts, the deceleration will appear entirely continuous to the rider. Nonetheless, the aggregate effect of microscopic graininess of energy transitions exert a profound influence on the macroscopic properties of everyday objects. Blackbody radiation is a prime example of this.

Once the quantum assumption was introduced, Planck was able to compute the flux spectrum of blackbody radiation with temperature T using standard thermodynamic methods. The answer is

$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \tag{3.1}$$

where k is the Boltzmann thermodynamic constant defined in Chapter 2. $B(\nu, T)$ is known as the *Planck function*. Note that the Planck function is independent of the direction of the radiation; this is because blackbody radiation is *isotropic*, i.e. equally intense in all directions. In a typical application of the Planck function, we wish to know the flux of energy exiting the surface of a blackbody through a small nearly flat patch with area dA, over a frequency band of width $d\nu$. Since energy exits through this patch at all angles, we must integrate over all directions. However, energy exiting in a direction which makes an angle θ to the normal to the patch contributes a flux $(BdAd\nu d\Omega)\cos\theta$ through the patch, since the component of flux parallel to the patch carries no energy through it. Further, using the definition of a steradian, $d\Omega = 2\pi \sin\theta d\theta$ for the set of all rays making angles between θ and $\theta + d\theta$ with the normal to the patch. Integrating from $\theta = 0$ to

$\theta = \pi/2$, and using the fact that B is independent of direction, we then find that the flux through the patch is $\pi B dA d\nu$. This is also the amount of electromagnetic energy in a frequency band of width $d\nu$ that would pass each second through a hoop enclosing area dA (from one chosen side to the other), placed in the interior of an ideal blackbody; an equal amount passes through the hoop in the opposite sense.

The way the angular distribution of the radiation is described by the Planck function is rather confusing, and requires a certain amount of practice to get used to. The following exercise will test the readers' comprehension of this matter.

Exercise 3.2.2 A radiation detector flies on an airplane a distance H above an infinite flat plain with uniform temperature T. The detector is connected to a watt-meter which reports the total radiant power captured by the detector. The detector is sensitive to rays coming in at angles $\leq \delta\theta$ relative to the direction in which the detector is pointed. The area of the aperture of the detector is δA. The detector is sensitive to frequencies within a small range $\delta\nu$ centered on ν_0.

If the detector is pointed straight down, what is the power received by the detector? What is the size of the "footprint" on the plain to which the detector is sensitive? How much power is emitted by this footprint in the detector's frequency band? Why is this power different from the power received by the detector?

How do your answers change if the detector is pointed at an angle of 45^o relative to the vertical, rather than straight down?

The Planck function depends on frequency only through the dimensionless variable $u = h\nu/(kT)$. Recalling that each degree of freedom has energy $\frac{1}{2}kT$ in the average, we see that u is half the ratio of the quantum of energy at frequency ν to the typical energy in a degree of freedom of the matter with which the electromagnetic energy is in equilibrium. When u is large, the typical energy in a degree of freedom cannot create even a single photon of frequency ν, and such photons can be emitted only by those rare molecules with energy far above the mean. This is the essence of the way quantization affects the blackbody distribution – through inhibition of emission of high-frequency photons. On the other hand, when u is small, the typical energy in a degree of freedom can make many photons of frequency ν, and quantization imposes less of a constraint on emission. The characteristic frequency kT/h defines the crossover between the classical world and the quantum world. Much lower frequencies are little affected by quantization, whereas much higher frequencies are strongly affected. At 300K, the crossover frequency is $6240 GigaHz$, corresponding to a wavenumber of $20814 m^{-1}$, or a wavelength of $48 microns$; this is in the far infrared range.

In terms of u, the Planck function can be rewritten

$$B(\nu, T) = \frac{2k^3T^3}{h^2c^2}\frac{u^3}{e^u - 1} \tag{3.2}$$

In the classical limit, $u \ll 1$, and $u^3/(\exp(u)-1) \approx u^2$. Hence, $B \approx 2kT\nu^2/c^2$, which is independent of h. In a classical world, where $h = 0$, this form of the spectrum would be valid for all frequencies, and the emission would increase quadratically with frequency without bound; a body with any nonzero temperature would emit infrared at a greater rate than microwaves, visible light at a greater rate than infrared, ultraviolet at a greater rate than visible, X-rays at a greater rate than ultraviolet, and so forth. Bodies in equilibrium would cool to absolute zero almost instantaneously through emission of a burst of gamma rays, cosmic rays and even higher frequency radiation. This is clearly at odds with observations, not least the existence of the Universe. We are saved from this catastrophe by the fact that h is nonzero, which limits the range of validity of the classical

form of B. At frequencies high enough to make $u \gg 1$, then $u^3/(\exp(u)-1) \approx u^3 exp(-u)$ and the spectrum decays somewhat more slowly than exponentially as frequency is increased. The peak of B occurs at $u \approx 2.821$, implying that the frequency of maximum emission is $\nu \approx (2.821k/h)T \approx 58.78 \cdot 10^9 T$. The peak of the frequency spectrum increases linearly with temperature. This behavior, first deduced empirically long before it was explained by quantum theory, is known as the *Wien Displacement Law*.

Because the emission decays only quadratically on the low frequency side of the peak, but decays exponentially on the high frequency side, bodies emit appreciable energy at frequencies much lower than the peak emission, but very little at frequencies much higher. For example, at one tenth the peak frequency, a body emits at a rate of 4.8% of the maximum value. However, at ten times the peak frequency, the body emits at a rate of only $8.9 \cdot 10^{-9}$ of the peak emission. The microwave emission from a portion of the Earth's atmosphere with temperature $250K$ (having peak emission in the infrared) is readily detectable by satellites, whereas the emission of visible light is not.

Since B is a density, one cannot obtain the corresponding distribution in wavenumber or wavelength space by simply substituting for ν in terms of wavenumber or wavelength in the formula for B. One must also take into account the transformation of $d\nu$. For example, to get the flux density in wavenumber space (call it B_n) we use $B(\nu,T)d\nu = B(n \cdot c, T)d(n \cdot c) = cB(n \cdot c, T)dn$, whence $B_n(n,T) = cB(n \cdot c, T)$. Thus, transforming to wavenumber space changes the amplitude but not the shape of the flux spectrum. The Planck density in wavenumber space is shown for various temperatures in Figure 3.2. Because the transformation of the density from frequency to wavenumber space only changes the labeling of the vertical axis of the graph, one can obtain the wavenumber of maximum emission in terms of the frequency of maximum emission using $n_{max} = \nu_{max}/c$. An important property of the Planck function, readily verified by a simple calculation, is that $dB/dT > 0$ for all wavenumbers. This means that the Planck function for a large temperature is strictly above one for a lower temperature, or equivalently, that increasing temperature increases the emission at each individual wavenumber.

If one transforms to wavelength space, however,

$$B(\nu,T)d\nu = B(c/\lambda,T)d(c/\lambda) = -\frac{c}{\lambda^2}B(c/\lambda,T)d\lambda = \frac{2k^5T^5}{h^4c^3}\frac{u^5}{e^u-1}d\lambda = B_\lambda d\lambda \tag{3.3}$$

where $u = kT/(h\nu) = kT\lambda/(hc)$, as before. Transforming to wavelength space changes the shape of the flux spectrum. B_λ has its maximum at $u \approx 4.965$, which is nearly twice as large as the value for the wavenumber or frequency spectrum.

Since the location of the peak of the flux spectrum depends on the coordinate used to measure position within the electromagnetic spectrum, this quantity has no intrinsic physical meaning, apart from being a way to characterize the shape of the curve coming out of some particular kind of measuring apparatus. A more meaningful quantity can be derived from the *cumulative flux spectrum*, value at a given point in the spectrum is the same regardless of whether we use wavenumber, wavelength, $\log\lambda$ or any other coordinate to describe the position within the spectrum. The cumulative flux spectrum is defined as

$$F_{cum}(\nu,T) = \int_0^\nu \pi B(\nu',T)d\nu' = \int_\infty^\lambda \pi B_\lambda(\lambda',T)d\lambda' \tag{3.4}$$

Note that in defining the cumulative emission we have included the factor π which results from integrating over all angles of emission in a hemisphere. $F_{cum}(\nu,T)$ thus gives the power emitted per square meter for all frequencies less than ν, or equivalently, for all wavelengths greater than c/ν.

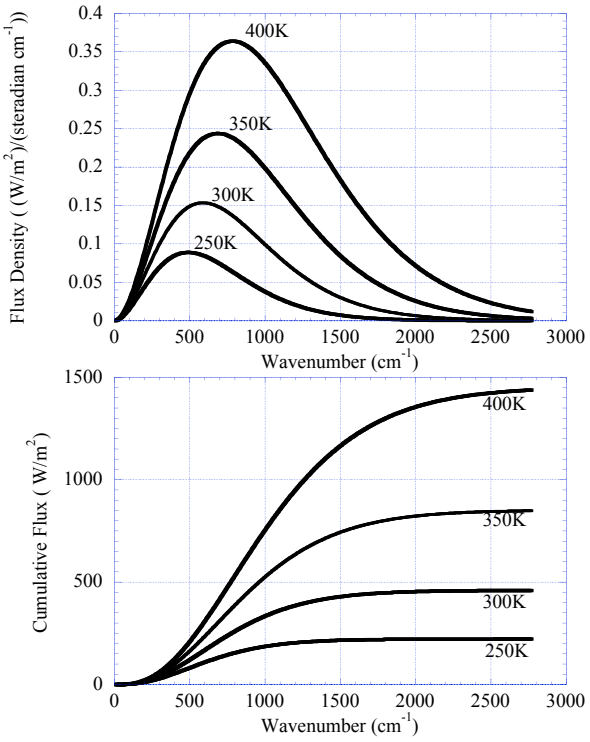

Figure 3.2: The spectrum of blackbody radiation for the various temperatures indicated on the curves. Upper Panel: The Planck density in wavenumber space. Lower Panel: The cumulative emission as a function of wavenumber. Note that the density has been tranformed such that the density times dn is the power per unit solid angle per unit area radiated in a wavenumber interval of width dn.

This function is shown for various temperatures in the lower panel of Fig. 3.2, where it is plotted as a function of wavenumber. The value of ν for which $F_{cum}(\nu, T)$reaches half the net emission $F_{cum}(\infty, T)$ provides a natural characterization of the spectrum. We will refer to this characteristic frequency as the *median emission frequency*. The median emission wavelength and wavenumber is defined analogously. Whether one uses frequency, wavelength or some other measure, the median emission is attained at $u \approx 3.503$. For any given coordinate used to describe the spectrum, the (angle-integrated) Planck density in that coordinate is the derivative of the cumulative emission with respect to the coordinate. Hence the peak in the Planck density just gives the point at which the cumulative emission function has its maximum slope. This depends on the coordinate used, unlike the point of median emission. Figure 3.1 shows the the portion of the spectrum in which blackbodies with various temperatures dominantly radiate. For example, a body with a temperature of around $4K$ radiates in the microwave region; this is the famous "Cosmic Microwave Background Radiation" left over from the Big Bang [1]. A body with a temperature of $300K$ radiates in the infrared, one with a temperature of a few thousand degrees radiates in the visible, and one with a temperature of some tens of thousands of degrees would radiate in the ultraviolet.

Next, we evaluate $F_{cum}(\infty, T)$, to obtain the total power F exiting from each unit area of the surface of a blackbody:

$$F = \int_0^\infty \pi B(\nu, T) d\nu = \int_0^\infty \pi B(u, T) \frac{kT}{h} du = [\frac{2\pi k^4}{h^3 c^2} \int_0^\infty \frac{u^3}{e^u - 1} du] T^4 = \sigma T^4 \tag{3.5}$$

where [2] $\sigma = 2\pi^5 k^4/(15c^2h^3) \approx 5.67 \cdot 10^{-8} W m^{-2} K^{-4}$. The constant σ is known as the *Stefan-Boltzmann constant*, and the law $F = \sigma T^4$ is the *Stefan-Boltzmann law*. This law was originally deduce from observations, and Boltzmann was able to derive the fourth-power scaling in temperature using classical thermodynamic reasoning. However, classical physics yields an infinite value for the constant σ. The formula for σ clearly reveals the importance of quantum effects in determining this constant, since σ diverges like $1/h^3$ if we try to pass to the classical limit by making h approach zero.

An important property of an ideal blackbody is that the radiation leaving its surface depends only on the temperature of the body. If a blackbody is interposed between an observer and some other object, all properties of the object will be hidden from the observer, who will see only blackbody radiation corresponding to the temperature of the blackbody. This remark allows us to make use of blackbody theory to determine the emission from objects whose temperature varies greatly from place to place, even though blackbody theory applies, strictly speaking, only to extensive bodies with uniform temperature. For example, the temperature of the core of the Earth is about $6000K$, but we need not know this in order to determine the radiation emitted from the Earth's surface; the outermost few millimeters of rock, ice or water at the Earth's surface contain enough matter to act like a blackbody to a very good approximation. Hence, the radiation emitted from the surface depends only on the temperature of this outer skin of the planet. Similarly, the temperature of the core of the Sun is about $16,000,000K$ and even at a distance from the center equal to 90% of the visible radius, the temperature is above $600,000K$. However, the Sun is encased in a layer a few hundred kilometers thick which is sufficiently dense to act like a blackbody, and which has a temperature of about $5780K$. This layer is known as the *photosphere*, because it is the source of most light exiting the Sun. Layers farther out from the center of the Sun can be considerably hotter than the photosphere, but they have a minimal effect on solar radiation

[1] What is remarkable about this observed cosmic radiation is not so much that it is in the microwave region, but that it has a blackbody spectrum, which says much about the interaction of radiation with matter in the early moments of the Universe.

[2] The definite integral $\int_0^\infty (u^3/(e^u - 1))du$ was determined by Euler, as a special case of his study of the behavior of the Riemann zeta function at even integers. It is equal to $6\zeta(4) = \pi^4/15$

because they are so tenuous. In Chapter 4 we will develop more precise methods for dealing with tenuous objects, such as atmospheres, which peter out gradually without having a sharply defined boundary.

An ideal blackbody would be opaque at all wavelengths, but it is a common situation that a material acts as a blackbody only in a limited range of wavelengths. Consider the case of window glass: It is transparent to visible light, but if you could see it in the infrared it would look as opaque as stone. Because it interacts strongly with infrared light, window glass emits blackbody radiation in the infrared range. At temperatures below a few hundred K, there is little blackbody emission at wavelengths shorter than the infrared, so at such temperatures the net power per unit area emitted by a pane of glass with temperature T is very nearly σT^4, even though it doesn't act like a blackbody in the visible range. Liquid water, and water ice, behave similarly. Crystalline table salt, and carbon dioxide ice, are nearly transparent in the infrared as well as in the visible, and in consequence emit radiation at a much lower rate than expected from the blackbody formula. (They would make fine windows for creatures having infrared vision). There is, in fact, a deep and important relation between absorption and emission of radiation, which will be discussed in Section 3.5.

3.3 Radiation balance of planets

As a first step in our study of the temperature of planets, let's consider the following idealized case:

- The only source of energy heating the planet is absorption of light from the planet's Sun.
- The *albedo*, or proportion of sunlight reflected, is spatially uniform.
- The planet is spherical, and has a distinct solid or liquid surface which radiates like a perfect blackbody.
- The planet's temperature is uniform over its entire surface.
- The planet's atmosphere is perfectly transparent to the electromagnetic energy emitted by the surface.

The uniform-temperature assumption presumes that the planet has an atmosphere or ocean which is so well stirred that it is able to rapidly mix heat from one place to another, smoothing out the effects of geographical fluctuations in the energy balance. The Earth conforms fairly well to this approximation. The equatorial annual mean temperature is only 4% above the global mean temperature of $286K$, while the North polar temperature is only 10% below the mean. The most extreme deviation occurs on the high Antarctic plateau, where the annual mean South polar temperature is 21% below the global mean. The surface temperature of Venus is even more uniform than that of Earth. That of Mars, which in our era, has a thin atmosphere and no ocean, is less uniform. Airless, rocky bodies like the Moon and Mercury do not conform at all well to the uniform temperature approximation.

Light leaving the upper layers of the Sun and similar stars takes the form of blackbody radiation. It is isotropic, and its flux and flux spectrum conform to the blackbody law corresponding to the temperature of the photosphere, from which the light escapes. Once the light leaves the surface of the star, however, it expands through space and does not interact significantly with

matter except where it is intercepted by a planet. Therefore, it is no longer blackbody radiation, though it retains the blackbody spectrum. In the typical case of interest, the planet orbits its star at a distance that is much greater than the radius of the star, and itself has a radius that is considerably smaller than the star and is hence yet smaller than the orbital distance. In this circumstance, all the rays of light which intersect the planet are very nearly parallel to the line joining the center of the planet to the center of its star; the sunlight comes in as a nearly *parallel beam*, rather than being isotropic, as would be the case for true blackbody radiation. The parallel-beam approximation is equivalent to saying that, as seen from the planet, the Sun occupies only a small portion of the sky, and as seen from the Sun the planet also occupies only a small portion of the sky. Even for Mercury, with a mean orbital distance of $58,000,000km$, the Sun (whose radius is $695,000km$) occupies an angular width in the sky of only about $2 \cdot 695,000/58,000,000$ radians, or 1.4^o.

The solar flux impinging on the planet is also reduced, as compared to the solar flux leaving the photosphere of the star. The total energy per unit frequency leaving the star is $4\pi r_\odot^2(\pi B(\nu, T_\odot))$, where $r_\odot$ is the radius of the star and $T_\odot$ is the temperature of its photosphere. At a distance r from the star, the energy has spread uniformly over a sphere whose surface area is $4\pi r^2$; hence at this distance, the energy flux per unit frequency is $\pi B r_\odot^2/r^2$, and the total flux is $\sigma T_\odot^4 r_\odot^2/r^2$. The latter is the flux seen by a planet at orbital distance r, in the form of a beam of parallel rays. It is known as the solar "constant" (even though it in fact depends on distance from the star), and will be denoted by $L_\odot$, or simply L where there is no risk of confusion with latent heat.

We are now equipped to compute the energy balance of the planet, subject to the preceding simplifying assumptions. Let a be the planet's radius. Since the cross-section area of the planet is πa^2 and the solar radiation arrives in the form of a nearly parallel beam with flux $L_\odot$, the energy per unit time impinging on the planet's surface is $\pi a^2 L_\odot$; the rate of energy absorption is $(1-\alpha)\pi a^2 L_\odot$, where α is the albedo. The planet loses energy by radiating from its entire surface, which has area $4\pi a^2$. Hence the rate of energy loss is $4\pi a^2 \sigma T^4$, where T is the temperature of the planet's surface. In equilibrium the rate of energy loss and gain must be equal. After cancelling a few terms, this yields

$$\sigma T^4 = \frac{1}{4}(1-\alpha)L_\odot \tag{3.6}$$

Note that this is independent of the radius of the planet. The factor $\frac{1}{4}$ comes from the ratio of the planet's cross-sectional area to its surface area, and reflects the fact that the planet intercepts only a disk of the incident solar beam, but radiates over its entire spherical surface. This equation can be readily solved for T. If we substitute for $L_\odot$ in terms of the photospheric temperature, the result is

$$T = \frac{1}{\sqrt{2}}(1-\alpha)^{1/4}\sqrt{\frac{r_\odot}{r}}T_\odot \tag{3.7}$$

Formula 3.7 shows that the blackbody temperature of a planet is much less than that of the photosphere, so long as the orbital distance is large compared to the stellar radius. From the displacement law, it follows that the planet loses energy through emission at a distinctly lower wavenumber than that at which it receives energy from its star. This situation is illustrated in Figure 3.3. For example, the energy received from our Sun has a median wavenumber of about 15000 cm^{-1}, equivalent to a wavelength of about .7 *microns*. An isothermal planet at Mercury's orbit would radiate to space with a median emission wavenumber of 1100 cm^{-1}, corresponding to a wavelength of 9 *microns*. An isothermal planet at the orbit of Mars would radiate with a median wavenumber of 550 cm^{-1}, corresponding to a wavelength of 18 *microns*.

Exercise 3.3.1 A planet with zero albedo is in orbit around an exotic hot star having a photo-

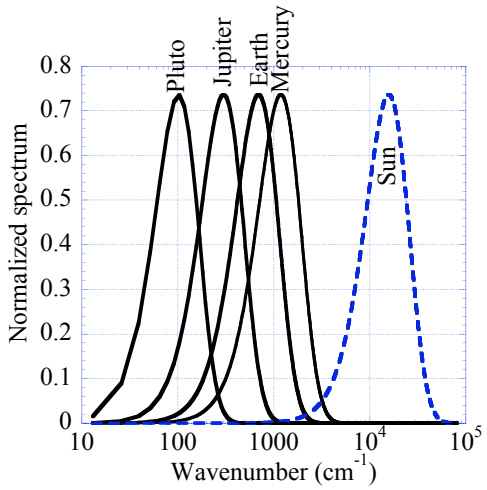

Figure 3.3: The Planck density of radiation emitted by the Sun and selected planets in radiative equilibrium with absorbed solar radiation (based on the observed shortwave albedo of the planets). The Planck densities are transformed to a logarithmic spectral coordinate, and all are normalized to unit total emission.

spheric temperature of $100,000K$. The ratio of the planet's orbit to the radius of the star is the same as for Earth (about 215). What is the median emission wavnumber of the star? In what part of the electromagnetic spectrum does this lie? What is the temperature of the planet? In what part of the electromagnetic spectrum does the planet radiate? Do the same if the planet is instead in orbit around a brown dwarf star with a photospheric temperature of $600K$.

The separation between absorption and emission wavenumber will prove very important when we bring a radiatively active atmosphere into the picture, since it allows the atmosphere to have a different effect on incoming vs. outgoing radiation. Since the outgoing radiation has longer wavelength than the incoming radiation, the flux of emitted outgoing radiation is often referred to as *outgoing longwave radiation*, and denoted by OLR. For a non-isothermal planet, the OLR is a function of position (e.g. latitude and longitude on an imaginary sphere tightly enclosing the planet and its atmosphere). We will also use the term to refer to the outgoing flux averaged over the surface of the sphere, even when the planet is not isothermal. As for the other major term in the planet's energy budget, we will refer to the electromagnetic energy received from the planet's star as the *shortwave* or *solar* energy. Our own Sun has its primary output in the visible part of the spectrum, but it also emits significant amounts of energy in the ultraviolet and near-infrared, both of which are shorter in wavelength than the OLR by which planets lose energy to space.

Formula 3.7 is plotted in Figure 3.4 for a hypothetical isothermal planet with zero albedo. Because of the square-root dependence on orbital distance, the temperature varies only weakly with distance, except very near the star. Neglecting albedo and atmospheric effects, Earth would have a mean surface temperature of about $280K$. Venus would be only 50K warmer than the Earth and Mars only $53K$ colder. At the distant orbit of Jupiter, the blackbody equilibrium temperature falls to $122K$, but even at the vastly more distant orbit of Neptune the temperature is still as

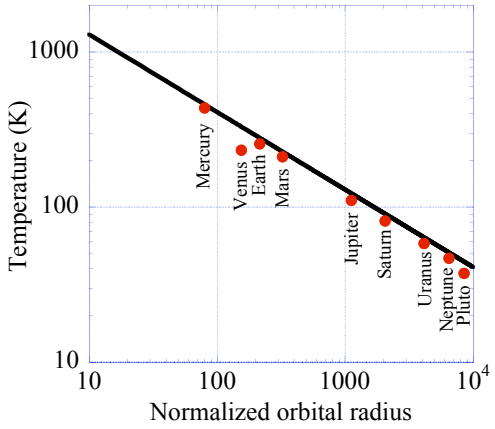

Figure 3.4: The equilibrium blackbody temperature of an isothermal spherical zero-albedo planet, as a function of distance from a Sun having a photospheric temperarature of 5800K. The orbital distance is normalized by the radius of the Sun. Dots show the equilibrium blackbody temperature of the Solar System planets, based on their actual observed albedos.

high as $50K$. The emission from all of these planets lies in the infrared range, though the colder planets radiate in the deeper (lower wavenumber) infrared. An exception to the strong separation between stellar and planetary temperature is provided by the "roasters" – a recently discovered class of extrasolar giant planets with $\frac{r}{r_\odot}$ as low as 5. Such planets can have equilibrium blackbody temperatures as much as a third that of the photosphere of the parent star. For these planets, the distinction between the behavior of incoming and outgoing radiation is less sharp.

It is instructive to compare the ideal blackbody temperature with observed surface temperature for the three Solar System bodies which have both a distinct surface and a thick enough atmosphere to enforce a roughly uniform surface temperature: Venus, Earth and Saturn's moon Titan. For this comparison, we calculate the blackbody temperature using the observed planetary albedos, instead of assuming a hypothetical zero albedo planet as in Fig. 3.4. Venus is covered by thick, highly reflective clouds, which raise its albedo to .75. The corresponding isothermal blackbody temperature is only $232K$ (as compared to $330K$ in the zero albedo case). This is far less than the observed surface temperature of $740K$. Clearly, the atmosphere of Venus exerts a profound warming effect on the surface. The warming arises from the influence of the atmosphere on the infrared emission of the planet, which we have not yet taken into account. Earth's albedo is on the order of .3, leading to a blackbody temperature of $255K$. The observed mean surface temperature is about 285K. Earth's atmosphere has a considerably weaker warming effect than that of Venus, but it is nonetheless a very important warming, since it brings the planet from subfreezing temperatures where the oceans would almost certainly become ice-covered, to temperatures where liquid water can exist over most of the planet. The albedo of Titan is .21, and using the solar constant at Saturn's orbit we find a black body temperature of $85K$. The observed surface temperature is about $95K$, whence we conclude that the infrared effects of Titan's atmosphere moderately warm the surface.

The way energy balance determines surface temperature is illustrated graphically in Figure 3.5. One first determines the way in which the mean infrared emission per unit area depends on the mean surface temperature T_s; for the isothermal blackbody calculation, this curve is simply σT_s^4. The equilibrium temperature is determined by the point at which the OLR curve intersects the curve giving the absorbed solar radiation (a horizontal line in the present calculation). In some sense, the whole subject of climate comes down to an ever-more sophisticated heirarchy of calculations of the curve $OLR(T_s)$; our attention will soon turn to the task of determining how the OLR curve is affected by an atmosphere. With increasing sophistication, we will also allow the solar absorption to vary with T_s, owing to changing clouds, ice cover, vegetation cover, and other characteristics.

We will now consider an idealized thought experiment which illustrates the essence of the way an atmosphere affects OLR. Suppose that the atmosphere has a temperature profile $T(p)$ which decreases with altitude,according to the dry or moist adiabat. Let p_s be the surface pressure, and suppose that the ground is strongly thermally coupled to the atmosphere by turbulent heat exchanges, so that the ground temperature cannot deviate much from that of the immediately overlying air. Thus, $T_s = T(p_s)$. If the atmosphere were transparent to infrared, as is very nearly the case for nitrogen or oxygen, the OLR would be σT_s^4. Now, let's stir an additional gas into the portion of the atmosphere between the ground and a pressure $p_{rad} < p_s$, and suppose that the gas is transparent to solar radiation, but interacts so strongly with infrared that it turns each portion of the atmosphere it is mixed with into a perfect blackbody. A gas which is fairly transparent to the incoming shortwave stellar radiation but which interacts strongly with the outgoing (generally infrared) emitted radiation is called a *greenhouse gas*. Carbon dioxide, water vapor and methane are some examples of greenhouse gases, and the molecular properties that make a substance a good greenhouse gas will be discussed in Chapter 4. If one imagines slicing the atmosphere into a number layers so thin that they are essentially isothermal, then each layer with pressure greater than or equal to p_{rad} radiates like an ideal blackbody at its own temperature, but it is only the topmost of these layers that determines the radiation loss to space, since radiation from all the others is absorbed before it reaches the topmost layer. Since the topmost layer has temperature $T(p_{rad})$ and higher altitude layers are assumed transparent to infrared, the OLR is $\sigma T(p_{rad})^4$, which is less than σT_s^4 to the extent that $p_{rad} < p_s$. As shown in Figure 3.5, a greenhouse gas acts like an insulating blanket, reducing the rate of energy loss to space at any given surface temperature. The equilibrium surface temperature of a planet with a greenhouse gas in its atmosphere must be greater than that of a planet without a greenhouse gas, in order to radiate away energy at a sufficient rate to balance the absorbed solar radiation.

In the real universe, greenhouse gases are continuously distributed in the atmosphere, rather than being confined to a single layer. Further, they increase the interaction of the atmosphere with infrared, but rarely so much so as to turn some upper portion of the atmosphere into an ideal blackbody. In reality,the infrared escaping to space is a blend of radiation emitted from a range of atmospheric levels, with some admixture of radiation from the planet's surface as well. The concept of an *effective radiating level* nonetheless has merit for real greenhouse gases. It does not represent a distinct physical layer of the atmosphere, but rather characterizes the mean depth from which infrared photons escape to space. As more greenhouse gas is added to an atmosphere, more of the lower parts of the atmosphere become opaque to infrared, preventing the escape of infrared radiation from those regions. This increases the altitude of the effective radiating level (i.e. decreases p_{rad}). From an observation of the actual OLR emitted by a planet, one can determine an equivalent blackbody radiating temperature T_{rad} from the expression $\sigma T_{rad}^4 = OLR$. This temperature is the infrared equivalent of the Sun's photospheric temperature; it is a kind of mean temperature of the regions from which infrared photons escape, and p_{rad} represents a mean pressure of these layers. For planets for which absorbed solar radiation is the only significant energy source,

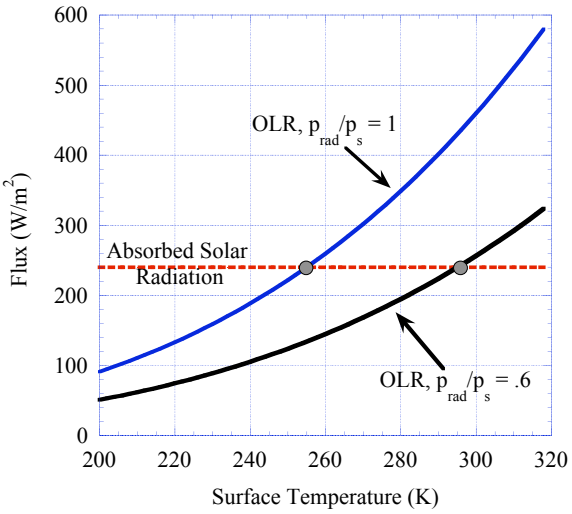

Figure 3.5: Determination of a planet's temperature by balancing absorbed solar energy against emitted longwave radiation. The horizontal line gives the absorbed solar energy per unit surface area, based on an albedo of .3 and a Solar constant of $1370 W/m^2$. The OLR is given as a function of surface temperature. The upper curve assumes the atmosphere has no greenhouse effect ($p_{rad} = p_s$), while the lower OLR curve assumes $p_{rad}/p_s = .6$, a value appropriate to the present Earth.

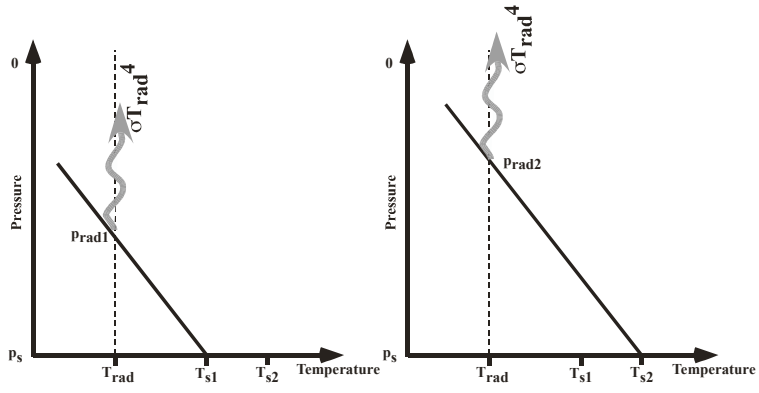

Figure 3.6: Sketch illustrating how the greenhouse effect increases the surface temperature. In equilibrium, the outgoing radiation must remain equal to the absorbed solar radiation, so T_{rad} stays constant. However, as more greenhouse gas is added to the atmosphere, p_{rad} is reduced, so one must extrapolate temperature further along the adiabat to reach the surface.

T_{rad} is equal to the ideal blackbody temperature given by Eq. 3.7. The arduous task of relating the effective radiating level to specified concentrations of real greenhouse gases will be taken up in Chapter 4.

Figure 3.7 illustrates the reduction of infrared emission caused by the Earth's atmosphere. At every latitude, the observed OLR is much less than it would be if the planet radiated to space at its observed surface temperature. At the Equator the observed OLR is $238 W/m^2$, corresponding to a radiating temperature of $255 W/m^2$. This is much less than the observed surface temperature of $298K$, which would radiate at a rate of $446 W/m^2$ if the atmosphere didn't intervene. It is interesting that the gap between observed OLR and the computed surface emission is less in the cold polar regions, and especially small at the Winter pole. This happens partly because, at low temperatures, there is simply less infrared emission for the atmosphere to trap. However, differences in the water content of the atmosphere, and differences in the temperature profile, can also play a role. These effects will be explored in Chapter 4.

Gases are not the only atmospheric constituents which affect OLR. Clouds consist of particles of condensed substance small enough to stay suspended for a long time. They can profoundly influence OLR. Gram for gram, condensed water interacts much more strongly with infrared than does water vapor. In fact, a mere 20 grams of water in the form of liquid droplets of a typical size is sufficient to turn a column of air $500m$ thick by one meter square into a very nearly ideal blackbody. To a much greater extent than for greenhouse gases, a water cloud layer in an otherwise infrared-transparent atmosphere really can be thought of as a discrete radiating layer. The prevalance of clouds in the high, cold regions of the tropical atmosphere accounts for the dip in OLR near the equator, seen in Figure 3.7. Clouds are unlike greenhouse gases, though, since they also strongly reflect the incoming solar radiation. It's the tendency of these two large effects to partly cancel that makes the problem of the influence of clouds on climate so challenging. Not all condensed substances absorb infrared as well as water does. Liquid methane (imporant on Titan) and CO2 ice (important on present and early Mars) are comparatively poor infrared absorbers.

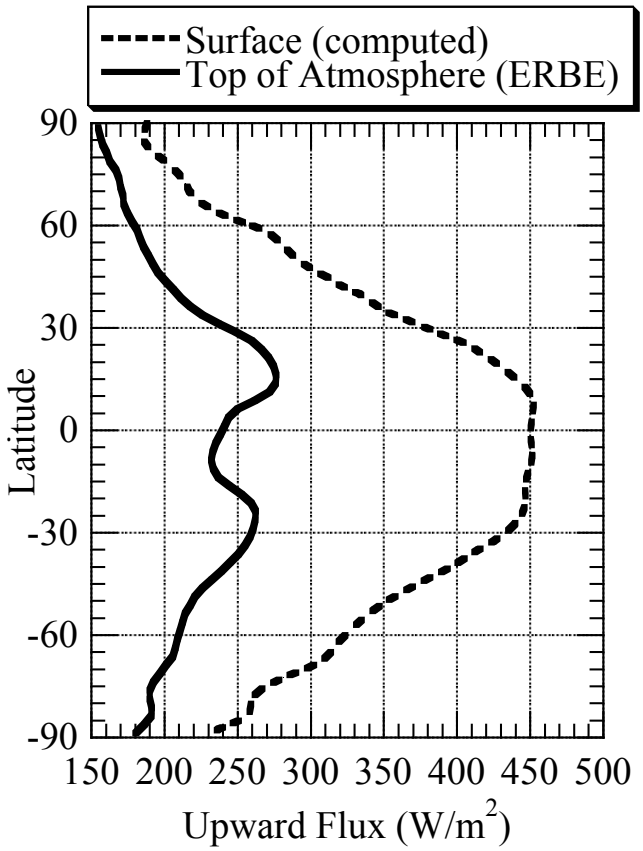

Figure 3.7: The Earth's observed zonal-mean OLR for January, 1986. The observations were taken by satellite instruments during the Earth Radiation Budget Experiment (ERBE), and are averaged along latitude circles. The figure also shows the radiation that would be emitted to space by the surface (σT_s^4) if the atmosphere were transparent to infrared radiation.

They affect OLR in a fundamentally different way, through reflection instead of absorption and emission. This will be discussed in Chapter 5.

In a nutshell, then, here is how the greenhouse effect works: From the requirement of energy balance, the absorbed solar radiation determines the effective blackbody radiating temperature T_{rad}. This is not the surface temperature; it is instead the temperature encountered at some pressure level in the atmosphere p_{rad}, which characterizes the infrared opacity of the atmosphere, specifically the typical altitude from which infrared photons escape to space. The pressure p_{rad} is determined by the greenhouse gas concentration of the atmosphere. The surface temperature is determined by starting at the fixed temperature T_{rad} and extrapolating from p_{rad} to the surface pressure p_s using the atmosphere's lapse rate, which is approximately governed by the appropriate adiabat. Since temperature decreases with altitude over much of the depth of a typical atmosphere, the surface temperature so obtained is typically greater than T_{rad}, as illustrated in Figure 3.6. Increasing the concentration of a greenhouse gas decreases p_{rad}, and therefore increases the surface temperature because temperature is extrapolated from T_{rad} over a greater pressure range. It is very important to recognize that greenhouse warming relies on the decrease of atmospheric temperature with height, which is generally due to the adiabatic profile established by convection. The greenhouse effect works by allowing a planet to radiate at a temperature colder than the surface, but for this to be possible, there must be some cold air aloft for the greenhouse gas to work with.

For an atmosphere whose temperature profile is given by the dry adiabat, the surface temperature is

$$T_s = (p_s/p_{rad})^{R/c_p} T_{rad}. \tag{3.8}$$

With this formula, the Earth's present surface temperature can be explained by taking $p_{rad}/p_s = .67$, whence $p_{rad} \approx 670mb$. Earth's actual radiating pressure is somewhat lower than this estimate, because the atmosperic temperature decays less strongly with height than the dry adiabat. The high surface temperature of Venus can be accounted for by taking $p_{rad}/p_s = .0095$, assuming that the temperature profile is given by the noncondensing adiabat for a pure CO_2 atmosphere. Given Venus' $93bar$ surface pressure, the radiating level is $880mb$ which, interestingly, is only slightly less than Earth's surface pressure. Earth radiates to space from regions quite close to its surface, whereas Venus radiates only from a thin shell near the top of the atmosphere. Note that from the observed Venusian temperature profile in Fig. 2.2, the radiating temperature ($253K$) is encountered at $p = 250mb$ rather than the higher pressure we estimated. As for the Earth, our estimate of the precise value p_{rad} for Venus is off because the ideal-gas noncondensing adiabat is not a precise model of the actual temperature profile. In the case of Venus, the problem most likely comes from the ideal-gas assumption, rather than condensation.

The concept of radiating level and radiating temperature also enables us to make sense of the way energy balance constrains the climates of gas giants like Jupiter and Saturn, which have no distinct surface. The essence of the calculation we have already done for rocky planets is to use the top of atmosphere energy budget to determine the parameters of the adiabat, and then extrapolate temperature to the surface along the adiabat. For a non-condensing adiabat, the atmospheric profile compatible with energy balance is $T(p) = T_{rad}(p/p_{rad})^{R/c_p}$. This remains the appropriate temperature profile for a (noncondensing) convecting outer layer of a gas giant, and the only difference with the previous case is that, for a gas giant, there is no surface to act as a natural lower boundary for the adiabatic region. At some depth, convection will give out and the adiabat must be matched to some other temperature model in order to determine the base of the convecting region, and to determine the temperature of deeper regions. There is no longer any distinct surface to be warmed by the greenhouse effect, but the greenhouse gas concentration of the atmosphere nonetheless affects $T(p)$ through p_{rad}. For example, adding some additional

	Observed OLR (W/m^2)	Absorbed Solar Flux (W/m^2)	T_{rad} (actual)	T_{rad} (Solar only)
Jupiter	14.3	12.7	126K	110K
Saturn	4.6	3.8	95K	81K
Uranus	.52	.93	55K	58K
Neptune	.61	.38	57K	47K

Table 3.1: The energy balance of the gas giant planets, with inferred radiating temperature. The solar-only value of T_{rad} is the radiating temperature that would balance the observed absorbed solar energy, in the absence of any internal heat source.

greenhouse gas to the convecting outer region of Jupiter's atmosphere would decrease p_{rad}, and therefore increase the temperature encountered at, say, the 1 bar pressure level.

The energy balance suffices to uniquely determine the temperature profile because the non-condensing adiabat is a one-parameter family of temperature profiles. The saturated adiabat for a mixture of condensing and noncondensing gases is also a one parameter family, defined by Eq. 2.30, and can therefore be treated similarly. If the appropriate adiabat for the planet had more than one free parameter, additional information beyond the energy budget would be needed to close the problem. On the other hand, a single component condensing atmosphere such as described by Eq. 2.24 yields a temperature profile with no free parameters that can be adjusted so as to satisfy the energy budget. The consequences of this quandary will be taken up as part of our discussion of the runaway greenhouse phenonenon, in Chapter 4.

Using infrared telescopes on Earth and in space, one can directly measure the OLR of the planets in our Solar System. In the case of the gas giants, the radiated energy is substantially in excess of the absorbed solar radiation. Table 3.1 compares the observed OLR to the absorbed solar flux for the gas giants. With the exception of Uranus, the gas giants appear to have a substantial internal energy source, which raises the radiating temperature to values considerably in excess of it would be if the planet were heated by solar absorption alone. Uranus is anomalous, in that it actually appears to be emitting less energy than it receives from the sun. Uncertainties in the observed OLR for Uranus would actually allow the emission to be in balance with solar absorption, but would still appear to preclude any significant internal energy source. This may indicate a profound difference in the internal dynamics of Uranus. On the other hand, the unusually large tilt of Uranus' rotation axis means that Uranus has an unusually strong seasonal variation of solar heating, and it may be that the hemisphere that has been observed so far has not yet had time to come into equilibrium, which would throw off the energy balance estimate.

Because it is the home planet, Earth's radiation budget has been very closely monitored by satellites. Very precise measurements show that the top of atmosphere radiation budget is currently out of balance, the Earth receiving about $1W/m^2$ more from Solar absorption than it emits to space as infrared. This is opposite from the imbalance that would be caused by an internal heating. It is a direct consequence of the rapid rise of CO_2 and other greenhouse gases, caused by the bustling activities of Earth's human inhabitants. The rapid greenhouse gas increase has cut down the OLR, but because of the time required to warm up the oceans and melt ice, the Earth's temperature has not yet risen enough to restore the energy balance.

Exercise 3.3.2 A typical well-fed human in a resting state consumes energy in the form of food at a rate of $100W$, essentially all of which is put back into the surroundings in the form of heat. An astronaut is in a spherical escape pod of radius r, far beyond the orbit of Pluto, so that it receives essentially no energy from sunlight. The air in the escape pod is isothermal. The skin of the escape pod is a good conductor of heat, so that the surface temperature of the sphere is identical to the

interior temperature. The surface radiates like an ideal blackbody.

Find an expression for the temperature in terms of r, and evaluate it for a few reasonable values. Is it better to have a bigger pod or a smaller pod? In designing such an escape pod, should you include an additional source of heat if you want to keep the astronaut comfortable?

How would your answer change if the pod were cylindrical instead of spherical? If the pod were cubical?

Bodies such as Mercury or the Moon represent the opposite extreme from the uniform-temperature limit. Having no atmosphere or ocean to transport heat, and a rocky surface through which heat is conducted exceedingly slowly, each bit of the planet is, to a good approximation, thermally isolated from the rest. Moreover, the rocky surface takes very little time to reach its equilibrium temperature, so the surface temperature at each point is very nearly in equilibrium with the instantaneous absorbed solar radiation, with very little day-night or seasonal averaging. In this case, averaging the energy budget over the planet's surface gives a poor estimate of the temperature, and it would be more accurate to compute the instantaneous equilibrium temperature for each patch of the planet's surface in isolation. For example, consider a point on the planet where the Sun is directly overhead at some particular instant of time. At that time, the rays of sunlight come in perpendicularly to a small patch of the ground, and the absorbed solar radiation per unit area is simply $(1-\alpha)L_\odot$; the energy balance determing the ground temperature is then $\sigma T^4 = (1-\alpha)L_\odot$, without the factor of $\frac{1}{4}$ we had when the energy budget was averaged over the entire surface of an isothermal planet. For Mercury, this yields a temperature of $622K$, based on the mean orbital distance and an albedo of .1. This is similar to the observed maximum temperature on Mercury, which is about $700K$ (somewhat larger than the theoretical calculation because Mercury's highly elliptical orbit brings it considerably closer to the Sun than the mean orbital position). The Moon, which is essentially in the same orbit as Earth and shares its Solar constant, has a predicted maximum temperature of $384K$, which is very close to the observed maximum. In contrast, the maximum surface temperature on Earth stays well short of $384K$, even at the hottest time of day in the hottest places. The atmosphere of Mars in the present epoch is thin enough that this planet behaves more like the no-atmosphere limit than the uniform-temperature limit. Based on a mean albedo of .25,the local maximum temperature should be $297K$, which is quite close to the observed maximum temperature.

More generally speaking, when doing energy balance calculations the temperature we have in mind is the temperature averaged over an appropriate portion of the planet and over an appropriate time interval, where what is "appropriate" depends on the response time and the efficiency of the heat transporting mechanisms of the planet under considerations. Correspondingly, the appropriate incident solar flux to use is the incident solar flux per unit of radiating surface, averaged consistently with temperature. We will denote this mean solar flux by the symbol S. For an isothermal planet $S = \frac{1}{4}L_\odot$, while at the opposite extreme $S = L_\odot$ for the instantaneous response at the subsolar point. In other circumstances it might be appropriate to average along a latitude circle, or over a hemisphere. A more complete treatment of geographical, seasonal and diurnal temperature variations will be given in Chapter 8.

Exercise 3.3.3 Consider a planet which is tide-locked to its Sun, so that it always shows the same face to the Sun as it proceeds in its orbit (just as the Moon always shows the same face to the Earth). Estimate the mean temperature of the day side of the planet, assuming the illuminated face to be isothermal, but assuming that no heat leaks to the night side.

Surface type	**Albedo**
Clean new H_2O snow	.85
Bare Sea ice	.5
Clean H_2O glacier ice	.6
Deep Water	.1
Sahara Desert sand	.35
Martian sand	.15
Basalt (any planet)	.07
Granite	.3
Limestone	.36
Grassland	.2
Deciduous forest	.14
Conifer forest	.09
Tundra	.2

Table 3.2: Typical values of albedo for various surface types. These are only representative values. Albedo can vary considerably as a function of detailed conditions. For example, the ocean albedo depends on the angle of the solar radiation striking the surface (the value given in the table is for near-normal incidence), and the albedo of bare sea ice depends on the density of air bubbles.

3.4 Ice-albedo feedback

Albedo is not a static quantity determined once and for all time when a planet forms. In large measure, albedo is determined by processes in the atmosphere and at the surface which are highly sensitive to the state of the climate. Clouds consist of suspended tiny particles of the liquid or solid phase of some atmospheric constituent; such particles are very effective reflectors of visible and ultraviolet light, almost regardless of what they are made of. Clouds almost entirely control the albedos of Venus, Titan and all the gas giant planets, and also play a major role in Earth's albedo. In addition, the nature of a planet's surface can evolve over time, and many of the surface characteristics are strongly affected by the climate. Table 3.2 gives the albedo of some common surface types encountered on Earth. The proportions of the Earth covered by sea-ice, snow, glaciers, desert sands or vegetation of various types are determined by temperature and precipitation patterns. As climate changes, the surface characteristics change too, and the resulting albedo changes feed back on the state of the climate. It is not a "chicken and egg" question of whether climate causes albedo or albedo causes climate; rather it is a matter of finding a consistent state compatible with the physics of the way climate affects albedo and the way albedo affects climate. In this sense, albedo changes lead to a form of climate *feedback*. We will encounter many other kinds of feedback loops in the climate system.

Among all the albedo feedbacks, that associated with the cover of the surface by highly reflective snow or ice plays a distinguished role in thinking about the evolution of the Earth's climate. Let's consider how albedo might vary with temperature for a planet entirely covered by a water ocean – a reasonable approximation to Earth, which is $\frac{2}{3}$ ocean. We will characterize the climate by the global mean surface temperature T_s, but suppose that, like Earth, the temperature is somewhat colder than T_s at the poles and somewhat warmer than T_s at the Equator. When T_s is very large, say greater than some threshold temperature T_o, the temperature is above freezing everywhere and there is no ice. In this temperature range, the planetary albedo reduces to the relatively low value (call it α_o) characteristic of sea water. At the other extreme, when T_s is very, very low, the whole planet is below freezing, the ocean will become ice-covered everywhere, and the

albedo reduces to that of sea ice, which we shall call α_i. We suppose that this occurs for $T_s < T_i$, where T_i is the threshold temperature for a globally frozen ocean. In general T_i must be rather lower than the freezing temperature of the ocean, since when the mean temperature $T_s = T_{freeze}$ the equatorial portions of the planet will still be above freezing. Between T_i and T_o it is reasonable to interpolate the albedo by assuming the ice cover to decrease smoothly and monotonically from 100% to zero. The phenomena we will emphasize are not particularly sensitive to the detailed form of the interpolation, but the quadratic interpolation

$$\alpha(T) = \begin{cases} \alpha_i & \text{for } T \leq T_i, \\ \alpha_o + (\alpha_i - \alpha_o)\frac{(T-T_o)^2}{(T_i-T_o)^2} & \text{for } T_i < T < T_o \\ \alpha_o & \text{for } T \geq T_o \end{cases} \tag{3.9}$$

qualitatively reproduces the shape of the albedo curve which is found in detailed calculations. In particular, the slope of albedo vs temperature is large when the temperature is low and the planet is nearly ice-covered, because there is more area near the Equator, where ice melts first. Conversely, the slope reduces to zero as the temperature threshold for an ice-free planet is approached, because there is little area near the poles where the last ice survives; moreover, the poles receive relatively little sunlight in the course of the year, so the albedo there contributes less to the global mean than does the albedo at lower latitudes. Note that this description assumes an Earthlike planet, which on average is warmest near the Equator. As will be discussed in Chapter 8, other orbital configurations could lead to the poles being warmer, and this would call for a different shape of albedo curve.

Ice albedo feedback of a similar sort could arise on a planet with land, through snow accumulation and glacier formation on the continents. The albedo could have a similar temperature dependence, in that glaciers are unlikely to survive where temperatures are very much above freezing, but can accumulate readily near places that are below freezing – *provided there is enough precipitation*. It is the latter requirement that makes land-based snow/ice albedo feedback much more complicated than the oceanic case. Precipitation is determined by complex atmospheric circulation patterns that are not solely determined by local temperature. A region with no precipitation will not form glaciers no matter how cold it is made. The present state of Mars provides a good example: its small polar glaciers do not advance to the Equator, even though the daily average equatorial temperature is well below freezing. Still, for a planet like Earth with a widespread ocean to act as a source for precipitation, it may be reasonable to assume that most continental areas will eventually become ice covered if they are located at sufficiently cold latitudes. In fairness, we should point out that even the formation of sea ice is considerably more complex than we have made it out to be, particularly since it is affected by the mixing of deep unfrozen water with surface waters which are trying to freeze.

Earth is the only known planet that has an evident ice/snow albedo feedback, but it is reasonable to inquire as to whether a planet without Earth's water-dominated climate could behave analogously. Snow is always "white" more or less regardless of the substance it is made of, since its reflectivity is due to the refractive index discontinuity between snow crystals and the ambient gas or vacuum. Therefore, a snow-albedo feedback could operate with substances other than water (e.g. nitrogen or methane). Titan presents an exotic possibility, in that its surface is bathed in a rain of tarry hydrocarbon sludge, raising the speculative possibility of "dark glacier" albedo feedbacks. Sea ice forming on Earth's ocean gets its high albedo from trapped air bubbles, which act like snowflakes in reverse. The same could happen for ices of other substances, but sea-ice albedo feedback is likely to require a water ocean. The reason is that water, alone among likely planetary materials, floats when it freezes. Ice forming on, say, a carbon dioxide or methane ocean would sink as soon as it formed, preventing it from having much effect on surface albedo.

Returning attention to an Earthlike waterworld, we write down the energy budget

$$(1 - \alpha(T_s))\frac{L_\odot}{4} = OLR(T_s) \tag{3.10}$$

This determines T_s as before, with the important difference that the Solar absorption on the left hand side is now a function of T_s instead of being a constant. Analogously to Fig. 3.5, the equilibrium surface temperature can be found by plotting the absorbed Solar radiation and the OLR vs. T_s on the same graph. This is done in Fig. 3.8, for four different choices of $L_\odot$. In this plot, we have taken $OLR = \sigma T^4$, which assumes no greenhouse effect [3]. In contrast with the fixed-albedo case, the ice-albedo feedback allows the climate system to have *multiple equilibria*: there can be more than one climate compatible with a given Solar constant, and additional information is required to determine which state the planet actually settles into. The nature of the equilibria depends on $L_\odot$. When $L_\odot$ is sufficiently small (as in the case $L_\odot = 1516W/m^2$ in Fig. 3.8) there is only one solution, which is a very cold globally ice-covered Snowball state, marked Sn_1 on the graph. Note that the Solar constant that produces a unique Snowball state exceeds the present Solar constant at Earth's orbit. Thus, were it not for the greenhouse effect, Earth would be in such a state, and would have been for its entire history. When *L_odot* is sufficiently large (as in the case $L_\odot = 2865W/m^2$ in Fig. 3.8) there is again a unique solution, which is a very hot globally ice-free state, marked H on the graph. However, for a wide range of intermediate $L_\odot$, there are three solutions: a Snowball state (Sn_2), a partially ice covered state with a relatively large ice sheet (e.g. A), and a warmer state (e.g. B) which may have a small ice sheet or be ice free, depending on the precise value of $L_\odot$. In the intermediate range of Solar constant, the warmest state is suggestive of the present or Pleistocene climate when there is a small ice-cap, and suggestive of Cretaceous-type hothouse climates when it is ice-free. In either case, the frigid Snowball state is available as an alternate possibility.

As the parameter $L_\odot$ is increased smoothly from low values, the temperature of the the Snowball state increases smoothly but at some point an additional solution discontinuously comes into being at a temperature far from the previous equilibrium, and splits into a pair as $L_\odot$ is further increased. As $L_\odot$ is increased further, at some point, the intermediate temperature state merges with the snowball state, and disappears. This sort of behavior, in which the behavior of a system changes discontinuously as some control parameter is continuously varied, is an example of a *bifurcation*.

Finding the equilibria tells only part of the story. A system placed exactly at an equilbrium point will stay there forever, but what if it is made a little warmer than the equilibrium? Will it heat up yet more, perhaps aided by melting of ice, and ultimately wander far from the equilibrium? Or will it cool down and move back toward the equilibrium? Similar questions apply if the state is made initially slightly cooler than an equilibrium. This leads us to the question of *stability*. In order to address stability, we must first write down an equation describing the time evolution of the system. To this end, we suppose that the mean energy storage per unit area of the planet's surface can be written as a function of the mean temperature; let's call this function $E(T_s)$. Changes in the energy storage could represent the energy required to heat up or cool down a layer of water of some characteristic depth, and could also include the energy needed to melt ice, or released by the freezing of sea water. For our purposes, all we need to know is that E is a monotonically increasing function of T_s. The energy balance for a time-varying system can then be written

$$\frac{dE(T_s)}{dt} = \frac{dE}{dT_s}\frac{dT_s}{dt} = G(T_s) \tag{3.11}$$

[3] Of course, this is an unrealistic assumption, since a waterworld would inevitably have at least water vapor – a good greenhouse gas – in its atmosphere

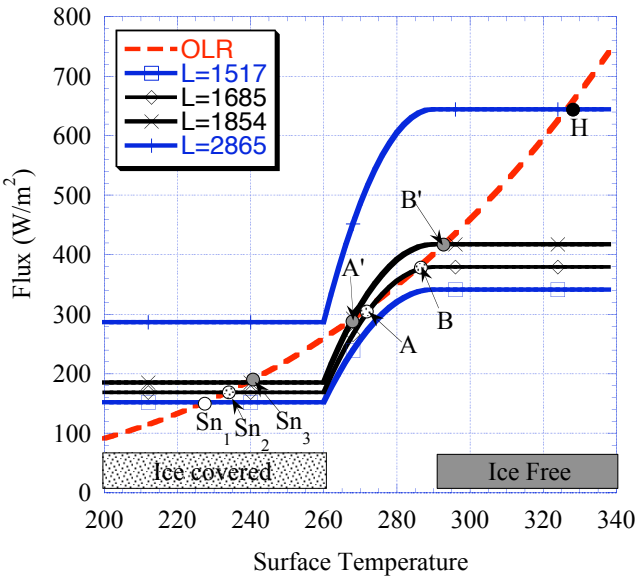

Figure 3.8: Graphical determination of the possible equilibrium states of a planet whose albedo depends on temperature in accordance with Eq. 3.9. The OLR is computed assuming the atmosphere has no greenhouse effect, and the albedo parameters are $\alpha_o = .1, \alpha_i = .6, T_i = 260K$ and $T_o = 290K$. The Solar constant for the various solar absorption curves is indicated in the legend.

where $G = \frac{1}{4}(1-\alpha(T_s))L_\odot - OLR(T_s)$. We can define the generalized heat capacity $\mu(T) = dE/dT$, which is positive by assumption. Thus,

$$\frac{dT_s}{dt} = \frac{G(T_s)}{\mu(T_s)} \tag{3.12}$$

By definition, $G = 0$ at an equilibrium point T_{eq}. Suppose that the slope of G is well-defined near T_{eq} – in formal mathematical language, we say that G is continuously differentiable at T_{eq}, meaning that the derivative of G exists and is a continous function for T_s in some neighborhood of T_{eq}. Then, if $dG/dT_s < 0$ at T_s, it will also be negative for some finite distance to the right and left of T_s. This is the case for points a and c in the net flux curve sketched in Fig. 3.9. If the temperature is made a little warmer than T_{eq} in this case, $G(T_s)$ and hence $\frac{dT_s}{dt}$ will become negative and the solution will move back toward the equilibrium. If the temperature is made a little colder than T_{eq}, $G(T_s)$ and hence $\frac{dT_s}{dt}$ will become positive, and the solution will again move back toward the equilibrium. In contrast, if $dG/dT_s > 0$ near the equilibrium, as for point c in the sketch, a temperature placed near the equilibrium moves away from it, rather than towards it. Such equilibria are *unstable*. If the slope happens to be exactly zero at an equilibrium, one must look to higher derivatives to determine stability. These are "rare" cases, which will be encountered only for very special settings of the parameters. If the d^2G/dT^2 is non zero at the equilibrium, the curve takes the form of a parabola tangent to the axis at the equilibrium. If the parabola opens upwards, then the equilibrium is stable to displacements to the left of the equilibrium, but unstable to displacements to the right. If the parabola opens downwards, the equilibrium is unstable to displacements to the left but stable to displacements to the right. Similar reasoning applies to the case in which the first non-vanishing derivative is higher order, but such cases are hardly ever encountered.

Exercise 3.4.1 Draw a sketch illustrating the behavior near marginal equilibria with $d^2G/dT^2 > 0$ and $d^2G/dT^2 < 0$. Do the same for equilibria with $d^2G/dT^2 = 0$, having $d^3G/dT^3 > 0$ and $d^3G/dT^3 < 0$

It is rare that one can completely characterize the behavior of a nonlinear system, but one dimensional problems of the sort we are dealing with are exceptional. In the situation depicted in Fig. 3.9, G is positive and dT/dt is positive throughout the interval between b and c. Hence, a temperature placed anywhere in this interval will eventually approach the solution c arbitrarily closely – it will be *attracted* to that stable solution. Similarly, if T is initially between a and b, the solution will be attracted to the stable equilibrium a. The unstable equilibrium b forms the boundary between the *basins of attraction* of a and c. No matter where we start the system within the interval between a and c (and somewhat beyond, depending on the shape of the curve further out), it will wind up approaching one of the two stable equilibrium states. In mathematical terms, we are able to characterize the *global behavior* of this system, as opposed to just the *local behavior* near equilibria.

At an equilibrium point, the curve of solar absorption crosses the OLR curve, and the stability criterion is equivalent to stating that the equilibrium is stable if the slope of the solar curve is less than that of the OLR curve where the two curves intersect. Using this criterion, we see that the intermediate-temperature large ice-sheet states, labeled A and A' in Fig. 3.8, are unstable. If the temperature is made a little bit warmer then the equilibrium the climate will continue to warm until it settles into the warm state (B or B') which has a small or nonexistent ice sheet. If the temperature is made a little bit colder than the equilibrium, the system will collapse into the snowball state (Sn_2 or Sn_3). The unstable state thus defines the boundary separating the basin of attraction of the warm state from that of the snowball state.

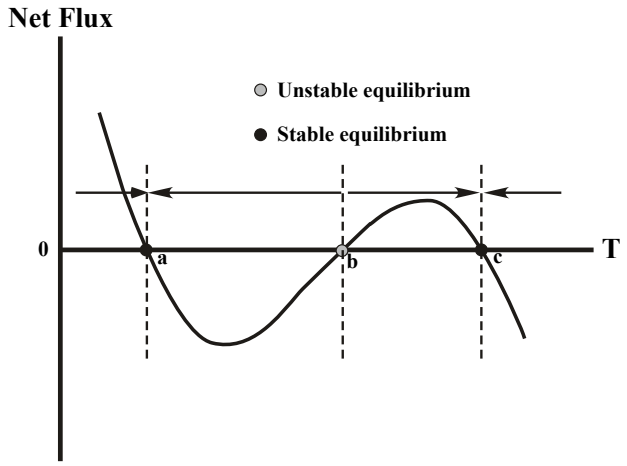

Figure 3.9: Sketch illustrating stable vs. unstable equilibrium temperatures.

Moreover, if the net flux $G(T)$ is continous and has a continuous derivative (i.e. if the curve has no "kinks" in it), then the sequence of consecutive equilibria always alternates between stable and unstable states. For the purpose of this theorem, the rare marginal states with $dG/dT = 0$ should be considered "wildcards" that can substitute for either a stable or unstable state. The basic geometrical idea leading to this property is more or less evident from Figure 3.9, but a more formalized argument runs as follows: Let T_a and T_b be equilibria, so that $G(T_a) = G(T_b) = 0$. Suppose that the first of these is stable, so $dG/dT < 0$ at T_a, and also that the two solutions are consecutive, so that $G(T)$ does not vanish for any T between T_a and T_b. Now if $dG/dT < 0$ at T_b, then it follows that $G > 0$ just to the left of T_b. The slope near T_a similarly implies that $G < 0$ just to the right of T_a. Since G is continuous, it would follow that $G(T) = 0$ somewhere between T_a and T_b. This would contradict our assumption that the two solutions are consecutive. In consequence, $dG/dT \geq 0$ at T_b. Thus, the state T_b is either stable or marginally stable, which proves our result. The proof goes through similarly if T_a is unstable. Note that we didn't actually need to make use of the condition that dG/dT be continuous everywhere: it's enough that it be continuous near the equilibria, so we can actually tolerate a few kinks in the curve.

A consequence of this result is that, if the shape of $G(T)$ is controlled continously by some parameter like $L_\odot$, then new solutions are born in the form of a single marginal state which, upon further change of $L_\odot$ splits into a stable/unstable or unstable/stable pair. The first member of the pair will be unstable if there is a pre-existing stable solution immediately on the cold side of the new one, as is the case for the Snowball states Sn in Fig. 3.8. The first member will be stable if there is a pre-existing unstable state on cold side, or a pre-existing stable state on the warm side (e.g. the state H in Fig. 3.8). What we have just encountered is a very small taste of the very large and powerful subject of *bifurcation theory*.

3.4.1 Faint Young Sun, Snowball Earth and Hysteresis

We we now have enough basic theoretical equipment to take a first quantitative look at the Faint Young Sun problem. To allow for the greenhouse effect of the Earth's atmosphere, we take $p_{rad} = 670mb$, which gives the correct surface temperature with the observed current albedo $\alpha = .3$.

How much colder does the Earth get if we ratchet the Solar constant down to $960W/m^2$, as it was 4.7 billion years ago when the Earth was new? As a first estimate, we can compute the new temperature from Eq. 3.8 holding p_{rad} and the albedo fixed at their present values. This yields $261K$. This is substantially colder than the present Earth. The fixed albedo assumption is unrealistic,however, since the albedo would increase for a colder and more ice-covered Earth, leading to a substantially colder temperature than we have estimated. In addition, the strength of the atmospheric greenhouse effect could have been different for the Early Earth, owing to changes in the composition of the atmosphere.

An attempt at incorporating the ice-albedo feedback can be made by using the energy balance Eq. 3.10 with the albedo parameterization given by Eq. 3.9. For this calculation, we choose constants in the albedo formula that give a somewhat more realistic Earthlike climate than those used in Figure 3.8. Specifically, we set $\alpha_o = .28$ to allow for the albedo of clouds and land, and $T_o = 295$ to allow a slightly bigger polar ice sheet. The position of the equilibria can be determined by drawing a graph like Fig. 3.8, or by applying a root-finding algorithm like Newton's method to Eq. 3.10. The resulting equilibria are shown as a function of $L_\odot$ in Figure 3.10, with p_{rad} held fixed at $670mb$. Some techniques for generating diagrams of this type are developed in Problem **??**. For the modern Solar constant, and $p_{rad} = 670mb$, the system has a stable equilibrium at $T_s = 286K$, close to the observed modern surface temperature, and is partially ice covered. However, the system has a second stable equilibrium, which is a globally ice-covered Snowball state having $T_s = 249K$. Even today, the Earth would stay in a Snowball state if it were somehow put there. The two stable equilibria are separated by an unstable equilibrium at $T_s = 270K$, which defines the boundary between the set of initial conditions that go to the "modern" type state, and the set that go to a Snowball state. The attractor boundary for the modern open-ocean state is comfortably far from the present temperature, so it would not be easy to succumb to a Snowball.

Now we turn down the Solar constant, and re-do the calculation. For $L_\odot = 960W/m^2$, there is only a single equilibrium point if we keep $p_{rad} = 670mb$. This is a stable Snowball state with $T_s = 228K$. Thus, if the Early Earth had the same atmospheric composition as today,leading to a greenhouse effect no stronger than the present one, the Earth would have inevitably been in a Snowball state. The open ocean state only comes into being when $L_\odot$ is increased to $1330W/m^2$, which was not attained until the relatively recent past. This contradicts the abundant geological evidence for prevalent open water throughout several billion years of Earth's history. Even worse, if the Earth were initially in a stable snowball state four billion years ago, it would stay in that state until $L_\odot$ increases to $1640W/m^2$, at which point the stable snowball state would disappear and the Earth would deglaciate. Since this far exceeds the present Solar constant, the Earth would be globally glaciated today. This even more obviously contradicts the data.

The currently favored resolution to the paradox of the Faint Young Sun is the supposition that the atmospheric composition of the early Earth must have resulted in a stronger greenhouse effect than the modern atmosphere produces. The prime candidate gases for mediating this change are CO_2 and CH_4. The radiative basis of the idea will be elaborated further in Chapter 4, and some ideas about why the atmosphere might have adjusted over time so as to maintain an equable climate despite the brightening Sun are introduced in Chapter 9. Fig. 3.11 shows how the equilibria depend on p_{rad}, with $L_\odot$ fixed at $960W/m^2$. Whichever greenhouse gas is the Earth's savior, if it is present in sufficient quantities to reduce p_{rad} to $500mb$ or less, then a warm state with an open ocean exists (the upper branch in Fig. 3.11). However, for $420mb < prad < 500mb$ a stable snowball state also exists, meaning that the climate that is actually selected depends on earlier history. If the planet had already fallen into a Snowball state for some reason, the early Earth would stay in a Snowball unless the greenhouse gases build up sufficiently to reduce p_{rad} below $420mb$ at some point.

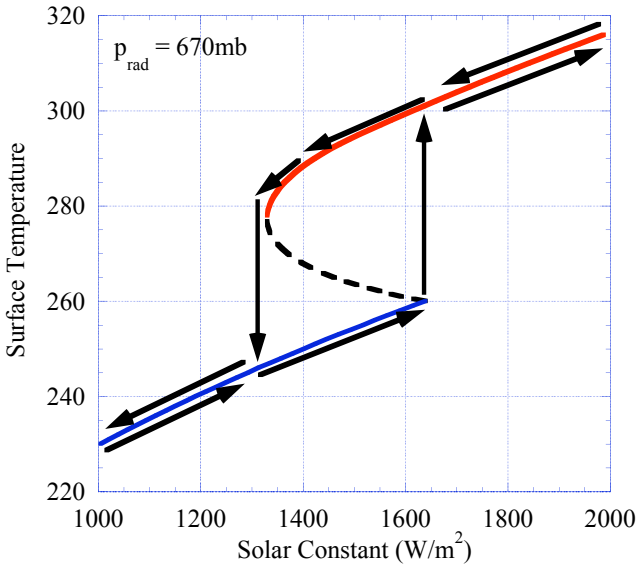

Figure 3.10: Hysteresis diagram obtained by varying $L_\odot$ with p_{rad}/p_s fixed at .67. Arrows indicate path followed by the system as $L_\odot$ is first increased, then decreased. The unstable solution branch is indicated by a dashed curve.

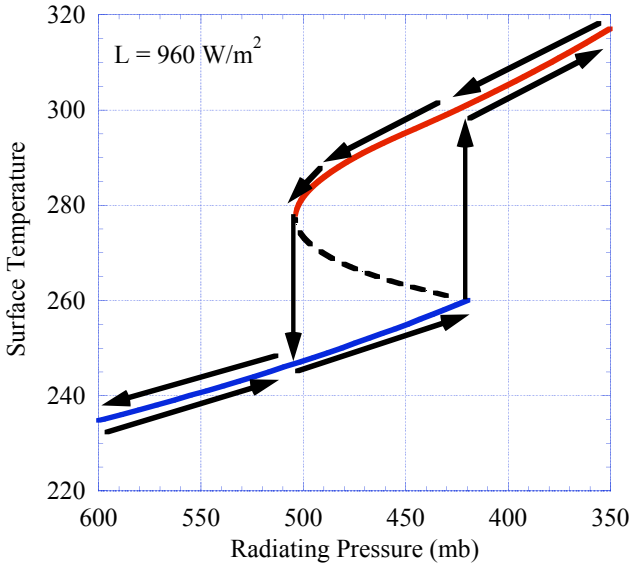

Figure 3.11: As in Fig. 3.10, but varying p_{rad} with $L_\odot = 960 W/m^2$.

Figures 3.10 and 3.11 illustrate an important phenomenon known as *hysteresis*: the state in which a system finds itself depends not just on the value of some parameter of the system, but the history of variation of that parameter. This is possible only for systems that have multiple stable states. For example, in 3.10 suppose we start with $L_\odot = 1000W/m^2$, where the system is inevitably in a Snowball state with $T = 230K$. Let's now gradually increase $L_\odot$. When $L_\odot$ reaches $1500W/m^2$ the system is still in a Snowball state, having $T = 254K$, since we have been following a stable solution branch the whole way. However, when $L_\odot$ reaches $1640W/m^2$, the Snowball solution disappears, and the system makes a sudden transition from a Snowball state with $T = 260K$ to the only available stable solution, which is an ice-free state having $T = 301K$. As $L_\odot$ increases further to $2000W/m^2$, we follow the warm, ice-free state and the temperature rises to $316K$. Now suppose we begin to gradually dim the Sun, perhaps by making the Solar system pass through a galactic dust cloud. Now, we follow the upper, stable branch as $L_\odot$ decreases, so that when we find ourselves once more at $L_\odot = 1500W/m^2$ the temperature is $294K$ and the system is in a warm, ice-free state rather than in the Snowball state we enjoyed the last time we were there. As $L_\odot$ is decreased further, the warm branch disappears at $L_\odot = 1330W/m^2$ and the system drops suddenly from a temperature of $277K$ into a Snowball state with a temperature of $246K$, whereafter the Snowball branch is again followed as $L_\odot$ is reduced further. The trajectory of the system as $L_\odot$ is increased then decreased back to its original value takes the form of an open loop, depicted in Fig. 3.10.

The thought experiment of varying $L_\odot$ in a hysteresis loop is rather fanciful, but many atmospheric processes could act to either increase or decrease the greenhouse effect over time. For the very young Earth, with $L_\odot = 960W/m^2$, the planet falls into a Snowball when p_{rad} exceeds $500mb$, and thereafter would not deglaciate until p_{rad} is reduced to $420mb$ or less (see Fig. 3.11). The boundaries of the hysteresis loop, which are the critical thresholds for entering and leaving the Snowball, depend on the solar constant. For the modern solar constant, the hysteresis loop operates between $p_{rad} = 690mb$ and $p_{rad} = 570mb$. It takes less greenhouse effect to keep out of the Snowball now than it did when the Sun was fainter, but the threshold for initiating a Snowball in modern conditions is disconcertingly close to the value of p_{rad} which reproduces the present climate.

The fact that the freeze-thaw cycle can exhibit hysteresis as atmospheric composition changes is at the heart of the Snowball Earth phenomenon. An initially warm state can fall into a globally glaciated Snowball if the atmospheric composition changes in such a way as to sufficiently weaken the greenhouse effect. Once the threshold is reached, the planet can fall into a Snowball relatively quickly – in a matter of a thousand years or less – since sea ice can form quickly. However, to deglaciate the Snowball, the greenhouse effect must be increased far beyond the threshold value at which the planet originally entered the Snowball state. Atmospheric composition must change drastically in order to achieve such a great increase, and this typically takes many millions of years. When deglaciation finally occurs, it leaves the atmosphere in a hyper-warm state, which only gradually returns to normal as the atmospheric composition evolves in such a way as to reduce the greenhouse effect. As discussed in Chapter 1, there are two periods in Earth's past when geological evidence suggests that one or more Snowball freeze-thaw cycles may have occurred. The first is in the Paleoproterozoic, around 2 billion years ago. At this time, $L_\odot \approx 1170W/m^2$, and the thresholds for initiating and deglaciating a Snowball are $p_{rad} = 600mb$ and $p_{rad} = 500mb$ in our simple model. For the Neoproterozoic, about 700 million years ago, $L_\odot \approx 1290W/m^2$ and the thresholds are at $p_{rad} = 650mb$ and $p_{rad} = 540mb$.

The boundaries of the hysteresis loop shift as the Solar constant increases, but there is nothing obvious in the numbers to suggest why a Snowball state should have occurred in the Paleoproterozoic and Neoproterozoic but not at other times. Hysteresis associated with ice-albedo

feedback has been a feature of the Earth's climate system throughout the entire history of the planet. Hysteresis will remain a possibility until the Solar constant increases sufficiently to render the Snowball state impossible even in the absence of any greenhouse effect (i.e. with $p_{rad} = 1000mb$). Could a Snowball episode happen again in the future, or is that peril safely behind us? These issues require an understanding of the processes governing the evolution of Earth's atmosphere, a subject that will be taken up in Chapter 9.

Exercise 3.4.2 Assuming an ice albedo of .6, how high does $L_\odot$ have to become to eliminate the possibility of a snowball state? Will this happen within the next five billion years? What if you assume there is enough greenhouse gas in the atmosphere to make $p_{rad}/p_s = .5$?

Note: The evolution of the Solar constant over time is approximately $L_\odot(t) = L_{\odot p} \cdot (.7 + (t/22.975) + (t/14.563)^2)$, where t is the age of the Sun in billions of years ($t = 4.6$ being the current age) and $L_{\odot p}$ is the present Solar constant. This fit is reasonably good for the first 10 billion years of Solar evolution.

The "cold start" problem is a habitability crisis that applies to waterworlds in general. If a planet falls into a Snowball state early in its history, it could take billions of years to get out if one needs to wait for the Sun to brighten. The time to get out of a Snowball could be shortened if greenhouse gases build up in the atmosphere, reducing p_{rad}. How much greenhouse gas must build up to deglaciate a snowball? How long would that take? What could cause greenhouse gases to accumulate on a Snowball planet? These important questions will be taken up in subsequent chapters.

Another general lesson to be drawn from the preceding discussion is that the state with a stable, small icecap is very fragile, in the sense that the planetary conditions must be tuned rather precisely for the state to exist at all. For example, with the present Solar constant, the stable small icecap solution first appears when p_{rad} falls below $690mb$. However, the icecap shrinks to zero as p_{rad} is reduced somewhat more, to $615mb$. Hence, a moderate strengthening in the greenhouse effect would, according to the simple energy balance model, eliminate the polar ice entirely and throw the Earth into an ice-free Cretaceous hothouse state. The transition to an ice-free state of this sort is continous in the parameter being varied; unlike the collapse into a snowball state or the recovery from a snowball, it does not result from a bifurcation. In light of its fragility, it is a little surprising that the Earth's present small-icecap state has persisted for the past two million years, and that similar states have occurred at several other times in the past half billion years. Does the simple energy-balance model exaggerate the fragility of the stable small-icecap state? Does some additional feedback process adjust the greenhouse effect so as to favor such a state while resisting the peril of the Snowball? These are largely unresolved questions. Attacks on the first question require comprehensive dynamical models of the general circulation, which we will not encounter in the present volume. We will take up, though not resolve, the second question in Chapter9. It is worth noting that small-icecap states like those of the past two million years appear to be relatively uncommon in the most recent half billion years of Earth's history, for which data is good enough to render a judgement about ice cover. The typical state appears to be more like the warm relatively ice-free states of the Cretaceous, and perhaps this reflects the fragility of the small-icecap state.

The simple models used above are too crude to produce very precise hysteresis boundaries. Among the many important effects left out of the story are water vapor radiative feedbacks, cloud feedbacks, the factors governing albedo of sea ice, ocean heat transports and variations in atmospheric heat transport. The phenomena uncovered in this exposition are general, however and can be revisited across a heirarchy of models. Indeed, the re-examination of this subject provides an unending source of amusement and enlightenment to climate scientists.

3.4.2 Climate sensitivity, radiative forcing and feedback

The simple model we have been studying affords us the opportunity to introduce the concepts of *radiative forcing, sensitivity coefficient* and *feedback factor*. These diagnostics can be applied across the whole spectrum of climate models, from the simplest to the most comprehensive.

Suppose that the mean surface temperature depends on some parameter Λ, and we wish to know how sensitive T is to changes in that parameter. For example, this parameter might be the Solar constant, or the radiating pressure. It could be some other parameter controlling the strength of the greenhouse effect, such as CO_2 concentration. Near a given Λ, the sensitivity is characterized by $dT/d\Lambda$.

Let G be the net top-of-atmosphere flux, such as used in Eq. 3.11. To allow for the fact that the terms making up the net flux depend on the parameter Λ, we write $G = G(T, \Lambda)$. If we take the derivative of the the energy balance requirement $G = 0$ with respect to Λ, we find

$$0 = \frac{\partial G}{\partial T}\frac{dT}{d\Lambda} + \frac{\partial G}{\partial \Lambda} \tag{3.13}$$

so that

$$\frac{dT}{d\Lambda} = -\frac{\frac{\partial G}{\partial \Lambda}}{\frac{\partial G}{\partial T}} \tag{3.14}$$

The numerator in this expression is a measure of the *radiative forcing* associated with changes in Λ. Specifically, changing Λ by an amount $\delta\Lambda$ will perturb the top-of-atmosphere radiative budget by $\frac{\partial G}{\partial \Lambda}\delta\Lambda$, requiring that the temperature change so as to bring the energy budget back into balance. For example, if Λ is the Solar constant L, then $\frac{\partial G}{\partial \Lambda} = \frac{1}{4}(1-\alpha)$. If Λ is the radiating pressure p_{rad}, then $\frac{\partial G}{\partial \Lambda} = -\frac{\partial OLR}{\partial p_{rad}}$. Since OLR goes up as p_{rad} is reduced, a reduction in p_{rad} yields a positive radiative forcing. This is a warming influence.

Radiative forcing is often quoted in terms of the change in flux caused by a standard change in the parameter, in place of the slope $\frac{\partial G}{\partial \Lambda}$ itself. For example, the radiative forcing due to CO_2 is typically described by the change in flux caused by doubling CO_2 from its pre-industrial value, with temperature and everything else is held fixed. This is practically the same thing as $\frac{\partial G}{\partial \Lambda}$ if we take $\Lambda = \log_2 pCO_2$, where pCO_2 is the partial pressure of CO_2. Similarly, the climate sensitivity is often described in terms of the temperature change caused by the standard forcing change, rather than the slope $\frac{dT}{d\Lambda}$. For example, the notation ΔT_{2x} would refer to the amount by which temperature changes when CO_2 is doubled.

The denominator of Eq. 3.14 determines how much the equilibrium temperature changes in response to a given radiative forcing. For any given magnitude of the forcing, the response will be greater if the denominator is smaller. Thus, the denominator measures the *climate sensitivity*. An analysis of ice-albedo feedback illustrates how a feedback process affects the climate sensitivity. If we assume that albedo is a function of temperature, as in Eq. 3.9, then

$$\frac{\partial G}{\partial T} = -\frac{1}{4}L\frac{\partial \alpha}{\partial T} - \frac{\partial OLR}{\partial T} \tag{3.15}$$

With this expression, Eq. 3.14 can be rewritten

$$\frac{dT}{d\Lambda} = -\frac{1}{1+\Phi}\left[\frac{\frac{\partial G}{\partial \Lambda}}{\frac{\partial OLR}{\partial T}}\right] \tag{3.16}$$

where

$$\Phi = \frac{1}{4}L\frac{\frac{\partial \alpha}{\partial T}}{\frac{\partial OLR}{\partial T}} \tag{3.17}$$

In writing this equation we primarily have ice-albedo feedback in mind, but the equation is valid for arbitrary $\alpha(T)$ so it could as well describe a variety of other processes. The factor in square brackets in Eqn. 3.16 is the sensitivity the system would have if the response were unmodified by the change of albedo with temperature. The first factor determines how the sensitivity is increased or decreased by the feedback of temperature on albedo. If $-1 < \Phi < 0$ then the feedback increases the sensitivity – the same radiative forcing produces a bigger temperature change than it would in the absence of the feedback. When $\Phi = -\frac{1}{2}$, for example, the response to the forcing is twice what it would have been in the absence of the feedback. The sensitivity becomes infinite as $\Phi \to -1$, and for $-2 < \Phi < -1$ the feedback is so strong that it actually reverses the sign of the response as well as increasing its magnitude. On the other hand, if $\Phi > 0$, the feedback reduces the sensitivity. In this case it is a *stabilizing feedback*. The larger Φ gets, the more the response is reduced. For example, when $\Phi = 1$ the response is half what it would have been in the absence of feedback. Note that the feedback term is the same regardless of whether the radiative forcing is due to changing L, p_{rad} or anything else.

As an example, let's compute the feedback parameter Φ for the albedo-temperature relation given by Eq. 3.9, under the conditions shown in Fig. 3.10. Consider in particular the upper solution branch, which represents a stable partially ice-covered climate like that of the present Earth. At the point $L = 1400W/m^2, T = 288K$ on this branch, we find $\Phi = -.333$. Thus, at this point the ice-albedo feedback increases the sensitivity of the climate by a factor of about 1.5. At the bifurcation point $L \approx 1330W/m^2, T \approx 277K$, $\Phi \to -1$ and the sensitivity becomes infinite. This divergence merely reflects the fact that the temperature curve is vertical at the bifurcation point. Near such points, the temperature change is no longer linear in radiative forcing. It can easily be shown that the temperature varies as the square root of radiative forcing near a bifurcation point, as suggested by the plot.

The ice-albedo feedback increases the climate sensitivity, but other feedbacks could be stabilizing. In fact Eq. 3.17 is valid whatever the form of $\alpha(T)$, and shows that the albedo feedback becomes a stabilizing influence if albedo increases with temperature. This could conceivably happen as a result of vegetation feedback, or perhaps dissipation of low clouds. The somewhat fanciful Daisyworld example in the Workbook section at the end of this chapter provides an example of such a stabilizing feedback.

The definition of the feedback parameter can be generalized as follows. Suppose that the energy balance function G depends not only on the control parameter Λ, but also on some other parameter R which varies systematically with temperature. In the previous example, $R(T)$ is the temperature-dependent albedo. We write $G = G(T, R(T), \Lambda)$. Following the same line of reasoning as we did for the analysis of ice-albedo feedback, we find

$$\Phi = \frac{\partial G}{\partial R}\frac{\partial R}{\partial T}\frac{\partial G}{\partial T} \tag{3.18}$$

For example, if R represents the concentration of water vapor on Earth, or of methane on Titan, and if R varies as a function of temperature, then the feedback would influence G through the OLR. Writing $OLR = OLR(T, R(T), \Lambda)$, then the feedback parameter is

$$\Phi = \frac{\frac{\partial OLR}{\partial R}\frac{\partial R}{\partial T}}{\frac{\partial OLR}{\partial T}} \tag{3.19}$$

assuming the albedo to be independent of temperature in this case. Now, since OLR increases with T and OLR decreases with R, the feedback will be destabilizing ($\Phi < 0$) if R increases with T. (One might expect R to increase with T because Clausius-Clapeyron implies that the saturation vapor pressure increases sharply with T, making it harder to remove water vapor by condensation,

all other things being equal). Note that in this case the water vapor feedback does not lead to a runaway, with more water leading to higher temperatures leading to more water in a never-ending cycle; the system still attains an equilibrium, though the sensitivity of the equilibrium temperature to changes in a control parameter is increased.

3.5 Partially absorbing atmospheres

The assumption underpinning the blackbody radiation formula is that radiation interacts so strongly with matter that it achieves thermodynamic equilibrium at the same temperature as the matter. It stands to reason, then, that if a box of gas contains too few molecules to offer much opportunity to intercept a photon, the emission will deviate from the blackbody law. Weak interaction with radiation can also arise from aspects of the structure of a material which inhibit interaction, such as the crystal structure of table salt or carbon dioxide ice. In either event, the deviation of emission from the Planck distribution is characterized by the *emissivity*. Suppose that $I(\nu, \hat{n})$ is the observed flux of radiation at frequency ν emerging from a body in the direction $\hat{n}$. Then the emissivity $e(\nu, \hat{n})$ is defined by the expression

$$I(\nu, \hat{n}) = e(\nu, \hat{n})B(\nu, T) \tag{3.20}$$

where T is the temperature of the collection of matter we are observing. Note that in assigning a temperature T to the body, we are assuming that the matter itself is in a state of thermodynamic equilibrium. The emissivity may also be a function of temperature and pressure. We can also define a mean emissivity over frequencies, and all rays emerging from a body. The mean emissivity is

$$\bar{e} = \frac{\int_{\nu,\Omega} e(\nu, \hat{n})B(\nu, T)\cos\theta d\nu d\Omega}{\sigma T^4} \tag{3.21}$$

where θ is the angle of the ray to the normal to the body's surface and the angular integration is taken over the hemisphere of rays leaving the surface of the body. With this definition, the net flux emerging from any patch of the body's surface is $F = \bar{e}\sigma T^4$. Even if e does not depend explicitly on temperature, $\bar{e}$ will be temperature dependent if e is frequency dependent, since the relative weighting of different frequencies, determined by $B(\nu, T)$ changes with temperature.

A blackbody has unit emissivity at all frequencies and directions. A blackbody also has unit absorptivity, which is just a restatement of the condition that blackbodies interact strongly with the radiation field. For a non-black body, we can define the absorptivity $a(\nu, \hat{n})$ by shining light at a given frequency and direction at the body and measuring how much is reflected and how much comes out the other side. Specifically, suppose that we shine a beam of electromagnetic energy with direction $\hat{n}$, frequency ν and flux F_{inc} at the test object. Then we measure the *additional* energy flux coming out of the object once this beam is turned on. This outgoing flux may come out in many different directions, because of scattering of the incident beam; in exotic cases, even the frequency could differ from the incident radiation. Let T and R be the transmitted and reflected energy flux, integrated over all angles and frequencies. Then, the absorptivity is defined by taking the ratio of the flux of energy left behind in the body to the incident flux. Thus,

$$a(\nu, \hat{n}) = \frac{F_{inc} - (T + R)}{F_{inc}} \tag{3.22}$$

The Planck function is unambigously the natural choice of a weighting function for defining the mean emissivity $\bar{e}$ for an object with temperature T. There is no such unique choice for defining the mean absorptivity over all frequencies and directions. The appropriate weighting function is

determined by the frequency and directional spectrum of the incident radiation which requires a detailed knowledge of its source. If the incident radiation is a blackbody with temperature T_{source} then $\bar{a}$ should be defined with a formula like Eq. 3.20, using $B(\nu, T_{source}$ as the weighting function. Note that the weighting function is defined by the temperature of the *source* rather than by the temperature of the the object doing the absorbing. As was the case for mean emissivity, the temperature dependence of the weighting function implies that $\bar{a}$ will vary with T_{source} even if $a = a(\nu)$ and is not explicitly dependent on temperature.

Absorptivity and emissivity might appear to be independent characteristics of an object, but observations and theoretical arguments reveal an intimate relation between the two. This relation, expressed by *Kirchhoff's Law of Radiation* is a profound property of the interaction of radiation with matter that lies at the heart of all radiative transfer theory. Kirchoff's Law states that the emissivity of a substance at any given frequency equals the emissivity measured at the same frequency. It was first inferred experimentally. The hard-working spectroscopists of the late nineteenth centuries employed their new techniques to measure the emission spectrum $I(\nu, \hat{n}, T)$ and absorptivity $a(\nu, \hat{n}, T)$ of a wide variety of objects at various temperatures. Kirchhoff found that, with the exception of a few phosphorescent materials whose emission was not linked to temperature, all the experimental data collapsed onto a single universal curve, independent of the material, once the observed emission was normalized by the observed absorptivity. In other words, virtually all materials fit the relation $I(\nu, \hat{n}, T)/a(\nu, \hat{n}, T) = f(\nu, T)$ with the same function f. If we take the limit of a perfect absorber – a perfectly "black" body – then $a = 1$ and we find that f is in fact what we have been calling the *Planck* function $B(\nu, T)$. In fact, it was this extrapolation to a perfect absorber that originally led to the formulation of the notion of blackbody radiation. Since $f = B$ and $I = eB$, we recover the statement of Kirchoff's law in the form $e/a = 1$.

The thought experiment sketched in Fig. 3.12 allows us to deduce Kirchhoff's law for the mean absorptivity and emissivity from the requirements of the Second Law of Thermodynamics. We consider two infinite slabs of a blackbody material with temperature T_o, separated by a gap. Into the gap, we introduce a slab of partially transparent material with mean absorptivity $\bar{a}(T_1)$ and mean emissivity $\bar{e}(T_1)$, where T_1 is the temperature of the test material. Note that this system is energetically closed. We next require that the radiative transfer between the blackbody material and the test object cause the system to evolve toward an isothermal state. In other words we are *postulating* that radiative heat transfers satisfy the Second Law. A *necessary* condition for radiative transfer to force the system to evolve towards an isothermal state is that the isothermal state $T_o = T_1$ be an equilibrium state of the system; if it weren't an initially isothermal state would spontaneously generate temperature inhomogeneities. Energy balance requires that $2\bar{a}(T_o)\sigma T_o^4 = 2\bar{e}(T_1)\sigma T_1^4$. Kirchhoff's law then follows immediately by setting $T_o = T_1$ in the energy balance, which then implies $\bar{a}(T_o) = \bar{e}(T_o)$. Note that the mean absorptivity in this statement is defined using the Planck function at the common temperature of the two materials as the weighting function.

A modification of the preceding argument allows us to show that in fact the emissivity and absorptivity should be equal at each individual frequency, and not just in the mean. To simplify the argument, we will assume that e and a are independent of direction. The thought experiment we employ is similar to that used to justify Kirchhoff's Law in the mean, except that this time we interpose frequency-selective mirrors between the test object and the blackbody material, as shown in Fig. 3.13. The mirrors allow the test object to exchange radiant energy with the blackbody only in a narrow frequency band $\Delta\nu$ around a specified frequency ν. The energy budget for the test object now reads $2e(\nu)B(\nu, T_1)\Delta\nu = 2a(\nu)B(\nu, T_o)\Delta\nu$. Setting $T_1 = T_o$ so that the isothermal state is an equilibrium, we find that $e(\nu) = a(\nu)$.

The preceding argument, presented in the form originally given by Kirchhoff, is the justifi-

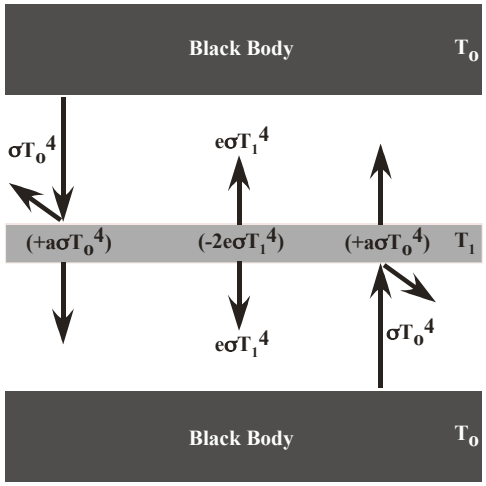

Figure 3.12: Sketch illustrating thought experiment for demonstating Kirchoff's Law in the mean over all wavenumbers. In the annotations on the sketch, $a = \bar{a}(T_o)$ and $e = \bar{e}(T_1)$.

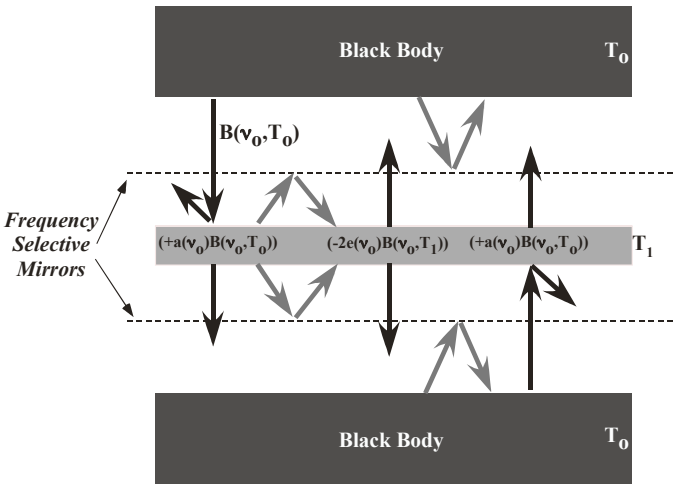

Figure 3.13: Sketch illustrating thought experiment for demonstrating Kirchoff's Law for a narrow band of radiation near frequency ν_0. The thin dashed lines represent ideal frequency-selective mirrors, which pass frequencies close to ν_0, but reflect all others without loss.

cation commonly given for Kirchhoff's Law. It is ultimately unsatisfying, as it applies equilibrium thermodynamic reasoning to a system in which the radiation field is manifestly out of equilibrium with matter; in the frequency-dependent form, it invokes the existence of mirrors with hypothetical material properties; worse, it takes as its starting point that radiative heat transfer will act like other heat transfers to equalize temperature, whereas we really ought to be able to demonstrate such a property from first principles of the interaction of radiation with molecules. The great mathematician David Hilbert, was among many who recognized these difficulties; in 1912 he presented a formal justification that eliminated the involvement of hypothetical ideal selective mirrors. The physical content of Hilbert's proof is that one doesn't need an ideal mirror, if one requires that a sufficient variety of materials with different absorbing and emitting properties will all come into an isothermal state at equilibrium. Hilbert's deriviation nonetheless relied on an assumption that radiation would come into equilibrium with matter at each individual wavelength considered separately. While Kirchoff did the trick with mirrors, Hilbert, in essence, did the trick with axioms instead, leaving the microscopic justification of Kirchoff's Law equally obscure. It is in fact quite difficult to provide a precise and concise statement of the circumstances in which a material will comply with Kirchoff's Law. Violations are quite commonplace in nature and in engineered materials, since it is quite possible for a material to store absorbed electromagnetic energy and emit it later, perhaps at a quite different frequency. A few examples that come to mind are phosphorescent ("glow in the dark") materials, fluorescence (e.g. paints that glow when exposed to ultraviolet, or "black" light), frequency doubling materials (used in making green laser pointers), and lasers themselves. In Nature, such phenomena involve insignificant amounts of energy, and are of no known importance in determining the energy balance of planets. We will content ourselves here with the statement that all known liquid and solid planetary materials, as well as the gases making up atmospheres, conform very well to Kirchoff's Law, except perhaps in the most tenuous outer reaches of atmospheres where the gas itself is not in thermodynamic equilibrium.

When applying Kirchhoff's law in the mean, careful attention must be paid to the weighting function used to define the mean absorptivity. For example, based on the incident Solar spectrum, the Earth has a mean albedo of about .3, and hence a mean absorptivity of .7. Does this imply that the mean emissivity of the Earth must be .7 as well? In fact, no such implication can be drawn, because Kirchhoff's Law only requires that the mean emissivity and absorptivity are the same when averaged over identical frequency weighting functions. Most of the Earth's thermal emission is in the infrared, not the visible. Kirchhoff's law indeed requires that the *visible* wavelength emissivity is .7, but the net thermal emission of the Earth in this band is tiny compared to the infrared, and contributes almost nothing to the Earth's net emission. Specifically, the Planck function implies that, at $255K$ the emission in visible wavelengths is smaller than the emission in infrared wavelengths by a factor of about 10^{-19}. Thus, if the infrared emission from some region were $100W/m^2$, the visible emission would be only $10^{-17}W/m^2$. Using $\Delta E = h\nu$ to estimate the energy of a photon of visible light, we find that this amounts to an emission of only 50 visible light photons each second, from each square meter of radiating surface. This tiny outgoing *thermal* emission of visible light should not be confused with the much larger outgoing flux of *reflected* solar radiation.

It is a corollary of Kirchhoff's law that $e \leq 1$. If the emissivity were greater than unity, then by Kirchhoff's Law, the absorptivity would also have to be greater than unity. In consequence, the amount of energy absorbed by the body per unit time would be greater than the amount delivered to it by the incident radiation. By conservation of energy, that would imply the existence of an internal energy source. However, any internal energy source would ultimately be exhausted, violating the assumption that the system is in a state of equilibrium which can be maintained indefinitely.

3.6 Optically thin atmospheres: The skin temperature

Since the density of an atmosphere always approaches zero with height, in accordance with the hydrostatic law, one can always define an outer layer of the atmosphere that has so few molecules in it that it will have low infrared emissivity. We will call this the *skin layer*. What is the temperature of this layer? Suppose for the moment that it is transparent to solar radiation, and that atmospheric motions do not transport any heat into the layer; thus, it is heated only by infrared upwelling from below. Because the emissivity of the skin layer is assumed small, little of the upwelling infrared will be absorbed, and so the upwelling infrared is very nearly the same as the OLR. The energy balance is between absorption and emission of infrared. Since the skin layer radiates from both its top and bottom, the energy balance reads

$$2e_{ir}\sigma T_{skin}^4 = e_{ir}OLR. \tag{3.23}$$

Hence,

$$T_{skin} = \frac{1}{2^{\frac{1}{4}}}(\frac{OLR}{\sigma})^{\frac{1}{4}} = \frac{1}{2^{\frac{1}{4}}}T_{rad} \tag{3.24}$$

where T_{rad} is defined as before. Thus, the skin temperature is colder than the blackbody radiating temperature by a factor of $2^{-\frac{1}{4}}$. The skin temperature is the natural temperature the outer regions of an atmosphere would have in the absence of *in situ* heating by solar absorption or other means. Note that the skin layer does not need any interior heat transfer mechanism to keep it isothermal, since the argument we have applied to determine T_{skin} applies equally well to any sublayer of the skin layer.

A layer that has low emissivity, and hence low absorptivity, in some given wavelength band is referred to as being *optically thin* in this band. A layer could well be optically thick in the infrared, but optically thin in the visible, which is in fact the case for strong greenhouse gases.

Now let's suppose that the entire atmosphere is optically thin, right down to the ground, and compute the pure radiative equilibrium in this system in the absence of heat transfer by convection. We'll also assume that the atmosphere is completely transparent to the incident Solar radiation. Let S be the incident Solar flux per unit surface area, appropriate to the problem under consideration (e.g. $\frac{1}{4}L_\odot$ for the global mean or $L_\odot$ for temperature at the subsolar point on a planet like modern Mars). Since the atmosphere has low emissivity, the heating of the ground by absorption of downwelling infrared emission coming from the atmosphere can be neglected to lowest order. Since the ground is heated only by absorbed Solar radiation, its temperature is determined by $\sigma T_s^4 = (1-\alpha)S$, just as if there were no atmosphere at all. In other words, $p_{rad} = p_s$ because the atmosphere is optically thin, so that the atmosphere does not affect the surface temperature no matter what its temperature structure turns out to be. Next we determine the atmospheric temperature. The whole atmosphere has small *but nonzero* emissivity so that the skin layer in this case extends right to the ground. The atmosphere is then isothermal, and its temperature T_a is just the skin temperature $2^{-1/4}T_s$.

The surface is thus considerably warmer than the air with which it is in immediate contact. There would be nothing unstable about this situation if radiative transfer were truly the only heat transfer mechanism coupling the atmosphere to the surface. In reality, the air molecules in contact with the surface will acquire the temperature of the surface by heat conduction, and turbulent air currents will carry the warmed air away from the surface, forming a heated, buoyant layer of air. This will trigger convection, mixing a deep layer of the atmosphere within which the temperature profile will follow the adiabat. The layer will grow in depth until the temperature at the top of the mixed layer matches the skin temperature, eliminating the instability. This situation is

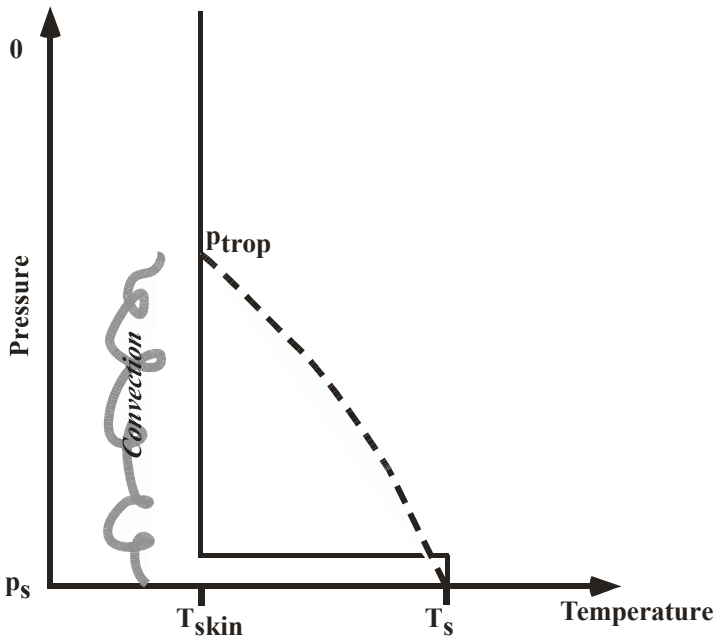

Figure 3.14: The unstable pure radiative equilibrium for an optically thin atmosphere (solid line) and the result of adjustment to the adiabat by convection (dashed line). The adjustment of the temperature profile leaves the surface temperature unchanged in this case, because the atmosphere is optically thin and has essentially no effect on the OLR.

depicted in Figure 3.14. The isothermal, stably stratified region above the mixed region is the stratosphere in this atmosphere, and the lower, adiabatic region is the troposphere; the boundary between the two is the tropopause. We have just formulated a theory of tropopause height for optically thin atmospheres. To make it quantitative, we need only require that the adiabat starting at the surface temperature match to the skin temperature at the tropopause. Let p_s be the surface pressure and p_{trop} be the tropopause pressure. For the dry adiabat, the requirement is then $T_s(p_{trop}/p_s)^{R/c_p} = T_{skin}$. Since $T_s = 2^{1/4}T_{skin}$, the result is

$$\frac{p_{trop}}{p_s} = 2^{-\frac{c_p}{4R}} \tag{3.25}$$

Note that the tropopause pressure is affected by R/c_p, but is independent of the insolation S.

The stratosphere in the preceding calculation differs from the observed stratosphere of Earth in that it is isothermal rather than warming with altitude. The factor we have left out is that real stratospheres often contain constituents that absorb solar radiation. To rectify this shortcoming, let's consider the effect of solar absorption on the temperature of the skin layer. Let e_{ir} be the infrared emissivity, which is still assumed small, and a_{sw} be the shortwave (mostly visible) absorptivity, which will also be assumed small. Note that Kirchhoff's Law does not require $e_{ir} = a_{sw}$, as the emissivity and absorptivity are at different wavelengths. The solar absorption of incident radiation is $a_{sw}S$. We'll assume that the portion of the solar spectrum which is absorbed by the

atmosphere is absorbed so strongly that it is completely absorbed before reaching the ground. This is in fact the typical situation for solar near-infrared and ultraviolet. In this case, one need not take into account absorption of the upwelling solar radiation reflected from the surface.

Exercise 3.6.1 Show that if the atmosphere absorbs uniformly throughout the solar spectrum, then the total absorption in the skin layer is $(1 + (1 - a_{sw})\alpha_g)a_{sw}S$, where α_g is the solar albedo of the ground. Show that the *planetary albedo* – i.e. the albedo observed at the top of the atmosphere – is $(1 - a_{sw})^2\alpha_g$.

The energy balance for the skin layer now reads

$$e_{ir}\sigma T^4 = e_{ir}OLR + a_{sw}S \tag{3.26}$$

Hence,

$$T = T_{skin}(1 + \frac{a_{sw}}{e_{ir}}\frac{S}{OLR})^{\frac{1}{4}} \tag{3.27}$$

where T_{skin} is the skin temperature in the absence of Solar absorption. The formula shows that Solar absorption always increases the temperature of the skin layer. The temperature increases as the ratio of shortwave absorption to infrared emissivity is made larger. So long as the temperature remains less than the Solar blackbody temperature, the system does not violate the Second Law of Thermodynamics, since the radiative transfer is still acting to close the gap between the cold atmospheric temperature and the hot Solar temperature. As the atmospheric temperature approaches that of the Sun, however, it would no longer be appropriate to use the infrared emissivity, since the atmosphere would then be radiating in the shortwave range. Kirchoff's Law would come into play, requiring $a/e = 1$. This would prevent the atmospheric temperature from approaching the photospheric temperature.

If the shortwave absorptivity is small, the skin layer can be divided into any number of sublayers, and the argument applies to determine the temperature of each one individually. This is so because the small absorptivity of the upper layers do not take much away from the Solar beam feeding absorption in the lower layers. We can then infer that the temperature of an absorbing stratosphere will increase with height if the absorption increases with height, making a_{sw}/e_{ir} increase with height.

Armed with our new understanding of the optically thin outer portions of planetary atmospheres, let's take another look at a few soundings. The skin temperature, defined in Eq 3.24, provides a point of reference. It is shown for selected planets in Table 3.3. Except for the Martian case, these values were computed from the global mean OLR, either observed directly (for Jupiter) or inferred from the absorbed Solar radiation. In the Martian case, the fast thermal response of the atmosphere and ocean makes the global mean irrelevant. Hence, assuming the atmosphere to be optically thin, we compute the skin temperature based on the upwelling infrared from a typical daytime and night-time surface temperature corresponding to the Martian soundings of Figure 2.2. The tropical Earth atmosphere sounding shown in Fig. 2.1 shows that the temperature increases sharply with height above the tropopause. This suggests that solar absorption is important in the Earth's stratosphere. For Earth, the requisite solar absorption is provided by ozone, which strongly absorbs Solar ultraviolet. This is the famous "ozone layer," which shields life on the surface from the sterilizing effects of deadly Solar ultraviolet rays. However, it is striking and puzzling that virtually the entire stratosphere is substantially colder than the skin temperature based on the global mean radiation budget. The minimum temperature in the sounding is $188K$, which is fully $26K$ below the skin temperature. If anything, one might have expected the tropical temperatures to exceed the global mean skin temperatures, because the local tropospheric temperatures are warmer

	Skin temperature
Venus	213.
Earth	214.
Mars (220K sfc)	185.
Mars (180K sfc)	151.
Jupiter	106.
Titan	72.

Table 3.3: Computed skin temperatures of selected planets.

than the global mean. A reasonable conjecture about what is going on is that high, thick tropical clouds reduce the local OLR, thus reducing the skin temperature. However, the measured tropical OLR In Fig. 3.7 shows that at best clouds reduce the tropical OLR to $240W/m^2$, which yields the same $214K$ skin temperature computed from the global mean budget. Apart from possible effects of dynamical heat transports, the only way the temperature can fall below the skin temperature is if the infrared emissivity becomes greater than the infrared absorptivity. This is possible, without violating Kirchoff's law, if the spectrum of upwelling infrared is significantly different from the spectrum of infrared emitted by the skin layer. We will explore this possibility in the next chapter.

Referring to Fig. 2.2 we see that the temperature of the Martian upper atmosphere declines steadily with height, unlike Earth; this is consistent with Mars' CO_2 atmosphere, which has only relatively weak absorption in the Solar near infrared spectrum. The Martian upper atmosphere presents the same quandary as Earth's though, in that the temperatures fall well below the skin temperature estimates. Just above the top of the Venusian troposphere, there is an isothermal layer with temperature $232K$, just slightly higher than the computed skin temperature. However, at higher altitudes, the temperature falls well below the skin temperature, as for Mars.

Between $500mb$ and $100mb$, just above Titan's troposphere, Titan has an isothermal layer with temperature of $75K$, which is very close to the skin temperature. Above $100mb$, the atmosphere warms markedly with height, reaching $160K$ at $10mb$. The solar absorption in Titan's stratosphere is provided mostly by organic haze clouds. Jupiter, like Titan, has an isothermal layer just above the troposphere, whose temperature is very close to the skin temperature. Jupiter's atmosphere also shows warming with height; its upper atmosphere becomes nearly isothermal at $150K$, which is $44K$ warmer than the skin temperature. This indicates the presence of solar absorbers in Jupiter's atmosphere as well, though the solar absorption is evidently more uniformly spread over height on Jupiter than it is on Earth or Titan.

We have been using the term "stratosphere" rather loosely, without having attempted a precise definition. It is commonly said, drawing on experience with Earth's atmosphere, that a stratosphere is an atmosphereic layer within which temperature increases with height. This would be an overly restrictive and Earth-centric definition. The dynamically important thing about a stratosphere is that it is much more stably stratified than the troposphere, i.e. that its temperature goes down less steeply than the adiabat appropriate to the planet under consideration. The stable stratification of a layer indicates that convection and other dynamical stirring mechanisms are ineffective or absent in that layer, since otherwise the potential temperature would become well mixed and the temperature profile would become adiabatic. An isothermal layer is stably stratified, because its potential temperature increases with height; even a layer like that of Mars' upper atmosphere, whose temperature decreases gently with height, can be stably stratified. We have shown that an optically thin stratosphere is isothermal in the absence of solar absorption. Indeed, this is often taken as a back-of-the envelope model of stratospheres in general, in simple calculations. In the next chapter, we will determine the temperature profile of stratospheres that

are not optically thin.

In a region that is well mixed in the vertical, for example by convection, will have a temperature that decreases with height. Dynamically speaking, such a mixed layer constitutes the troposphere. By contrast the stratosphere may be defined as the layer above this, within which vertical mixing plays a much reduced role. Note, however, that the temperature minimum in a profile need not be coincident with the maximum height reached by convection; will revisit this matter in Chapter 7. Yet a further complication is that, in midlatitudes, large scale winds associated with storms are probably more important than convection as the stirring mechanism establishing the tropopause.

We conclude this chapter with a few comparisons of observed tropopause heights with the predictions of the optically thin limit. We'll leave Venus out of this comparison, since its atmosphere is about as far from the optically thin limit as one could get. On Mars, using the dry adiabat for CO_2 and a $6mb$ surface pressure puts the tropopause at $2.8mb$, which is consistent with the top of the region of steep temperature decline seen in the daytime Martian sounding in Fig. 2.2. For Titan, we use the dry adiabat for N_2 and predict that the tropopause should be at $816mb$,which is again consistent with the sounding. If we use the methane/nitrogen moist adiabat instead of the nitrogen dry adiabat, we put the tropopause distinctly higher, at about $440mb$. Because the moist adiabatic temperature decreases less rapidly with height than the dry adiabat, one must go to greater elevations to hit the skin temperature (as in Fig. 3.14). The tropopause height based on the saturated moist adiabat is distinctly higher than seems compatible with the sounding, from which we infer that the low levels of Titan must be undersaturated with respect to Methane. Using $R/c_p = \frac{2}{7}$ for Earth air and $1000mb$ for the surface pressure, we find that the Earth's tropopause would be at $545mb$ in the optically thin, dry limit. This is somewhat higher in pressure (lower in altitude) that the actual midlatitude tropopause, and very much higher in pressure than the tropical tropopause. Earth's real atmosphere is not optically thin, and the lapse rate is less steep than the dry adiabat owing to the effects of moisture. The effects of optical thickness will be treated in Chapter 7, but we can already estimate the effect of using the moist adiabat. Using the computation of the water-vapor/air moist adiabat described in Chapter 2, the tropopause rises to $157mb$, based on a typical tropical surface temperature of $300K$ and the skin temperature estimated in Table 3.3. This is much closer to the observed tropopause (defined as the temperature minimum in the sounding), with the remaining mismatch being accounted for by the fact that the minimum temperature is appreciably colder than the skin temperature.

Chapter 4

Radiative transfer in temperature-stratified atmospheres

4.1 Overview

Our objective in this chapter is to treat the computation of a planet's energy loss by infrared emission in sufficient detail that the energy loss can be quantitatively linked to the actual concentration of specific greenhouse gases in the atmosphere. Unlike the simple modelof the greenhouse effect described in the preceding chapter, the infrared radiation in a real atmosphere does not all come from a single level; rather, a bit of emission is contributed from each level (each having its own temperature), and a bit of this is absorbed at each intervening level of the atmosphere. The radiation comes out in all directions, and the rate of emission and absorption is strongly dependent on frequency. Dealing with all these complexities may seem daunting, but in fact it can all be boiled down to a conceptually simple set of equations which suffice for a vast range of problems in planetary climate.

It was shown in Chapter 3 that there is almost invariably an order of magnitude separation in wavelengths between the shortwave spectrum at which a planet receives stellar radiation and the longwave (generally infrared) spectrum at which energy is radiated to space. This is true throughout the Solar system, for cold bodies like Titan and hot bodies like Venus, as well as for bodies like Earth that are habitable for creatures like ourselves. The separation calls for distinct sets of approximations in dealing with the two kinds of radiation. Infrared is both absorbed and emitted by an atmosphere, at typical planetary temperatures. However, the long infrared wavelengths are not appreciably scattered by molecules or water clouds, so scattering can be neglected in many circumstances. One of the particular challenges of infrared radiative transfer is the intricate dependence of absorption and emission on wavelength. The character of this dependence is linked to the quantum transitions in molecules whose energy corresponds to infrared photons; it requires an infrared-specific description.

In contrast, planets do not emit significant amounts of radiation in the shortwave spectrum, though shortwave scattering by molecules and clouds is invariably significant; absorption of shortwave radiation arises from quite different molecular processes than those involved in infrared

absorption, and its wavelength-dependence has a correspondingly different character. Moreover, solar radiation generally reaches the planet in the form of a nearly parallel beam, whereas infrared from thermal emission by the planet and it's atmosphere is more nearly isotropic. The approximations pertinent to shortwave radiation will be taken up in Chapter 5, where we will also consider the effects of scattering on thermal infrared.

We'll begin with a general formulation of the equations of plane-parallel radiative transfer without scattering, in Section 4.2. Though we will be able to derive certain general properties of the solutions of these equations, the equations are not very useful in themselves because of the problem of wavelength dependence. To gain further insight, a detailed examination of an idealized model with wavelength-independent infrared emissivity will be presented in Section 4.3. A characterization of the wavelength dependence of the absorption of real gases, and methods for dealing with that dependence, will be given in Sections 4.4 and 4.5.

4.2 Basic Formulation of Plane Parallel Radiative Transfer

We will suppose that the properties of the radiation field and the properties of the medium through which it travels are functions of a single coordinate, which we will take to be the pressure in a hydrostatically balanced atmosphere. (Recall that in such an atmosphere there is a one to one correspondence between pressure and altitude). This is the *plane-parallel* assumption. Although the properties of planetary atmospheres vary geographically with horizontal position within the spherical shell making up the atmosphere, in most cases it suffices to divide up the sphere into patches of atmosphere which are much larger in the horizontal than they are deep, and over which the properties can be considered horizontally uniform. In this case, vertical radiative transfer is much more important than horizontal transfer, and the atmosphere can be divided up into a large number of columns that act independently, insofar as radiative transfer is concerned.

In this section, we will develop an approximate form of the equations of plane parallel radiative transfer. The errors introduced in this approximation are small enough that the resulting equations are sufficiently accurate to form basis of the infrared radiative transfer component of virtually all large scale climate models. These equations will certainly be good enough for addressing the broad-brush climate questions that are our principal concern.

4.2.1 Optical thickness and the Schwarzschild equations

Although the radiation field varies in space only as a function of pressure, p, its intensity depends also on direction. Let $I(p, \hat{n}, \nu)$ be the flux density of electromagnetic radiation propagating in direction $\hat{n}$, measured at point p. This density is just like the Planck function $B(\nu, T)$, except that we allow it to depend on direction and position. Now we suppose that the radiation propagates through a thin layer of atmosphere of thickness Δp as measured by pressure. The absorption of energy at frequency ν is proportional to the number of molecules of absorber encountered; assuming the mixing ratio of the absorber to be constant within the layer for small Δp, the number of molecules encountered will be proportional to Δp, in accord with the hydrostatic law. By Kirchoff's law, the absorptivity and emissivity of the layer are the same; we'll call the value $\Delta\tau$, and keep in mind that in general it will be a function of ν. Let θ be the angle between the direction of propagation $\hat{n}$ and the vertical, as shown in Figure 4.1. Now, let $\Delta\tau^*$ be the emissivity (and absorptivity) of the layer for radiation propagating in the direction $\theta = 0$. We may define the proportionality between emissivity and pressure through the relation $\Delta\tau^* = -\kappa\Delta p/g$ where g

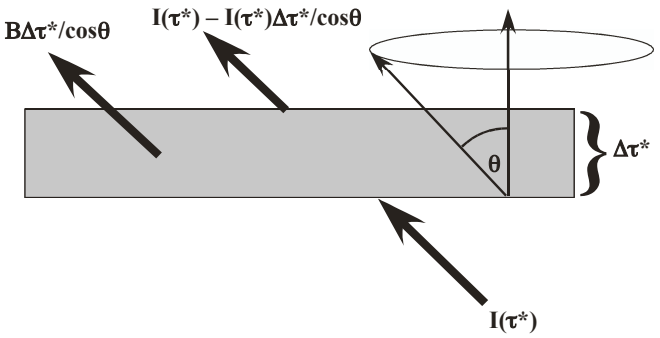

Figure 4.1: Sketch of the radiative energy balance for a slab of atmosphere illuminated by incident radiation from below.

is the acceleration of gravity and κ is an absorption coefficient. It has units of area per unit mass, and can be thought of as an absorption cross-section per unit mass – in essence, the area taken out of the incident beam by the absorbers contained in a unit mass of atmosphere. In general κ is a function of frequency, pressure, temperature and the mixing ratios of the various greenhouse gases in the atmosphere. Passing to the limit of small Δp, we can define an *optical thickness* coordinate through the differential equation

$$\frac{d\tau^*}{dp} = -\frac{1}{g}\kappa \tag{4.1}$$

Since pressure decreases with altitude, τ^* increases with altitude. Radiation propagating at an angle θ relative to the vertical acts just like vertically propagating radiation, except that the thickness of each layer through which the beam propagates, and hence the number of absorbing molecules encountered, is increased by a factor of $1/\cos\theta$. Hence, the optical thickness for radiation propagating with angle θ is simply $\tau = \tau^*/\cos\theta$. From now on, we will use τ^* in place of p as the vertical coordinate, and only transform back to pressure coordinates when necessary.

The specific absorption cross section κ depends on the number of molecules of each greenhouse gas encountered by the beam and the absorption properties characteristic to each kind of greenhouse gas molecule. Letting q_i be the mass-specific concentration of greenhouse gas i, we may write

$$\kappa(\nu, p, T) = \sum_{i=0}^{n} \kappa_i(\nu, p, T) q_i(p) \tag{4.2}$$

The specific concentrations q_i depend on p because we are using pressure as the vertical coordinate, and the concentration of the gas may vary with height;a well-mixed greenhouse gas would have constant q_i. The dependence of the coefficients κ_i on p and T arises from certain aspects of the physics of molecular absorption, to be discussed in Section 4.4.

Consider now the situation illustrated in Fig. 4.1, in which radiation at a given frequency and angle is incident on slab of atmosphere from below. In general, part of the incident radiation is scattered into other directions. However, for infrared and longer wave radiation interacting with gases, such scattering is negligible; scattering is also negligible for infrared interacting with condensed cloud particles made of substances such of water, which are strong absorbers. Here, we shall neglect scattering, though it will be brought back into the picture in Chapter 5. The radiation at the same angle which comes out the top of the slab is then the incident flux minus

the small amount absorbed in the slab, plus the small amount emitted. Thus

$$I(\tau^* + \Delta\tau^*, \hat{n}, \nu) = (1 - \frac{\Delta\tau^*}{\cos\theta})I(\tau^*, \hat{n}, \nu) + B(\nu, T(\tau^*))\frac{\Delta\tau^*}{\cos\theta} \tag{4.3}$$

or, passing to the limit of small $\Delta\tau^*$,

$$\frac{d}{d\tau^*}I(\tau^*, \hat{n}, \nu) = -\frac{1}{\cos\theta}[I(\tau^*, \hat{n}, \nu) - B(\nu, T(\tau^*))] \tag{4.4}$$

For a precise solution, one needs to solve this equation separately for each θ and then integrate over angles to get the net upward and downward fluxes. The angular distribution of radiation changes with distance from the source, since radiation propagating near the direction $\theta = 0$ or $\theta = \pi$ decays more gradually than radiation with θ nearer to $\pi/2$. Hence, radiation that starts out isotropic at the source (as is the blackbody emission) tends to become more forward-peaked as it propagates. For some specialized problems, it is indeed necessary to solve for the angular distribution explicitly in this fashion, which is rather computationally demanding. Fortunately, the isotropy of the blackbody source term tends to keep longwave radiation isotropic enough to allow one to make do with a much more economical approximate set of equations.

We can derive an equation for the net upward flux per unit frequency, I_+, by multiplying Eq. 4.4 by $\cos\theta$ and integrating over all solid angles in the upward-pointing hemisphere. Integrating over the downward hemisphere yields the net downward flux I_-. However, because of the factor $1/\cos\theta$ on the right hand side of Eq. 4.4, the hemispherically averaged intensity appearing on the right hand side is not I_+. Instead, it is $\int I(\tau^*, \hat{n}, \nu)d\Omega$, or equivalently $\int_0^{\pi/2} 2\pi I(\tau^*, \theta, \nu)\sin\theta d\theta$. One cannot proceed further without some assumption about the angular distribution. If we assume that the distribution remains approximately isotropic, by virtue of the isotropic source B, then $I(\tau^*, \theta, \nu)$ is independent of θ, and hence the problematic integral becomes $2\pi I\int_0^{\pi/2}\sin\theta d\theta$ which is equal to $2I_+$ under the assumption of isotropy. This result yields a closed equation for I_+. It states that, if the radiation field remains approximately isotropic, the decay rate is the same as for unidirectional radiation propagating with an angle $\bar{\theta}$ such that $\cos\bar{\theta} = \frac{1}{2}$, i.e. $\bar{\theta} = 60^o$. We will absorb this factor of $1/\cos\bar{\theta}$ into a redefined optical depth $\tau_\nu = \tau^*/\cos\bar{\theta}$. In some cases, one can do better than the pure isotropy assumption by choosing $\bar{\theta}$ to optimize the fit to an angularly resolved calculation, but that is a refinement we shall not pursue.

In terms of the redefined τ_ν, the equations for the upward and downward flux are

$$\begin{aligned}\frac{d}{d\tau_\nu}I_+ &= -I_+ + \pi B(\nu, T(\tau_\nu)) \\ \frac{d}{d\tau_\nu}I_- &= I_- - \pi B(\nu, T(\tau_\nu))\end{aligned} \tag{4.5}$$

These are known as the *two-stream equations*, and will serve as the basis for all subsequent discussion of radiative transfer in this book, save that we will incorporate the neglected scattering term in Chapter 5.

Because of the neglect of scattering, the equations for I_+ and I_- are uncoupled, and each consists of a linear, inhomogeneous first order differential equation. The solution can be obtained by substituting $I_+ = A(\tau_\nu)\exp(-\tau_\nu)$, and similarly for I_-, which reduces the problem to evaluation of a definite integral for A. The result is

$$\begin{aligned}I_+(\tau_\nu, \nu) &= I_+(0)e^{-\tau_\nu} + \int_0^{\tau_\nu} \pi B(\nu, T(\tau'_\nu))e^{-(\tau_\nu - \tau'_\nu)}d\tau'_\nu \\ I_-(\tau_\nu, \nu) &= I_-(\tau_\infty)e^{-(\tau_\infty - \tau_\nu)} + \int_{\tau_\nu}^{\tau_\infty} \pi B(\nu, T(\tau'_\nu))e^{-(\tau'_\nu - \tau_\nu)}d\tau'_\nu\end{aligned} \tag{4.6}$$

where τ'_ν is a dummy variable. Note that τ_∞ depends on ν in general, though we have suppressed the subscript for the sake of readability. The physical content of these equations is simple: $I_+(\tau_\nu, \nu)$ consists of two parts. The first is the portion of the emission from the ground which is transmitted by the atmosphere (the first term in the expression for I_+). The second is the radiation emitted by the atmosphere itself, which appears as an exponentially-weighted average (the second term in the expression for I_+) of the emission from all layers below τ_ν, with more distant layers given progressively smaller weights. Similarly, $I_-(\tau_\nu, \nu)$ is an exponentially-weighted average of the emission from all layers above τ_ν, plus the transmission of incident downward flux. The atmospheric emission to space will be most sensitive to temperatures near the top of the atmosphere. This emission will dominate the OLR when the atmosphere is fairly opaque to the radiation emitted from the ground, whereas the transmitted ground emission will dominate when the atmosphere is fairly transparent. The downward radiation into the ground will be most sensitive to temperatures nearest the ground.

In the long run, it will save us some confusion if we introduce special notation for temperatures and fluxes at the boundaries; this will prove especially important when there is occasion to switch back and forth between pressure and optical thickness as a vertical coordinate. The temperature at the top of the atmosphere ($p = 0$ or $\tau = \tau_\infty$) will be denoted by T_∞,and the temperature of the air at the bottom of the atmosphere ($p = p_s$ or $\tau = 0$) will be called T_{sa}. For planets with a solid or liquid surface this is the temperature of the gas in immediate contact with the surface. For such planets, one must distinguish the temperature of the air from the temperature of the surface (the "ground") itself, which will be called T_g. The outgoing and incoming fluxes at the top of the atmosphere will be called $I_{+,\infty}(\nu)$ and $I_{-,\infty}(\nu)$ respectively, while the upward and downward fluxes at the bottom of the atmosphere will be called $I_{+,s}(\nu)$ and $I_{-,s}(\nu)$

For planets with a liquid or solid surface, we require that $I_{+,s}(\nu)$ be equal to the upward flux emitted by the ground, which is $e(\nu)B(\nu, T_g)$, where $e(\nu)$ is the emissivity of the ground. Continuity of the fluxes is required because, the air being in immediate contact with the ground, there is no medium between the two which could absorb or emit radiation, nor is there any space where radiation "in transit" could temporarily reside. We generally assume that there is no infrared radiation incident on the top of the atmosphere, so that the upper boundary condition is $I_{-,\infty} = 0$. The incident solar radiation does contain some near-infrared, but this is usually treated separately as part of the shortwave radiation calculation (see Chapter 5). For planets orbiting stars with cool photospheres, such as red giant stars, it might make sense to allow $I_{-,\infty}$ to be nonzero and treat the incoming infrared simultaneously with the internally generated thermal infrared. Since the radiative transfer equations are linear in the intensities, it is a matter of taste whether to treat the incoming stellar infrared in this way, or as part of the calculation dealing with the shorter wave part of the incoming stellar spectrum.

The weighting function appearing in the integrands in Eq. 4.6 is the *transmission function*. Written as a function of pressure, it is

$$\mathfrak{T}_\nu(p_1, p_2) = e^{-|\tau_\nu(p_1) - \tau_\nu(p_2)|} \tag{4.7}$$

$\mathfrak{T}_\nu(p_1, p_2)$ is the proportion of incident energy flux at frequency ν which is transmitted through a layer of atmosphere extending from p_1 to p_2; whatever is not transmitted is absorbed in the layer. Note that $\mathfrak{T}_\nu(p, p')d\tau'_\nu = d\mathfrak{T}_\nu$ (with p held constant), if $p < p'$, and $\mathfrak{T}_\nu(p, p')d\tau'_\nu = -d\mathfrak{T}_\nu$ if $p > p'$. Using this result Eq. 4.6 can be re-written

$$\begin{aligned} I_+(\tau_\nu, \nu) &= I_{+,s}(\nu)\mathfrak{T}_\nu(p, p_s) - \int_{p'=p}^{p_s} \pi B(\nu, T(p'))d\mathfrak{T}_\nu(p, p') \\ I_-(\tau_\nu, \nu) &= I_{-,\infty}(\nu)\mathfrak{T}_\nu(0, p) + \int_{p'=0}^{p} \pi B(\nu, T(p'))d\mathfrak{T}_\nu(p, p') \end{aligned} \tag{4.8}$$

In the integrals above, the differential of $\mathfrak{T}_\nu$ is meant to be taken with p held fixed. Integration by parts then yields the following alternate form of the solution to the two-stream equations:

$$\begin{aligned} I_+(p,\nu) &= \pi B(\nu,T(p)) + (I_{+,s}(\nu) - \pi B(\nu,T_{sa}))\mathfrak{T}_\nu(p,p_s) + \int_p^{p_s} \pi \mathfrak{T}_\nu(p,p')dB(\nu,T(p')) \\ I_-(p,\nu) &= \pi B(\nu,T(p)) + (I_{-,\infty}(\nu) - \pi B(\nu,T_\infty))\mathfrak{T}_\nu(0,p) - \int_0^{p} \pi \mathfrak{T}_\nu(p,p')dB(\nu,T(p')) \end{aligned} \tag{4.9}$$

Neither of these forms of the solution is particularly convenient for analytic work, but either one can be used to good advantage when carrying out approximate integrations via the trapezoidal rule (see Section 4.4.5). For analytical work, and some kinds of numerical integration, it helps to rewrite the integrand using $dB = (dB/dT)(dT/dp')dp'$. The result is In rewriting the integrand, we have used $dB = (dB/dT)(dT/d\tau'_\nu)d\tau'_\nu = (dB/dT)(dT/dp')dp'$

$$\begin{aligned} I_+(p,\nu) &= \pi B(\nu,T(p)) + (I_{+,s}(\nu) - \pi B(\nu,T_{sa}))\mathfrak{T}_\nu(p,p_s) + \int_p^{p_s} \pi \mathfrak{T}_\nu(p,p')\frac{dB}{dT}|_{T(p')}\frac{dT}{dp'}dp' \\ I_-(p,\nu) &= \pi B(\nu,T(p)) + (I_{-,\infty}(\nu) - \pi B(\nu,T_\infty))\mathfrak{T}_\nu(0,p) - \int_0^{p} \pi \mathfrak{T}_\nu(p,p')\frac{dB}{dT}|_{T(p')}\frac{dT}{dp'}dp' \end{aligned} \tag{4.10}$$

A considerable advantage of any of the forms in Eq. 4.8, 4.9 or 4.10 is that the integration variable p' is no longer dependent on frequency. This will prove particularly useful when we come to consider real gases, for which the optical thickness has an intricate dependence on frequency. The first two terms in the expression for the fluxes in either Eq. 4.9 or 4.10 give the exact result for an isothermal atmosphere; in each case, the first of the two terms represents the contribution of the local blackbody radiation, whereas the second accounts for the modifying effect of the boundaries. The boundary terms vanish at points far from the boundary, where $\mathfrak{T}$ is small. Note that the boundary term for I_+ vanishes identically if the upward flux at the boundary has the form of blackbody radiation with temperature equal to the surface air temperature. For a planet with a solid or liquid surface, this would be the case if the ground temperature equals the surface air temperature and the ground has unit emissivity.

The main reason for dealing with radiative transfer in the atmosphere is that one needs to know the amount of energy deposited in or withdrawn from a layer of atmosphere by radiation. This is the radiative heating rate (with negative heating representing a cooling). It is obtained by taking the derivative of the net flux, which gives the difference between the energy entering and leaving a thin layer. The heating rate per unit optical thickness, per unit frequency, is thus

$$\mathfrak{H}_\nu = -\frac{d}{d\tau_\nu}(I_+(\tau_\nu,\nu) - I_-(\tau_\nu,\nu)) \tag{4.11}$$

This must be integrated over all frequencies to yield the net heating rate. For making inferences about climate, one ordinarily requires the heating rate per unit mass rather than the heating rate per unit optical depth. This is easily obtained using the definition of optical depth, specifically,

$$H_\nu = g\frac{d}{dp}(I_+ - I_-) = g\frac{d\tau_\nu}{dp}\frac{d}{d\tau_\nu}(I_+ - I_-) = \frac{\kappa}{\cos\bar{\theta}}\mathfrak{H}_\nu \tag{4.12}$$

When integrated over frequency this heating rate has units W/kg. One can convert into a temperature tendency K/s by dividing this value by the specific heat c_p.

4.2.2 Some special solutions of the Two-Stream equations

Beer's law

Suppose that the atmosphere is too cold to radiate significantly at the frequency under consideration. In that case, $B(\nu, T) \approx 0$ and the internal source vanishes. This would be the case, for example, if ν is in the visible light range and the temperature of the atmosphere is Earthlike. In this case, the solutions are simply $I_+ = I_+(0)\exp(-\tau_\nu)$ and $I_- = I_-(\tau_\infty)exp(\tau_\nu - \tau_\infty)$. The exponential attenuation of radiation is known as *Beer's Law*. Here we've neglected scattering, but in Chapter 5 we'll see that a form of Beer's law still applies even if scattering is taken into account.

Infinite isothermal medium

Consider next an unbounded isothermal medium. In this case, it is readily verified that $I_+ = I_- = \pi B(\nu, T)$ is an exact solution to 4.6. The right hand sides of the equations vanish, but the derivatives on the left hand sides vanish also, because T is independent of τ_ν. Hence, in an unbounded isothermal medium, the radiation field reduces to uniform blackbody radiation.

Since the fluxes are independent of τ_ν, the radiative heating rate vanishes, from which we recover the fact that blackbody radiation is in equilibrium with an extended body of isothermal matter.

Exercise 4.2.1 Derive this result from Eq. 4.6; from Eq. 4.10.

Finite-thickness isothermal slab

Now let's consider an isothermal layer of finite thickness, embedded in an atmosphere which is completely transparent to radiation at frequency ν. We suppose further that there is no radiation at this frequency incident on the layer from either above or below. We are free to define $\tau_\nu = 0$ at the center of the layer, so that $\tau_\nu = \frac{1}{2}\tau_\infty$ at the top of the layer and $\tau_\nu = -\frac{1}{2}\tau_\infty$ at the bottom. The boundary conditions corresponding to no incident flux are $I_- = 0$ at the top of the layer and $I_+ = 0$ at the bottom of the layer. The solution $I_+ = I_- = \pi B$ is still a *particular* solution within the layer, since the layer is isothermal, but it does not satisfy the boundary conditions. A homogeneous solution must be added to each flux in order to satisfy the boundary conditions. The homogeneous solutions are just exponentials, and so we easily find that the full solution within the layer is

$$\begin{aligned} I_+(\tau_\nu, \nu) &= [1 - \exp(-(\tau + \frac{\tau_\infty}{2}))]\pi B \\ I_-(\tau_\nu, \nu) &= [1 - \exp(+(\tau - \frac{\tau_\infty}{2}))]\pi B \end{aligned} \tag{4.13}$$

Exercise 4.2.2 Derive this result from Eq. 4.10.

The radiation emitted out the top of the layer is $I_+(\tau_\infty, \nu)$, or $(1 - \exp(-\tau_\infty))\pi B$, which reduces to the blackbody value πB when the layer is optically thick for the frequency in question, i.e. $\tau_\infty(\nu) \gg 1$. The same applies for the emission out of the bottom of the layer, *mutatis mutandum*. Note that in the optically thick limit, $I_+ = I_- = \pi B$ through most of the layer, and

the inward-directed intensities only fall to zero in the two relatively thin skin layers near the top and bottom of the slab.

In the opposite extreme, when the slab is optically thin, both τ and τ_∞ are small. Using the first order Taylor expansion of the exponentials, we find that the emission out the top of the layer is $\tau_\infty \pi B$, and similarly for the bottom of the layer. Hence, τ_∞ in this case is just the bulk emissivity of the layer. This is consistent with the way we constructed the Schwarzschild equations, which can be viewed as a matter of stacking a great number of individually optically thin slabs upon each other.

Substituting into Eq. 4.11, the heating rate for this solution is

$$\mathfrak{H}_\nu = -[\exp(-\tau) + \exp(\tau)] \exp(-\frac{\tau_\infty}{2}) \pi B \tag{4.14}$$

In the optically thick case, the heating rate is nearly zero in the interior of the slab, but there is strong radiative cooling within about a unit optical depth of each surface. In this case the radiation drains heat out of a thin skin layer near each surface, causing intense cooling there. In the optically thin limit, the cooling is distributed uniformly throughout the slab.

It turns out that condensed water is a much better infrared absorber than the same mass of water vapor. Hence, an isolated absorbing layer such as we have just considered can be thought of as a very idealized model of a cloud. The following slight extension makes the connection with low lying stratus clouds, such as commonly found over the oceans, more apparent.

Exercise 4.2.3 Instead of being suspended in an infinite transparent medium, suppose that the cloud is in contact with the ground, and that the ground has the same temperature as the cloud. We still assume that the air above the cloud is transparent to radiation at the frequency under consideration. Compute the upward and downward fluxes, and the radiative heating rate, in this case.

This exercise shows that convection in boundary layer stratus clouds can be driven by cooling at the top, rather than heating from below. This is rather important, since the reflection of sunlight by the cloud makes it hard to warm up the surface. Entrainment of dry air due to top-driven convection is one of the main mechanisms for dissipating such clouds.

Optically thick limit

We now depart from the assumption of constant temperature. While allowing T to vary in the vertical, we assume the atmosphere to be *optically thick* at frequency ν. This means that a small change in pressure p amounts to a large change in the optical thickness coordinate τ_ν. Referring, to Eq. 4.1, we see that the assumption of optical thickness is equivalent to the assumption that $\kappa \delta p / g \gg 1$, where δp is the typical amount by which one has to change the pressure in order for the temperature to change by an amount comparable to its mean value. For most atmospheres, it suffices to take δp to be the depth of the whole atmosphere, namely p_s, so that the optical thickness assumption becomes $\tau_\infty = \kappa p_s / g \gg 1$ Since κ depends on frequency, an atmosphere may be optically thick near one frequency, but optically thin near another.

The approximate form of the fluxes in the optically thick limit can be most easily derived from the integral expression in the form given in Eq. 4.6. Consider first the expression for I_+. Away from the immediate vicinity of the bottom boundary, the boundary term proportional to $\mathfrak{T}_\nu(p, p_s)$ is exponentially small and can be dropped. To simplify the integral, we note that $\mathfrak{T}_\nu(p, p')$

is very small unless p' is close to p. Therefore, as long as the temperature gradient is a continuous function of p', it varies little over the range of p' for which the integrand contributes significantly to the integral. Hence dT/dp' can be replaced by its value at p, which can then be taken outside the integral. Likewise, dB/dT can be evaluated at $T(p)$, so that this term can also be taken outside the integral. Finally, if one is not too close to the bottom boundary,

$$\int_p^{p_s} \mathfrak{T}_\nu(p', p)dp' \approx \int_p^{\infty} \mathfrak{T}_\nu(p', p)dp' = \frac{g\cos\bar{\theta}}{\kappa}\int_\tau^{\infty} \mathfrak{T}_\nu(\tau', \tau)d\tau' = \frac{g\cos\bar{\theta}}{\kappa}, \tag{4.15}$$

whence the upward flux in the optically thick limit becomes

$$I_+(\nu, p) = \pi B(\nu, T(p)) + \pi\frac{g\cos\bar{\theta}}{\kappa}\frac{dB}{dT}|_{T(p)}\frac{dT}{dp} \tag{4.16}$$

Near the bottom boundary, the neglected boundary term would have to be added to this expression. In addition, Eq. 4.15 would need to be corrected to allow for the fact that there is not room for $\int \mathfrak{T}$ to integrate out to its asymptotic value for an infinitely thick layer.

Using identical reasoning, the downward flux becomes

$$I_-(\nu, p) = \pi B(\nu, T(p)) - \pi\frac{g\cos\bar{\theta}}{\kappa}\frac{dB}{dT}|_{T(p)}\frac{dT}{dp} \tag{4.17}$$

so long as one is not too near the top of the atmosphere. Near the top of the atmosphere, the neglected boundary term becomes significant.

In both expressions the second term, proportional to the temperature gradient, becomes progressively smaller as κ is made larger and the atmosphere becomes more optically thick. To lowest order, then, the upward and downward fluxes are both equal to the blackbody radiation flux at the local temperature. In this sense, the optically thick limit looks "locally isothermal." The term proportional to the temperature gradient represents a small correction to the locally isothermal behavior. In the expression for I_+, for example, if $dT/dp > 0$ the correction term makes the upward flux somewhat greater than the local blackbody value. This makes sense, because a small portion of the upwelling radiation comes from lower layers where the temperature is warmer than the local temperature. Note that the correction term depends on ν through the frequency dependence of κ, as well as through the frequency dependence of dB/dT.

The radiation exiting the top of the atmosphere ($I_{+,\infty}$) is of particular interest, because it determines the rate at which the planet loses energy. In the optically thick approximation, we find that as long as dT/dp is finite at $p = 0$, $I_{+,\infty}$ becomes close to $\pi B(\nu, T_\infty)$ as the atmosphere is made more optically thick. Hence, at frequencies where the atmosphere is optically thick, the planet radiates to space like a blackbody with temperature equal to that of the upper regions of the atmosphere – the regions "closest" to outer space.

Similarly, the downward radiation ($I_{-,s}(\nu)$) from the atmosphere into the ground – sometimes called the *back radiation* – is of interest because it characterizes the radiative effect of the atmosphere on the surface energy budget. In the optically thick limit, $I_{-,s}(\nu) = \pi B(\nu, T_{sa})$ to lowest order, so that the atmosphere radiates to the ground like a blackbody with temperature equal to the low level air temperature. If $dT/dp > 0$ at the ground, as is typically the case, then the correction term slightly reduces the downward radiation, because some of the radiation into the ground comes from higher altitudes where the air is colder. Suppose now that the surface temperature T_g is equal to the air temperature T_{sa}, and that the surface has unit emissivity at the frequency under consideration. In that case, the net radiative heating of the ground is

$$I_{-,s}(\nu) - B(\nu, T_g) = I_{-,s}(\nu) - B(\nu, T_{sa}) = -\pi\frac{g\cos\bar{\theta}}{\kappa}\frac{dB}{dT}|_{T_{sa}}\frac{dT}{dp}|p_s \tag{4.18}$$

at frequencies where the solar flux is negligible. This is negative when $dT/dp > 0$, representing a radiative cooling of the ground. The radiative cooling vanishes in the limit of large κ. In the optically thick limit, then, the surface cannot get rid of heat by radiation unless the ground temperature becomes larger than the low level air temperature. Remember, though, that the radiative heating of the ground is but one term in the surface energy budget coupling the surface to the atmosphere. Turbulent fluxes of moisture and heat also exchange energy between the surface and the atmosphere, and these become dominant when the radiative term is weak.

In the optically thick limit, the net flux is

$$I_+ - I_- = 2\pi \frac{g \cos\bar{\theta}}{\kappa} \frac{dB}{dT} \frac{dT}{dp} \tag{4.19}$$

whence the radiative heating rate is

$$H_\nu = \frac{d}{dp}[D(\nu, p) \frac{dT}{dp}] \tag{4.20}$$

where

$$D(\nu, p) = 2\pi \frac{g^2 \cos\bar{\theta}}{\kappa} \frac{dB}{dT}|_{T(p)} \tag{4.21}$$

Hence, in the optically thick limit, the heating and cooling caused by radiative transfer acts just like a thermal diffusion in pressure coordinates, with the diffusivity given by $D(\nu, p)$. Since $dB/dT > 0$, the radiative diffusivity is always positive. It becomes weak as κ becomes large. Note that the diffusive approximation to the heating is is only valid when one is not too close to the top and bottom of the atmosphere. Near the boundaries, the neglected boundary terms contribute an additional heating which is exponentially trapped near the top and bottom of the atmosphere. The effect of the boundary terms is explored in Problem **??**

Consider an atmosphere which is transparent to solar radiation, and within which heat is redistributed only by infrared radiative transfer. Eq 4.12 then requires that the net upward flux $I_+ - I_-$ must be independent of altitude when integrated over all wavenumbers. This constant flux is nonzero, since the infrared flux through the system is set by the rate at which infrared escapes from the top of the atmosphere – namely, the OLR. Integrating Eq. 4.19 over the infrared yields an expression for dT/dp in terms of the OLR and the frequency-integrated diffusivity; because both OLR and diffusivity are positive, it follows that $dT/dp > 0$ for an optically thick atmosphere in pure infrared radiative equilibrium – that is, the temperature decreases with altitude. The more optically thick the atmosphere becomes, the smaller is D, and hence the stronger is the temperature variation in equilibrium. Pure radiative equilibrium will be discussed in detail in Sections 4.3.4 and 4.6, and the optically thick limit is explored in Problem **??**.

Optically thin limit

The *optically thin* limit is defined by $\tau_\infty \ll 1$. Since $\tau_\nu \leq \tau_\infty$ and $\tau'_\nu \leq \tau$ in Eq. 4.6, all the exponentials in the expression for the fluxes are close to unity. Moreover, the integral is carried out over the small interval $[0, \tau_\nu]$, and hence is already of order τ_∞ or less. It is thus a small correction to the first term, and we may set the exponentials in the integrand to unity and still have an expression that is accurate to order τ_∞. The boundary terms are not integrated, though, so we must retain the first two terms in the Taylor series expansion of the exponential to achieve

the same accuracy. With these approximations, the fluxes become

$$\begin{aligned} I_+(\tau_\nu) &= (1-\tau_\nu)I_{+,s} + \int_0^{\tau_\nu} \pi B(\nu, T(\tau'_\nu))d\tau'_\nu \\ I_-(\tau_\nu) &= (1-(\tau_\infty-\tau_\nu))I_{-,\infty} + \int_{\tau_\nu}^{\tau_\infty} \pi B(\nu, T(\tau'_\nu))d\tau'_\nu \end{aligned} \tag{4.22}$$

In this case, the upward flux is the sum of the upward flux from the boundary (diminished by the slight atmospheric absorption on the way up) with the sum of the unmodified blackbody emission from all the layers below the point in question. The downward flux is interpreted similarly.

In order to discuss the radiation escaping the top of the atmosphere and the back-radiation into the ground, we introduce the mean emission temperature $\bar{T}_\nu$, defined by solving the relation

$$B(\bar{T}_\nu, \nu) = \frac{1}{\tau_\infty}\int_0^{\tau_\infty} B(\nu, T(\tau'_\nu))d\tau'_\nu \tag{4.23}$$

With this definition, the boundary fluxes are

$$\begin{aligned} I_{+,\infty} &= (1-\tau_\infty)I_+(0) + \tau_\infty \pi B(\bar{T}_\nu, \nu) \\ I_{-,s} &= (1-\tau_\infty)I_{-,\infty}) + \tau_\infty \pi B(\bar{T}_\nu, \nu) \end{aligned} \tag{4.24}$$

According to this expression, an optically thin atmosphere acts precisely like an isothermal slab with temperature $\bar{T}_\nu$ and (small) emissivity τ_∞. It is only in the optically thin limit that the radiative effect of the atmosphere mimics that of an isothermal slab.

Substituting the approximate form of the fluxes into the expression for radiative heating rate, we find

$$H_\nu = \frac{\kappa}{\cos\bar{\theta}} \cdot [(I_{+,s} + I_{-,\infty}) - 2\pi B(\nu, T(p))] \tag{4.25}$$

This is small, because κ is small in the optically thin limit. The first pair of terms are always positive, and represent heating due to the proportion of incident fluxes which are absorbed in the atmosphere. The second term is always negative, and represents cooling by blackbody emission of the layer of air at pressure p. In contrast to the general case or the optically thick case, the cooling term is purely a function of the local temperature; radiation emitted by each layer escapes directly to space or to the ground, without being significantly captured and re-emitted back by any other layer.

Typical greenhouse gases are optically thin in some spectral regions and optically thick in others. We have seen that the infrared heating rate becomes small in both limits. From this result, we deduce the following general principle: *The infrared heating rate of an atmosphere is dominated by the spectral regions where the optical thickness is order unity.* If an atmosphere is optically thick throughout the spectrum, the heating is dominated by the least thick regions; if it is optically thin throughout, it is dominated by the least thin regions.

4.3 The Grey Gas Model

We will see in Section 4.4 that for most atmospheric gases κ, and hence the optical thickness, has an intricate dependence on wavenumber. This considerably complicates the solution of the radiative transfer equations, since the fluxes must be solved for individually on a very dense grid of wavenumbers, and then the results integrated to yield the net atmospheric heating, which is

the quantity of primary interest. The development of shortcuts that can improve on a brute-force integration is an involved business, which in some respects is as much art as science, and leads to equations whose behavior can be difficult to fathom. The radiative transfer equations become much simpler if the optical thickness is independent of wavenumber. This is known as the *grey gas approximation*. For grey gases, the Schwartzschild equations can be integrated over wavenumber, yielding a single differential equation for the net upward and downward flux. More specifically, we shall assume only that the optical thickness is independent of wavenumber within the infrared spectrum, and that the temperature of the planet and its atmosphere is such that essentially all the emission of radiation lies in the infrared. Instead of integrating over all wavenumbers, we integrate only over the infrared range, thus obtaining a set of equations for the net infrared flux. Because of the assumption regarding the emission spectrum, the integrals of the Planck function $\pi B(\nu, T)$ over the infrared range can be well approximated by σT^4.

With the exception of clouds of strongly absorbing condensed substances like water, the grey gas model yields a poor representation of radiative transfer in real atmospheres, for which the absorption is typically strongly dependent on wavenumber. Nonetheless, a thorough understanding of the grey gas model provides the starting point for any deeper inquiry into atmospheric radiation. Here, we can find many of the fundamental phenomena laid bare, because one can get much farther before resorting to detailed numerical computations. Further, grey gas radiation has proved valuable as a placeholder radiation scheme in theoretical studies involving the coupling of radiation to fluid dynamics, when one wants to focus on dynamical phenomena without the complexity and computational expense of real gas radiative transfer. Sometimes, a simple scheme which is easy to understand is better than an accurate scheme which defies comprehension.

The grey gas versions of the two-stream Schwarzschild equations are obtained by making τ_ν independent of frequency and integrating the resulting equations over all frequencies. The result is

$$\begin{aligned} \frac{d}{d\tau} I_+ &= -I_+ + \sigma T(\tau)^4 \\ \frac{d}{d\tau} I_- &= I_- - \sigma T(\tau)^4 \end{aligned} \tag{4.26}$$

Grey gas versions of the solutions given in the previous section can similarly be obtained by integrating the relations over all frequencies, taking into account that τ is now independent of ν. The expressions have precisely the same form as before, except that I_+ and I_- now represent total flux integrated over all longwave frequencies, and every occurence of πB is replaced by σT^4. To avoid unnecessary proliferation of notation, when the context allows little possibility of confusion we will use the same symbols I_+ and I_- to represent the longwave-integrated flux as we used earlier to represent the frequency-dependent flux spectrum. When we need to emphasize that a flux is a frequency dependent spectrum, we will include the dependence explicitly (as in "$I_+(\nu)$" or "$I_+(\nu, p)$"; when we need to emphasize that a flux represents the longwave-integrated net flux, we will use an overbar (as in $\bar{I}_+$).

4.3.1 OLR and back-radiation for an optically thin grey atmosphere

The OLR and surface back-radiation for an optically thin grey atmosphere are obtained by integrating Eq. 4.24 over all frequencies. The result is

$$\begin{aligned} I_{+,\infty} &= (1 - \tau_\infty) I_+(0) + \tau_\infty \sigma \bar{T}^4 \\ I_{-,s} &= (1 - \tau_\infty) I_{-,\infty}) + \tau_\infty \sigma \bar{T}^4 \end{aligned} \tag{4.27}$$

where the mean atmospheric emission temperature is given by

$$\bar{T}^4 = \frac{1}{\tau_\infty}\int_0^{\tau_\infty} T^4 d\tau \tag{4.28}$$

The first term in the expression for $I_{+,\infty}$ is the proportion of upward radiation from the ground which escapes without absorption by the intervening atmosphere, while the second is the emission to space added by the atmosphere. In the expression for $I_{-,s}$ the first term is the proportion of incoming infrared flux which reaches the surface without absorption, while the second is the downward emission from the atmosphere. Note that for an optically thin atmosphere the atmospheric emission to space is identical to the atmospheric emission to the ground; in this regard the atmosphere radiates like an isothermal slab with temperature $\bar{T}$. According to Eq. 4.24, a nongrey atmosphere behaves similarly, if it is optically thin *for all frequencies*.

4.3.2 Radiative properties of an all-troposphere dry atmosphere

Let's consider an atmosphere for which the convection is so deep that it establishes a dry adiabat throughout the depth of the atmosphere. Thus, $T(p) = T_s(p/p_s)^{R/c_p}$ all the way to $p = 0$. We wish to compute the OLR for this atmosphere, which is done by substituting this $T(p)$ into the grey-gas form of Eq. 4.6 and evaluating the integral for I_+ at $\tau = \tau_\infty$, i.e. the top of the atmosphere. Since the temperature is expressed as a function of pressure, it is necessary to substitute for pressure in terms of optical thickness in order to carry out the integral. We'll suppose that κ is a constant, so that $\tau_\infty - \tau = \kappa p/g = \tau_\infty p/p_s$. Using this to eliminate pressure from $T(p)$, the integral for OLR becomes

$$\begin{aligned} OLR &= I_{+,s}e^{-\tau_\infty} + \int_0^{\tau_\infty} \sigma T_s^4 (\frac{\tau_\infty - \tau'}{\tau_\infty})^{4R/c_p} e^{-(\tau_\infty - \tau')} d\tau' \\ &= I_{+,s}e^{-\tau_\infty} + \int_0^{\tau_\infty} \sigma T_s^4 (\frac{\tau_1}{\tau_\infty})^{4R/c_p} e^{-\tau_1} d\tau_1 \\ &= I_{+,s}e^{-\tau_\infty} + \sigma T_s^4 \tau_\infty^{-4R/c_p} \int_0^{\tau_\infty} \tau_1^{4R/c_p} e^{-\tau_1} d\tau_1 \end{aligned} \tag{4.29}$$

The second line is derived by introducing a new dummy variable $\tau_1 = \tau_\infty - \tau'$. This is the the optical depth measured relative to the top of the atmosphere, and the re-expressed integral is computed by integrating from the top down, rather than from the ground up. The first term on the right hand side of Eq. 4.29 represents the proportion of the upward surface radiation which survives absorption by the atmosphere and reaches space. The second term is the net emission from the atmosphere. In the optically thin limit, the integral becomes small and the exponential in the first term approaches unity; thus, the OLR approaches the emission from the ground, $I_{+,s}$. As the atmosphere is made more optically thick, the boundary term becomes exponentially small, and the integral becomes more and more dominated by the emission from the upper reaches of the atmosphere. However, to obtain the optically thick limit, we cannot use the grey gas form of Eq 4.16, since dT/dp becomes infinite at $p = 0$ when the dry adiabat extends all the way to the top of the atmosphere.

In the optically thick limit, $\tau_\infty \gg 1$, the first term becomes exponentially small and the upper limit of the integral can be replaced by ∞, yielding the expression

$$OLR = \sigma T_s^4 \tau_\infty^{-4R/c_p} \Gamma(1 + \frac{4R}{c_p}) \tag{4.30}$$

where Γ is the Gamma function, defined by $\Gamma(s) \equiv \int_0^\infty \zeta^{s-1} exp(-\zeta) d\zeta$. Using integration by parts, $\Gamma(s) = (s-1)\Gamma(s-1)$, while $\Gamma(1) = 1$, so $\Gamma(n) = (n-1)!$. For Earth air, $4R/c_p = \frac{8}{7}$ so $\Gamma(1+4R/c_p)$ will be close to $\Gamma(2)$, which is unity; in fact it is approximately 1.06. For any of the gases commonly found in planetary atmospheres, $\Gamma(1+4R/c_p)$ will be an order unity constant. As the atmosphere is made more optically thick, the OLR goes down algebraically like τ_∞^{-4R/c_p}, becoming much less than the value σT_s^4 prevailing for a transparent atmosphere. The OLR approaches zero as τ_∞ is made large because the temperature vanishes at the top of the atmosphere, and as the atmosphere is made more optically thick, the OLR is progressively more dominated by the emission from the cold upper reaches of the atmosphere.

The calculation can be related to the conceptual greenhouse effect model introduced in the previous chapter by computing the effective radiating pressure p_{rad}. Recall that $\sigma T_{rad}^4 = OLR$, so

$$\sigma T_s^4 (\frac{p_{rad}}{p_s})^{4R/c_p} = OLR = \sigma T_s^4 \tau_\infty^{-4R/c_p} \Gamma(1 + \frac{4R}{c_p}) \tag{4.31}$$

whence $(p_{rad}/p_s) = \tau_\infty^{-1}(\Gamma(1+4R/c_p))^{c_p/4R}$ This formula implies that the radiation to space comes essentially from the top unit optical depth of the atmosphere. If an atmosphere has optical depth $\tau_\infty = 100$, then it is only the layer between roughly the top of the atmosphere ($\tau = 100$) and $\tau = 99$ which dominates the OLR. For the all-troposphere model, the maximum temperature of the top unit optical depth approaches zero as the atmosphere is made more optically thick, because this entire layer corresponds to pressures approaching zero ever more closely as κ is made larger.

If S is the absorbed solar radiation per unit area of the planet's surface, then the surface temperature in balance with S is obtained by setting the OLR equal to S. Solving for the surface temperature, we find that in the optically thick all-troposphere limit, the surface temperature is

$$T_s = (S/\sigma)^{\frac{1}{4}} \Gamma(1 + \frac{4R}{c_p})^{-\frac{1}{4}} \cdot \tau_\infty^{R/c_p} \tag{4.32}$$

The first term is the temperature the planet would have in the absence of any atmosphere. As τ_∞ increases, the surface becomes warmer without bound. This constitutes our simplest quantitative model of the greenhouse effect for a temperature-stratified atmosphere. Note that the greenhouse warming depends on the lapse rate. For an isothermal atmosphere ($R/c_p = 0$) there is no greenhouse warming. For fixed optical depth, the greenhouse warming becomes larger as the R/c_p, and hence the lapse rate, becomes large. For Venus, the absorbed solar radiation is approximately $163 W/m^2$, owing to the high albedo of the planet. For a pure CO_2 atmosphere $R/c_p \approx .2304$, for which $\Gamma \approx .969$. Then, the $737K$ surface temperature of Venus can be explained accounted for if $\tau_\infty = 156$, which is a very optically thick atmosphere. This is essentially the calculation used by Carl Sagan to infer that the dense CO_2 atmosphere of Venus could give it a high enough surface temperature to account for the then-mysterious anomalously high microwave radiation emitted by the planet (microwaves being directly emitted to space by the hot surface without significant absorption by the atmosphere).

Exercise 4.3.1 This exercise illustrates the fact that if the Earth's atmosphere acted like a grey gas, then a doubling of CO_2 would make us toast! Using Eq. 4.32, find the τ_∞ that yields a surface temperature of $285K$ for the Earth's absorbed solar radiation (about $270 W/m^2$ allowing a crude correction for net cloud effects). Now suppose we double the greenhouse gas content of the atmosphere. If the Earth's greenhouse gases were grey gases, this would imply doubling the value of τ_∞ from the value you just obtained. What would the resulting temperature be? Note that this rather alarming temperature doesn't even fully take into account the amplifying effect of water vapor feedback.

An examination of the radiative heating rate profile for the all-troposphere case provides much insight into the processes which determine where the troposphere leaves off and where a stratosphere will form. We'll assume that $I_{-,\infty} = 0$ and that the turbulent heat transfer at the ground is efficient enough that $T_{sa} = T_g$. Consider first the optically thin limit, for which the grey gas version of Eq. 4.25 is

$$H_\nu = \frac{\kappa}{\cos\bar{\theta}} \cdot [\sigma T_g^4 - 2\sigma T(p)^4] \tag{4.33}$$

assuming the stated boundary conditions. Since the radiative heating rate is nonzero, the temperature profile will not be in a steady state unless some other source of heating and cooling is provided to cancel the radiative heating. According to Eq. 4.33, the atmosphere is cooling at low altitudes, where $T > T_g/2^{\frac{1}{4}}$, i.e. where the local temperature is greater than the skin temperature. The cooling will make the atmosphere's potential temperature lower than the ground temperature, which allows the air in contact with the ground to be positively buoyant. The resulting convection brings heat to the radiatively cooled layer, allowing a steady state to be maintained if the convection is vigorous enough. However, in the upper atmosphere, where $T < T_g/2^{\frac{1}{4}}$ the atmosphere is being *heated* by upwelling infrared radiation, and there is no obvious way that convection could provide the cooling needed to make this region a steady state. Instead, the atmosphere in this region is expected to warm until a stratosphere in pure radiative equilibrium forms. Indeed, the tropopause as estimated by the boundary between the region of net heating and net cooling is located at the point where $T(p)$ equals the skin temperature; this is precisely the same result as we obtained in the steady state model of the tropopause for an optically thin atmosphere, as discussed in Section 3.6.

In the optically thick limit it is easiest to infer the infrared heating profile from an examination of the expression for net infrared flux, which becomes

$$I_+ - I_- = 2\frac{g\cos\bar{\theta}}{\kappa}(4\sigma T^3)\frac{dT}{dp} = 8\sigma T_g^4 \frac{R}{c_p}\frac{g^2\cos\bar{\theta}}{\kappa p_s}(\frac{p}{p_s})^{4R/c_p - 1} \tag{4.34}$$

in the all-troposphere grey-gas case. Recall that this expression breaks down in thin layers within roughly a unit optical depth of the bottom and top boundaries. The formula shows that whether the bulk of an optically thick atmosphere is heating or cooling depends on the lapse rate. If $4R/c_p > 1$ the optically thick net flux decreases with height, and most of the atmosphere is *heated* by infrared radiation, and hence we expect a deep stratosphere and shallow troposphere. If $4R/c_p < 1$, corresponding to a weaker temperature lapse rate, most of the atmosphere instead experiences infrared radiative cooling, so we expect a deep troposphere. These expectations will be confirmed in Chapter 7. Real gases are typically optically thick at some wavenumbers but optically thin at others, so the behavior of the radiative heating profile proves more subtle, as does the question of tropopause height.

Figure 4.2 shows numerically computed profiles of net infrared flux ($I_+ - I_-$) for a range of optical thicknesses, with $R/c_p = \frac{2}{7}$. In this case, $4R/c_p > 1$, and we expect deep heating in the optically thick limit. For $\tau_\infty = 50$ the profile does follow the optically thick approximate form over most of the atmosphere, and exhibits a decrease in flux with height, implying deep heating. There is a thin layer of cooling near the ground, where the optically thick formula breaks down. When $\tau_\infty = 10$, the flux only conforms to the optically thick limit near the center of the atmosphere; there is a region of infrared cooling that extends from the ground nearly to 70% of the surface pressure. The case $\tau_\infty = 1$ looks quite like the optically thin limit, with the lower half of the atmosphere cooling and the upper half heating.

The troposphere is defined as the layer stirred by convection, and since hot air rises, buoyancy driven convection transports heat upward where it is balanced by radiative cooling. Therefore, at

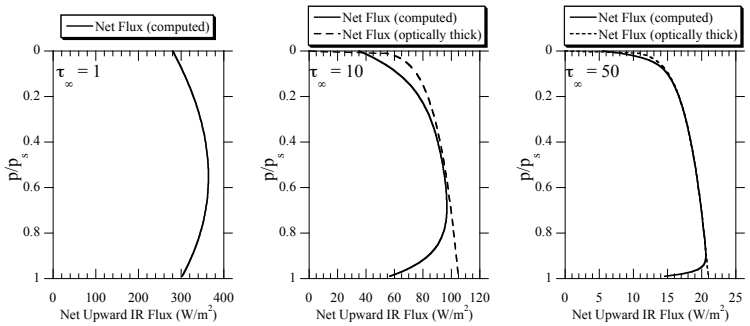

Figure 4.2: Net infrared flux $(I_+ - I_-)$ for the all-troposphere grey-gas model, for $\tau_\infty = 1, 10$ and 50. In the latter two cases, the dashed line gives the result of the optically thick approximation. The surface temperature is fixed at $300K$, and the temperature profile is the dry adiabat with $R/c_p = \frac{2}{7}$.

least the upper region of a troposphere invariably experiences radiative cooling. In the calculation discussed above, the layer with cooling, fated to become the troposphere, occurs in the lower portion of the atmosphere. In Figure 4.2 one notices that the radiative cooling decreases as the atmosphere is made more optically thick, suggesting that tropospheric convection becomes more sluggish in an optically thick atmosphere, there being less radiative cooling to be offset by convective heating. However, one should note also that this sequence of calculations is done with fixed surface temperature, and that the OLR decreases as optical thickness is made larger. Hence, in the optically thick cases, it takes less absorbed solar radiation to maintain the surface temperature of the planet. There is less flux of energy through the system, and correspondingly less convection. Mars, at a more distant orbit than Earth, receives less solar energy; if Mars were given an atmosphere with enough greenhouse effect to warm it up to Earthlike temperatures, one would expect the radiative cooling in its troposphere to be less than Earth's, and one would expect the convection to be more sluggish.

The presence of a stratosphere causes the OLR to exceed the values implied by the all-troposphere calculation, since the upper portions of an atmosphere with a stratosphere are warmer than the all-troposphere model would predict. If the stratosphere is optically thin, it has a minor effect on the OLR; in essence, the all-troposphere OLR formula provides a good estimate if the effective radiating level is below the tropopause. If the stratosphere becomes optically thick, then the OLR is in fact determined by the stratospheric structure. Problem **??** explores some aspects of the effect of an optically thick stratosphere on OLR. Puzzling out the effect of the stratosphere on OLR is rather tricky, because the tropopause height itself depends on the optical thickness of the atmosphere. An optically thin atmosphere obviously can't have an optically thick stratosphere, but an optically thick atmosphere can nevertheless have an optically thin stratosphere if the tropopause height increases rapidly enough with τ_∞. The grey gas radiative cooling profiles discussed above suggest that the stratosphere becomes optically thick when $4R/c_p > 1$. In contrast, for $4R/c_p < 1$ the radiatively cooled layer extends toward the top of the atmosphere in the optically thick limit, and hence the stratosphere could remain optically thin. The full problem will be taken up in detail in Chapter 7.

4.3.3 A first look at the runaway greenhouse

We have seen in Chapter 2 that the mass of an atmosphere in equilibrium with a reservoir of condensed substance (e.g. a water ocean) is not fixed. It increases with temperature in accordance with the dictates of the Clausius-Clapeyron relation. If the condensible substance is a greenhouse gas, then the optical thickness τ_∞ increases with temperature. This tends to reduce the OLR, offsetting or even reversing the tendency of rising temperature to increase the OLR. What are the implications of this for the dependence of OLR on surface temperature, and for planetary energy balance? The resulting phenomena are most commonly thought about in connection with the effects of a water ocean on evolution of a planet's climate, but the concept generalizes to any condensible greenhouse gas in equilibrium with a large condensed reservoir. We'll take a first look at this problem here, in the context of the grey-gas model.

In the general case, we'd like to consider an atmosphere in which the condensible greenhouse gas is mixed with a noncondensible background of fixed mass (which may also have a greenhouse effect of its own). This is the case for water vapor in the Earth's atmosphere, for methane on Titan, and probably also for water vapor in the early atmosphere of Venus. It could also have been the case for mixed nitrogen-CO_2 atmospheres on Early Mars, with CO_2 playing the role of the condensible component. We will eventually take up such atmospheres, but the difficulty in computing the moist adiabat for a two-component atmosphere introduces some distractions which get in the way of grasping the key phenomena. Hence, we'll start with the simpler case in which the atmosphere consists of a pure condensible component in equilibrium with a reservoir (an "ocean," or perhaps a glacier). In this case, the saturated moist adiabat is given by the simple analytic formula Eq. 2.24, obtained by solving the simplified form of the Clausius-Clapeyron relation for temperature in terms of pressure. We've already seen in Chapter 2 that a mixed atmosphere is dominated by the condensible component at large temperatures, so if we are primarily interested in the large-temperature behavior, the use of the one-component condensible atmosphere is not at all a bad approximation.

We write $T(p) = T_o/(1 - \frac{RT_o}{L}\ln\frac{p}{p_o})$, where (p_o, T_o) are a *fixed* reference temperature and pressure on the saturation curve, such as the triple point temperature and pressure. If the surface pressure is p_s, then the surface temperature is $T_s = T(p_s)$. Hence, specifying surface pressure is equivalent to specifying surface temperature in this problem. To keep the algebra simple, we'll assume a constant specific absorption κ. Then $\tau_\infty = \kappa p_s/g$, which increases as T_s is made larger. Further, for constant specific absorption, $p/p_o = (\tau_\infty - \tau')/\tau_o$ where $\tau_o = \kappa p_o/g$. Now, the choice of the reference temperature and pressure (p_o, T_o) is perfectly arbitrary, and we'll get the same answer now matter what choice we make (within the accuracy of the approximate form of Clausius-Clapeyron we are using). Hence, we are free to set $p_o = g/\kappa$ so that $\tau_o = 1$. T_o then implicitly depends on κ, and becomes larger as κ gets smaller. T_o is the temperature at the level of the atmosphere where the optical depth measured relative to the top of the atmosphere is unity.

Substituting the one-component $T(p)$ into the integral giving the solution to the Schwarzschild equation, and substituting for pressure in terms of optical thickness, we find

$$\begin{aligned} OLR &= I_+(0)e^{-\tau_\infty} + \int_0^{\tau_\infty} \sigma \frac{T_o^4}{(1 - \frac{RT_o}{L}\ln\frac{p}{p_o})^4} e^{-(\tau_\infty - \tau')} d\tau' \\ &= I_+(0)e^{-\tau_\infty} + \sigma T_o^4 \int_0^{\tau_\infty} \frac{1}{(1 - \frac{RT_o}{L}\ln\tau_1)^4} e^{-\tau_1} d\tau_1 \end{aligned} \tag{4.35}$$

where we have in the second line defined a new dummy variable $\tau_1 = \tau_\infty - \tau'$ as before. The surface temperature enters the expression for OLR only through τ_∞, which is proportional to

surface pressure. In the optically thin limit, the integral on the right hand side of the expression is small (because τ_∞ is small). This happens at low surface temperatures, because p_s is small when the surface temperature is small. The OLR then reduces to the first term, which is approximately $I_+(0)$,i.e. the unmodified upward radiation from the surface. In the optically thick limit, which occurs for high surface temperatures, the term proportional to $I_+(0)$ is negligible, and the second term dominates. This term consists of the flux σT_o^4 multiplied by a non-dimensional integral. Recall that T_o is a constant dependent on the thermodynamic and infrared optical properties of the gas making up the atmosphere; it does not change with surface temperature. Because of the decaying exponential in the integrand, the integral is dominated by the contribution from the vicinity of $\tau_1 = 0$, and will therefore become independent of τ_∞ for large τ_∞ [1]. In the optically thick (high temperature) limit, then, the integral is a function of RT_o/L alone. From this we conclude that the OLR becomes independent of surface temperature in the limit of large surface temperature (and hence large τ_∞). This limiting OLR is known as the *Kombayashi-Ingersoll limit*. It was originally studied in connection with the long-term history of water on Venus, using a somewhat different argument than we have presented here. We shall use the term to refer to a limiting OLR arising from the evaporation of any volatile greenhouse gas reservoir, whether computed using a grey gas model or a more realistic radiation model.

It is readily verified that the integral multiplying σT_o^4 approaches unity as RT_o/L approaches zero. In fact, for typical atmospheric gases L/R is a very large temperature, on the order of several thousand Kelvins. Hence, unless the specific absorption is exceedingly small, RT_o/L tends to be small, typically on the order of .1 or less. For $RT_o/L = .1$, the integral has the value of .905. Thus, the limiting OLR is essentially σT_o^4. Recalling that T_o is the temperature of the moist adiabat at one optical depth unit down from the top of the atmosphere, we see that the limiting OLR behaves very nearly as if all the longwave radiation were emitted from a layer one optical depth unit from the top of the atmosphere.

Figure 4.3 shows some results from a numerical evaluation of the integral in Eq. 4.35. For small surface temperatures, there is little atmosphere, and the OLR increases like σT_s^4. As the surface temperature is made larger, the atmosphere becomes thicker and the OLR eventually asymptotes to a limiting value, as predicted. In accordance with the argument given above, the limiting OLR should be slightly less than the blackbody flux corresponding to the temperature T_o found one optical depth down from the top of the atmosphere. This temperature depends on g/κ, which is the pressure one optical depth down from the top. For $g/\kappa = 100Pa$, solving the simplified Clausius-Clapeyron relation for T at $100Pa$ yields $T_0 = 250.3K$, whence $\sigma T_o^4 = 222.6W/m^2$; for $g/\kappa = 1000Pa$, $T_o = 280.1K$ and $\sigma T_o^4 = 349.0W/m^2$. These values are consistent with the numerical results shown in the graph. Note that for a given atmospheric composition (which determines κ) the Kombayashi-Ingersoll limit depends on the acceleration of gravity. A planet with stronger surface gravity will have a higher Kombayashi-Ingersoll limit than one with weaker gravity.

We are now prepared to describe the runaway greenhouse phenomenon. Let S be the absorbed solar radiation per unit surface area of the planet, and let the limiting OLR computed above be OLR_{max}. If $S < OLR_{max}$ the planet will come to equilibrium in the usual way, warming up until it loses energy by infrared radiation at the same rate as it receives energy from its star. But what happens if $S > OLR_{max}$? In this case, as long as there is still an ocean or other condensed reservoir to feed mass into the atmosphere, the planet cannot get rid of all the solar energy it receives no matter how much it warms up; hence the planet continues to warm until the surface temperature becomes so large that the entire ocean has evaporated into the atmosphere.

[1] Technically, the integral diverges at extremely large τ_∞, because the denominator of the integrand can vanish. This is an artifact of assuming a constant latent heat and has no physical significance.

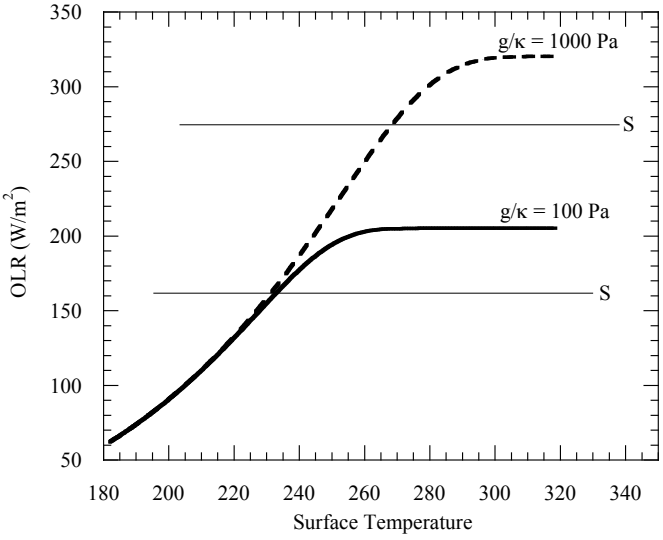

Figure 4.3: OLR vs surface temperature for a one-component grey gas condensible atmosphere in equilibrium with a reservoir. Calculations were done for thermodynamic parameters L and R corresponding to water vapor. Results are shown for two different values of g/κ, where κ is the specific cross section of the gas and g is the acceleration of gravity.

The temperature at this point depends on the mass and composition of the volatile reservoir. For example, the Earth's oceans contain enough mass to raise the surface pressure to about $100 bars$ if dumped into the atmosphere in the form of water vapor. The ocean has been exhausted when the saturation vapor pressure reaches this value. Using the simplified exponential form of the Clausius-Clapeyron relation to extrapolate the vapor pressure from the sea level boiling point (1 bar at $373.15K$), we estimate that this vapor pressure is attained at a surface temperature of about $550K$. This estimate is inaccurate, because the latent heat of vaporization varies appreciably over the range of temperatures involved. A more exact value based on measurements of properties of steam is $584K$, but the grim implications for survival or emergence of life as we know it are largely the same.

At temperatures larger than that at which the ocean is depleted, the mass of the atmosphere becomes fixed and no longer increases with temperature. The greenhouse gas content of the atmosphere – which in the present case is the entire atmosphere – no longer increases with temperature. As a result, the OLR is once more free to increase as the surface becomes warmer, and the planet will warm up until it reaches an equilibrium at a temperature warmer than that at which the ocean is depleted. The additional warming required depends on the gap between the Kombayashi-Ingersoll limit and the absorbed solar radiation. Once the ocean is gone, the lower atmosphere is unsaturated and air can be lifted some distance before condensation occurs. The resulting atmospheric profile is on the dry adiabat in the lower atmosphere, transitioning to the moist adiabat at the altitude where condensation starts. The situation is identical to that depicted for CO_2 in Fig. 2.6. Rain will still form in the condensing layer. Much of it will evaporate in the lower noncondensing layer; some of it may reach the ground, but the resulting puddles would tend to rapidly evaporate back into the highly undersaturated lower atmosphere. As surface temperature is made larger, the altitude where condensation sets in moves higher, until at very large

temperatures the atmosphere behaves like a noncondensing dry system (albeit one where the entire atmosphere may consist of water vapor).

The runaway greenhouse phenomenon may explain how Venus wound up with such a radically different climate from Earth, despite having started out in a rather similar state. The standard story goes something like this: Venus started with an ocean, and with most of it's CO_2 bound up in rocks as is the case for Earth. However, it was just enough closer to the Sun to trigger a runaway greenhouse. Once the entire ocean had evaporated into the atmosphere, there was so much water vapor in the upper atmosphere that it could be broken apart by energetic solar ultraviolet rays, whereafter the light hydrogen could escape to space. The highly reactive oxygen left behind would react to form minerals at the surface. Once there was no more liquid water in play, the reactions that bind up carbon dioxide in rocks could no longer take place (as will be explained in Chapter 9), so all the planet's CO_2 outgassed from volcanism and stayed in the atmosphere, leading to the hot, dry super-dense atmosphere of modern Venus.

Assuming habitability to require a reservoir of liquid water, the Kombayashi-Ingersoll limit for water determines the inner orbital limit for habitability, since if the Solar constant exceeds the limiting flux a runaway will ensue and any initial ocean will not persist. It also determines how long it takes before the planet's Sun gets bright enough to trigger a runaway, and thus sets the lifetime of a water-dependent biosphere (Earth's included). Accurate calculations of the Kombayashi-Ingersoll limit are therefore of critical importance to understanding the limits of habitability both in time and orbital position. The grey gas model is not good enough to determine the value of p_o appropriate to a given gas, and so cannot be used for accurate evaluations of the runaway greenhouse threshold. We can at least say that, all other things being equal, a planet with larger surface gravity will be less susceptible to the runaway greenhouse. This is so because $p_o = g/\kappa$, whence larger g implies larger p_o, which implies in turn larger $T(p_o)$ and hence a larger limiting OLR. This observation may be relevant to the class of extrasolar planets known as "Large Earths."

We will revisit the runaway greenhouse using more realistic radiation physics in Section 4.5.1. The effects of the stratosphere and of clouds will be brought into the picture in Chapter 7.

4.3.4 Pure radiative equilibrium for a grey gas atmosphere

For the temperature profiles discussed in Sections 4.3.2 and 4.3.3, the net infrared radiative heating computed from Eq. 4.11 is nonzero at virtually all altitudes; generally it is negative, and acts to cool the atmosphere. These temperature profiles can be maintained in equilibrium only if some other heat transport mechanism makes up the deficit. It is implicit in these solutions that convection will supply whatever heating is necessary to keep the atmosphere in a steady state. The atmosphere/surface system is in radiative equilibrium as a unit, but it need not be in radiative equilibrium at each altitude individually, since convection can redistribute heat in the vertical so as to maintain equilibrium Now, we'll investigate solutions for which, in contrast, the net radiative heating vanishes individually at each altitude. Such solutions are in *pure radiative equilibrium*, as apposed to *radiative-convective equilibrium*. First we'll consider the case in which the only radiative heating is supplied by infrared; later we'll bring heating by atmospheric solar absorption into the picture.

Assuming the atmosphere to be transparent to solar radiation, pure radiative equilibrium requires that the frequency-integrated longwave radiative heating $\mathfrak{H}$ vanish for all τ. From the grey gas version of Eq. 4.11, we then conclude that $I_+ - I_-$ is independent of τ. Applying the upper boundary condition, we find that this constant is $I_+(\tau_\infty)$, which is the OLR. Now, by taking the

difference between the equations for I_+ and I_- we find

$$0 = \frac{d}{d\tau}(I_+ - I_-) = -(I_+ + I_-) + 2\sigma T^4 \tag{4.36}$$

which gives us the temperature in terms of $(I_+ + I_-)$. Next, taking the sum of the equations for I_+ and I_- yields

$$\frac{d}{d\tau}(I_+ + I_-) = -(I_+ - I_-) \tag{4.37}$$

This is easily solved by noting that $-(I_+ - I_-) = const = -OLR$. In consequence,

$$2\sigma T^4 = (I_+ + I_-) = (1 + \tau_\infty - \tau)OLR \tag{4.38}$$

where we have again used the boundary condition at τ_∞. This expression gives us the pure radiative equilibrium temperature profile $T(\tau)$. In pure radiative equilibrium, the temperature always approaches the skin temperature at the top of the atmosphere, where $\tau = \tau_\infty$. This recovers the result obtained in the previous chapter, in Section 3.6. When the atmosphere is optically thin, $\tau_\infty - \tau$ is small throughout the atmosphere, and the entire atmosphere becomes isothermal with temperature equal to the skin temperature. When the atmosphere is not optically thin, the temperature decreases gently with height, approaching the skin temperature as the top of the atmosphere is approached.

Eq. 4.38 also gives us the upward and downward fluxes, since we now know both $I_+ - I_-$ and $I_+ + I_-$ at each τ. In particular, the downward flux into the ground is

$$I_-(0) = \frac{1}{2}((I_+ + I_-) - (I_+ - I_-)) = \frac{1}{2}((1 + \tau_\infty)OLR - OLR) = \frac{1}{2}\tau_\infty OLR \tag{4.39}$$

For an optically thin atmosphere, the longwave radiation returned to the ground by the atmosphere iss only a small fraction of that emitted to space. A the atmosphere becomes optically thick, the radiation returned to the ground becomes much greater than that emitted to space, because the radiative equilibrium temperature near the ground becomes large and the optical thickness implies that the radiation into the ground is determined primarily by the low level temperature. If we assume the planet to be in radiative equilibrium with the absorbed solar radiation $(1-\alpha)S$, where α is the albedo of the ground, then $OLR = (1-\alpha)S$ and the radiative energy budget of the ground is

$$\sigma T_s^4 = (1 - \alpha)S + I_-(0) = (1 - \alpha)S \cdot (1 + \frac{1}{2}\tau_\infty) \tag{4.40}$$

where T_s is the surface temperature. This, together with the temperature profile determined by Eq. 4.38, determines what the thermal state of the system would be in the absence of heat transport mechanisms other than radiation. For an optically thin atmosphere, the surface temperature is only slightly greater than the no-atmosphere value. As τ_∞ becomes large, the surface temperature increases without bound. Note that, while this formula yields a greenhouse warming of the surface, the relation between surface temperature and τ_∞ is different from that given by the all-troposphere radiative convective calculation in Eq. 4.32, because the pure radiative equilibrium temperature profile is different from the adiabat which would be established by convection.

Let's now compare the surface temperature with the temperature of the air in immediate contact with the surface. From Eq. 4.38 we find that the low level air temperature is determined by $\sigma T(0)^4 = (1 - \alpha)S \cdot (\frac{1}{2} + \frac{1}{2}\tau_\infty)$ Taking the ratio,

$$\frac{T(0)}{T_s} = (\frac{\frac{1}{2} + \frac{1}{2}\tau_\infty}{1 + \frac{1}{2}\tau_\infty})^{1/4} \tag{4.41}$$

Thus, the surface is always warmer than the overlying air in immediate contact with it. In the previous chapter, we saw that this was the case for pure radiative equilibrium in an optically thin atmosphere, but now we have generalized it ot arbitrary optical thickness. In the optically thin limit, the formula reduces to our earlier result, $T(0) = 2^{-1/4}T_s$. In the optically thick limit, $T_s - T(0) = \frac{1}{4}T_s/\tau_\infty$, whence the temperature jump (relative to surface temperature) falls to zero as the atmosphere is made more optically thick. As we already discussed in Section 3.6, cold air immediately above a warmer surface constitutes a very unstable situation. Under the action of diffusive or turbulent heat transfer between the surface and the nearby air, a layer of air near the surface will heat up to the temperature of the surface, whereafter it will be warmer than the air above it. Being buoyant, it will rise and lead to convection, which will stir up some dept of the atmosphere and establish an adiabat – creating a troposphere.

In pure radiative equilibrium, the surface heating inevitably gives rise to convection. However, it is also possible that the temperature profile in the interior of the atmosphere may become unstable to convection, even without the benefit of a surface. This is a particularly important possibility to consider for gas giant planets, which have no distinct surface to absorb solar radiation and stimulate convection. To determine stability, we must compute the lapse rate dT/dp in radiative equilibrium, and see if it is steeper than that of the adiabat (moist or dry) appropriate to the atmosphere. Taking the derivative of 4.38 with respect to optical thickness, we find

$$8\sigma T^3 \frac{dT}{d\tau} = -OLR \tag{4.42}$$

whence, using $d/dp = (d\tau/dp)(d/d\tau)$, we find

$$\frac{1}{T}\frac{dT}{dp} = -\frac{1}{4}\frac{1}{(1+\tau_\infty - \tau)}\frac{d\tau}{dp} \tag{4.43}$$

Note that this slope, and hence the stability criterion, is independent of the OLR. Given that the OLR is equal to the energy sources driving the planet's atmosphere, this result says that the degree of instability is independent of the magnitude of the planet's energy source. Plugging in the expression for the idea gas dry adiabat, the atmosphere is found to be stable where

$$\frac{R}{c_p} \geq -\frac{1}{4}p\frac{d\tau}{dp}\frac{1}{(1+\tau_\infty - \tau)} \tag{4.44}$$

In the case of constant absorption coefficient κ, we have $pd\tau/dp = -\kappa p/g\cos\bar{\theta}$, which is just $\tau - \tau_\infty$. Thus, the stability condition becomes

$$\frac{R}{c_p} \geq \frac{1}{4}\frac{\tau_\infty - \tau}{(1+\tau_\infty - \tau)} \tag{4.45}$$

The right hand side has its maximum at the ground $\tau = 0$, and the maximum value is $\frac{1}{4}\tau_\infty/(1+\tau_\infty)$. The more optically thick the atmosphere is, the more unstable it is near the ground. For large optical thickness, the stability criterion becomes the remarkably simple statement $4R/c_p \geq 1$. Dry Earth air, with $R/c_p = 2/7$, just misses being unstable by this criterion.

Generally speaking, a sharp decrease of absorption coefficient with height tends to destabilize the atmosphere, particularly if it occurs in a place where the atmosphere is optically thick. The easiest way to get such a sharp decrease is for the concentration of longwave absorbers to decrease sharply with height.

4.3.5 Effect of atmospheric solar absorption on pure radiative equilibrium

Now we will examine how the absorption of solar radiation within an atmosphere affects the temperature structure of the atmosphere in radiative equilibrium. The prime application of this calculation is to understand the thermal structure of stratospheres. Under what circumstances does the temperature of a stratosphere increase with height? The effect of solar absorption on gas giant planets like Jupiter is even more crucial. There being no distinct surface to absorb sunlight, *all* solar driving of the atmosphere for gas giants comes from deposition of solar energy within the atmosphere. In this case, the profile of absorption determines in large measure where, if anywhere, the radiative equilibrium atmosphere is unstable to convection, and therefore where a troposphere will tend to form. The answer determines whether convection on gas giants is driven in part by solar heating as opposed to ascent of buoyant plumes carrying heat from deep in the interior of the planet.

In the Earth's stratosphere, solar absorption is largely due to the absorption of ultraviolet by ozone. On Earth as well as other planets having appreciable water in their atmospheres, absorption of solar near-infrared by water vapor and water clouds is important. CO_2 also has significant near-infrared absorption, which is relatively unimportant at present-day CO_2 concentrations on Earth, but becomes significant on the Early Earth when CO_2 concentrations were much higher; solar near-infrared absorption by CO_2 is of course important in the CO_2 dominated atmospheres of Mars (present and past) and Venus. Solar absorption by dust is important to the Martian thermal structure throughout the depth of the atmosphere. On Titan, it is solar absorption by organic haze clouds that control the thermal structure of the upper stratosphere. Solar absorption is also crucial to the understanding of the influence of greenhouse gases like CH_4 and SO_2, which strongly absorb sunlight in addition to being radiatively active in the thermal infrared. Strong solar absorption also would occur in the high-altitude dust and soot cloud that would be lofted in the wake of a global thermonuclear war or asteroid impact (the "Nuclear Winter" problem).

Since the Schwartzschild equations in this chapter are used to describe the infrared flux alone, the addition of solar heating does not change these equations. The heating due to solar absorption only alters the condition for local equilibrium, which now involves the deposition of solar as well as infrared flux. We write the solar heating rate per unit optical depth in the form $Q_\odot = dF_\odot/d\tau$, where $F_\odot$ is the net downward solar flux as a function of infrared optical depth. At the top of the atmosphere, $F_\odot = (1-\alpha)S$, where α is the planetary albedo – that is, the albedo measured at the top of the atmosphere. Since atmospheres at typical planetary temperatures do not emit significantly in the solar spectrum, there is no internal source of solar flux and therefore $F_\odot$ must decrease monotonically going from the top of the atmosphere to the ground.

The net radiative heating at a given position is now the sum of the infrared and solar term, i.e.

$$-\frac{d}{d\tau}(I_+ - I_-) + \frac{d}{d\tau}F_\odot = 0 \tag{4.46}$$

Integrating this equation and requiring that the top of atmosphere energy budget be in balance with the local absorbed solar radiation, we find

$$(I_+ - I_-) - F_\odot = 0 \tag{4.47}$$

At the top of the atmosphere, this reduces to $OLR - (1-\alpha)S = 0$, which is the requirement for top of atmosphere energy balance. Because the solar absorption does not change the infrared Schwartzschild equations, Eq. 4.37 is unchanged from the case of pure radiative equiibrium without

solar absorption. Substituting Eq 4.47 and integrating, we obtain

$$I_+ + I_- = \int_\tau^{\tau_\infty} F_\odot(\tau')d\tau' + (1-\alpha)S \tag{4.48}$$

In writing this expression we have made use of the boundary condition $I_+ - I_- = OLR = (1-\alpha)S$ at the top of the atmosphere. The heat balance equation 4.36 needs to be slightly modified, since the infrared cooling now balances the solar heating, instead of being set to zero. Thus,

$$\frac{d}{d\tau}F_\odot = \frac{d}{d\tau}(I_+ - I_-) = -(I_+ + I_-) + 2\sigma T^4 \tag{4.49}$$

from which we infer

$$2\sigma T^4 = \frac{d}{d\tau}F_\odot + \int_\tau^{\tau_\infty} F_\odot(\tau')d\tau' + (1-\alpha)S \tag{4.50}$$

This gives the vertical profile of temperature in terms of the vertical profile of the solar flux; the previous case (without solar absorption) can be recovered by setting $F_\odot = const = (1-\alpha)S$. At the top of the atmosphere, the integral in Eq. 4.49 vanishes, and the temperature becomes identical to the temperature of a skin layer heated by solar absorption, derived in Chapter 3 (Eq. 3.27).

Taking the derivative with respect to τ yields

$$\frac{d}{d\tau}2\sigma T^4 = \frac{d^2}{d\tau^2}F_\odot - F_\odot \tag{4.51}$$

This equation provides a simple criterion determining when the solar absorption causes the temperature to increase with height. When there is no absorption, $F_\odot$ is a constant and since it is positive the temperature decreases with height.

By way of illustration, let's suppose that the net downward solar flux decays exponentially as it penetrates the atmosphere. Specifically, let $F_\odot = (1-\alpha)S\exp(-(\tau_\infty - \tau)/\tau_S)$, where τ_S is a constant. τ_S is the decay rate of solar radiation, measured in infrared optical depth units. When τ_S is large, solar absorption is weak compared to infrared absorption, and one must go a great distance before the solar beam is appreciably attenuated. Conversely, when τ_S is small, solar absorption is strong and the solar beam decays to zero over a distance so short that infrared is hardly attenuated at all. With the assumed form of the solar flux, the temperature profile is given by

$$2\frac{\sigma T^4}{(1-\alpha)S} = 1 + \tau_S + (\frac{1}{\tau_S} - \tau_S)e^{-(\tau_\infty - \tau)/\tau_S} \tag{4.52}$$

If $\tau_S > 1$ the temperature decreases with height, and if $tau_S < 1$ the temperature increases with height. Defining the skin temperature as $T_{skin} \equiv (\frac{1}{2\sigma}(1-\alpha)S))^{1/4}$ the temperature at the top of the atmosphere is $(1+1/\tau_S)T_{skin}$, which reduces to the skin temperature when τ_S is large and becomes much greater than the skin temperature when τ_S is small. If the atmosphere is deep enough that essentially all solar radiation is absorbed before reaching the ground, then the exponential term vanishes in the deep atmosphere and the deep atmosphere becomes isothermal with temperature $(1+\tau_S)T_{skin}$. Thus, when τ_S is small, all the solar radiation is absorbed within a thin layer near the top of the atmosphere. The temperature increases rapidly with height in this layer, but the bulk of the atmosphere below is approximately isothermal at the skin temperature. The strong solar absorption causes the deep atmosphere, and the ground (if there is one) , to be colder than it would have been in the absence of an atmosphere. This *anti-greenhouse effect* arises because the deep atmosphere is heated only by downwelling infrared emitted by the solar-absorbing layer.

This downward radiation equals the upward radiation loss to space, which must equal $(1-\alpha)S$ to satisfy the top of atmosphere balance. The deep atmosphere falls to the skin temperature because it is being illuminated by this flux from one side, but is radiating from both its top and bottom.

It can happen that the atmosphere is deep enough to absorb all solar radiation before it reaches the ground, even if the rate of solar absorption is weak and $\tau_S >> 1$. This would happen if the atmosphere is so optically thick in the infrared that $\tau_/infty/\tau_S >> 1$ despite τ_S being large. In this case, the deep atmosphere is still isothermal, but it becomes much hotter than the skin temperature – indeed it becomes hotter without bound as τ_S is made larger. In this case, it is the top of the atmosphere which equilibrates at the relatively cold skin temperature, while the deep atmosphere exhibits a strong greenhouse effect. Because the deep atmosphere is isothermal, it is stable and will not generate a troposphere.

The general lesson to take away from this discussion is that solar absorption near the top of the atmosphere stabilizes the atmosphere, reduces the greenhouse effect, and cools the lower portion of the atmosphere and also the ground. This is important in limiting the effectiveness of greenhouse gases like CH_4 and SO_2, which significantly absorb solar radiation when their concentration becomes very high. In contrast, solar absorption concentrated near the ground has an effect which is not much different from simply reducing the albedo of the ground itself.

4.4 Real gas radiation: Basic principles

4.4.1 Overview: OLR through thick and thin

It would be exceedingly bad news for planetary habitability if real greenhouse gases were grey gases (see Exercise 4.3.1). Greenhouse gas concentrations would have to be tuned exceedingly accurately to maintain a planet in a habitable temperature range, and there would be little margin for error. Thus, it is of central importance that, for real gases, OLR varies much more gradually with greenhouse gas concentration than it would for an idealized grey gas [2]. This is another area in which the quantum nature of the Universe directly intervenes in macroscopic phenomena governing planetary climate.

Infrared Radiative tranfer is a very deep and complex subject, and mastery of the material in this section will still not leave the reader prepared to write state-of-the-art radiation codes. Nor will we cover the myriad engineering tricks large and small which are needed to make a radiation code fast enough to embed in a general circulation model, where it will need to be invoked a dozen times per model day at each of several thousand grid points. We do aspire to provide enough of the basic physics to allow the student to understand why OLR is less sensitive to the concentration of a typical real gas than to a grey gas, and to help the student develop some intuition about the full possible range of behaviors of greenhouse gases on Earth and other planets, now and in the distant past or future. Such an understanding should extend even to greenhouse gases that are not at present commonly considered in the context of climate, or implemented in standard "off the shelf" radiation models. What would you do, for example, if you found yourself wondering whether SO_2 or H_2S significantly affected the climate of Early Earth or Mars? The grey gas model does not provide an adequate first attack on such problems. We thus aspire to provide enough of the basic algorithmic equipment to allow the student to build simplified radiative models from scratch, that

[2] Lest there be any misunderstanding, we must emphasize at this point that "less sensitive" does not mean "insensitive." If CO_2 were a grey gas, then doubling it's concentration, as we are poised to do within the century, would be unquestionably lethal. Because CO_2 is not in fact a grey gas, the results may be merely catastrophic.

get the OLR and infrared heating profiles roughly correct.

Even though we will have recourse to a "professionally" written radiation code in Section 4.5, we'd like to at least draw back the curtain a little bit, so that the reader will not be left with the all-too-common notion that radiation routines are black boxes, the internal workings of which can only be understood by the high priesthood of radiative transfer. Hopefully, this will also open the door to entice more people into innovative work on the subject.

Since the main point is to understand how the wavenumber dependence of absorptivity affects the sensitivity of OLR to greenhouse gas concentration, we'll begin with a discussion of the spectrum of outgoing longwave radiation in an idealized case. Let's consider a planet whose surface radiates like an ideal blackbody in the infrared, having an atmosphere whose air temperature at the surface T_{sa} is equal to the ground temperature T_g. The temperature $T(p)$ is monotonically decreasing with height in the troposphere, and is patched continuously to an isothermal stratosphere having temperature T_{strat}. The atmosphere consists of mostly of infrared-transparent N_2 and O_2 with a surface pressure of $10^5 Pa$, like Earth. Unlike Earth, the only greenhouse gas is a mythical substance (call it Oobleck), which is a bit like CO_2, but much simpler to think about. It has the same molecular weight as CO_2, but it's absorption coefficient $\kappa_{Oob}(\nu)$ has an absorption band centered on wavenumber $\nu_o = 600 cm^{-1}$. Within $200 cm^{-1}$ of ν_o, κ_{Oob} has the constant value κ_o. Outside of this limited range of wavenumbers, Oobleck is transparent to infrared, i.e. $\kappa_{Oob} = 0$. To make life even simpler for the atmospheric physicists of this planet, κ_{Oob} is independent of both temperature and pressure. Like real CO_2, the specific concentration of Oobleck (q_{Oob}) is constant throughout the depth of the atmosphere.

What does the spectrum of OLR look like for this planet? The answer is shown in the left panel of Figure 4.4. In this figure, we have assumed that the Oobleck molecule has an absorptivity of $1 m^2/kg$. Then, with a molar concentration of $300 ppmv$ (like CO_2 in the 1960's), the specific concentration is $4.6 \cdot 10^{-4}$ and the optical thickness $\kappa_o q_{Oob} p_s/(g \cos\bar{\theta}$ is 9.4 within the absorption band. Since the atmosphere is optically thick in this wavenumber region, infrared radiation in this part of the spectrum exits the atmosphere with the temperature of the stratosphere. This is exactly what we see in the graph. Outside the absorption band, the atmosphere is transparent, and hence infrared leaves the top of the atmosphere at the much higher temperature of the ground. The overall appearance of the OLR spectrum is that the greenhouse gas has "dug a ditch" in the spectrum of OLR, or perhaps "taken a bite" out of it. The ditch in the spectrum reduces the total OLR of the planet, but not so much so as if the absorption were strong throughout the spectrum, as would be the case for a grey gas. This is the typical way that real greenhouse gases work: they make the atmosphere optically thick in a limited part of the spectrum, while leaving it fairly transparent elsewhere. The strength of the greenhouse effect is not so much a matter of how deep the ditch, but how wide.

Oobleck is a very contrived substance, but the above exercise gives us a fair idea of what to look for when interpreting real observations of the spectrum of OLR. Figure 4.5, giving the OLR spectrum of Mars observed at two times of day by the TES instrument on Global Surveyor, is a case in point. Mars has an essentially pure CO_2 atmosphere complicated only be optically thin ice clouds and dust clouds (which can be very thin between major dust storms). The planet thus provides perhaps the purest illustration of the CO_2 radiative effect available in the Solar system. In Figure 4.5 a CO_2 "ditch" centered on about $650 cm^{-1}$ is evident both in the afternoon and sunset spectra. At the trough of this ditch, the radiation exits the atmosphere with a radiating temperature of about $170 K$ both in the afternoon and sunset cases. This temperature is somewhat warmer than the coldest temperature encountered in the upper atmosphere of Mars (see Fig 2.2), but is still compatible with the strong decrease of temperature with height seen in the soundings. Away from CO_2 ditch, the atmosphere appears transparent, and the emission resembles the blacbody

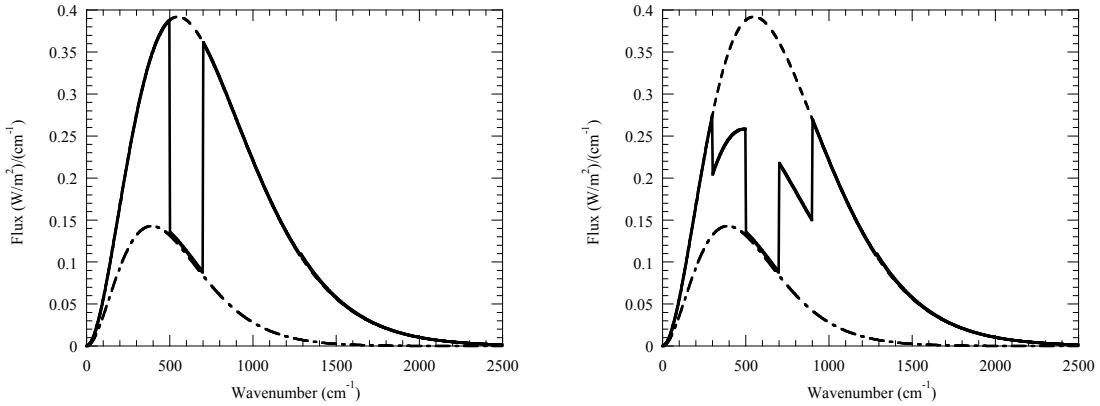

Figure 4.4: **CAPTION. The calculation was carried out for $T_s = 280K$ and $T_{strat} = 200K$. **LeftPanel **RightPanel

emission from a land surface having temperature $265K$ in the afternoon case and $212K$ in the sunset case. These numbers are compatible with the observed range of ground temperature on Mars, cross-checked by near-surface data from landers.

In a situation like that shown in the left panel of Figure 4.4, the OLR is as low as it is going to get. Increasing the greenhouse gas concentration q_G cannot lower the OLR further since, in the spectral region where the gas is radiatively active the atmosphere is already radiating at the coldest available temperature.

From Eq. 4.8, 4.9 or 4.10, if we know the transmission function, we can carry out the integral needed to obtain the radiative fluxes. As we shall see shortly, in most cases the dependence of κ on wavenumber is so intricate that solving the problem by doing a brute-force integral over wavenumber is prohibitive if one aims to do the calculation enough times to gain some insight from modelling a climate (even in a single dimension). In any event, doing the calculation with enough spectral resolution to directly resolve all the wiggles in $\kappa(\nu)$ provides much more information about spectral variability than is needed in most cases. What we really want is to understand something about the properties of the transmission function averaged over a finite sized spectral region of width Δ, centered on a given frequency ν. Specifically, let's choose Δ to be small enough that the Planck function B and its derivative dB/dT are both approximately constant over the spectral interval of width Δ. In that case, when the solution for the flux given in Eq. 4.9 or its alternat forms is averaged over Δ, B can be treated as nearly independent of ν and taken outside the average. In consequence, the resulting band-averaged equations have precisely the same form as the original ones, save that the fluxes are replaced by average fluxes like

$$\bar{I}_+(\nu, p) = \frac{1}{\Delta}\int_{\nu-\Delta/2}^{\nu+\Delta/2} I_+(\nu', p)d\nu' \tag{4.53}$$

and the transmission function is replaced by

$$\bar{\mathfrak{T}}_\nu(p, p') = \frac{1}{\Delta}\int_{\nu-\Delta/2}^{\nu+\Delta/2} \mathfrak{T}_{\nu'}(p, p')d\nu' \tag{4.54}$$

We need to learn how to derive properties of $\bar{\mathfrak{T}}_\nu(p, p')$. The essential challenge is that the nonlinear exponential function stands between the statistics of κ_ν and the statistics of $\mathfrak{T}_\nu$.

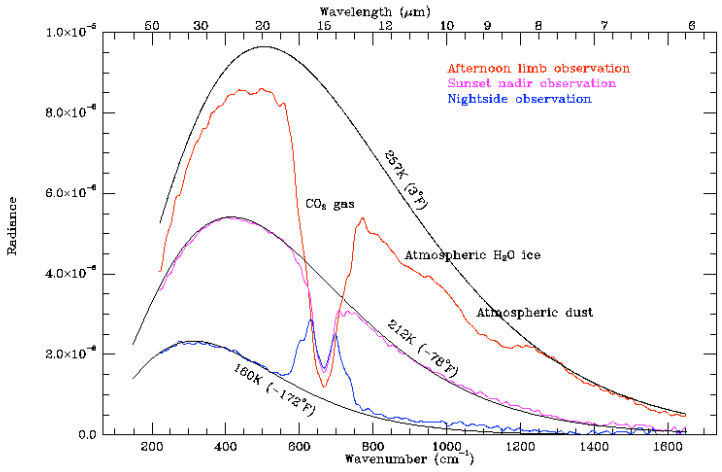

Figure 4.5: **Mars TES. **PLACEHOLDER** (replace with redrawn TES figure)

The transmission function satisfies the *multiplicative property*, that $\mathfrak{T}_\nu(p_1, p_2) = \mathfrak{T}_\nu(p_1, p')\mathfrak{T}_\nu(p', p_2)$ if p' is between p_1 and p_2. The multiplicative property means that the transmission along a path through the atmosphere can be obtained by taking the product of the transmissions along any number of constitutent parts of the path. The band-average transmission loses this valuable property, because for two general functions f and g, $\int f(\nu)g(\nu)d\nu \neq (\int f(\nu)d\nu)(\int g(\nu)d\nu)$. The equality holds only if the two functions are uncorrelated, which is not generally the case for the transmission in two successive parts of a path. In the first part of the path, the strongly absorbed frequencies are used up first, and are no longer available for absorption in the second part of the path. The system has memory, and one can thing of the light as becoming "tired," or depleted more and more in the easily absorbed frequencies the longer it travels, with the result that the absorption in the latter parts of the path are weaker than they would be if fresh light were being absorbed.

4.4.2 The absorption spectrum of real gases

We will now take a close look at the absorption properties of CO_2, in order to introduce some general ideas about the nature of the absorption of infrared radiation by molecules in a gas. Continuing to use CO_2 as an example, these ideas will be developed in Sections **??**,**??** and **??** into a computationally efficient means of calculating infrared fluxes in a real-gas atmosphere. A survey of the spectral characteristics of selected other greenhouse gases will be given in Sections **??** and **??**.

Figure 4.6 shows the absorption coefficient of CO_2 as a function of wavenumber, for pure CO_2 gas at a pressure of $1bar$ and a temperature of $296K$. In some spectral regions, e.g. 1700-1800 cm^{-1}, CO_2 at this temperature and pressure is essentially transparent. This is a *window region* through which infrared can easily escape to space if no other greenhouse gas intervenes. For a $285K$ blackbody, $60W/m^2$ can be lost through this window. There are two major bands in which absorption occurs. For Earthlike temperatures, the lower wavenumber band, from about 450 to 1100 cm^{-1} is by far the most important. At $285K$ the blackbody emission in this band

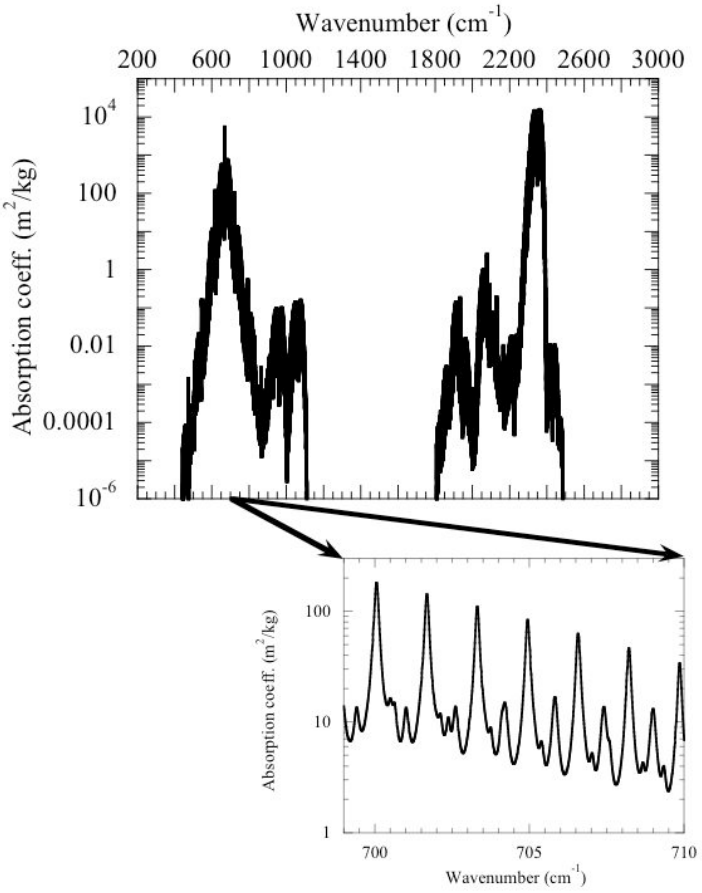

Figure 4.6: **CAPTION

is $218 W/m^2$ out of a total of $374 W/m^2$, so the absorption in this band is well tuned to intercept terrestrial infrared and to thus reduce OLR. The blackbody emission in the higher wavenumber band, from 1800 to 2500 cm^{-1}, is only $6 W/m^2$. This band has a minor effect on OLR for Earth, but it can become important for much hotter planets like Venus, and even for Earth is important for the absorption of solar near-infrared. Within either band, the absorption coefficient varies by more than eight orders of magnitude.

The absorption does not vary randomly. It is arrange around six peaks (three in each major band), with the overall envelope of the absorption declining approximately exponentially with distance from the peak. However, there is a great deal of fine-scale variation within the overall envelope. Zooming in on a typical region in the inset to Figure 4.6 we see that the absorption can vary by an order of magnitude over a wavenumber range of only a few tenths of a cm^{-1}. Most significantly, the absorption peaks sharply at a discrete set of frequencies.

Why does the absorption peak at preferred frequencies? In essence, molecules are like little radio receivers, tuned to listen to light only at certain specific frequencies. Since energy is conserved,

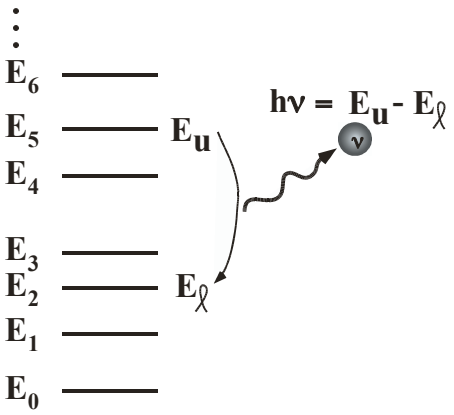

Figure 4.7: Schematic of emission of a photon by transition from a higher energy state to a lower energy state

the absorption or emission of a photon must be accompanied by a change in the internal energy state of the molecule. It is a consequence of quantum mechanics that the internal energy of a molecule can only take on values drawn from a finite set of possible energy states, the distribution of which is determined by the structure of the molecule. If there are N state, there are $N(N-1)/2$ possible transitions, and each one leads to a possible absorption/emission line as illustrated in Figure 4.7. Transitions between different energy states of a molecule's electron configuration almost invariably correspond to visible or ultraviolet frequencies. The energy states involved in infrared absorption and emission are connected with displacement of the nuclei in the molecule, and take the form of vibrations or rotations. Every molecule has an equilibrium configuration, in which each nucleus is placed so that the sum of the electromagnetic forces from the other nuclei and from the electron cloud sum up to zero. A displacement of the nuclear positions will result in a restoring force that brings the system back toward equilibrium, leading to vibrations. The nucleii can be thought of as being connected with quantum-mechanical springs (one between each pair of nucleii) of different spring constants, and the vibrations can be thought of as arising from a set of coupled quantum-mechanical oscillators. Rigid molecules, held together by rigid rods rather than springs, would have rotational states but not vibrational states. The fact that molecules are not rigid causes the rotational states to couple to the vibrational states, through the coriolis and centrifugal forces.

Noble gases (He, Ar, etc.) are monatomic, have only electron transitions, and are not active in the infrared. A diatomic molecule (Fig. 4.8) has a set of energy levels associated with the oscillation caused by pulling the nuclei apart and allowing them to spring back; it also has a set of energy levels associated with rotation about either of the axes perpendicular to the line joining the nuclei. Centrifugal force couples the stretching to the rotation. Triatomic molecules (Fig. 4.9) have an even richer set of vibrations and rotations, especially if their equilibrium state is bent rather than linear (Fig. 4.10). Polyatomic molecules like CH_4, NH_3, SF_6 and the chlorofluorcarbons (e.g. CFC-12, which is CCl_2F_2) have yet more complex modes of vibration and rotation. As the set of energy states becomes richer and more complex, the set of differences between states fills in more and more of the spectrum, making the molecule a better infrared absorber.

For a molecule to be a good infrared absorber and emitter, it is not enough that it have transitions whose energy corresponds to the infrared spectrum. In order for a photon to be absorbed

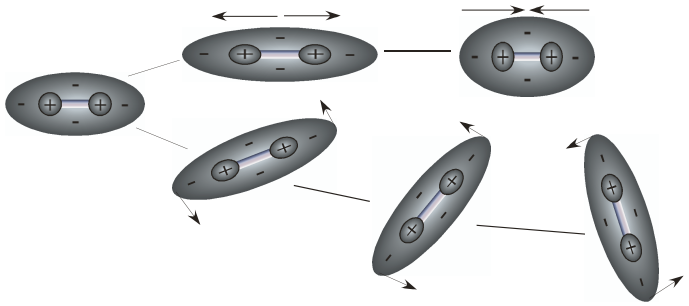

Figure 4.8: Vibration and rotation modes of a diatomic molecule made of a pair of identical atmos, with associated charge distributions

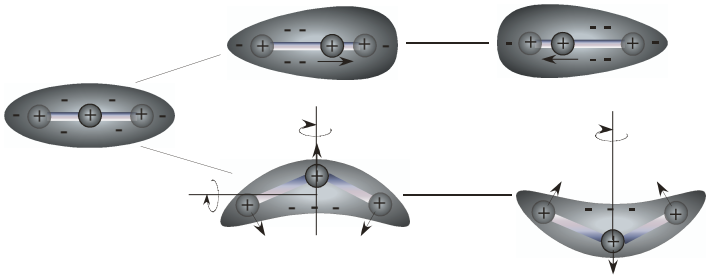

Figure 4.9: Vibration and rotation modes of a linear symmetric triatomic molecule (like CO_2), with associated charge distributions

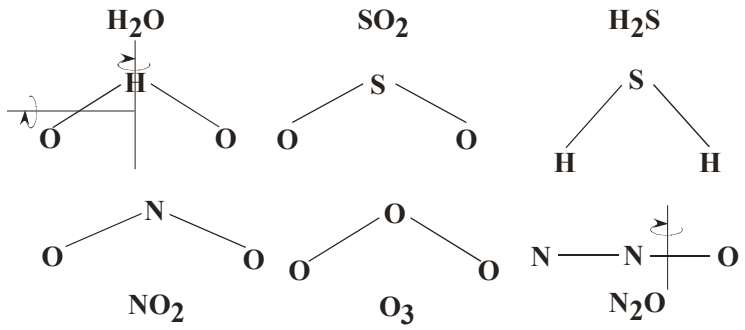

Figure 4.10: Some polar triatomic molecules. Two different modes of rotation are indicated for the H_2O molecule. There is a third mode of rotation about an axis perpendicular to the page.

or emitted, the associated molecular motions must also couple strongly to the electromagnetic field. Although the quantum nature of radiation is crucial for many purposes, when it comes to the interaction of infrared or longer wavelength radiation with molecules, one can productively think of the interaction in semiclassical terms. The reason is that the wavelength of infrared is on the order of 10 microns, which is two to three orders of magnitude larger than the size of the molecules we will be considering. Thus, one can think of the infrared light as providing a large scale fluctuating electric and magnetic field which alters the environment in which the molecule finds itself, and exerts a force on the constituent parts of the molecule. This force displaces the nuclei and electron cloud, and excites vibration or rotation. Conversely, a vibrating or rotating molecule creates a moving charge distribution, which classically radiates an electromagnetic wave. While one must fully take into account quantum effects in describing molecular motion, one need not for our purposes confront the much harder problem of quantizing the electromagnetic field as well (the problem of "quantum field theory"). The only way in which we make use of the quantum nature of the electromagnetic field is in converting the energy difference $E_u - E_\ell$ into a frequency of light, via $\Delta E = h\nu$.

The strongest interaction is between an electromagnetic field and a particle with a net charge. A charged particle will experience a net force when subjected to an electric field, which will cause the particle to accelerate. However, ions are extremely rare throughout most of a typical planetary atmosphere. The molecules involved in determining a planet's energy balance are almost invariably electrically neutral. The next best thing to having a net charge is to have a disproportionate part of the molecule's negatively charged electron cloud bunched up on one side of the molecule, while a compensating excess of positive charged nuclei are at the other side. This creates a *dipole moment*, which experiences a net torque when placed in an electric field, causing the dipole axis to try to align with the field.

Many common atmospheric molecules have no dipole moment in their unperturbed equilibrium state. Such *nonpolar* molecules can nonetheless couple strongly to the electromagnetic field. They do so because vibration and rotation can lead to a dipole moment through distortion of the equilibrium positions of the electron cloud and the nucleii. As illustrated in Figure 4.8, diatomic molecules made of two identical atoms, do not acquire a dipole moment under the action of either rotation or stretching. This is why N_2, O_2 and H_2 are essentially transparent to infrared radiation. An important exception to this rule is the *continuum absorption* to be discussed in Section 4.4.8, in which molecules in dense (and preferably also cold) atmospheres can form associations that are long-lasting enough to act like larger asymmetric molecules. On Titan, continuum absorption is practically the only kind of absorption there is.

CO_2 is a linear molecule with the two oxygens symmetrically disposed about the central carbon, as illustrated in Figure 4.9. A uniform stretch of such a molecule does not create a dipole moment, but a vibrational mode which displaces the central atom from one side to the other does. In addition, bending modes of CO_2 have a fluctuating dipole moment, which can in turn be further influenced by rotation. Both these modes are illustrated schematically in Figure 4.9. Modes of this sort make CO_2 a very good greenhouse gas - the more so because the typical energies of the transitions involved happen to correspond to frequencies near the peak of the Planck function for Earthlike temperatures.

Some molecules – called *polar* have a dipole moment even in their undisturbed state. Most common diatomic gases made of two different elements – notably HF and HCl – are polar, and their vibrational and rotational modes cause fluctuations in the dipole which make them quite good infrared absorbers. They are not commonly thought of as greenhouse gases, because they are highly chemically reactive and do not appear in radiatively significant quantities in any known planetary atmosphere. However, one must keep an open mind about such things. Most triatomic

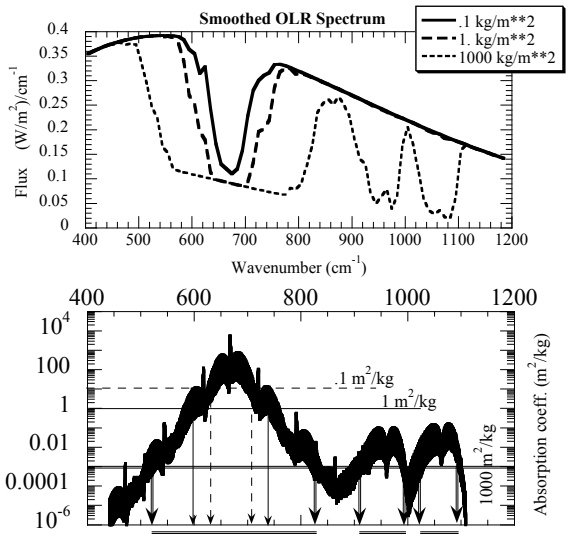

Figure 4.11: **CAPTION [**OLR smoothed over 10 cm**-1]

atmospheric gases (H_2O, SO_2,O_3, NO_2 and H_2S, among others) are polar. CO_2, a symmetric linear molecule with the carbon at the center, is a notable exception. Ammonia (NH_3) is also polar, having its three hydrogens sticking out on one side like legs of a tripod attached to the nitrogen atom at the other side. Polar molecules couple strongly to the electromagnetic field, and their asymmetry also gives them a rich set of coupled rotation and vibration modes with many opportunities for transitions corresponding to the infrared spectrum. The spectrum is enriched because rotation about the axis with the largest moment of inertia (shown as the vertical axis for the water molecule in Figure 4.10) causes the wing molecules to fling outwards, changing the bond angle and the dipole moment. The molecule can also rotate about an axis perpendicular to the plane of the Figure, leading to distinct set of energy levels. Further, energy can be stored in rotations about the axis with minimum moment of inertia (shown as horizonal in the Figure). For a linear molecule like CO_2, rotation about the corresponding axis has essentially no energy.

With the notable exception of the collision-induced continuum discussed in Section 4.4.8, the absorption spectrum of a gas is built by summing up the contributions of the thousands of spectral lines from each of the radiatively active constituents of the gas. To proceed further, then, we must look more deeply into the nature of the lines and how they are affected by pressure and temperature.

4.4.3 I walk the line

An individual spectral line is described by a line *position* (i.e. the wavenumber at the center), a line *shape*, a line *strength* (or *intensity*), and a line *width*. The line shape is described by a nondimensional function of nondimensional argument, $f(x)$, normalized so that the total area under the curve is unity. The contribution of a single spectral line to the absorption coefficient for

substance G can then be written

$$\kappa_G(\nu, p, T) = \frac{S}{\gamma} f(\frac{\nu - \nu_c}{\gamma}) \tag{4.55}$$

where ν_c is the frequency of the center of the line, S is the line intensity and γ is the line width. Note that $\int \kappa_G d\nu = S$. As a line is made broader, the area remains fixed, so that the absorption in the wings increases at the expense of decreased absorption near the center.

The pressure and temperature dependence of κ_G enters almost entirely through the pressure and temperature dependence of S and γ. The line center ν_c can be regarded as independent of pressure and temperature for the purposes of computation of planetary radiation balance. At very low pressures (below $1000Pa$), one may also need to make the line shape dependent on pressure.

Every line has an intrinsic width determined by the charactistic time for spontaneous decay of the higher energy state (analogous to a radioactive half-life). This width is far too narrow to be of interest in planetary climate problems. In addition, the lines of a molecule in motion will experience *Doppler broadening*, associated with the fact that a molecule moving towards a light source will see the frequency shifted to higher values, and conversely for a molecule moving away. For molecules in thermodynamic equilibrium, the velocities have a Gaussian distribution, and so the line shape becomes $f(x) = exp(-x^2)/\sqrt{\pi}$. The width is $\gamma = \gamma(T) = \nu_c \frac{v}{c}$, where $v = \sqrt{2RT}$, R being the gas constant for the molecule in question. v is a velocity, which is essentially the typical speed of a molecule at temperature T. For CO_2 at $250K$, the Doppler line width for a line with center $600cm^{-1}$ is only about $.0006cm^{-1}$.

The type of line broadening of primary interest in planetary climate problems is *collisional broadening*, alternatively called *pressure broadening*. Collisional broadening arises because the kinetic energy of a molecule is not quantized, and therefore if a molecule has experienced a collision sufficiently recently, energy can be borrowed from the kinetic energy in order to make up the difference between the photon's energy and the energy needed to jump one full quantum level. The theory of this process is exceedingly complex, and in many regards incomplete. There is a simple semi-classical theory that predicts that collision-broadened lines should have the *Lorentz line shape* $f(x) = 1./(\pi \cdot (1 + x^2))$, and this shape seems to be supported by observations, at least within a hundred widths or so of the line center. For the Lorentz shape, absorption decays rather slowly with distance from the center; 10 half-widths γ from the center, the Lorentz absorption has decayed to only $\frac{1}{101}$ of its peak value, whereas the Gaussian doppler-broadened line has decayed to less that 10^{-43} of its peak. There are indications that the very far tails of collision broadened lines may die off somewhat faster than predicted by the Lorentz shape, but such issues (important though they be) are beyond the level of sophistication we aspire to here.

In the simplest theories leading to the Lorentz line shape, the width of a collision-broadened line is proportional to the mean collision frequency, i.e. the reciprocal of the time between collisions. For many common planetary gases the width is on the order of a tenth of a cm^{-1} when the pressure is 1 *bar* and the temperature is around $300K$. For fixed temperature, the collision frequency is directly proportional to pressure, and laboratory experiment shows that the implied proportionality of line width to pressure is essentially exact. Holding pressure fixed, the density goes down in inverse proportion to temperature while the mean molecular velocity goes up like the square root of temperature. This should lead to a collision frequency and line width that scales like $1/\sqrt{T}$. Various effects connected with the way the collision energy affects the partial excitation of the molecule lead to the measured temperature exponent differing somewhat from its ideal value of $\frac{1}{2}$. Putting both effects together, if the width is known at a standard state (p_o, T_o), then it can be

extrapolated to other states using

$$\gamma(p,T) = \gamma(p_o,T_o)\frac{p}{p_o}(\frac{T_o}{T})^n \tag{4.56}$$

where n is a line-dependent exponent derived from quantum mechanical calculations and laboratory measurements. It is tabulated along with standard-state line widths in spectral line databases. One must typically go to very low pressures before Doppler broadening starts to become important. For example, for a collision-broadened line with width $.1cm^{-1}$ at 1 bar, the width doesn't drop to values comparable to the Doppler width until the pressure falls to $6mb$ – comparable to the middle stratosphere of Earth or the surface pressure of Mars. Even then, the collision broadening dominates the absorption when one is not too close to the line center, because the Lorentz shape tails fall off so much more gradually than the Gaussian.

Standard spectral databases tabulate the self-broadened and air-broadened widths at standard temperature and pressure, but if one were interested in, say, broadening of NH_3 by collisions with H_2, one would have to either find specialized laboratory experiments or extrapolate based on molecular weights and hope for the best.

The line intensities are independent of pressure, but they do increase with temperature. For temperatures of interest in most planetary atmospheres, the temperature dependence of the line intensity is well described by

$$S(T) = S(T_o)(\frac{T}{T_o})^n \exp(-\frac{h\nu_\ell}{k}(\frac{1}{T}-\frac{1}{T_o})) \tag{4.57}$$

where n is the line-width exponent defined above and $h\nu_\ell$ is the energy of the lower energy state in the transition that gives rise to the line. This energy is tabulated in standard spectroscopic databases, and is usually stated as the frequency ν_ℓ. Determination of the lower state energy is a formidable task, since it means that one must assign an observed spectral line to a specific transition. When such an assignment cannot be made, one cannot determine the temperature dependence of the strength of the corresponding line.

Now let's compute the average transmission function associated with a single collision-broadened spectral line in a band of wavenumbers of width Δ. We'll assume that the line is narrow compared to Δ, so that the absorption coefficient can be regarded as essentially zero at the edges of the band. Without loss of generality, we can then situate the line at the center of the band. The mean transmission function is

$$\bar{\mathfrak{T}}(p_1,p_2) = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2}[\exp(-\frac{1}{g\pi}\int_{p_1}^{p_2}\frac{S(T)\gamma q}{\nu'^2+\gamma^2}dp)]d\nu' \tag{4.58}$$

where $\nu' = \nu - \nu_c$. The argument of the exponential is just the optical thickness of the layer between p_1 and p_2, and to keep the notation simple we will assume the integral to be taken in the sense that makes it positive. The double integral and the nonlinearity of the exponential make this a hard beast to work with, but there are two limits in which the result becomes simple. When the layer of atmosphere between p_1 and p_2 is optically thin even at the center of the line, where absorption is strongest, the line is said to be in the *weak line regime*. All lines are in this regime in the limit $p_2 \to p_1$, though if the line is very narrow or the intensity is very large, the atmospheric layer might have to be made exceedingly small before the weak line limit is approached. For weak

lines the exponential can be approximated as $\exp(-\delta\tau) \approx 1. - \delta\tau$, whence

$$\begin{aligned}\bar{\mathfrak{T}}(p_1,p_2) &\approx 1 - \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \frac{1}{g\pi\cos\bar{\theta}} \int_{p_1}^{p_2} \frac{S(T)\gamma q}{\nu'^2+\gamma^2} dp) d\nu' \\ &= 1 - \frac{1}{\Delta}\frac{1}{g\cos\bar{\theta}}\int_{p_1}^{p_2} S(T) q dp \\ &= 1 - \frac{1}{\Delta} S(\bar{T})\ell_w\end{aligned} \tag{4.59}$$

where $\bar{T} \equiv (T(p_1) + T(p_2))/2$ and the *weighted path* for strong lines is

$$\ell_w(p_1,p_2) \equiv \frac{1}{g\cos\bar{\theta}}\int_{p_1}^{p_2} \frac{S(T(p))}{S(\bar{T})} q(p) dp \tag{4.60}$$

Note that for weak lines, the averaged transmission is independent of the line width. From the expression for $\bar{\mathfrak{T}}$ we can define the *equivalent width* of the line, $W \equiv S(\bar{T})\ell_w$. To understand the meaning of the equivalent width, imagine that absorption takes *all* of the energy out of the incident beam within a range of wavenumbers of width W, leaving the rest of the spectrum. undisturbed. The equivalent width W is defined such that the amount of energy thus removed is equal the amount removed by the actual absorption, which takes just a little bit of energy out of each wavenumber throughout the spectrum.

When the layer of atmosphere between p_1 and p_2 is optically thick at the line center, the transmission is reduced to nearly zero there. This defines the *strong line* limit. For strong lines, there is essentially no transmission near the line center; all the transmission occurs out on the wings of the lines. Since essentially nothing gets through near the line centers anyway, there is little loss of accuracy in replacing the line shape by it's far-tail form, $\pi^{-1}S\gamma/\nu'^2$. With this approximation to the line shape, the band-averaged transmission may be written:

$$\begin{aligned}\bar{\mathfrak{T}}(p_1,p_2) &\approx \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} [\exp(-\frac{1}{\nu'^2}\frac{1}{g\pi\cos\bar{\theta}}\int_{p_1}^{p_2} S(T)\gamma q dp)] d\nu' \\ &= \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} \exp(-\frac{X}{\nu'^2}) d\nu' \\ &= \frac{1}{2\zeta_m}\int_{-\zeta_m}^{\zeta_m} \exp(-\frac{1}{\zeta^2}) d\zeta\end{aligned} \tag{4.61}$$

where $X \equiv \sqrt{S(\bar{T})\gamma(p_o,T_o)\ell_s}/\pi$, and the weighted path for strong lines is

$$\ell_s \equiv \frac{1}{g\cos\bar{\theta}}\int_{p_1}^{p_2} \frac{S(T(p))}{S(\bar{T})}\frac{p}{p_o} q(p) dp \tag{4.62}$$

The third line in the expression for $\bar{\mathfrak{T}}$ comes from introducing the rescaled dummy variable $\zeta \equiv \nu'/\sqrt{X}$; the limit of integration then becomes $\zeta_m = \Delta/(2\sqrt{X})$ Unless the path is enormous, ζ_m will be very large, because the averaging interval Δ is invariably taken to be much larger than the typical line width (otherwise there would be little point in averaging). For $\zeta_m >> 1$, the integral in the last line can be evaluated analytically, and is

$$\int_0^{\zeta_m} \exp(-\frac{1}{\zeta^2}) d\zeta \approx \zeta_m - \sqrt{\pi} \tag{4.63}$$

(see Problem **??**). Therefore, substituting for X, the expression for $\bar{\mathfrak{T}}$ in the strong limit becomes

$$\bar{\mathfrak{T}}(p_1, p_2) \approx 1 - \frac{1}{\Delta} 2\sqrt{S(\bar{T})\gamma(p_o)\ell_s} \tag{4.64}$$

For strong lines the equivalent width is $W \equiv 2\sqrt{S(\bar{T})\gamma(p_o)\ell_s}$. In this case, the width of the chunk taken out of the spectrum increases like the square root of the path because the absorption coefficient decreases like $1/\nu'^2$ with distance from the line center, implying that the width of the spectral region within which the atmosphere is optically thick scales like the square root of the path. Unlike weak lines, strong lines really do take almost all of the energy out of a limited segment of the spectrum. The multiplicative property for transmission is equivalent to an additive property for equivalent width. The nonlinearity of the square root linking path to equivalent width in the strong line case thus means that the band-averaged transmission has lost the multiplicative property. As in our earlier general discussion of this property, the loss stems from the progressive depletion of energy in parts of the spectrum near the line center.

The pressure-weighting of the strong-line path reflects the fact that, away from the line centers, the atmosphere becomes more optically thick as pressure is increased and the absorption is spread over a greater distance around each line.

In solving radiative transfer problems related to planetary climate, one typically takes the bandwidth Δ large enough that the band contains a great many lines. For example, there are about 600 CO_2 lines in the band between 600 and 625 cm^{-1}. In the weak line limit the transmission is linear in the absorption coefficient, so one can simply sum the equivalent widths of all the lines in the band to obtain the total equivalent width $W = \sum W_i$. For strong lines, the situation is a bit more complicated, because of the nonlinearity of the exponential function. For the same reasons one loses the multiplicative property of transmission upon band averaging, one generally loses the additive property of equivalent widths. There is one important case in which additivity of equivalent widths is retained, however. If the lines are *non-overlapping*, in the sense that they are far apart compared to the width over which each one causes significant absorption, then the absorption from each line behaves almost as if the line were acting in isolation. In this case, each line essentially takes a distinct chunk out of the spectrum, and the equivalent widths can be summed up to yield the net transmission.

The additivity of strong-line equivalent widths breaks down at large paths. Since each W_i increases like the square root of the path, eventually the sum exceeds Δ, leading to the absurdity of a negative transmission. What is going wrong is that, as the equivalent widths become large, the absorption regions associated with each line start to overlap. One is trying to take away the same chunk of the spectrum more than once. This doesn't work for spectra any more than it works for ten hungry people trying to eat an eighth of a pizza each.

$$\bar{\mathfrak{T}} = \exp(-\frac{1}{\Delta}\sum W_i) \tag{4.65}$$

Note that when the sum of the equivalent widths is small compared to Δ, this expression reduces to the previous expression given for individual or non-overlapping lines.

4.4.4 Behavior of the band-averaged transmission function

Although the absorption spectrum has very complex behavior, the band-averaged transmission function averages out most of the complexity. The definition of the transmission guarantees that it decays monotonically as $|p_1 - p_2|$ increases and the path increases, but in addition the decay is

invariably found to be smooth, proceeding without erratic jumps, kinks or other complex behavior. This smoothness is what makes computationally economical radiative transfer solutions possible, and the various schemes for carrying out the calculation of fluxes amount to different ways of exploiting the smoothness of the band-averaged transmission function.

By way of example, the band-averaged transmission function for CO_2 is shown for three different bands in Fig. 4.12. The calculation of $\bar{\mathfrak{T}}_\nu(p_1, p_2)$ was carried out using a straightforward – and very time consuming – integration of the transmission over frequency; at each frequency in the integrand, one must do an integral of $\kappa_{CO2}(\nu, p)$ over pressure, and each of those κ must be evaluated as a sum over the contributions of up to several hundred lines. Temperature was held constant at $296K$ and a constant mass-specific concentration of .0005 ($330ppmv$) of CO_2 mixed with air was assumed. The pressure p_1 was held fixed at $100mb$, while p_2 was varied from $100mb$ to $1000mb$. This plot thus gives an indication of the upward flux transmitted from each layer of the atmosphere, as seen looking down from the Earth's tropical tropopause. The results are plotted as a function of the pressure-weighted strong-line path, which for constant q and T is $q \cdot (p_2^2 - p_1^2)/(2gp_o \cos \bar{theta})$, where the reference pressure p_o is taken to be $10^5 Pa$. Plotting the results this way makes it easier to compare them with theoretical expectations, and also makes it easier to generalize the results to transmission between different pairs of pressure levels, which will have different amounts of pressure broadening. The rationale for using the strong-line path is that the lines are narrow enough that almost all parts of the spectrum are far from the line centers in comparison to the width, and in such cases the collision-broadened absorption coefficient increases linearly with pressure almost everywhere. This behavior is incorrect near the line centers, but the error in the transmission introduced by this shortcoming is minimal, since the absorption is so strong there the contribution to the transmission is essentially zero anyway. This reasoning – based directly on what we have learned from the strong-line limit – is at the basis of most representations of pressure-broadening effects in radiative calculations. Here, we only are using it as a graphical device, since the transmission itself is computed without approximation. Note that the strong line path becomes proportional to the (weak-line) mass path $q \cdot (p_2 - p_1)/(g \cos \bar{theta})$ when $p_2 \to p_1$, with proportionality constant p_1/p_o. In the present calculation, when p_2 is at its limit of $1000mb$, the path is about $5kg/m^2$, which is about half the unweighted mass path over the layer. This reflects the fact that the lower pressure over most of the layer weakens the absorption relative to the reference value at $p = p_o$.

Apart from noticing that the transmission function is indeed smooth, we immediately remark that the transmission first declines sharply, as portions of the spectrum with the highest absorption coefficient are absorbed. At larger paths, the spectrum becomes progressively more depleted in easily-absorbed wavenumbers, and the decay becomes slower. For the two strongly absorbing bands in the left panel, the transmission curve becomes nearly vertical at small paths, as suggested by the square-root behavior of the strong line limit. There is guaranteed to be a weak-line region at sufficiently small paths, where the slope becomes finite, but in these bands the region is so tiny it is invisible. In fact, the strong line transmission function in Eq. 4.64 fits the calculated transmission in the 575-600 cm^{-1} band almost exactly throughout the range of paths displayed, when used with the random-overlap modification in Eq. 4.65. For the more strongly absorbing 600-625 cm^{-1} band the fit is very good out to paths of 1.5 kg/m^2, but thereafter the actual transmission decays considerably more rapidly than the strong-line form. This mismatch occurs because the derivation of the strong-line transmission function assumes that the absorption coefficients within the band approach zero arbitrarily closely: as more and more radiation is absorbed, there is always some region where the absorption coefficient is arbitrarily close to zero, which leads to ever-slower decay. In reality, overlap between the skirts of the lines leads to finite-depth valleys between the peaks (see the inset of Fig. **??**), and the absorption is bounded below by a finite positive value. The decay of the transmission at large paths is determined by the local minima in the valleys, and will tend

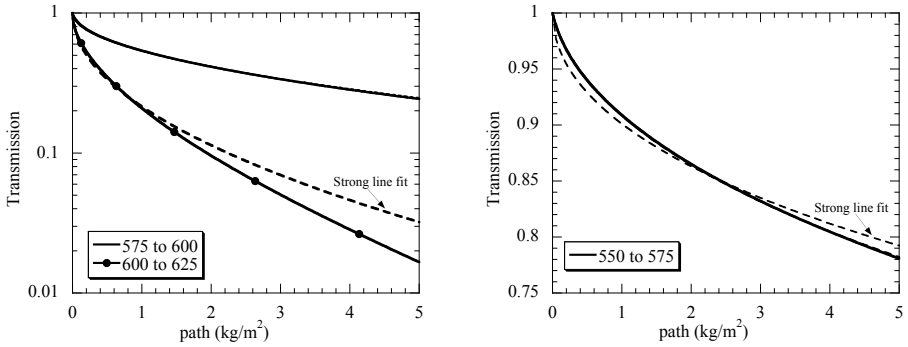

Figure 4.12: The band averaged transmission as a function of path, for the three different bands, as indicated. In each case, the transmission is computed between a fixed pressure $p_1 = 100mb$ and a higher pressure p_2 ranging from $100mb$ to $1000mb$. Calculations were carried out assuming the temperature to be constant at $296K$, with a constant CO_2 specific concentration of q = .0005, and assuming a mean propagation angle $\cos th\bar{e}ta = \frac{1}{2}$. Results are plotted as a function of the pressure-weighted path for strong lines, $q \cdot (p_2^2 - p_1^2)/(2gp_o \cos th\bar{e}ta)$, where $p_o = 1000mb$. In the left panel, the best fit to the strong-line transmission function is shown as a dashed curve; the fit is essentially exact for the $575 - 600cm^{-1}$ band, so the fitted curve isn't visible. For the weaker absorption band in the right panel, fits are shown both for the strong line and the Malkmus transmission function, but the Malkmus fit is essentially exact and can't be distinguished.

toward exponential decay, rather than the slower decay predicted by the strong line approximation.

For the weakly absorbing band shown in the right panel of Fig. 4.12, a hint of weak-line behavior can be seen at small values of the path, with the result that the behavior diverges noticeably from the best strong-line fit. The representation of the transmission can be improved by adopting a two-parameter fit tailored to give the right answer in both the weak and strong limits. The *Malkmus model* is a handy and widely-used example of this approach. It is defined by

$$\sum W_i = 2\frac{R^2}{S}\frac{p_1}{p_o}\left(\sqrt{1 + \frac{S^2}{R^2}\left(\frac{p_o}{p_1}\right)^2 \ell_s} - 1\right) \tag{4.66}$$

where R and S are the parameters of the fit [3]. The parameters can be identified with characteristics of the absorption spectrum in the band by looking at the weak line (small ℓ_s) and strong line (large ℓ_s) limits. For small ℓ_s, the sum of the equivalent widths is $S \cdot (p_o/p_1)\ell_s = S\ell_w$, so by comparing with Eq. 4.59 we identify S as the sum of the line intensities. For large ℓ, the sum is $2\sqrt{R^2\ell}$, whence on comparison with Eq. 4.64 we identify R^2 as the sum of $\gamma_i(p_o)S_i$ for all the lines in the band. The parameters R and S can thus be determined directly from the database of line intensities and widths, though in some circumstances it can be advantageous to do a direct fit to the results of a line-by-line calculation like that in 4.12 instead. One uses the Malkmus equivalent-width formula with the random-overlap transformation given in Eq. 4.65, so as to retain validity at large paths. With the Malkmus model, the transmission function in the weakly absorbing 550-575 cm^{-1} band can be fit almost exactly. Since the Malkmus model reduces to the strong ling form at large paths,

[3]The factor p_1/p_o deals with the difference between the strong line and weak line paths, and is necessary so that the limits work out properly for small and large path. There is some flexibility in defining this factor. It is common to use $\frac{1}{2}(p_1 + p_2)/p_o$ to make things look more symmetric in $p_1 and p_2$. This slightly changes the way the function interpolates between the weak and strong limits, without changing the endpoint behavior

it fits the transmission functions in the left panel of Fig. 4.12 at least as well as the strong line curve did. However, it does nothing to improve the fit of the strongly absorbing case at large paths, since that mismatch arises from a failure of the strong-line assumption itself.

The Malkmus model is a good basic tool to have in one's radiation modelling toolkit. It works especially well for CO_2, and does quite well for a range of other gases as well. There are other fits which have been optimized to the characteristics of different greenhouse gases (e.g. *Fels-Goody* for water vapor), and fits with additional parameters. Most of the curve-fit families have troubles getting the decay of the transmission right when very large paths are involved, though if the trouble only appears after the transmission has decayed to tiny values, the errors are inconsequential.

Empirical fits to the transmission function are a time-honored and effective means of dealing with infrared radiative transfer. This approach has a number of limitations, however. We have already seen some inadequacies in the Malkmus model when the path gets large; patching up these problems leads to fits with more parameters, and finding fits that are well-tailored to the characteristics of some new greenhouse gas one wants to investigate can be quite involved. It also complicates the implementation of the algorithm to have to use different classes of fits for different gases, and maybe even according to the band being considered. A more systematic and general approach is called for. The one we shall pursue now, known as *exponential sums*, has the additional advantage that it can be easily generalized to allow for the effects of scattering, which is not possible with band-averaged fits like the Malkmus model. As a gentle introduction to the subject, let's consider the behavior of the integral

$$\bar{\mathfrak{T}}(\ell) = \frac{1}{\Delta}\int_{\nu_o-\Delta/2}^{\nu_o+\Delta/2} e^{-\kappa_G(\nu)\ell} d\nu \tag{4.67}$$

where κ_G is the absorption coefficient for a greenhouse gas G and ℓ is a mass path. This would in fact be the exact expression for the band-averaged transmission for a simplified greenhouse gas whose absorption coefficient is independent of pressure and temperature. In this case, the path ℓ between pressure p_1 and p_2 is simply the unweighted mass path $|\int qdp|/(g\cos\bar{\theta})$, which reduces to $q|p_1 - p_2|/(g\cos\bar{\theta})$ if the concentration q is constant.

The problem we are faced with is the evaluation of the integral of a function $f(x)$ which is very rapidly varying as a function of x. The ordinary way to approximate the integral is as a Riemann-Stieltjes sum, dividing the interval up into N sub-intervals $[x_j, x_j + 1]$ and summing the areas of the rectangles, i.e.

$$\int_0^1 f(x)dx \approx \sum_1^N f(\frac{x_j + x_{j+1}}{2})(x_{j+1} - x_j) \tag{4.68}$$

The problem with this approach is that a great many rectangles are needed to represent the complex area under the curve $f(x)$. Instead, we may define the function $H(a)$, which is the sum of the lengths of the intervals for which $f(x) \leq a$, as illustrated in Fig. 4.13. Now, the integral can be approximated instead by the sum

$$\int_0^1 f(x)dx = \int_{f=f1}^{f_2} f dH(f) \approx \sum_1^M \frac{f_j + f_{j+1}}{2}(H(f_{j+1} - H(f_j)) \tag{4.69}$$

where we have divided the range of the function f (i.e. $[f_1, f_2]$) into M partitions. This representation can be very advantageous if $H(f)$ is a much more smoothly varying function than $f(x)$.

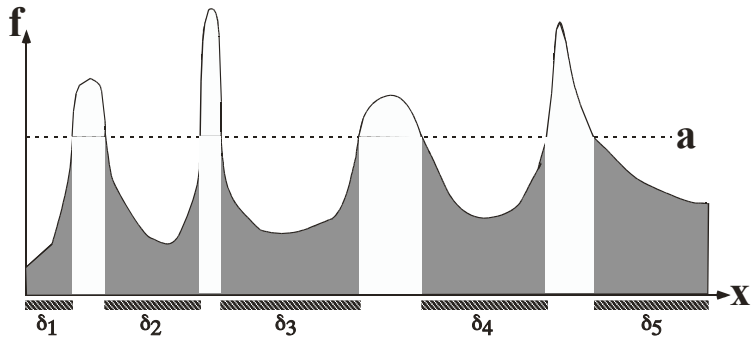

Figure 4.13: **CAPTION

The idea is to apply the Lebesgue integration technique to the integral to the transmission function defined in Eq 4.67, with the absorption coefficient κ_G playing the role of f and the frequency ν playing the role of x. Thus, if $H(a)$ is the sum of the lengths of the frequency intervals in the band for which $\kappa_G \leq a$, then $H(a) = 0$ when a is less than the minimum of κ_G and $H(a)$ approaches the bandwidth Δ when a approaches the maximum of κ_G. The transmission function can then be written approximately as

$$\bar{\mathfrak{T}}(\ell) = \int_{\kappa_1}^{\kappa_2} e^{-\kappa_G \ell} dH(\kappa_G) \approx \sum_{1}^{M} e^{-(\kappa_{j+1}+\kappa_j)\ell/2}(H(\kappa_{j+1}) - H(\kappa_j)) \tag{4.70}$$

This is the exponential sum formula. It can be regarded as an M term fit to the transmission function, much as the Malkmus model is a two-parameter fit.

Because the absorption coefficient varies over such an enormous range, it is more convenient to work with $H(\ln \kappa)$ rather than $H(\kappa)$. A typical result for CO_2 is shown in Fig. 4.14, computed for two bands at a pressure of $100mb$. The function is quite smooth, and can be reasonably well characterized by ten points or less. In contrast, given that the typical line width at $100mb$ is only $.01cm^{-1}$, evaluation of the transmission integral in the Riemann form, Eq. 4.68, would require at least 25000 points in a band of width $25cm^{-1}$. Thus, the exponential sum approach is vastly more economical of computer time than a direct line-by-line integration would be.

4.4.5 A homebrew radiation model

We have now laid out all the ingredients that go into a real gas radiation model, and are ready to begin assembling them. The ingredients are and the recipe is:

- **Divide spectrum into bands
- **Prepare in advance: Malkmus or H for each band for gas or gases
- **Program up a function to compute band avg transmission function in each band
- **Use it in a numerical approx to Eq. ** for each band, to get the band avgd fluxes
- Sum up the fluxes in each band to get the total flux
- Serve and enjoy

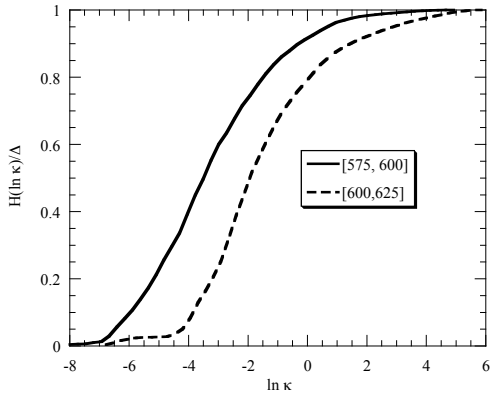

Figure 4.14: Cumulative probability function of the natural log of the absorption coefficient for CO_2. Results are given for the 600-625 cm^{-1} and 575-600 cm^{-1} bands, and were computed at a pressure of $100mb$ and a temperature of $296K$.

In the typical climate simulation application, one is given a list of values of temperature and greenhouse gas concentrations tabulated on a finite grid of pressure levels p_j for $j = 0, ...N$, and one must compute the fluxes based on this information. Either of Eq. 4.8 or 4.9 provides a suitable basis for numerical evaluation when one is working from atmospheric data tabulated on a grid. In writing down the approximate expressions for the flux, we will adopt the convention that $j = 0$ at the top of the atmosphere and that $j = N$ represents the ground. We shall use the superscript (k) to refer to quantities averaged or integrated over the band k, centered on frequency $\nu^{(k)}$ and having width $\Delta^{(k)}$. Let's define the gridded quantities:

$$\begin{aligned} B_j &\equiv B(\nu^(k), \frac{1}{2}(T_j + T_{j+1}))\Delta^(k) \\ \bar{\mathfrak{T}}_{ij}^{(k)} &\equiv \bar{\mathfrak{T}}^(k)(p_i, p_j) \\ e_{ij} &\equiv \bar{\mathfrak{T}}_{ij}^{(k)} - \bar{\mathfrak{T}}_{i(j+1)}^{(k)} \end{aligned} \tag{4.71}$$

The trapezoidal-rule approximation to the the expression for upward flux in band (k), based on Eq. 4.8, is then simply

$$I_+^{(k)}(p_i) = I_{+,s}^(k)\bar{\mathfrak{T}}_{iN}^{(k)} + \sum_{j=i}^{N} B_j e_{ij} \tag{4.72}$$

The expression for the downward flux follows a similar form.

4.4.6 CO_2 and Planetary Climate

4.4.7 Spectroscopic properties of selected greenhouse gases

Carbon Dioxide

Water Vapor

Methane

4.4.8 Collision-induced continuum

4.5 OLR results from comprehensive real-gas radiation models

Numerical results using the CCM3 IR radiation code. For the workbook, students will be provided with a python-wrapped version of this radiation code, and will use it for exercises. (For an example of one use of the python-wrapped code, see the web page for the web-driven version, at `geosci.uchicago.edu/ archer/PS134`, in the lab manual section. We will also provide a simple exponential sum code for a single gas, which will illustrate the general principle of how IR radiation codes are constructed.

OLR(T), with and without water vapor feedback. The general concept of water vapor feedback. The importance of UTH. Linear behavior of OLR(T). Effect of WV feedback on the slope. (Graph: OLR(rh,T) for various rh, at fixed CO2)) In this section we stop short of the runaway greenhouse, which will be revisited in Section **.

OLR as a function of CO2 or methane. Logarithmic behavior. Overlap effects. Importance of logarithmic behavior for planetary temperature regulation. We will make sure to give methane equal play with CO2, because of its possible role in the Early Earth pre-oxygen climate.

Use of polynomial OLR fits to derive simplified energy-balance climate models. Application to the zero-D model of ice-albedo feedback.

Trade-off between methane and CO2 in determining OLR. Graph of OLR vs. CO2 and OLR vs CH4 (each in isolation), using NCAR model (verify against MODTRAN or something like that). Figure: Contour plot of OLR for mean surf temp, as a function of CH4 and CO2. (shows effect of band overlap). Effect of overlap with moisture. (could do first two graphs with and without 50dependent on the temperature chosen). (Note that some of this discusssion can be put in the atmospheric evolution chapter instead, but probably the basic contour plot belongs here).

4.5.1 Another look at the runaway greenhouse

4.6 Pure radiative equilibrium for real gas atmospheres

Re-visit same issues we looked at in the grey gas case, but now we need to rely more on numerical calculations. For Earth, this mostly uses CCMrad, but for Mars and idealized calculations we use our homebrew code.

Two main issues to look at: (1) Temperature of the stratosphere in absence of solar absorption. How the temperature can be below the grey-gas skin temperature. Skin temperature for one-band and two-band Oobleck. Skin temperature for one band Malkmus (nonoverlapping lines). How this makes the stratospheric thermal structure dependent on the troposphere (via the spectrum, in contrast to the grey-gas case, for which we only need to know the OLR; this affects the way we solve the full radiative-convective problem (future chapter).

(2)Temperature jump at the surface. This is still there in a dry atmosphere with CO2 as GHG, but in a moist atmosphere, the temperature jump is replaced by a thin layer with a very large temperature gradient.

4.7 Condensed substances: Clouds

IR effect of clouds. Water and water-ice clouds act like grey bodies. Methane and CO2 clouds are different, because IR scattering is important.

IR radiative effect of cloud layers, for water-like clouds (i.e. clouds that are good IR absorbers/emitters, but not scatterers). Absorption spectrum of liquid water (find data). Emissivity of clouds as a function of particle size and condensate mass. (Assume each particle acts like a perfect blackbody, though there ought to be a size cutoff).

Computation of cloud effect on OLR using CCMrad

Chapter 5

Scattering

If one only needed to be concerned with Rayleigh scattering of atmospheres, one could well dispense with this chapter and just make some reference to the effect of atmospheres on planetary albedo. However, clouds play a critical role in planetary climate, and a certain amount of treatment of scattering is needed if one is to understand things like the importance of particle size. I would prefer not to have to incorporate a chapter on scattering, since it is already well treated in many books on radiative transfer, but in order to keep this course self-contained, it is necessary to have this chapter.

This chapter will be kept fairly concise. It will be limited to two-stream approximations to the scattering equations (which have been shown to be sufficient for most climate purposes), including the widely-used δ-eddington formulation for clouds. It would cover Rayleigh scattering and have a resumé of Mie theory.

For planetary climates (notably Titan and Early Mars), it is necessary to simultaneously treat gaseous absorption of IR and scattering of IR by clouds. This is a topic that is little-discussed in existing texts, and I will give a fair amount of attention to it here, concentrating on exponential sum approaches. I will also present some Monte-Carlo simulations highlighting the essential difficulty of the problem.

Aside from the discussion of IR scattering, this chapter would go beyond what is available in other texts, because of the general computational philosophy of the entire book, which will be adhered to here: All algorithms used in results discussed in the chapter will be provided as Python scripts which the student can use for further exploration, in conjunction with the Workbook.

**Solar absorption on Jupiter and Saturn. What drives convection/tropopause

Chapter 6

The Surface Energy Balance

6.1 Formulation of the surface flux problem

This results of this chapter are pertinent to a planet with a distinct surface, which may defined as an interface across which the density increases substantially and discontinously. The typical interface would be between a gaseous atmosphere and a solid or liquid surface. In the Solar system, there are only three examples of bodies having both a distinct surface and a thick enough atmosphere to significantly affect the surface temperature. These are Venus, Earth, Titan and Mars; among these, the present Martian atmosphere is so thin that it only marginally affects the surface temperature, though this situation was probably different early in the planet's history when the atmosphere may have been thicker. Although thin atmospheres have little effect on the surface temperature, the atmosphere itself can still have interesting behavior, and the flux of energy from the surface to the atmosphere provides a crucial part of the forcing which drives the atmospheric circulation. This is the case for example, for the thin Nitrogen atmosphere of Neptune's moon Triton. Apart from the examples we know, it is worth thinking of the surface balance in general terms, because of the light it sheds on the possible nature of the climates of extrasolar planets already detected or awaiting discovery.

The exchange of energy between the surface and the overlying atmosphere determines the surface temperature relative to the air temperature It also turns out that it determines the exchange of mass between the surface and the atmosphere (as in sublimation from a glacier or evaporation from an ocean, lake or swamp). Because outer space is essentially a vacuum, the only energy exchange terms at the top of the atmosphere are radiative. At the surface, energy can be exchanged by means of fluid motions as well as by radiation.

The atmospheric gas in direct contact with the surface must have the same velocity as the surface; because the surface material is so much denser (and in the case of a solid so much more rigid) than the atmosphere, the atmospheric flow must typically adjust to the presence of the surface over a rather short distance. The resulting strong shears lead to random-seeming complex turbulent motions sustained by the kinetic energy of the shear flow near the boundary. We may subdivide the atmosphere into the *free atmosphere* – which is sufficiently far above the surface to be little affected by turbulence stirred up at the surface, and the *planetary boundary layer* (*PBL*, for short) where the transfer of heat, chemical substances, and momentum is strongly affected by surface-driven turbulence. We may further identify the *surface layer*, which is the thin portion of the PBL near the ground within which all the vertical fluxes may be considered independent of

height.

Given that the whole troposphere is created by convection – which is a form of buoyancy-driven turbulence – it is not at once clear why the PBL should exist as a distinct entity from the troposphere in general. The main reason one can typically distinguish the PBL is that mechanically driven turbulence is more trapped near the surface than is buoyancy driven turbulence, and also has distinct time and space scales. On the present Earth, the effect of moisture is also important in maintaining the distinction, since moisture gives deep convection an intermittent character: most of the troposphere-forming mixing takes place in rare convective events, while most of the troposphere remains quiescent most of the time. Because dry (i.e. noncondensing) convection is typically shallower than moist convection, in planets which have both forms the dry convection can often be treated as part of the boundary layer. This is the case for Earth, and likely for other planets with a surface and an atmosphere in which latent heat release is important (Titan and perhaps Early Mars being the only other known examples so far). For planets like Present Mars or Venus, where dry convection is the *only* form of convection, it is less clear that the PBL can be productively distinguished from the troposphere in general. Even in such cases, though, one can identify a constant-flux surface layer.

As in previous chapters, we let T_g be the temperature of the planet's surface. Previously, we used T_{sa} to denote the temperature of the air in immediate contact with the ground, but now we modify the definition somewhat, and allow T_{sa} to be the temperature at the top of the surface layer, assuming the air at the bottom of the surface layer (which is in contact with the ground) has the same temperature as the ground itself. A model of the PBL is necessary to connect T_{sa} to the temperature of the lowest part of the free troposphere. For many purposes, we can dispense with the PBL and patch the surface layer directly to the free troposphere. We shall adhere to this expedient in most of the following discussion.

Now let's discuss, in general terms, how the surface budget affects the climate. The state of the atmosphere and the ground must adjust so that the top-of-atmosphere and surface budgets are simultaneously satisfied. If the atmosphere is optically thick in the longwave spectrum, the top-of-atmosphere budget becomes decoupled from the surface budget, since radiation from the ground and lower portions of the atmosphere is absorbed before it escapes to space. In this case, the determination of T_g can be decomposed into two stages carried out in sequence. First one determines T_{sa} by adjusting this temperature until the top of atmosphere balance is satisfied, assuming that the rest of the troposphere is related to T_{sa} through the appropriate dry or moist adiabat. Then, once T_{sa} is known, one makes use of a model of the surface flux terms to determine the value of T_g which balances the surface budget with T_{sa} fixed at the previously determined value. This can be done without reference to the top-of-atmosphere budget, since the OLR is independent of T_g in the optically thick limit.

If the atmosphere is very optically thin in the longwave spectrum, the OLR is determined entirely by the ground temperature and ground emissivity. Further, since an optically thin atmosphere radiates very little, the only way the atmosphere itself loses energy is through turbulent exchange with the surface. Suppose first that the atmosphere is transparent to solar radiation. In that case, *in equilibrium* the net turbulent exchange between atmosphere and surface must vanish, since otherwise the atmospheric temperature would rise or fall, there being nothing to balance a net exchange. In consequence, the ground temperature will be just what it would have been without an atmosphere despite the presence of turbulence. In this case, one determines the ground temperature as if the planet were in a vacuum, the top of atmosphere budget is automatically satisfied, and then, once T_g is known, the surface budget is used to determine T_{sa}, and (via an adiabat) the rest of the atmospheric structure. It is exactly the inverse of the process used in the optically thick case. In fact, the basic picture is little altered even if the atmosphere absorbs

solar radiation. In that case, the requirement that the atmosphere be in equilibrium implies that any solar radiation absorbed in the atmosphere be passed on to the surface by turbulent fluxes. The result is much the same as if the solar radiation were absorbed directly by the surface; one does the ground temperature calculation as before, but simply remembers to add the atmospheric absorption to the solar energy directly absorbed by the ground. It should be kept in mind that these considerations apply only in equilibrium. Even an optically thin atmosphere can affect the transient behavior of the surface (e.g. in the diurnal or seasonal cycle), as will be discussed in Chapter 8.

In the intermediate case, where the atmosphere is neither optically thick nor thin, one must solve for T_{sa} and T_g simultaneously, so as to find the values that satisfy both the top-of-atmosphere and surface energy budgets. We'll do this crudely in the present chapter through the indroduction of atmospheric transparency factors, and with more elaboration in Chapter 7 where we discuss full radiative-convective models. Generally speaking, though, when the atmosphere is not too optically thin, the surface budget will affect the temperature of the ground. For Earth this temperature is of interest because the ground is where people live and where much of the biosphere resides as well; for a broad range of planets actual or hypothetical the ground temperature also affects chemical processes which determine atmospheric composition, as well as the melting of ices at the surface. We shall see, however, that it is a fairly common circumstance that the surface fluxes effectively constrain the ground temperature to be nearly equal to the overlying air temperature, so that the climate can be determined without detailed reference to how the surface balance works out.

6.2 Radiative exchange

6.2.1 Shortwave radiation

The surface receives radiant energy in the form of shortwave (solar) and longwave (thermal infrared) flux. The shortwave flux incident on the surface is equal to the shortwave flux incident at the top of the atmosphere, diminished by whatever proportion is absorbed in the atmosphere or scattered back to space. We will call the shortwave flux incident on the ground S_g. The shortwave flux absorbed at the surface is then $(1-\alpha_g)S_g$, where α_g is the albedo of the ground. S_g is affected by clouds, atmospheric absorption and atmospheric Rayleigh scattering.

6.2.2 The behavior of the longwave back-radiation

The longwave radiation striking the surface is the infrared *back radiation* emitted by the atmosphere, which was discussed in Chapter 4. The back radiation depends on both the greenhouse gas content of the atmosphere – which determines its emissivity – and the temperature profile. When the atmosphere is optically thick in the infrared, most of the back radiation comes from the portions of the atmosphere near the ground, whereas in an optically thinner atmosphere the back radiation comes from higher – and generally colder – parts of the atmosphere, and is correspondingly weaker. If the atmosphere is very optically thin, the back radiation will be weak regardless of the atmospheric temperature profile, simply because an optically thin atmosphere radiates very little. As in Chapter 4, $I_{-,s}$ will denote the back radiation integrated over all longwave frequencies. The absorbed infrared flux is then $e_g I_{-,s}$, where e_g is the longwave emissivity of the ground. The ground loses energy by upward radiation at a rate $e_g \sigma T_g^4$. Thus, the net infrared cooling of the ground is

$$F_{g,ir} = e_g \cdot (\sigma T_g^4 - I_{-,s}) \tag{6.1}$$

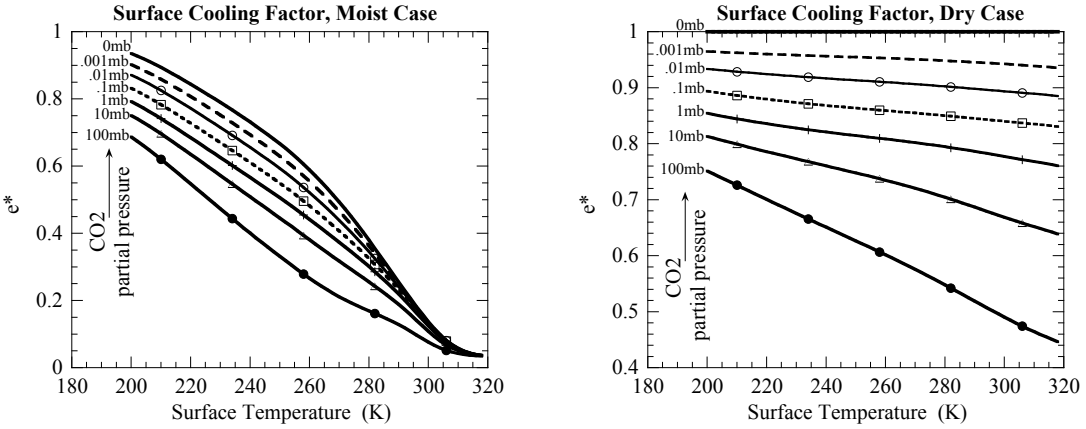

Figure 6.1: Surface cooling factor e^* for a $1bar$ Nitrogen-Oxygen atmosphere with water vapor and CO_2. The surface gravity is that for Earth. In the left panel, the calculations were done with free tropospheric relative humidity set to 50%, and low-level relative humidity set to 80%. Results in the right panel are for a dry atmosphere (zero relative humidity, but with the temperature profile kept the same as in the moist case). In both cases, the numbers on the curves indicate the partial pressure of CO_2 in mb.

According to Eq. 4.18, $I_{-,s}$ approaches σT_{sa}^4 when the atmosphere is optically thick throughout the infrared. In order to characterize the optical thickness of the atmosphere, we introduce the effective low level atmospheric emissivity e_a, defined so that $I_{-,s} = e_a \sigma T_{sa}^4$. e_a depends on the temperature profile as well as the optical thickness, as illustrated by Eq. 4.18 in the optically thick limit. When $T_g = T_{sa}$ the surface cooling becomes $e_g \cdot (1 - e_a)\sigma T_g^4$, which vanishes in the optically thick limit where $e_a \to 1$. Let $e^* = (1 - e_a)$; this is the effective emissivity of the ground when the air temperature equals the ground temperature. If the air temperature is not too different from the ground temperature, we may linearize the term σT_g^4 about $T_g = T_{sa}$, which results in

$$F_{g,ir} = e_g \cdot e^* \sigma T_{sa}^4 + (4\sigma T_g^3 e_g)(T_g - T_{sa}) \tag{6.2}$$

From this equation we can define the infrared coupling coefficient, $b_{ir} = 4\sigma T_g^3 e_g$. When b_{ir} is large, a small temperature difference leads to a large radiative imbalance, and it is correspondingly hard for the ground temperature to differ much from the overlying air temperature. Later, we will derive analogous coupling coefficients for the turbulent transfers.

Figure 6.1 shows how e^* varies with temperature for an Earthlike atmosphere in which the only greenhouse gases are water vapor and CO_2, with the water vapor relative humidity held fixed as temperature is changed. In the moist case (left panel), e^* rapidly approaches zero as the temperature increases; this is because of the increasing optical thickness caused by the increase of water vapor content with temperature (owing to the fixed *relative* humidity). Increasing the CO_2 content also increases the optical thickness, correspondingly reducing e^*. At low temperatures, the CO_2 effect dominates, because there is little water in the atmosphere. However, by the time Earthlike tropical temperatures ($300K$) are reached, water vapor is sufficient to make e^* essentially zero all on its own without any help from CO_2. To underscore the relative role of CO_2 and water vapor, results for a dry atmosphere are given in the right hand panel of Figure 6.1. e^* still goes down with temperature, because temperature affects the opacity of CO_2; however the decline is much less pronounced than it is in the moist case. Even with $100mb$ of CO_2 in the atmosphere, e^*

falls only to about .4 at $320K$, and significant infrared cooling of the surface is possible. In sum, CO_2 by itself is relatively ineffective at limiting surface cooling, but the opacity of water vapor can practically eliminate surface infrared cooling at temperatures above $300K$, unless the ground temperature significantly exceeds the air temperature.

Though the results of Fig. 6.1 were were computed for Earth conditions, they give a fair indication of the extent of surface radiative cooling on other planets whose atmospheres consist of an infrared-transparent background gas mixed with CO_2 and with water vapor fed through exchange with a condensed reservoir. Through the hydrostatic relation, the surface gravity g affects the mass of greenhouse gas represented by a given partial pressure; the lower the g the greater the mass (and hence the greater the optical thickness), and conversely. This is especially important in the case of water vapor, since in that case the partial pressure is set by temperature, through the Clausius-Clapeyron relation. Thus, for a "large Earth" with high g, it takes a higher temperature to make the lower atmosphere optically thick. For example, calculations of the sort used to make Figure 6.1 show that with $1mb$ of CO_2 in a moist atmosphere having temperature $280K$, increasing g to $100m/s^2$ increases e^* to .507 (vs .303 for $g = 10m/s^2$). In the same atmospheric conditions, e^* falls to .102 for a "small Earth" with $g = 1m/s^2$. Increasing the pressure of the transparent background gas makes the greenhouse gases more optically thick through pressure broadening. With $g = 10m/s^2$, increasing the background air pressure to $10bar$ has a very profound effect, lowering e^* to .094. Reducing the air pressure below $1000mb$ should in principle increase e^*, but in fact it is found to very slightly reduce it, to .299. It appears that the reduction in opacity from less pressure broadening is offset by the changes in the moist adiabat that occur when the air pressure is reduced: the latent heat of condensation is spread over less background gas, so the temperature aloft is greater and hence the air aloft contains more water.

Without water vapor, it takes an enormous amount of CO_2 to make the lower atmosphere optically thick. This case is relevant to Venus and Venus-like planets, which may be defined to be planets having a dry rocky surface and a thick, dry CO_2 atmosphere. It takes a very carefully written radiation calculation to handle this regime properly, because many minor absorption bands of CO_2 become important when the atmosphere is very thick. Calculations performed with one of these industrial-strength codes reveals the following, for a dry planet with the gravity of Venus. With surface temperature of $300K$, $e^* = .3$ for a $1bar$ atmosphere, falling further to .014 at $10bar$ and .001 at $100bar$ (somewhat greater than Venus today). However, if the calculation is done with a surface temperature of $700K$ instead, which is close to that of modern Venus, the higher temperature shifts the surface emission to higher wavenumbers where it is not as effectively absorbed and re-emitted by the CO_2 in the atmosphere. In consequence, the values of e^* are uniformly greater: .34 at $1bar$, .11 at $10bar$ and .06 at $100bar$.

In the opposite extreme, atmospheres like the thin Martian atmosphere have very little effect on the surface radiative cooling. For a Martian CO_2 atmosphere on the dry adiabat with $7mb$ of surface pressure, $e^* = .9$ at $220K$, falling only modestly to .86 at $280K$. Recall that, per square meter of surface, Mars actually has vastly more CO_2 in it's atmosphere than the Earth has at present; allowing for the difference in gravity, a $7mb$ pure CO_2 atmosphere on Mars has as much CO_2 per unit area as an Earth atmosphere with a CO_2 partial pressure of $18.5mb$ at the ground. In comparison, the present Earth's atmosphere has a partial CO_2 pressure of a mere $.38mb$ (in 2006). The weak emission of the Martian atmosphere is due to the low total pressure, which yields little collisional broadening of the emission lines. If the same amount of CO_2 on Mars at present were mixed into a $1bar$ atmosphere of N_2, the effective surface emissivity e^* falls to .75 at $230K$ and .69 at $280K$.

Clouds of an infrared-absorbing substance like water act just like a very effective greenhouse gas in making the lower atmosphere optically thick (making e_a close to unity). It takes very little

cloud water to make the lower atmosphere act essentially like a blackbody. Infrared-scattering clouds in the surface layer, like those made of methane or CO_2, have a very different effect on the back-radiation. First, they shield the surface from back-radiation coming down from the upper atmosphere by reflecting it, rather than absorbing it; hence the shielding is accomplished without the cloud layer heating up in response to absorption. More importantly, the downwelling radiation from a reflective cloud is determined by the upwelling ground radiation incident upon it; the resulting back radiation is then determined by the ground temperature, and is independent of the cloud temperature. As a result, the surface cannot increase its longwave cooling by warming up until it is substantially warmer than the atmosphere. This gives a scattering cloud great potency to increase the ground temperature, if it allows sufficient solar radiation to get through to the ground. Either IR-reflecting or absorbing clouds are different from a greenhouse gas, in that they also strongly increase the shortwave albedo.

6.2.3 Radiatively driven ground-air temperature difference

Now we consider the equilibrium temperature difference between the ground and the overlying air that would be attained in the absence of turbulent heat exchange. This temperature difference is important in determining the extent to which convection is driven from below, by positive buoyancy generated near the ground. We have already discussed this issue for the case in which the atmosphere itself is in pure radiative equilibrium (See 3.6,4.3.4 and 4.6). Our concern now is with what happens once convection has set in and altered the atmospheric temperature profile.

If the only heat exchange is radiative, the surface budget reads

$$(1-\alpha_g)S_g + \sigma e_a e_g T_{sa}^4 = \sigma e_g T_g^4 \tag{6.3}$$

Since the second term on the left hand side is positive, the infrared back-radiation always drives T_g to exceed its no-atmosphere value. However, this value might be more or less than T_{sa}. To examine this difference, we linearize the surface radiation budget about T_{sa}, which results in

$$(1-\alpha_g)S_g = \sigma e^* e_g T_{sa}^4 + b_{ir}\cdot(T_g - T_{sa}) \tag{6.4}$$

The linearized form can be immediately solved for the ground-air temperature difference. Substituting the expression for b_{ir}, we find

$$\begin{aligned}(T_g - T_{sa}) &= \frac{1}{4}\frac{(1-\alpha_g)S_g}{\sigma e_g T_{sa}^4}T_{sa} - \frac{1}{4}e^* T_{sa} \\ &= \frac{1}{4}((\frac{T_o}{T_{sa}})^4 - e^*)T_{sa}\end{aligned} \tag{6.5}$$

where T_o is the no-atmosphere ground temperature, which satisfies $\sigma e_g T_o^4 = (1-\alpha_g)S_g$. For planets with an optically thick lower atmosphere, the ground temperature can get extraordinarily hot relative to the air temperature if there are no turbulent fluxes to help carry away the heat. The first term on the right hand side of Eq. 6.5 is large in tropical Earth conditions. For $(1-\alpha_g)S_g = 300W/m^2$ and $T_{sa} = 300K$ with $e_g = 1$, it has the value $49K$. But in tropical Earth conditions, e^* is on the order of .1, so the second term subtracts little ($15K$ for $T_{sa} = 300K$). Thus, the ground temperature is $34K$ warmer than the overlying air temperature, or $334K$. In reality, the sea surface temperature hardly ever gets more than a few degrees warmer than the free-air temperature in the Earth's tropics.

Ironically, for planets which have such a strong greenhouse effect that the low level air temperature is much larger than the no-atmosphere value, $T_g - T_{sa}$ can be quite small even if the

lower atmosphere is optically thick enough to make $e^* \approx 0$, and even in the absence of turbulent heat fluxes. This conclusion is readily deduced from the factor multiplying T_{sa} in the second line of Eq. 6.5. For example, Venus has a small T_o because of the highly reflective clouds which keep sunlight from reaching the surface, yet has a high T_{sa} because of its strong greenhouse effect. In consequence, this factor is only .0024 for Venus in the limit $e^* = 0$, whence $T_g - T_{sa} \approx 1.8K$. Taking $e^* = .06$, as we have estimated for Venus, would actually cause the ground to be slightly cooler than the overlying air. For planets like Venus, the surface radiation budget is dominated by infrared back-radiation, and the comparatively feeble sunlight has little power to drive the ground temperature to values much greater than the overlying air temperature. It is situations like the Earth's tropics, which combine an optically thick lower atmosphere (due to water vapor in our case) with a rather modest greenhouse effect, where the radiation budget tends to drive the ground temperature to large values relative to that of the overlying air.

When the lower atmosphere is optically thin, as in the case of present Mars, the ground-air temperature difference cannot be determined without considering the top-of-atmosphere balance simultaneously with the surface balance. For an optically thin atmosphere, Eq. 6.3 tells us that T_g is just slightly greater than its no-atmosphere value, but it does not by itself tell us how T_g relates to T_{sa}. The general idea for an optically thin atmosphere is that the ground temperature is close to what it would be without an atmosphere, while the atmosphere cools down until the energy it loses by emission is equal to the energy gained by absorption of infrared upwelling from the ground (plus atmospheric solar absorption, if there is any). This generally leaves the low level air temperature much colder than the ground, since the atmosphere loses energy by radiating out of *both* it's top and its bottom. The most straightforward way to make this more precise is to consider the radiative energy budget of the atmosphere, which is the difference between top-of-atmosphere and surface energy budget.

The net infrared radiative flux into the bottom of the atmosphere is $e_g \sigma T_g^4 - e_a \sigma T_{sa}^4$, while the infrared flux out of the top of the atmosphere is the OLR. As discussed in Chapter 4, the OLR is the sum of the emission from the atmosphere itself and the portion of the upward emission from the ground which is transmitted by the atmosphere. Let a_+ be the proportion of upward radiation from the ground which is absorbed by the full depth of the atmosphere, and express the upward atmospheric emission escaping the top of the atmosphere in the form $e_{a,top}\sigma T_{sa}^4$. Then $OLR = e_{a,top}\sigma T_{sa}^4 + (1 - a_+)e_g \sigma T_g^4$. Let's assume for the moment that the atmosphere does not absorb any solar radiation. Then, in the absence of turbulent heat fluxes the atmospheric energy budget reads

$$0 = OLR - (e_g \sigma T_g^4 - e_a \sigma T_{sa}^4) = a_+ e_g \sigma T_g^4 - (e_{a,top} + e_a)\sigma T_{sa}^4 \tag{6.6}$$

whence

$$T_{sa} = (\frac{a_+ e_g}{e_{a,top} + e_a})^{\frac{1}{4}} T_g \tag{6.7}$$

Note that we have not yet made use of the assumption that the atmosphere is optically thin. For an optically thick atmosphere with a very strong greenhouse effect (like Venus), $a_+ \approx e_a \approx 1$ and $e_{a,top} \approx 0$, and so we recover our previous result that $T_{sa} \approx T_g$ for such an atmosphere,provided the emissivity of the ground is close to unity. For an optically thin atmosphere, a_+ $e_{a,top}$ and e_a are all small, so one needs to know precisely how small the absorption coefficient is relative to the two emission coefficients. For an isothermal atmosphere –whether grey or not– Eq. 4.6 implies $e_{a,top} = e_a$. For a grey atmosphere, it follows in addition that $a_+ = e_{a,top} = e_a$. In this case $T_{sa} = T_g/2^{1/4}$, reproducing the result of Section 3.6. When the atmosphere is not grey, the absorption coefficient differs somewhat from atmospheric emission coefficient, because the spectrum of the upwelling radiation from the ground is different from that of the atmospheric emission (by virtue of the difference between ground temperature and air temperature). However, the deviation

from the grey gas result is typically modest for an isothermal atmosphere. For example, a $7mb$ Marslike pure CO_2 atmosphere with a uniform temperature of $230K$ has $a_+ = e_{a,top} = e_a \approx .14$

However, introduction of a vertical temperature gradient strongly affects the relative magnitude of the three coefficients. If we take the same Marslike atmosphere with the same ground temperature and pressure, but stipulate that the temperature is on the dry adiabat rather than isothermal, then a_+ and e_a are reduced slightly (to .116 and .106, respectively), but are still approximately equal. In contrast, $e_{a,top}$ is substantially reduced, to .043. In consequence, the temperature jump at the ground is $T_{sa} = T_g/1.28^{1/4}$ – substantially weaker than the isothermal case, but still quite unstable. Results for a dry Earth, with 300 $ppmv$ of CO_2 in a $1bar$ N_2/O_2 atmosphere having $300K$ surface temperature, are similar: $a_+ \approx e_a \approx .14$ while $e_{a,top} \approx .04$. What is happening in both cases is that the atmosphere appears optically thin when averaged over all wavenumbers, but is really quite optically thick in a narrow band of wavenumbers near the principal CO_2 absorption band. The optical thickness in this range introduces a strong asymmetry in the upward and downward radiation, and also weights the absorption towards the bottom of the atmosphere (which is also where a disproportionate amount of the infrared back-radiation is coming from. A rule of thumb for such cases is that a_+ and e_a will have similar magnitudes, while $e_{a,top}$ will be smaller (but, in the optically thin case, still non-negligable); it follows that the surface temperature jump is weaker than the isothermal case, but still unstable. For an optically thin grey gas the situation is different. In that case, $e_a = e_{a,top}$ and both are less than a_+; nonetheless, the relative magnitudes are such that an unstable temperature jump can generally be sustained at the surface even if the lower atmosphere is on a dry adiabat (see Problem **??**).

The upshot of the preceding discussion is that, in the absence of atmospheric solar absorption, the radiative balance in an optically thin atmosphere almost always drives the surface to be notably warmer than the overlying atmosphere, even if convection has established an adiabat in the atmosphere. This provides a source of buoyancy that can maintain the convection which stirs the troposphere and maintains the adiabat. A moist adiabat is more isothermal than the dry adiabat, so our conclusion is even firmer in that case. Atmospheric solar absorption, on the other hand, would warm the atmosphere relative to the surface, weakening or even eliminating the unstable surface jump.

Moving on, let's consider the temperature the ground of a planet would have in radiative equilibrium at night-time, when $S_g = 0$. In this case, there is little to be gained by linearizing the surface budget, as it reduces to simply $\sigma T_g^4 = e_a \sigma T_{sa}^4$, whence $T_g/T_{sa} = (e_a)^{1/4}$. For an optically thick lower atmosphere, the infrared back radiation keeps the ground temperature nearly equal to the air temperature. However, when the lower atmosphere is not optically thick, the ground temperature plummets at night, or would do so if it had time to reach equilibrium. Cold climates tend to be comparatively optically thin because they cannot hold much water vapor even in saturation. For example, using the moist case in Figure 6.1, we find that when $T_{sa} = 240K$, we find $e_a \approx .3$ with $.1mb$ of CO_2 in the atmosphere. This implies that at night the ground temperature plunges toward the fearsomely cold value $T_g = 177K$. Liquid surfaces like oceans cannot generally cool down rapidly enough to approach the night-time equilibrium temperature, because turbulent motions in the fluid bring heat to the surface which keeps it warm. Solid surfaces like snow, ice, sand or rock can cool down very quickly, though, and do indeed plunge to very low temperatures at night. This situation applies to Snowball Earth and to the present-day Arctic and Antarctic. Very cold climates are of necessity dry, because of the limitations imposed by Clausius-Clapeyron. However, even relatively warm climates can be dry if the moisture source is lacking. This is why deserts can go from being unsurvivably hot in the daytime to uncomfortably cold at night. Turbulent fluxes can bring additional heat to the ground and moderate the night-time cooling somewhat, but these fluxes tend to be weak in the situation just described, because

turbulent eddies must expend a great deal of energy to lift cold dense air from the ground to the outer edge of the surface layer (a matter taken up in more detail in Section 6.5)

The preceding discussion technically applies whether or not T_{sa} itself drops substantially at night, but is most meaningful in the situation where the atmosphere cools slowly enough that the atmosphere remains relatively warm as the night-time ground temperature drops. This is a fair description of the situation in the massive atmospheres of Titan, Earth and Venus, except to some extent during the long polar night on Titan and Earth. The tenuous atmosphere of present Mars, in contrast, cools substantially throughout its depth during the night, even at midlatitudes. In this situation, the relative temperature of air and ground at night is determined by the relative rates of cooling of the two media, rather than radiative equilibrium. We will take up the issue of thermal response time in detail in Chapter 8.

6.3 Basic models of turbulent exchange

Anybody who has watched dry leaves or dust blow around on a windy day has noticed that where the air comes up against the surface there arises a complex mass of turbulent eddies. In comparison, the interior of planetary atmospheres are fairly quiescent places, except in the immediate vicinity of rapidly rising buoyant plumes and active cloud systems. The turbulent fluid motions near the planetary surface exchange energy between the surface and the atmosphere, both in the form of sensible heat (energy corresponding to the change of temperature in a mass) and latent heat (energy associated with the change of phase of a condensible substance, with fixed temperature). Representing the effects of turbulence is not like representing radiation, where we can write down some basic physical principles then proceed through a set of systematic approximations until we arrive at a set of equations we can solve. When it comes to turbulence, the state of physics is not yet up to that challenge, and may never be. Instead, one must take a largely empirical approach from the outset, constrained by some fairly broad principles such as conservation of energy.

In this section we will derive the so-called *bulk exchange* formulae describing the flux of a quantity from the surface to the overlying atmosphere. The general idea is the same whether the quantity is a chemical tracer, sensible heat (associated with temperature fluctuation) or latent heat, so we will first present the formulae for a general tracer.

Let c be the specific concentration of some substance, and c' be the fluctuating or "turbulent" part, usually thought of as a deviation from a time or space mean over some suitable interval. Further, let w' be the fluctuating vertical velocity at the top of the surface layer. Then, the flux of the substance, in kg/m^2, is

$$F_c = \overline{\rho w' c'} \approx \rho_s \overline{w' c'} \tag{6.8}$$

where the overbar represents a time or space average and ρ is the total density of the gas making up the atmosphere. We assume further that the surface layer is thin enough that the variation in pressure and temperature across it is small enough that the variations in density can be neglected. Thus, the density factor can be replaced by a constant typical surface density, ρ_s, and taken outside the average. The ideal gas law states that $\rho = p/RT$. If the surface layer has a thickness of a few tens of meters or less, then the hydrostatic law typically guarantees that the contribution of pressure to the density variations is small. It is not inconceivable, however, that the temperature difference across the surface layer could reach 10% of the mean, leading to corresponding changes in the density. With a little more work, the effect of these fluctuations can be brought into the

picture, but we will not pursue this refinement as the effects are probably overwhelmed by the uncertainties in the representation of turbulence itself.

Next, we must estimate the correlation $\overline{w'c'}$. We build this estimate from a typical vertical velocity δw, a typical concentration fluctuation δc, and a non-dimensional factor $0 < a < 1$ describing the degree of correlation. Thus, we write $\overline{w'c'} = a \cdot \delta w \cdot \delta c$. Next, we assume that δw is proportional to the mean horizontal wind speed U at the top of the surface layer, so $\delta w = s \cdot U$. The constant of proportionality s can be thought of as a typical slope characterizing the turbulent eddies, which is in turn roughly related to the roughness of the surface. Note that U is the wind *speed*, and is therefore positive. We then assume that the typical concentration fluctuation scales with the concentration difference between the air in contact with the ground and the edge of the surface layer, so $\delta c = f \cdot (c_g - c_{sa})$, where c_{sa} is the concentration at the edge of the surface layer, c_g is that at the ground, and f is a nondimensional constant of proportionality. Putting it all together and lumping the proportionality constants into the *drag coefficient* $C_D \equiv a \cdot s \cdot f$, we write

$$F_c = \rho_s C_D U (c_g - c_{sa}) \tag{6.9}$$

C_D is called the *drag coefficient* because when c is taken to be the turbulent velocity itself, the flux formula gives the flux of momentum, and hence the drag force on the surface. In writing the flux in the form of Eq. 6.9, we have adopted the convention that a positive flux represents a transfer of substance from the ground to the atmosphere. The turbulent flux acts like a diffusion, transferring substance from regions of higher concentration to regions of lower concentration. It is like a bucket-brigade, with partly empty buckets being handed downstairs from the top of the surface layer to the ground, where they are filled and sent back upstairs again (or with full buckets sent downstairs to be partly dumped out on the ground). The mass of substance in a bucket being carried upstairs is proportional to $\rho_s c_g$, while the mass of substance in a bucket going downstairs is proportional to $\rho_s c_{sa}$, while $C_D U$ gives the rate at which buckets are being handed up or down the stairs.

6.3.1 Sensible heat flux

To obtain the sensible heat flux, we take $c_p T$ to be our tracer. This is essentially the dry static energy (see Eq. 2.20), since the surface layer is thin enough that the height z can be taken to be nearly constant. With this choice of tracer, Eq. 6.9 becomes

$$F_{sens} = c_p \rho_s C_D U (T_g - T_{sa}) \tag{6.10}$$

If the ground is warmer than the air, heat is carried away from the ground at a rate proportional to the temperature difference. If the ground is cooler than the air, the sensible heat flux instead acts to warm the ground.

If C_D is independent of temperature, then F_{sens} is exactly linear in the difference between the ground temperature and air temperature. Hence the coupling coefficient b_{sens} – analogous to b_{ir} – is simply $b_{sens} = c_p \rho_s C_D U$.

Note that the sensible heat flux becomes small when the atmosphere has low density. The "wind-chill" factor on present Mars would be exceedingly weak! Conversely, very dense atmospheres like those of Venus or Titan can very effectively exchange heat between the surface and the atmosphere. With $C_D = .001$,$U = 10 m/s$ and $T_g - T_{sa} = 1K$ the sensible heat flux is $.13 W/m^2$ on present Mars, $11 W/m^2$ on Earth, $55 W/m^2$ on Titan, and a whopping $540 W/m^2$ on Venus. It is for similar reasons that immersion in near-freezing water is far more life-threatening than walking about scantily clad in air of the same temperature – water is about 1000 times denser than Earth

air. One must take care to distinguish thickness of an atmosphere (in terms of density) from optical thickness. An atmosphere can be thick (i.e. dense) while being optically thin, and conversely a thin (low density) atmosphere can nonetheless by optically thick if the greenhouse gas it is made of is sufficiently effective.

Now let's suppose that the sensible heat flux dominates the surface energy budget. By "dominates," we mean that the sensible heat flux due to a small departure from equilibrium (considering the sensible heat flux alone) overwhelms the other terms in the surface energy balance. This would be true if the wind speed and density were large, provided that the ground and atmosphere are dry enough that evaporation remains small. Sensible heat flux vanishes when $T_g = T_{sa}$, so this is the state that the system is driven to when sensible heat flux dominates. Taking the radiative and latent fluxes into account would cause a small deviation from this limit.

6.3.2 Latent heat flux

Whatever the condensed substance making up the surface, some of the condensed substance will transform into the vapor phase in the atmosphere contacting the surface, until it reaches the saturation vapor pressure determined by Clausius-Clapeyron. If the winds then carry away this vapor-laden air and replace it with unsaturated air, more mass will evaporate or sublimate from the surface. Since the phase change involves latent heat, a flux of mass away from the surface cools the surface by carrying away latent heat. Conversely, a flux of mass from vapor into the condensed surface will warm the atmosphere where condensation occurs. All substances will evaporate or sublimate to some extent, and whether the latent heat flux is significant is a matter of how big the saturation vapor pressure is at the typical temperature of the surface. For water ice on Titan at $95K$, the vapor pressure is under $10^{-15}Pa$, so the latent heat flux of water is utterly negligable. The situation is the same for basalt at $300K$ on Earth, or even at $750K$ on Venus. However, the vapor pressure of CO_2 on present Mars, of liquid water or water ice on Earth, and of methane on Titan are all high enough to allow substantial latent heat flux. Whatever the condensible substance in question we will use terms like "humidity" by analogy with the archetypal case of water vapor on Earth. Also, for the sake of verbal economy we will often refer simply to "evaporation" in situations where the actually process might be either evaporation or sublimation.

In dealing with latent heat flux, it is more convenient to deal with the mass mixing ratio of the condensible to dry air, rather than specific humidity. This makes it somewhat easier to treat cases where the condensible makes up a substantial part of the total mass. Thus, we use the mass mixing ratio r_w as the tracer in Eq. 6.9. If ρ_a is the density of dry air in the surface layer, then the mass of condensible per unit volume is $\rho_a r_w$ and this mass carries a latent heat $L\rho_w r_w$. we can write the mixing ratio r_{sa} at the edge of the surface layer as $h_{sa} r_{sat}(T_{sa})$, where where h_{sa} is the relative humidity at the outer edge of the surface layer and $r_{sat}(T)$ is the saturation mass mixing ratio. In terms of saturation vapor pressure, the saturation mass mixing ratio is $(M_w/M_a)(p_{sat}(T)/p_a$, with p_a being the partial pressure of dry air in the surface layer. Now suppose that at the ground there is a reservoir of a condensed phase of the substance "w" – an ocean, lake, swamp, snow field, glacier or the surface of an icy moon. In this case, the vapor pressure in the air in contact with the surface must be in equilibrium with the condensed phase, and must therefore follow the Clausius-Clapeyron relation evaluated at the temperature of the ground. Equivalently, we can say that $r_g = r_{sat}(T_g)$. Using the two mixing ratios, the latent heat flux becomes

$$F_L = L\rho_a C_D U(r_{sat}(T_g) - h_{sa} \cdot r_{sat}(T_a)) \tag{6.11}$$

Alternately, using the definition of the mixing ratios and assuming the partial pressure of dry air

to be approximately constant within the boundary layer, Eq. 6.11 can be written

$$F_L = \frac{L}{R_w T_{sa}} C_D U (p_{sat}(T_g) - h_{sa} \cdot p_{sat}(T_{sa})) \tag{6.12}$$

The latter form of the latent heat flux demonstrates that the flux is in fact unaffected by the presence of the dry air. Assuming temperature and wind to be held constant, the evaporation from the Earth's ocean would remain unchanged even if all the N_2 were taken out of the atmosphere. This conclusion would no longer be valid if the gases in question had substantial non-ideal behavior, for then the law of partial pressures would no longer hold.

Exercise 6.3.1 Derive Eq. 6.12. What do you have to assume about the air temperature within the surface layer?

From Eq. 6.12 we observe that latent heat flux carries heat away from the ground when the saturation mixing ratio at the ground is less than the mixing ratio of the surface layer. Since typically $h_{sa} < 1$, this can happen even if the ground is colder than the overlying air. We also note that the latent heat flux becomes insignificant at sufficiently cold temperatures, since both saturation vapor pressures in the equation become small in that limit.

Sensible and radiative heat transport carry no mass away from the surface, but latent heat transport is of necessity accompanied by mass transfer. The mass flux into or out of the ground is simply F_L/L. The mass flux is needed for calculating the rate of ablation of glaciers by sublimation, the drying out of lakes or soil by evaporation, and the rate of salinity change at the surface of an ocean (since evaporation carries away the condensible but not the solute).

Now let's look at how the fluxes behave when the temperature difference between the ground and the outer edge of the surface layer is small. Carrying out a Taylor series expansion of the flux about $T_g = T_{sa}$, as we did for the infrared cooling case, we write

$$F_L = E_o + b_L \cdot (T_g - T_{sa}) \tag{6.13}$$

Defining the characteristic flux $F_L^* \equiv C_D U p_{sat}(T_{sa})$, we find

$$E_o = (1 - h_{sa}) \frac{L}{R_w T_{sa}} F_L^*, b_L = \frac{1}{T_{sa}} (\frac{L}{R_w T_{sa}})^2 F_L^* \tag{6.14}$$

where R_w is the gas constant for the condensible. The Clausius-Clapeyron relation has been used to substitute for dp_{sat}/dT in the expression for b_L. E_o is the heat flux due to evaporation or sublimation that would occur with $T_g = T_{sa}$; it vanishes if the surface layer is saturated ($h_{sa} = 1$), but is positive otherwise. Both E_o and b_L are proportional to the characteristic flux F_L^*, which vanishes vanishes as $T_{sa} \to 0$, since the saturation vapor pressure vanishes like $\exp(-L/R_w T)$ in this limit. As one might expect, latent heat flux becomes negligible at sufficiently low temperatures. How low one must go for this to be the case depends on the gas in question. As temperature increases, the characteristic flux becomes large, and hence E_o and b_L become large as well. The increase is abetted by the fact that $L/R_w T$ is a large number at typical planetary temperatures (e.g. 18.06 for water vapor at $300K$, or 10.3 for methane at $95K$). For temperatures high enough that b_L becomes large, a modest ground-air temperature difference leads to a very large increase in latent heat flux. This tends to make it hard for the ground temperature to differ much from the free air temperature in such cases.

Table 6.1 gives some typical values of E_o and b_L for water, carbon dioxide and methane. In all three cases, we see that the latent heat flux rises very strongly with temperature. For water,

	H_2O	H_2O	H_2O	H_2O	CO_2	CO_2	CH_4	CH_4
T_{sa} (K)	230	273	300	320	150	160	80	95
E_o (W/m^2)	.72	40.8	193.3	557.8	52.5	182.1	93.2	640.0
b_L (W/m^2K)	.28	11.2	38.6	98.0	24.4	74.4	55.6	243.

Table 6.1: Latent heat flux coefficients for various gases at selected temperatures T_{sa}. Computed with $U = 10m/s$, $C_D = .001$ and boundary layer relative humidity $h_{sa} = 70\%$.

latent heat flux is insignificant at temperatures of $230K$ or lower. The feeble latent flux of a Watt per square meter or so would be utterly dominated by infrared cooling of the surface, or by the sensible heat flux arising from a ground-air temperature difference of as little as $1K$. This corresponds to the situation in the Antarctic night of the present Earth, or to the daily average tropical temperatures on a Snowball Earth. However, even at the freezing point of water, the latent heat flux is quite substantial. With a $5K$ ground-air temperature difference, the flux would be nearly $100W/m^2$, which is almost half of the typical midlatitude absorbed solar radiation in the ocean, and roughly equal to the typical absorbed solar radiation in ice. The latent heat flux is also comparable to the typical infrared cooling of the surface at such temperatures (inferred from Figure 6.1). As temperature is increased further to values characteristic of the modern tropics, the flux increases dramatically; it would take about 90% of the supply of absorbed solar energy going into the ocean in order to sustain the evaporation arising from just a $2K$ ground-air temperature difference. At these temperatures, the latent flux is considerably in excess of the surface infrared cooling.

For the other gases in the table, the latent heat flux becomes substantial at much lower temperatures. At temperatures comparable to the Martian polar Spring, the latent heat flux due to CO_2 sublimation is comparable to the water vapor values for Earth's midlatitudes or tropics (assuming the same degree of boundary layer saturation). These fluxes are particularly consequential in light of weak supply of solar radiation on Mars, relative to Earth. Alternately one may compare the latent flux to the infrared cooling of the surface in the thin Martian atmosphere (σT_g^4, or $37W/m^2$ at $160K$). Either way, we conclude that latent heat flux plays a key role in determining surface temperature at places on Mars where seasonal CO_2 frost is sublimating or being deposited. At Titan temperatures, latent heat flux due to methane evaporation is enormous; the solar radiation reaching Titan's surface is well under $5W/m^2$, which is two orders of magnitude less than the Methane evaporation one gets under the conditions of Table 6.1. Somehow or other, conditions near Titan's surface must adjust until the evaporation is reduced to the point where it can be balanced by the supply of energy to the surface, but the numbers in the table tell us that methane latent heat flux is the dominant constraint on the adjusted state. Ironically, Titan, at $95K$ is like an extreme form of the Earth's tropics, in that evaporation dominates the surface energy budget to an even greater extent than it does in Earth's tropics. If the temperature of the Earth's tropics were raised to $320K$, as might happen in the high CO_2 world following deglaciation of a Snowball Earth, then E_o on Earth, too would greatly exceed the available solar energy, though not to such an extent as it does on Titan. The way the surface conditions adjust to accomodate this state of affairs will be taken up in the Section 6.4.

When the surface is sufficiently cold relative to the air, vapor from the air can be deposited on the surface in the form of dew or frost. In this case the latent heat flux is negative, and carries energy from the atmosphere to the ground. If the boundary layer is saturated ($h_{sa} = 1$) then frost or dew deposition occurs whenever $T_g < T_{sa}$. If the boundary layer is unsaturated deposition won't occur until the ground temperature is made sufficiently cold that the saturation vapor pressure there falls below the partial pressure of the condensible in the overlying atmosphere (a temperature

known as the "dew point" or "frost point"). When latent heat is being carried to the surface – as it is during he seasonal polar CO_2 frost formation on Mars – the rate of condensation is limited by the rate at which the surface can get rid of the deposited latent heat. Since the surface is colder than the atmosphere during deposition, sensible heat flux carries heat the wrong way to balance the budget, so it is only infrared cooling of the surface that can sustain frost or dew. Otherwise, the surface will simply warm in response to the deposited latent heat until it is no longer cold enough for frost or dew to form.

Over land, there are two further complications that must be considered. The first is that land, unlike a deep ocean or lake or a thick glacier, can dry out. If the land surface is a mix of condensible and (essentially) noncondensible substance, the latent heat flux can exhaust the supply of condensible, whereafter the boundary condition $r_g = r_{sat}(T_g)$ is no longer appropriate. In the absence of further supply of condensible at the ground, the latent heat flux must fall to zero. In such a case, one must keep track of the mass of the condensible reservoir at the ground, and zero out the latent heat flux when the reservoir is exhausted. This would be the case for thin snow cover, scattered puddles, or soil moisture on Earth, for CO_2 frost layers on Mars and for liquid methane swamps on Titan. For soil moisture, a common simple model is the *bucket model*, in which each square meter of soil surface is treated as a bucket whose capacity is determined by its porosity and depth. The bucket is filled by rainfall, and emptied by evaporation. Once the bucket is full, any additional rainfall is assumed to run off into rivers (which may or may not be tracked, according to the level of sophistication of the model). As long as the bucket has some water in it evaporation is sustained, but when the bucket is empty latent heat flux is zeroed out and only radiative and sensible heat transfers at the ground are allowed. The bucket model may serve also as a model of conditions at Titan's surface, which may consist not only of liquid methane puddles but also bogs consisting of beds of granular water ice sand or pebbles whose pores are saturated with liquid methane.

The second complication over land concerns the effect of land plants. At present, Earth's climate provides the only example where this must be taken into account. Plants actively pump water from deep storage, at rates determined by their own physiological requirements. This is known as *transpiration*, and given that moisture flux over vegetated land is always some mix of transpiration and evaporation, the joint process is called *evapotranspiration*. In this case, the moisture boundary condition at the ground may be more appropriately represented as a flux condition determined by plant physiology rather than setting the moisture mixing ratio at the ground. The moisture flux may be limited by rate at which trees pump moisture, and not by rate at which turbulence carries it away. The mixing ratio at the ground still cannot exceed saturation, so when the transpiration becomes strong enough to saturate the air in contact with the ground, one can revert to the previous model of conventional evaporation. Yet a further complication in vegetated terrain is the very notion of ground and ground temperature. Is "the ground" the forest surface or the elevated leaf canopy? Is the ground temperature that of the leaf surface or the soil? How do we take into acount the mix of illuminated hot leaves and relatively cool leaves in shade? A proper treatment of these factors requires a detailed model of the microclimate in the vegetation layer, which is beyond what we aspire to in this book. One need not abandon all hope of estimating conditions over vegetated terrain, however. As a rule of thumb, dense forests that get enough rainfall to survive in the long term tend to act more or less like the ocean, save for an eleveated C_D caused by greater surface roughness. Grasslands, shrub, tundra and prairie can be crudely modeled using the bucket model.

When evaporation dominates the surface budget, equilibrium requires $F_L = 0$, or equivalently $p_{sat}(T_g) = h_{sa}p_{sat}(T_{sa})$. Since p_{sat} is monotonically increasing in temperature, this relation requires $T_g < T_{sa}$ if the boundary layer air is unsaturated $h_{sa} < 1$. Thus, evaporation or sublima-

tion drives the ground temperature to be *colder* than the overlying air temperature. However, the ground and surface could also achieve equilibrium by transferring enough moisture to the surface layer that it becomes saturated ($h_{sa} = 1$), in which case $T_g = T_{sa}$ in equilibrium, as for the case of sensible heat flux. The extent to which equilibrium is attained by adjusting temperature vs. humidity depends on the competition between the rate at which moisture is supplied to the boundary layer and the rate at which dry air from aloft is entrained into the boundary layer. Observed boundary layers on Earth and Titan are significantly undersaturated, leading to the conclusion that the ground temperature would be considerably less than the air temperature, if other fluxes did not intervene. Using the linearized form of the latent heat, the equilibrium ground-air temperature difference is $T_g - T_{sa} \approx -E_o/b_L$. For the conditions of Table 6.1, this is $-2.6K$ for Titan at $95K$. For a hot Earth at $320K$, the difference is about $-5.7K$. There are currently no observations of the state of saturation over the sublimating Martian CO_2 frost cap, but given the saturation assumed in the table the equilibrium occurs with $T_g - T_{sa} \approx -2.4K$ when the air temperature is $260K$. Thus, even when evaporation dominates, the equilibrium ground temperature does not differ greatly from the overlying air temperature. This was also found to be the case when the surface budget is dominated by sensible heat flux. It is only the radiative terms that can drive the ground temperature to be substantially different from the overlying air temperature.

6.4 Joint effect of the fluxes on surface conditions

Including turbulent heat fluxes, the surface energy budget can be written

$$0 = F_{rad} - F_{sens} - F_L \tag{6.15}$$

where F_{rad} is the net radiative flux into the surface, given by

$$F_{rad} = (1 - \alpha_g)S_g + \sigma e_a e_g T_{sa}^4 - \sigma e_g T_g^4 \tag{6.16}$$

Without turbulent fluxes, the surface budget would be $F_{rad} = 0$. F_{rad} in isolation can drive the ground temperature to be either larger or smaller (and perhaps *much* larger or smaller) than the air temperature, according to the circumstances discussed in Section 6.2. Sensible heat flux always drives the ground temperature and air temperature to become identical, whereas latent heat flux drives the ground temperature to be colder than the air temperature, by an amount that depends on the boundary layer relative humidity. When all three fluxes act in concert the resulting behavior depends on the relative importance of the fluxes.

We'll begin our tour of the range of possible behaviors by discussing how the surface balance is accomplished for typical conditions in the Earth's tropical oceans. Take $T_{sa} = 300K$, $C_D = .0015$, $U = 5m/s$ and $h_{sa} = 80\%$. We'll assume the absorbed solar radiation $(1-\alpha_g)S_g$ is $320W/m^2$, which is typical of clear-sky conditions over the tropical ocean. To determine the back-radiation, we need e_a. At tropical temperatures in the moist case, this coefficient is not very sensitive to CO_2, and has a value of about .9. The terms making up the surface balance are shown in the left panel of Figure 6.2. As noted previously, the equilibrium ground temperature would be exceedingly large without turbulent heat flux. In the figure, the no-turbulence equilibrium occurs where F_{rad} crosses zero, at around $336K$. Adding sensible heat flux to the budget makes the slope of the flux curve more negative, and brings the equilibrium ground temperature down to $316K$. Adding in evaporation steepens the curve yet more, and brings the ground temperature down to $303K$, which is only slightly warmer than the $300K$ temperature of the overlying air. At the equilibrium point, the dominant balance is between the evaporation ($206W/m^2$) and the absorbed solar radiation ($320W/m^2$), leaving only $114W/m^2$ to be balanced by the other terms. The sensible heat flux is

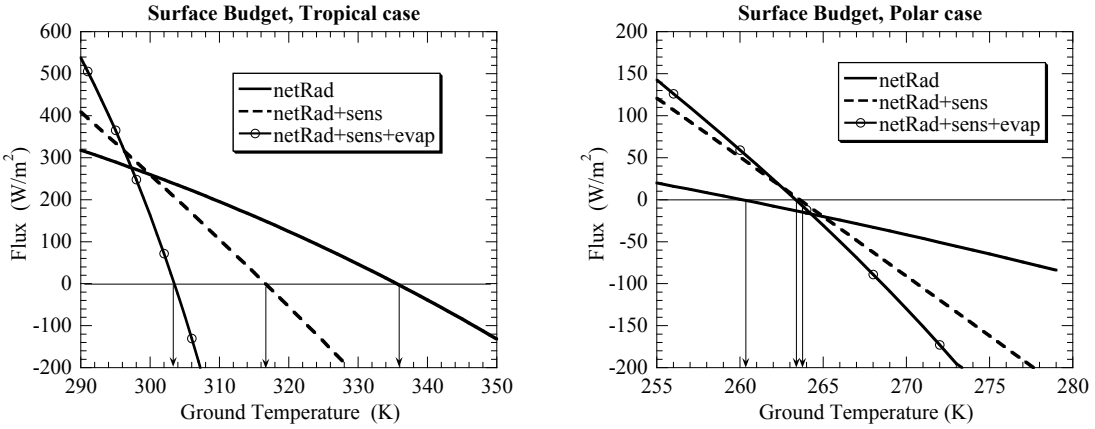

Figure 6.2: **CAPTION

weak because the ground temperature and air temperature are nearly identical, which also makes the net infrared cooling of the surface weak given that $e_a \approx 1$.

Next we'll discuss a typical set of Earth polar or midlatitude winter conditions. We set the absorbed solar flux $(1-\alpha_g)S_g$ to $100W/m^2$, taking a low value on account of the high albedo of snow or ice and the reduced solar flux received at high latitudes. We'll set $T_{sa} = 265K$, in which case $e_a \approx .6$ with $300ppmv$ of CO_2 in the atmosphere. The remaining parameters are held at the same values used in the tropical case. The main differences from the tropical case are that in the cold case the latent heat flux and the infrared back-radiation are weaker – the latter doubly so because of the lower air temperature and the lower e_a. The right panel of Figure 6.2 shows that because of the weak solar radiation and the weak back radiation, the radiative equilibrium surface temperature is nearly $5K$ colder than T_{sa}, in contrast to the tropical case. The situation here is a less extreme version of the night-time radiative equilibrium temperature considered in Section 6.2.3. Since e_a is fairly small the temperature plummets at night when $(1-\alpha_g)S_g = 0$. In the present case, S_g doesn't vanish, but its weak value is insufficient to warm up the ground temperature to the point where it exceeds the air temperature. This is the typical daytime condition in high latitude winter over ice and snow. Warm air imported from low latitudes helps to keep T_{sa} from getting too cold in the polar and midlatitude winter, but the weak sunlight and weak back-radiation leave the ground colder.

Since the radiative equilibrium ground temperature in the cold case is colder than the air temperature, adding in sensible heat flux conveys heat from the atmosphere to the ground, warming the ground up to just over $263K$. The sublimation is weak at such cold temperatures, and causes little additional change in the surface temperature. While the dominant balance in the tropical case was between solar heating and evaporative cooling, the dominant balance in the cold case is radiative, with slight modifications due to sensible heat flux. For any given air temperature, the amount by which the ground temperature departs from the air temperature depends on the absorbed solar radiation, but the sensible heat flux always pulls the ground temperature back towards equality with air temperature. For example, at higher latitudes or deeper in the winter or near sundown, we might take $S = 50W/m^2$. In this case the radiation-only ground temperature is $246.6K$, which is substantially below the air temperature; however, addition of sensible heat flux brings the ground temperature back up to $260K$. Nearer to noon, or as summer approaches, we might have $S = 150W/m^2$. In this case, the radiation-only ground temperature is $271.8K$; again

addition of sensible heat flux brings the ground temperature closer to air temperature, in this case by cooling the ground to $267.1K$, rather than warming it.

Next let's estimate the maximum daytime temperature over a subtropical desert on Earth. Solid surfaces like sand or rock take little time to reach equilibrium, and so the maximum temperature can be estimated by computing the equilibrium temperature at local solar noon. Using the present Earth solar constant and a relatively high albedo of .35 (typical of Sahara desert sand), the absorbed flux is about $890W/m^2$. Over the interior of a dry desert, there should be little moisture in the boundary layer, so set $e_a = .72$ corresponding to a boundary layer relative humidity of 20%. Finally, we take $T_{sa} = 300K$. In these circumstances the radiative equilibrium ground temperature is a torrid $383K$ – hot enough to boil water. When sensible heat flux is added into the budget, heat is transferred from the ground to the air, moderating the surface temperature. Taking a relatively high drag coefficient $C_D = .003$ on account of the roughness of land surfaces, the equilibrium ground temperature is brought down to $330K$ if the surface layer wind speed is $5m/s$. The temperature approaches the radiative temperature as the wind is made weaker; for example when the wind is reduced to $2.5m/s$ the temperature increases to $349K$. Consistent with these estimates, the hottest satellite-observed ground temperatures do indeed occur in subtropical deserts, and are near $340K$. With a wind of $5m/s$, making the ground moist and turning on evaporation brings the equilibrium temperature down from $330K$ to $306K$. The general lesson is that dry surfaces heat up greatly during the daytime. Their maximum temperature can greatly exceed the overlying air temperature, especially when the wind is light. This can contribute to the *urban heat island* effect, since constructed environments often replace moisture-holding surfaces with low albedo impermeable surfaces like asphalt, which hold little water and dry out quickly. The surface heating also leads to amplified climate change over land, in circumstances where a formerly moist soil becomes dry, or *vice versa*.

$$\Delta T \equiv T_g - T_{sa} = \frac{(1-\alpha_g)S_g - e_g \cdot e^* \sigma T_{sa}^4 - E_o}{b_{ir} + b_{sens} + b_L} \tag{6.17}$$

The numerator in this expression is the energy imbalance the surface would have if the ground temperature were equal to the overlying air temperature. It can be either positive or negative and its sign determines the sign of ΔT, since all three terms in the denominator are positive.

6.5 Monin-Obukhov theory

This is here mainly so we can deal with suppression of turbulence in stably stratified boundary layers, which is quite important at night-time and over ice or snow. For moist systems, we point out the importance of water-vapor buoyancy (and its generalizations) in destabilizing boundary layers with the surface colder than the free atmosphere.

6.6 Mass balance and melting

6.7 Precipitation-temperature relations

Here we discuss the factors constraining precipitation. The discussion follows the lines of my discussion of the subject in my Nature article on deep-time hydrological cycle problems. The general

theme is the importance of the surface budget constraint for precipitation, when the temperature is high.

6.8 Simple models of sea ice in equilibrium

Chapter 7

Radiative-convective models

7.1 Dry grey-gas atmospheres

Complete solution for the vertical structure of an atmosphere whose troposphere is on the dry adiabat, and which is patched to a radiative-equilibrium stratosphere. Issue of why there is no stable temperature jump at the tropopause. Behavior of the tropopause height in the optically thin limit (easy; was already done in Chapter 3). Behavior of the tropopause height in the optically thick limit (hard). Reappearance of the $4R/c_p$ rule. Effect of lapse rate on the tropopause height. Diagnosing the convective heat flux.

How to do convection in the time-dependent case (approach to equilibrium). Need to conserve dry static energy.

Effect of upper-atmospheric solar absorption. Nuclear winter. Post-bolide dust layer. Stratospheric ozone. Tholin clouds on Titan and Early Earth. Note that Mars also has substantial internal heating, even in the troposphere, owing to dust. In the Mars case, the heating has a profound effect because of the thin atmosphere, which implies that the solar absorption is spread over a relatively small mass of atmosphere.

Grey gas models of water vapor feedback. Generalization to feedback by similar substances (condensible CO2, or methane).

7.2 Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature?

Discuss why factors that affect the TOA radiation budget have much more leverage over the surface temperature than things that affect the surface budget. It's true that TOA changes work their way into the surface budget, because the low level air temp. changes, and this changes all fluxes between the atmosphere and surface. However these surface flux changes are a cause, not an effect. This can be clarified by using a grey-gas model coupled to a simple surface turbulent exchange scheme. First look at what happens if we make the atmosphere more optically thick aloft, throwing off the TOA budget. Look at how surface adjusts, for various flux coupling coefficients. Then do a case where we increase the downward radiation to the surface by increasing the low

level optical thickness (e.g. add a low cloud), without affecting the OLR. What does this do to surface temperature? Note that when the upper atmosphere is already optically thick, increasing the low level opacity doesn't change the atmospheric cooling, since increased emission to ground is offset by decreased loss of downwelling IR from aloft, and increased emission to upper atmosphere; this does change the amount of convective heat transport, though. What happens if the upper atmosphere is optically thin, though? (e.g. if we have a pure nitrogen atmosphere and introduce a surface cloud layer?) In that case, the increased emissivity could change the OLR (indirectly, through its effect on surface temperature). This is a good case to analyze closely.

Could also look at what happen if we put in a cloud that reduces surface solar absorption, but leaves TOA budget unchanged. (Do equilibrium case only).

Bottom line is that, with enough surface moisture, surface budget perturbations affect precip more than temperature. Exception is dry land, where surface balance can add appreciable warming to the general atmospheric warming.

Note that some aspects of this problem (e.g. the buffering of surface budget changes by "stiffness" provided by evaporative heat transfer) where already discussed in the surface budget chapter.

7.3 Real gas atmospheres

Numerical solutions for radiative-convective equilibrium for real gases.

7.4 Sensitivity of climate to CO_2 changes

Glacial-interglacial cycles. Cretaceous hothouse. Anthropogenic global warming. CO_2 threshold for Neoproterozoic snowball.

In this section, we revisit the ice-albedo feedback bifurcation diagram, with an improved OLR model, including real gas CO2, and also the water vapor feedback. Discuss what climate would be like with just water vapor feedback and no CO2. How close would orbit have to be in order to have an equable climate without CO2.

These estimates will be based on clear-sky radiation, with a few remarks about how clouds complicate the picture.

7.5 Methane-dominated greenhouse

7.6 The atmosphere as a heat engine

In this section, we use the results of Chapter 2 to do vertically integrated budgets of atmospheric dry static energy and of entropy. The consequences of these budgets are discussed. An important aspect of atmospheric operation is that there is a pressure difference between where energy is put into the system (near the surface) and aloft (where energy is lost by IR emission).

The climate is an open system, energetically. It is true that the rate of energy flowing in is

equal to the rate flowing out, but the behavior of the system is strongly influenced by the fact that energy flows *through* the system – specifically that energy comes in in the form of Solar radiation, and goes out in the form of longer wavelength radiation (typically infrared).

Conventional thermodynamic analysis uses pressure and volume. It is possible to do this for the atmosphere, but volume isn't a very natural thermodynamic variable for atmospheric work. Discuss how to avoid explicit use of volume by using the dry (or moist) static energy. (height Z replaces volume).

Idealized atmospheric Carnot cycle: Put heat in by solar heating of air in contact with ground at p_s. Lift adiabatically to p_{rad} and lose heat by IR radiative cooling there. Compress adiabatically back to surface pressure and start over. Note that this process is not a closed loop: the surface temperature you wind up with is lower than what you started with, so you can sustain a bit of mechanical work with that temperature difference. (Estimate this). This process is not quite like the classic Carnot cycle. Note that with latent heat, the atmosphere is like a steam engine (or, following Pauluis, maybe more like a dehumidifier).

7.7 Effect of the diurnal cycle on tropopause height

7.8 The runaway greenhouse revisited

Radiative-convective model calculations of runaway greenhouse. Role of the stratosphere and the tropopause height.

Alternate stratospheric mechanism (OLR limit through moistening of stratosphere).

Role of cloud effects in inhibiting true runaway state.

7.9 Mars, present, past and future

Greenhouse effect of the present thin Martian atmosphere.

Simple models of the interaction between temperature change and the sublimation of polar CO_2 deposits. (This will be based on an assumed pole-equatuator temperature gradient, as for our simple models of ice-albedo feedback. A more precise treatment requires results from the chapter on Seasonal Cycles). Analogy between CO_2 sublimation effects and water vapor feedback on Earth. "Runaway dry-icehouse" due to loss of atmosphere by condensation.

What happens to Mars as the Sun continues to warm? Will Mars become habitable? If it does, will it stay habitable?

Early Mars. Problems with the gaseous CO_2 greenhouse as an explanation (owing to CO_2 condensation). Need for consideration of CO_2 clouds, which will be taken up in next chapter.

7.10 Titan

The Methane greenhouse effect on Titan. Methane condensation feedback (analogous to water vapor feedback). Note that Nitrogen also condenses at Titan temperatures, so that mass of atmo-

sphere changes with the climate.

Since Methane clouds are also infrared scatterers rather than absorbers, a full treatment of Titan's climate also requires IR scattering theory, developed in next chapter. On Titan (and perhaps Early Earth), high altitude tholin clouds are also important.

7.11 Gas Giants

Two features distinguish the problem for gas giants like Jupiter or Saturn: (1) There is no surface to absorb solar radiation, so solar energy is deposited entirely by internal absorption in the Atmosphere; (2) Interior fluid motions can efficiently transport heat from the deep interior to the upper layers, and can yield a heat source comparable to solar absorption. The latter effect is weak on solid planets like Earth, since fluid motions in the very viscous mantle are sluggish.

This section explores the consequence of the above features, in radiative-convective models. The use of planetary energy balance observations in determining the relative role of interior and Solar heat sources was already discussed in Chapter 3.

Chapter 8

Variation of temperature with season and latitude

Why is the Earth generally hotter near the Equator than at the poles? Why is it generally hotter in Summer than in Winter, especially outside the tropics? Would this be true on other planets as well? How would the pattern change over time, as features of the planet's orbit vary? More generally, in this section we seek to understand the features of a planet that determine the magnitude and pattern of geographic and seasonal variations in temperature.

8.1 A few observations of the Earth

First, let's take a look at how the Earth's surface temperature varies with the seasons. Figure 8.1 shows the zonal-mean air temperature near the surface for representative months in each of the four seasons. The first thing we note is that the temperature is fairly uniform in the tropics (30S-30N), but declines sharply as the poles are approached. The temperature difference between the Equator and 60N is 39K in the Winter but only 12K in the Summer. The Southern Hemisphere has a much weaker seasonal cycle, except over the Antarctic continent: The temperature difference between the Equator and 60S is 26K in the Winter and 22K in the Summer. However, over Antarctica, poleward of 60S the seasonal cycle is extreme. Noting that the Northern Hemisphere has more land than the Summer, the data imply that the oceans have a strong moderating effect on the seasonal cycle. The temperature patterns in Figure 8.1 are what we seek to explain in terms of the response of climate to the geographically and seasonally varying Solar forcing.

An even better appreciation of the effect of land masses on the seasonal cycle can be obtained by examining the map of July-January temperature differences, shown in Figure 8.1. This map shows that the strongest seasonal temperature contrast occurs in the interior of large continents, and that the ocean temperature varies by at most a few degrees over the year – and even less in the Tropics. The strong seasonal cycle of the Northern Hemisphere continents extends very little beyond the coastlines, and the seasonal cycle of the Northern oceans has similar magnitude to that of the more extensive Southern oceans.

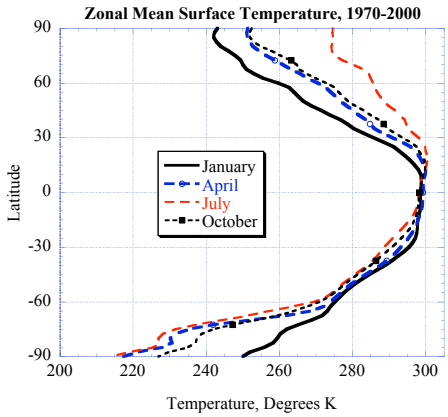

Figure 8.1: Observed zonal mean surface air temperatures for January, April, July and October. Computed from NCEP data for 1970-2000.

8.2 Distribution of incident solar radiation

The geographical variations of temperature are driven by variations in the amount of sunlight falling on each square meter of surface, and also by variations in albedo. Seasonal variations are driven by changes in the geographical distribution of absorbed sunlight as the planet proceeds through its orbit. Therefore, the starting point for any treatment of seasonal and geographical variation must be the study of how the light of a planet's sun is distributed over the spherical surface of the planet. This section dealso only with the distribution of incident sunlight, or *insolation*. The geographical distribution of the amount of sunlight absorbed is affected also by the distribution of the albedo. The albedo variations can also affect the seasonal distribution of solar forcing through seasonal variations in ice,snow, cloud and vegetation cover.

It will help to first consider an airless planet, so that we don't at once have to deal with the possible effects of scattering of the solar beam by the atmosphere. If our planet is far from its Sun, as compared to the radius of the Sun, the sunlight encountering the planet comes in as a beam of parallel rays with flux L. Even if the surface of the planet is perfectly absorbing, the sunlight the planet intercepts is not spread uniformly over its surface; per unit area, parts of the planet where the sun is directly overhead receive a great deal of energy, whereas parts where the Sun grazes the surface at a shallow angle receive little, because the small amount of sunlight intercepted is spread over a comparatively large area, as shown in Figure 8.2. The night side of the planet, of course, receives no solar energy at all.

To obtain a general expression for the distribution of incident solar radiation per unit of surface area, we may divide up the surface of the planet into a great many small triangles, and consider each one individually. The solar energy intercepted by a triangle is determined by the area of the shadow that would be cast by the triangle on a screen oriented perpendicular to the solar beam. To compute this area, suppose that one of the vertices of the triangle is located at

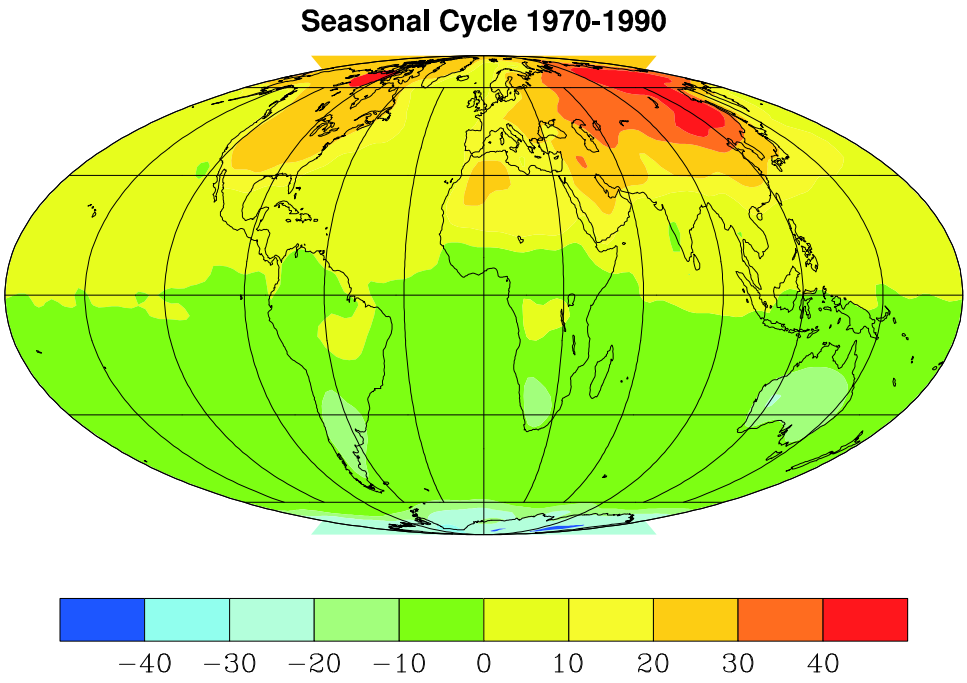

Figure 8.2: Map of July-January surface air temperature difference.

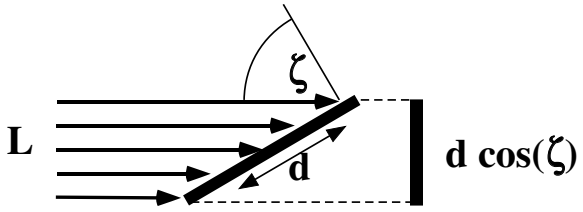

Figure 8.3:

the origin, and that the two sides coming from this vertex are given by the vectors $\vec{r_1}$ and $\vec{r_2}$ By the definition of the cross product, the area of the triangle is given by $2A\hat{n} = (\vec{r_1} \times \vec{r_2})$ where $\hat{n}$ is the unit normal to the plane containing the triangle. To obtain the area of the shadow cast by the triangle, we apply the cross product to the projection of the vectors $\vec{r_1}$ and $\vec{r_2}$ onto the plane. These projections are given by $\vec{r_1} - \hat{z}\vec{r_1} \cdot \hat{z}$ and $\vec{r_2} - \hat{z}\vec{r_2} \cdot \hat{z}$, where $\hat{z}$ is the unit vector pointing in the direction of the Sun. The cross product of these two vectors is

$$(\vec{r_1} - \hat{z}\vec{r_1} \cdot \hat{z}) \times (\vec{r_2} - \hat{z}\vec{r_2} \cdot \hat{z}) = \vec{r_1} \times \vec{r_2} - (\hat{z} \times \vec{r_2})(\vec{r_1} \cdot \hat{z}) - (\vec{r_1} \times \hat{z})(\vec{r_2} \cdot \hat{z}) \tag{8.1}$$

Now, the cross product of two vectors in the xy plane must point in the direction of the z axis. Hence, we can obtain the magnitude of the above vector by taking its dot product with $\hat{z}$. This is very convenient, since the dot product of $\hat{z}$ with the second two terms vanishes, leaving us with

$$2A_\perp = \hat{z} \cdot (\vec{r_1} \times \vec{r_2}) = 2A\hat{z} \cdot \hat{n} = 2A\cos(\zeta) \tag{8.2}$$

where $A_\perp$ is the area of the shadow and ζ is the angle between the normal to the patch of surface and the direction of the sun. This is known as the *zenith angle*. When the the zenith angle is zero,

the Sun is directly overhead, and when it is 90^o the sunlight comes in parallel to the surface and leaves no energy behind. Zenith angles greater than 90^o are unphysical, since they represent light that would have to pass through the solid body of the planet in order to illuminate the underside of the surface; these are on the night side of the planet. If one draws a line from the center of the planet to the center of the Sun, the zenith angle will be zero where the line intersects the surface of the planet; this is the *subsolar point*. At any given instant, the curves of constant zenith angle make a set of concentric circles centered on the subsolar point, with a zenith angle of 90^o along the great circle which at the given instant separates the dayside of the planet from the nightside. If the surface were in equilibrium with the instantaneous incident solar flux, the subsolar point would be the hottest spot on the planet, with temperature falling to zero with distance away from the hot spot. As the planet rotates through its day/night cycle, a given point of the surface is swept through a range of distances from the hot spot, leading to a diurnal temperature variation. As the planet proceeds through its orbit in the course of the year, the diurnal cycle will change as the orientation of the planet's rotation axis changes relative to the Sun. Insofar as the surface actually takes a finite amount of time to heat up or cool down, the diurnal cycle will be attenuated to one extent or another.

As the next step toward realism, let's now consider a rapidly rotating planet whose axis of rotation is perpendicular to the line connecting the center of the planet to the center of its Sun. If the axis of rotation is in fact perpendicular to the plane of the orbit, this situation prevails all year round; otherwise, the condition is met only at the equinoxes, and indeed the condition defines the equinoxes. We assume that the planet is rotating rapidly enough that the day-night difference in solar radiation is averaged out and the corresponding temperature fluctuations are small. In other words, the length of the day is assumed to be short compared to the characteristic thermal response time of the planet's surface, a concept which will be explored quantitatively in Section 8.3. Consider a small strip of the planet's surface near a latitude ϕ, of angular width $d\phi$. If a is the planet's radius, then the area of this strip is $2\pi a^2 \cos(\phi) d\phi$, if angles are measured in radians. The cross section area of the strip seen edge-on looking from the Sun determines the amount of solar flux intercepted by the strip. This area is $2a^2 \cos^2(\phi) d\phi$ when $d\phi$ is small. In consequence, the incident solar radiation per unit area at latitude ϕ is $L \cos(\phi)/\pi$. At the Equator, the solar radiation per unit area is L/π, which is somewhat greater than the value $L/4$ which we obtained in Chapter 3 by averaging solar radiation over the *entire* surface of the planet. If the planet has no atmosphere to transport heat or create a greenhouse effect, the equilibrium temperature as a function of latitude is

$$T = (\frac{L \cos(\phi)}{\pi \sigma})^{\frac{1}{4}} \tag{8.3}$$

The temperature has its maximum at the Equator, and falls to zero at the poles.

Exercise 8.2.1 For the geometric situation described above, derive an expression for the cosine of the zenith angle as a function of latitude and longitude. Re-derive the expression for the daily-average distribution of solar absorption by averaging the cosine of the zenith angle along latitude circles.

Now we turn to the general case, in which the axis of rotation of the planet is not perpendicular to the plane containing the orbit. The angle between the perpendicular to the orbital plane, and the planet's axis of rotation, is known as the *obliquity*, and we shall call it γ. It can be regarded as constant over the course of a planet's year, though there are longer term variations which will be of interest to us later. The task now is to determine the solar zenith angle as a function of latitude, position along the latitude circle, and time of year.

Let the point P be the center of the planet, and S be the center of the sun. If we draw a line from P to S, it will intersect the surface of the planet at a latitude δ, which is called the *latitude of the sun*, or sometimes the *subsolar latitude*. It is a function of the orientation of the planet's axis alone, and serves as a characterization of where we are in the march of the seasons. If the obliquity of the planet is γ, then δ ranges from γ at the Northern Hemisphere summer solstice to $-\gamma$ at the Southern Hemisphere summer solstice. Let Q be a point on the planet's surface, characterized by its latitude ϕ and its "hour angle" h, which is the longitude relative to the longitude at which local noon (the highest sun position) is occurring throughout the globe. For radiative purposes, we just need to compute the zenith angle ζ, defined previously. To get the zenith angle, we only need to take the vector dot product of the vector $\vec{QS}$ and the vector $\vec{PQ}$. To do this, it is convenient to introduce a local Cartesian coordinate system centered at P, with the z-axis coincident with the axis of rotation, the x-axis lying in the plane containing the rotation axis and $\vec{PS}$, and the y-axis orthogonal to the other two, chosen to complete a right-handed coordinate system.

First, note that by the definition of the dot product,

$$\cos(\zeta) = \frac{\vec{PQ} \cdot \vec{QS}}{|PQ||QS|} \tag{8.4}$$

Further, $\vec{PQ} + \vec{QS} = \vec{PS}$, so

$$\cos(\zeta) = \frac{\vec{PQ} \cdot \vec{PQ}}{|PQ||QS|} + \frac{\vec{PQ} \cdot \vec{PS}}{|PQ||QS|} \approx \frac{\vec{PQ} \cdot \vec{PS}}{|PQ||QS|} \tag{8.5}$$

where we drop the first term based on the assumption that the radius of the planet is a small fraction of its distance from the Sun. For the same reason, $|QS|$ in the denominator can with good approximation be replaced by $|PS|$, leaving the expression in the form of a dot product between two unit vectors. Letting $\hat{n}_1 = \vec{PQ}/|PQ|$ and $\hat{n}_2 = \vec{PS}/|PS|$, the unit vectors have the following components in the local Cartesian coordinate system.

$$\hat{n}_1 = (\cos(\phi)\cos(h), \cos(\phi)\sin(h), \sin(\phi)), \hat{n}_2 = (\cos(\delta), 0, \sin(\delta)) \tag{8.6}$$

whence

$$\cos(\zeta) = \cos(\phi)\cos(\delta)\cos(h) + \sin(\phi)\sin(\delta) \tag{8.7}$$

When $\cos\zeta < 0$ the sun is below the horizon.

The cosine of the zenith angle attains a maximum value $\cos(\phi - \delta)$ when $h = 0$, and a minimum value $-\cos(\phi + \delta)$ when $h = \pm\pi$. Both values are above the horizon when $|\phi| > |\pi/2 - \delta|$, corresponding to the perpetual polar summer day. Both values are below the horizon when $|\phi| > |\pi/2 + \delta|$, corresponding to the perpetual polar winter night. At the solstices, δ takes on its extreme values of $\pm\gamma$. Therefore, perpetual day or night are experienced at some time of year for latitudes poleward of $\pi/2 - \gamma$. These circles are known as the *Arctic* and *Antarctic* circles on Earth. Apart from the case of perpetual day or night, there is a *terminator* which separates the illuminated from the dark side of the planet. The position of the terminator is given by

$$\cos h_t = -\tan(\phi)\tan(\delta) \tag{8.8}$$

If Ω is the angular velocity of rotation of the planet, so that the planet's day is $T_{day} = 2\pi/\Omega$ time units, then the number of time units of daylight is $2h_t/\Omega = (h_t/\pi)T_{day}$. We shall adopt the convention $h_t = 0$ in the case of perpetual night, and $h_t = \pm\pi$ for perpetual day.

Exercise 8.2.2 For a given latitude ϕ, what δ yields the least hours of daylight? What δ yields the most hours of daylight? The Earth's present obliquity is 23.5 degrees, and it's length of day is 23.94hours. Sketch a plot of the maximum and minimum hours of daylight vs. latitude for the Earth.

The diurnal variations of the zenith angle lead to hot days and cold nights. Where the thermal response time is long enough to average out an appreciable portion of the diurnal temperature variation, the daily mean incident solar flux is an informative statistic. Since the incident solar flux per square meter of surface is $L\cos\zeta$, where L is the solar constant in W/m^2, one can obtain the daily mean flux by averaging $\cos\zeta$ over a rotation period of the planet. This results in a nondimensional flux factor f, by which one multiplies the solar constant in order to obtain the daily mean solar radiation incident on each square meter of the planet's spherical surface. The daily average can be performed analytically, resulting in

$$\begin{aligned} f(\phi,\delta) &= \frac{1}{2\pi}\int_{-h_t}^{h_t} cos(\zeta)dh \\ &= \frac{1}{\pi}[\cos(\phi)\cos(\delta)\sin(h_t) + \sin(\phi)\sin(\delta)h_t] \end{aligned} \tag{8.9}$$

where h_t is determined by Eq. 8.8. This derivation of the daily average assumes that the length of the day is much less than the length of the year, so that δ may be regarded as constant over the course of the day. If the length of the day is a significant fraction of the length of the year, as is be the case for nearly tide-locked planets like Mercury or Venus, the expression still gives the correct average along the latitude circle, but this average is no longer identical to the time average over a day.

During the equinoxes, $\delta = 0$ and $f = \cos(\phi)/\pi$, independant of the obliquity. This agrees with the result we obtained earlier by direct geometrical reasoning. At other times of year, the daily mean flux is governed by two competing factors: the varying length of day, which tends to produce higher fluxes near the summer pole, and the average zenith angle, which tends to produce high fluxes near the subsolar latitude (which remains near the Equator if the obliquity is not too large). The latitude where the maximum daily mean insolation occurs is always between the subsolar latitude δ and the summer pole. For $\delta = 0$ the maximum occurs at the Equator, and a little numerical experimentation shows that the latitude of the maximum increases to about 43.4^o when $\delta = 23.4^o$ (and similarly, with reversal of signs, in the Southern hemisphere). For larger δ, the length-of-day effect wins out over the slant angle effect at the pole, and the maximum occurs at the summer pole itself. This state of affairs just barely happens at the solstice for the present obliquity of Earth and Mars; as a result, the summer hemisphere solstice insolation is fairly uniform in these two cases. It is also useful to note that the daily mean insolation at the summer pole exceeds the daily mean insolation at the Equator when $|\delta| > 17.86^o$.

To obtain a general appreciation of the seasonal cycle, recall that δ varies from $-\gamma$ during the southern hemisphere summer solstice to γ at the northern hemisphere summer solstice, taking on a value of zero at the equinox which lies between the two solstices. Consider a planet with uniform albedo, so that the absorbed solar radiation is determined by the distribution of incident solar radiation. Suppose further that the thermal response time is long enough to average out the diurnal cycle, but short compared to the length of the year. If the obliquity is below 23.4^o, the "hot spot" starts some distance poleward of the Equator in the Southern hemisphere, moves to the Equator as the equinox is approached, and then migrates a similar distance into the Northern hemisphere as the Northern summer solstice is aproached. If the obliquity is greater than 23.4^o, the hot spot starts at the South Pole, discontinously jumps to -43.4^o at the point in the season

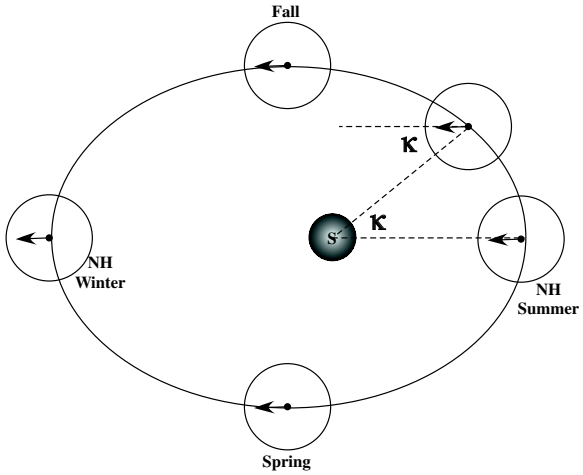

Figure 8.4:

where the subsolar latitude crosses -23.4^o, smoothly migrates throught the Equator and on to 43.4^o when the subsolar latitude approaches 23.4^o, and then discontinuously jumps to the North Pole. Note that in either case, the hot spot crosses the Equator twice per year, at the equinoxes; the two solstices are the coldest times at the Equator. The climate at the Equator has a periodicity that is *half* the planet's year.

It only remains to express δ as a function of the position of the planet in its orbit. The planet is spinning like a top, and if there are no torques acting on the planet (an assumption we will relax later) its angular momentum is conserved. Hence the rotation axis keeps a fixed orientation relative to the distant stars throughout the year. This is why Polaris is the Northern Hemisphere pole star all year around. Let κ be the angle describing the position of the planet, as shown in Figure 8.4. We shall adopt the convention that $\kappa = 0$ occurs at the Northern Hemisphere summer solstice. We shall refer to κ as the *season angle*, but it is more commonly (and more obscurely) referred to as the *longitude of the sun*. In our case, we have defined the longitude of the sun relative to the Northern Hemisphere summer solstice, but other choices are also common, for example defining it relative to the Northern winter solstice or the Spring equinox. When discussing the progression through the seasonal cycle on planets other than Earth, the season angle is almost universally used to describe where the planet is in its cycle, since this description obviates the need to make up names for months for each planet. If we project the rotation axis onto the plane of the ecliptic (i.e. the plane containing the planet's orbit), then the angle made by this vector with $\vec{PS}$ is equal to κ. The rotation axis projected onto the plane of the ecliptic acts like the hand of a clock, which rotates around the clock face once per year, though at a non-uniform rate if the orbit is not perfectly circular.

Let $\hat{n}$ be the unit normal vector to the plane of the ecliptic, and $\hat{n}_a$ be the unit vector in the direction of the rotation axis. Introduce a new cartesian coordinate system with x pointing along $\vec{PS}$, z pointing along $\hat{n}$, and y perpendicular to the two in a right-handed way. Then $\hat{n}_a = (\cos(\kappa)\sin(\gamma), \sin(\kappa)\sin(\gamma), \cos(\gamma))$ and the latitude of the sun is the complement of the angle between $\hat{n}_a$ and the x axis, whence

$$\sin(\delta) = \cos(\frac{\pi}{2} - \delta) = \cos(\kappa)\sin(\gamma) \tag{8.10}$$

In the limit of small obliquity, this equation reduces to $\delta = \gamma\cos\kappa(t)$. For a circular orbit, $\kappa(t) = \Omega t$,

where Ω is the orbital angular velocity (2π divided by the orbital period). In this special case, the subsolar latitude varies cosinusoidally over the year, with amplitude given by the obliquity. This is actually not a bad approximation even for the roughly 23^o current obliquity of Earth and Mars, agreeing with the true value to two decimal places. At the opposite extreme, when $\gamma = 90^o$, the subsolar latitude is given by $\delta = \pi/2 - \kappa$, which is not at all sinusoidal.

Exercise 8.2.3 Compute the length of day as a function of the time of year for the latitude at which you are currently located. Compare with data for the current day, either observed yourself or presented in the newspaper weather report. Compute the length of a shadow that would be cast by a tall, thin skyscraper of height 100m, as a function of the time of day and time of year at your latitude.

Contour plots of the flux factor for various obliquities are shown in Figure 8.5. These plots assume the orbit to be perfectly circular, so that there is no variation in distance from the Sun in the course of the year. Over the course of the year, the hot spot moves from south of the Equator to North of the Equator, and back again, passing over the Equator at the equinoxes. The amplitude of the excursion increases with obliquity, and goes all the way from pole to pole for sufficiently large obliquity. Earth, Mars, Saturn,Titan, and Neptune with present-day obliquities of 23.5^o, 24^o,26.7^o,26.7^o,and 29.6^o respectively, are qualitatively like the 20^o case. The pattern of variation of incident solar radiation which forces the seasonal cycle is similar in all these cases. However, the nature of the seasonal cycle will differ amongst these planets because the differing nature of the atmospheres and planetary surfaces will lead to different thermal response times. In the case of gas giant planets, another variable is the proportion of energy received from solar energy vs. the that received by transport from the interior of the planet. Insofar as the latter becomes dominant, the role of solar heating, and hence the prominence of the seasonal cycle, becomes less. Jupiter has a low obliquity (3.1^o), which, compounded by a high proportion of internal heating (** per cent) should lead to a minimal seasonal cycle. At the opposite extreme is Uranus, which has an obliquity of nearly 90^o, and a small proportion (about ** percent) of internal heating. Venus is so slowly rotating that its obliquity is of little interest. Obliquity is not constant in time; it varies gradually over many thousands of years. We will see in Section 8.5.1 that relatively slight variations in the Earth's obliquity are believed to contribute to the coming and going of the ice ages. The obliquity of Mars varies more dramatically, and perhaps with greater consequence; at various times in the past it could have reached values as high as 50^o and as low as 15^o.

If the thermal response time of the planet is a year or more, then a considerable part of the seasonal cycle is averaged out and the annual mean insolation becomes an informative statistic. It will be seen in the next section that this is the case for watery planets like the Earth. The annual mean flux factor is shown in Figure 8.6. When obliquity is small, the poles receive hardly any radiation. As obliquity is increased, the polar regions receive more insolation, at the expense of the equatorial regions. For Earthlike obliquity, the maximum insolation occurs at the Equator, which is why this region of Earth's surface tends to be warmest. When the obliquity exceeds 53.9^o, the annual mean polar insolation becomes greater than the annual mean equatorial insolation. For such a planet, the poles will be warmer than the tropics, provided that the thermal response time is long enough to average out most of the seasonal cycle. Consider a planet with 20^o obliquity, zero albedo, and a very long thermal response time. If the planet were put in Earth's orbit about the Sun, the Solar constant would be $1370W/m^2$, yielding equatorial insolation of $422W/m^2$ and polar insolation of $149W/m^2$, based on the flux factors given in Figure 8.6. In the absence of any greenhouse effect or lateral energy transport by atmospheres or oceans, the equatorial temperature would be 294K and the polar temperature would be 226K. If one takes into account the clear-sky greenhouse effect of an Earthlike atmosphere with 300ppm CO_2 and 50%

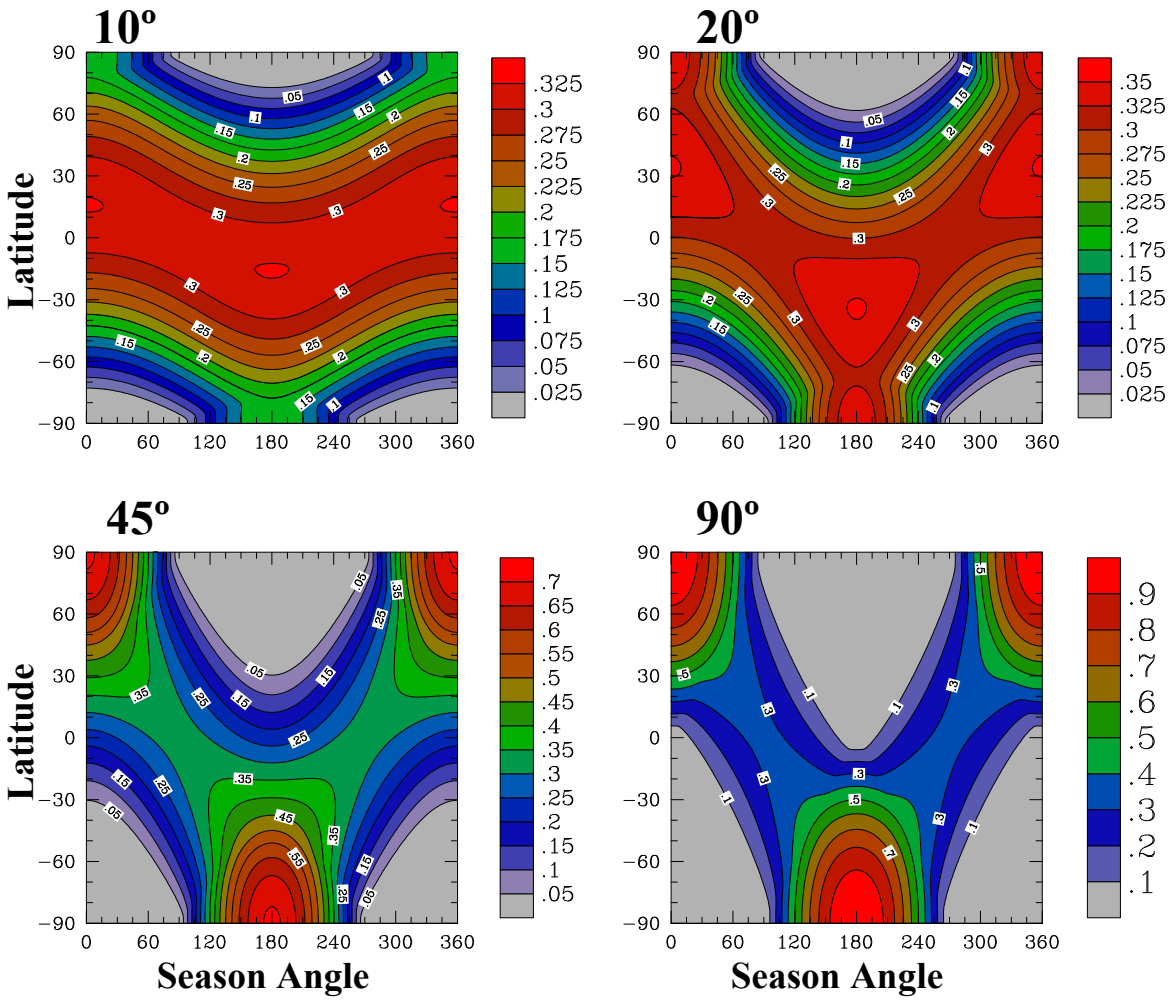

Figure 8.5: The seasonal and latitudinal distribution of daily-mean flux factor for four different values of the obliquity. In these plots, a circular orbit has been assumed. To obtain the daily mean energy flux incident on each square meter of the planet's surface, one multiplies the flux factor by the solar constant. For example, if the solar constant is $1000 W/m^2$, the incident solar flux at the pole during the Summer solstice is about $700 W/m^2$ if the obliquity is 45^o.

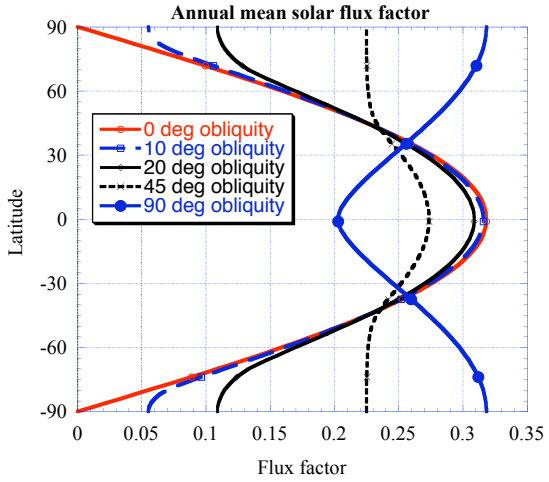

Figure 8.6: The annual mean flux factor for various obliquities, assuming a circular orbit.

relative humidity using the OLR results given in Chapter 4, the polar temperature rises to 237K, but the equatorial temperature becomes problematic: The annual mean equatorial solar flux is near or above the runaway greenhouse threshold discussed in Chapter 7, leading to extremely high or even unbounded equatorial temperatures. Lowering the relative humidity to 20% to reflect the fact that much of the tropical troposphere is very dry (recall Chapter **??**) still leaves the tropics with temperatures in excess of 350K. Part of the problem lies in the neglect of albedo. Simply using the observed planetary albedo in the tropics gives the wrong answer, because almost all of the cloud albedo is offset by the cloud greenhouse effect in the present climate (Chapter **??**). Using an albedo of .15, based on the observed tropical clear-sky albedo, reduces the equatorial solar absorption to $360 W/m^2$, which is in balance with a tropical temperature of 318K assuming a relative humidity of 20%. This is still well in excess of the observed tropical temperature. In the real atmosphere, heat transports due to large scale atmospheric and oceanic motions remove some of the heat from the tropics and deposit it at high latitudes, reducing the tropical temperatures and increasing the polar temperatures. Since incorporation of ice-albedo effects would reduce the polar temperatures below the estimates given above, such transports are also needed to bring the polar temperatures up into the observed range. Some elementary models of heat transport will be discussed in Chapter 10, though a proper treatment of the subject must be deferred until Volume II, where the necessary fluid dynamical background will be developed.

If we put the same planet at the orbit of Mars the temperatures become 238K and 183K without any greenhouse effect. Since there is little water vapor feedback at such low temperatures, the greenhouse effect is less dramatic in this case. Addition of an Earthlike atmosphere with 300ppmv CO_2 would increase the equatorial and polar temperatures to 256K and 192K, respectively, based on a linear OLR fit in the range 200K to 250K (specifically, $OLR = 80.6 + 1.83(T - 200)$). Thus, a waterworld placed at the orbit of Mars would require a much stronger greenhouse effect than the Earth's to avoid succumbing to a snowball state.

The effects of obliquity on the seasonal and latitudinal pattern of insolation may be summed up as follows. Increasing obliquity increases the intensity of the seasonal cycle at mid to high

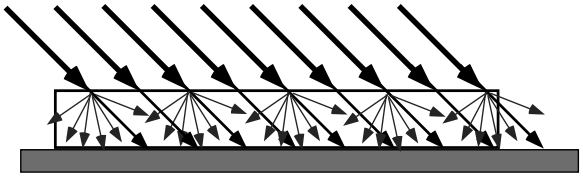

Figure 8.7: Schematic effect of atmospheric scattering on insolation

latitudes. The summer insolation gets steadily higher relative to the global mean, and a greater area of the winter hemisphere exposed to cold perpetual night or low insolation. Increasing obliquity also increases the annual mean polar insolation, though the way this affects polar climate depends on the thermal response time of the atmosphere-surface system. The increase might show up as very hot summers and bitterly cold winters, or as year-round warming, accordingly as the response time is fast or slow.

The preceding results on incident solar radiation have been derived in the absence of an atmosphere, but can still be used if there is in intervening atmosphere which may absorb or scatter solar radiation before it reaches the surface. The general geometry is illustrated in Figure 8.7. In this case, one suspends an imaginary sphere at an altitude above which the atmosphere is too thin to have a signficant effect on the solar radiation. The preceding results then give the solar flux entering each square meter of the surface of this sphere, and the angle at which the light enters the atmosphere. This is all that is needed as input to one-dimensional scattering models of the sort discussed in Chapter 5. One simply divides up the atmosphere into a series of patches near each latitude and longitude point, within which the properties are considered uniform, and applies a one-dimensional column model to each of these patches. As illustrated in 8.7, if the horizontal size of the patches is large compared to the depth of the atmosphere, the energy loss by horizontal scattering from one patch to another can be neglected, and each patch can be considered energetically closed, so far as radiation is concerned. The atmosphere has three effects, which can be inferred from the column radiation model: (1) Some of the incident solar radiation reaches the surface in the form of diffuse radiation at a continuous distribution of angles, rather than at the zenith angle, (2) Some of the solar radiation is absorbed in the atmosphere, rather than at the surface, and (3) some of the incident solar radiation is reflected back to space instead of entering the climate system. Of the three effects, it is the last – the effect of the atmosphere, including its clouds, on planetary albedo – that is most important for determining the climate. Diffuse radiation and atmospheric absorption do not change the amount of energy entering a column, but only the place and angle with which it enters. Often, this is of little consequence, so one can get a good estimate of the planet's temperature if one can obtain an estimate of the planetary albedo from one means or another.

The above reasoning can even to some extent apply to gas giant planets which have no surface. One can still define the imaginary sphere through which radiation enters the system, as before, but the problem comes in defining a characteristic depth scale. For the purposes of solar radiation, it suffices to consider the depth of atmosphere over which most of the solar radiation is absorbed, in effect a "photic" zone. This is typically shallow compared to the prodigious size of the gas giants. The full problem, including internal heat sources and dynamical motions, might require consideration of a deeper layer. Whatever the depth of this "active layer," the preceding reasoning applies provided one can sensibly model the large scale aspects of the climate on the basis of averaging over patches of horizontal extent that is large compared to the characteristic depth scale. For giant planets, as for the Earth or any other planet, the essential difficulty is that clouds, temperature, water vapor and other climate variables are manifestly not uniform over

length scales comparable to or longer than the depth of the atmosphere. One makes progress by boldly assuming that one can represent the effects of these fluctuating quantities by their large scale averages. It is an assumption that is difficult or impossible to justify mathematically, and in some cases may not even be true. With the present state of the art, one can only make progress by proceeding on the basis of the averaging assumption, and seeing how things work out.

8.3 Thermal Inertia

At several points in the preceding discussion, we have needed to make reference to the thermal response time of the system. In the present section, we shall make this notion precise. The heat storage in the planet's solid or liquid surface, and in its atmosphere means that it takes time for the system to heat up or cool down. The strength of this effect, known as *thermal inertia*, determines the extent to which the seasonal and diurnal fluctuations are averaged out in the climate response.

8.3.1 Thermal inertia for a mixed-layer ocean

The concept of thermal inertia is well illustrated by consideration of heat storage in the mixed layer of an ocean. Consider a layer of incompressible fluid with density ρ and specific heat c_p, which is well-mixed by turbulence to a depth H. The assumption of well-mixedness implies that any heating or cooling applied to the surface is distributed instantaneously throughout the depth of the mixed layer, whose temperature thus remains uniform. Let $S(t)$ be the solar flux heating the mixed layer, and assume that the cooling of the mixed layer (by infrared radiation or other means) can be written as a function of temperature, which we shall call $F(T)$. For example, if the atmosphere above the ocean has no greenhouse effect and carries no heat away from the surface by turbulent transport, the cooling is the radiative cooling $F(T) = \sigma T^4$. The energy balance equation for the mixed layer is then

$$\frac{d}{dt}\rho c_p H T = S(t) - F(T) \tag{8.11}$$

Exercise 8.3.1 Consider a planet with a 50m deep mixed layer water ocean ($c_p = 4218 J/kg$, $\rho = 1000 kg/m^3$). Suppose that the atmosphere for some reason has no effect whatsoever on the surface energy budget. (Why would this situation be hard to arrange, even for a pure N_2 atmosphere?) Hence $F(T) = \sigma T^4$. Suppose that the temperature of the polar ocean is $300K$ when the sun sets and the long polar night begins. Find a solution to Eq. 8.11 for this situation, and use it to determine how long it takes for the ocean to fall to the freezing point (about 271K for salt water)?

We may define a thermal inertia coefficient $\mu = \rho c_p H$ for the mixed layer ocean. If an amount of energy ΔE is added to or removed from the system, the corresponding temperature change is $\Delta E/\mu$. For a 50m mixed layer water ocean, $\mu = 2.1 \cdot 10^8 J/m^2$, so that an energy flux of $100 W/m^2$ out of the surface would lead to a cooling rate of $100/\mu = 4.74 \cdot 10^{-7} K/s = .04 K/day$. Clearly, a rather shallow layer of well-mixed water can buffer a considerable surface flux imbalance. The Earth's ocean is several kilometers deep, but it is only the upper few tens of meters that are well mixed on short time scales; 50m is in fact a reasonable approximation to the overall mixed layer depth of Earth's ocean, though there are geographical variations. Most other liquids would do about as well as water at storing heat. It is primarily the mixing depth that determines the thermal buffering effect of a planet's ocean.

Atmospheres also have thermal inertia, which can be considered in a fashion analogous to a mixed layer. The entire mass of the troposphere is well-mixed, and this generally makes up most of the mass of an atmosphere. In that case, the energy per square meter needed to raise the temperature of an atmosphere by 1K is $\mu = c_p p_s/g$, where p_s is the surface pressure, g the acceleration of gravity and c_p the mean specific heat of the gas making up the atmosphere. It is convenient to express this value in terms of the depth of a water mixed layer ocean, H_{eq} which would have the same thermal inertia coefficient. For the Earth atmosphere, $H_{eq} = 2.4m$, which is insignificant in comparison to the mixed layer depth of the ocean. Hence, one expects the Earth's atmosphere to come into equilibrium much more quickly than the ocean. The current 6mb CO_2 atmosphere of Mars has $H_{eq} = .03m$, while the massive atmosphere of Venus has H_{eq} in excess of $155m$. Neither Mars nor Venus has an ocean to buffer the seasonal cycle, but the Venus atmosphere alone can be expected to have a considerable moderating effect, whereas present Mars should behave more or less as if each point of the globe is in instantaneous equilibrium. Early Mars (circa 4 billion years ago) may have had a 2 bar CO_2 atmosphere, which would translate into a $10m$ equivalent mixed layer. This is considerably greater than that of Earth's atmosphere, but still not enough to have much moderating effect, in view of the fact that Mars' year is about twice as long as Earth's. Titan has a mostly N_2 atmosphere with a surface pressure only slightly in excess of Earth, but it's weak gravitational acceleration of $1.35m/s^2$ means that this pressure translates into a much greater mass of atmosphere per square meter of planetary surface. Thus,the Titan atmosphere has an equivalent mixed layer depth of about $24m$. Given the low temperature of Titan, and consequent low rate of energy loss by infrared emission, this value is expected to yield a very considerable buffering effect on Titan's seasonal cycle, regardless of whether there is a liquid ocean at the surface. For example, based on a typical surface temperature of 90K, blackbody emission would cool the planet only at a rate of about 1K per 300 Earth days, if insolation were completely shut off.

Exercise 8.3.2 The specific heat of liquid Methane is 3450.J/K. How deep would a well-mixed methane ocean on Titan have to be for it to have thermal inertia comparable to Titan's atmosphere?

We shall now consider some simple solutions to the mixed layer model, keeping in mind that this model applies to atmospheres as well as oceans, with a suitable choice of the equivalent mixed layer depth. At this point we assume that $\rho c_p H$ is constant, though models with a time-varying mixed layer depth are possible. Without any loss of generality we may write the insolation and temperature in the form

$$S = S_o + S'(t), T = T_o + T'(t) \tag{8.12}$$

where S_o and T_o are the time means of S and T and the deviations have zero time mean. Now, suppose that $T' << T_o$ for whatever reason; this need not require that $S' << S_o$, since the temperature fluctuations might be small by virtue of a slow response time of the system. Because the temperature fluctuations are small, the surface cooling can be expanded about T_o and approximated by a linear function:

$$F(T) = F(T_o) + aT'(t), b = \frac{dF}{dT}(T_o) \tag{8.13}$$

Now we choose T_o to be the equilibrium temperature corresponding to the mean insolation S_o, i.e. $F(T_o) = S_o$. With these assumptions the equation for the temperature fluctuation becomes

$$\frac{dT'}{dt} = \frac{1}{\rho c_p H}(S'(t) - bT') \tag{8.14}$$

or equivalently, if we define the relaxation time $\tau = (\rho c_p H)/b$,

$$\frac{dT'}{dt} + \frac{T'}{\tau} = \frac{1}{\rho c_p H} S'(t) \tag{8.15}$$

We can distinguish two limiting cases for Equation 8.15. When the time scale over which S' varies is slow compared to τ, then the first term on the left hand side is negligable compared to the second, whence the solution becomes $T' = \tau S'(t)/(\rho c_p H) = S'(t)/a$. In other words, the system acts as if it's in equilibrium with the instantaneous solar radiation at each time. In the opposite limit, the time scale of the solar fluctuation is rapid compared to τ, in which case it is the second term on the left hand side that may be neglected. Thus,

$$T'(t) = \frac{1}{\rho c_p H} \int_0^t S'(t')dt' \tag{8.16}$$

In this case, the temperature is out of phase with the heating, and represents a time average of the fluctuating heating. The peak temperature occurs later than the peak solar heating, since it takes time for the mixed layer to respond to the accumulating heating. Further, in this case, the seasonal temperature fluctuation becomes small as the mixed layer depth is made large, since the mixed layer becomes more and more efficient at averaging out the seasonal fluctuations of solar flux.

The variations in solar radiation over the course of a year are not sinusoidal, but we can nonetheless gain some further insight into the seasonal cycle by writing $S' = S_1 \cos(\omega t)$. For this form of forcing, Eq. 8.15 can be solved most easily by using complex exponentials. Since $S' = S_1 Real(\exp(-i\omega t))$, the solution may be written $T' = Real(A \exp(-i\omega t))$. Substituting this form of solution into Eq. 8.15 we find

$$A = \frac{S_1}{\rho c_p H} \frac{1/\tau + i\omega}{1/\tau^2 + \omega^2} = |A| e^{i\Delta} \tag{8.17}$$

where the phase and amplitude are

$$\Delta = \arctan(\omega\tau), |A| = \frac{S_1}{\rho c_p H} \frac{1}{\sqrt{(}1/\tau^2 + \omega^2)} \tag{8.18}$$

With these definitions, the solution can be written

$$T'(t) = |A| \cos(\omega t - \Delta) \tag{8.19}$$

The character of the response depends on the period of the forcing relative to the characteristic response time of the system. This determines both the amplitude of the fluctuation and the phase shift relative to the forcing. For $\omega\tau << 1$ we have $\Delta = 0$ and $|A| = S_1/a$. For $\omega\tau >> 1$ we have $\Delta = \pi/2$ and $|A| = S_1/(\rho c_p H \omega)$. Note that in this case the temperature fluctuation becomes weak in inverse proportion to the frequency of the solar forcing fluctuation. These are special cases of the limits discussed previously, but we now have the further advantage of an explicit formula showing how the phase and amplitude of the seasonal cycle vary between the two extreme cases.

So far, we have not specified the flux which is to be used for the heat loss term $F(T)$ in Eq. 8.11, or for the damping coefficient b in the linearized form of the equation. One candidate for this flux is the top-of-atmosphere infrared heat loss. The other is the combined turbulent and infrared heat exchange between the planetary surface and the atmosphere, discussed in Chapter 6. There are two circumstances in which the top-of-atmosphere flux (the OLR) is the appropriate

one to use. If the time scale under consideration is long enough that the surface budget can come into equilibrium, then the net solar flux absorbed at the surface is equal to the net turbulent and infrared flux passing from the surface into the atmosphere. In this case, we may consider the energy budget of the surface-atmosphere system as a whole, whence the OLR gives the heat loss from the system. The thermal inertia is provided by the atmosphere, and one uses he atmosphere's equivalent mixed layer depth in the mixed layer model equations. Alternately, if the response time of the atmosphere is short enough compared to the time scale under consideration, the energy budget of the atmosphere comes into equilibrium. In this case, the net energy exchange between the surface and the atmosphere must equal the OLR, since otherwise the atmosphere would warm up or cool down until equilibrium is achieved. Hence, one can use the OLR for the heat loss term in the surface energy budget, obviating the need to know the detailed physics behind the surface to atmosphere energy transfer. In this case, the thermal inertia is provided by the heat capacity of the mixed layer ocean.

In either case, one can compute $OLR(T)$ using a radiation model and some assumption linking the temperature and humidity profile to surface temperature, or one can use one of the linear or polynomial fits to the OLR curve discussed in Section 4. For example, with a linear fit to the OLR curve for a terrestrial atmosphere with 300ppmv CO_2 and 50% relative humidity, b is about $2(W/m^2)/K$ in the range 250K to 310K. The corresponding relaxation time τ is 1200 days for a 50 meter mixed layer, or 60 days for the 2.4m mixed layer which is equivalent to the thermal inertia of the Earth's atmosphere. In consequence, the seasonal cycle is expected to be strongly attenuated on the ocean- covered parts of the Earth (apart from coastal effects). The atmosphere alone does not have enough thermal inertia to damp out the seasonal cycle, but it does have enough thermal inertia to keep the atmospheric temperature roughly constant in the course of the diurnal cycle. Colder temperatures tend to make the relaxation time longer. For example, in an Earthlike atmosphere with 300ppm CO_2, the relaxation time roughly doubles at 160K. As noted earlier, Titan has a very long relaxation time owing to its thick atmosphere and low temperature; now we can make the statement more precise. Ignoring the greenhouse effect and setting $b = 4\sigma T^3, T = 90K$ we find a relaxation time of 20 Earth years, based on the equivalent 24m mixed layer depth of Titan's atmosphere. Since Titan's year (which is the same as Saturn's year) is about 30 Earth years, the seasonal cycle on Titan is expected to be considerably damped, though not so much so as the seasonal cycle over the Earth's oceans. The weak greenhouse effect from methane in Titan's atmosphere would somewhat enhance the damping. In contrast, a similar calculation for the thin atmosphere of present Mars gives a relaxation time of only .8 Earth days, based on $T = 200K$. Since a Mars day is approximately the same as an Earth day, the thermal inertia of the Martian atmosphere at present has relatively little damping effecton the diurnal cycle.

The thermal relaxation process is different if the time scale under consideration is short compared to the response time of the atmosphere, but long compared to the response time of the surface. In this case, the atmospheric temperature remains approximately constant while the surface temperature fluctuates. This is the way the diurnal cycle works on Earth over ice or land. The relaxation time of surface temperature is then determined using the turbulent and radiative surface-atmosphere flux formulae discussed in Chapter 6, rather than the OLR. The situation of present Mars is not like this, since the atmosphere has little thermal inertia. There, the diurnal cycle affects the entire depth of the atmosphere, and the diurnal response is approximately governed by the OLR and the thermal inertia of the surface, much as for the Earth's seasonal cycle.

In the general case, where neither the atmosphere nor the ocean are in equilibrium, one must write a separate mixed layer model for each of these two components, coupling them through the surface exchange flux formulae, and allowing the atmosphere to lose energy through it's top

via OLR. The exploration of this case will be left to the reader.

8.3.2 Thermal inertia of a solid surface

Heat diffuses slowly through a non-metallic solid, so when the underlying surface is solid it is typically necessary to consider the continous distribution of temperature as a function of depth within the solid. To a good approximation, heat flux within a solid is proportional to the temperature gradient; the proportionality constant is called the *thermal conductivity*, which we shall call κ_T. Balancing the rate of change of heat content against the convergence of heat flux yields the *diffusion equation*

$$\partial_t \rho c_p T = \partial_z \kappa_T \partial_z T \tag{8.20}$$

In this equation it is assumed that there are no internal heat sources. The surface heat budget enters the problem through the boundary condition at the surface ($z = 0$), which states that the diffusive heat flux into the surface equals the net heating of the surface by insolation and radiative and turbulent heat transfers. Using the same notation as we employed for the mixed layer case, this boundary condition reads

$$\kappa_T \partial_z T|_{z=0} = S(t) - F(T) \tag{8.21}$$

When S is a constant S_o, the problem is solved with a constant temperature T_o satisfying $S_o = F(T_o)$, just as for the mixed layer case. Linearizing the boundary condition about T_o and substituting the complex exponential form for S' yields

$$\kappa_T \partial_z T'|_{z=0} = S_1 e^{-i\omega t} - bT' \tag{8.22}$$

If ρc_p is constant, this boundary condition can be satisfied by a solution of the diffusion equation of the form

$$T' = Ae^{i(kz-\omega t)}, k = \sqrt{\frac{\omega}{D}}\frac{1-i}{\sqrt{2}} \tag{8.23}$$

where A is a constant and D is the diffusivity $\kappa_T/(\rho c_p)$. The complex vertical wavenumber k has been determined by substitution of the exponential form of T' into the diffusion equation. A will be determined by substitution of the solution into the boundary condition, but before doing so it is worth pausing to make some remarks on the solution Eq. 8.23. This solution was first obtained by Fourier, in his study of diurnal and seasonal variations of temperatures in the interior of the Earth. Eq. 8.23 shows that the characteristic depth to which temperature fluctuations penetrate is $\sqrt{(D/\omega)}$. Low frequency fluctuations penetrate to a greater depth than high frequency fluctuations, because heat has a longer time to diffuse before the surface temperature reverses. Note also that the phase lag of the time of maximum temperature with depth also reflects the time required for the surface conditions to penetrate to the interior. For the diffusivity of water ice (Table 8.1) the characteristic depth is 12 cm for the diurnal period, 2.4m for the annual period, 24m for a century and 76m for a millennium. Solid rock yields similar numbers. Hence, the temperature profile within ice or rock still contains information about temperatures centuries or even millennia in the past, albeit in a rather smoothed and degraded form. This fact has been exploited in reconstructions of past temperatures.

Exercise 8.3.3 You are designing a lunar colony to be placed at a Lunar latitude where the sun is directly overhead at noon. The moon has an albedo close to zero, and the response time of the surface is rapid, so that the noontime surface temperature is close to the instantaneous equilibrium temperature of 394K (re-derive this temperature yourself). At night, the equilibrium temperature would be absolute zero, but there is not enough time to reach equilibrium; still the night-time

	$\rho c_p (J/m^3)$	Conductivity $(Wm^{-1}K^{-1})$	Diffusivity (m^2/s)
Water Ice	$1.93 \cdot 10^6$	2.24	$1.16 \cdot 10^{-6}$
Fresh Snow	$.21 \cdot 10^6$	.08	$.38 \cdot 10^{-6}$
Old Snow	$1.0 \cdot 10^6$	.42	$.05 \cdot 10^{-6}$
Sandy Soil	$1.28 \cdot 10^6$	.3	$.24 \cdot 10^{-6}$
Clay Soil	$1.42 \cdot 10^6$	.25	$.18 \cdot 10^{-6}$
Peat Soil	$.575 \cdot 10^6$	.06	$.1 \cdot 10^{-6}$
Rock	$2.02 \cdot 10^6$	2.9	$1.43 \cdot 10^{-6}$
Lunar Regolith	$1 \cdot 10^6$	.01	$.01 \cdot 10^{-6}$

Table 8.1: Thermal properties of some common surface materials

temperature plummets to 100K. Since the Moon is tide-locked to the Earth, the Lunar day is 28 Earth days. The diffusivity of the Lunar regolith ("soil") is about $10^{-8} m^2/s$.

Approximate the day-night temperature variation by a sinusoidal curve. What would be the constant temperature far below the surface (neglecting internal heat sources)? How deeply would the colony habitat have to be buried in order for the ambient diurnal temperature fluctuations to be less than 1K?

NB: Given the low diffusivity of the regolith, your main difficulty is likely to be getting rid of the heat generated by energy use (biological and otherwise) within the colony.

Now we substitute Eq. 8.23 into the boundary condition (8.22). The result is

$$\begin{aligned} A &= \frac{S_1}{b + \rho c_p \sqrt{\omega D} \frac{1+i}{\sqrt{2}}} \\ &= \frac{S_1}{b} \frac{1}{1 + \sqrt{\omega \tau_D} \frac{1+i}{\sqrt{2}}} \\ &= \frac{S_1}{\rho c_p \sqrt{D/\omega}} \frac{1}{\frac{1}{\tau_1} + \frac{1+i}{\sqrt{2}} \omega} \end{aligned} \tag{8.24}$$

where $\tau_D = (\rho c_p)^2 D/b^2$ and $\tau_1 = \rho c_p \sqrt{D/\omega}/b$. Upon comparison of the third line of this equation with the solution for the mixed layer model, it is seen that the solid case acts somewhat like a mixed layer model with frequency dependent layer depth $\sqrt{D/\omega}$. For low frequency forcing, $\omega \tau_D \ll 1$, the surface temperature follows the instantaneous equilibrium, $A = S_1/b$, just as for the mixed layer case. For high frequency forcing, the amplitude of the surface temperature fluctuation decays like $1/\sqrt{(\omega)}$. This is slower than was the case for the fixed-depth mixed layer, since the layer determining the thermal inertia now gets thinner as frequency is increased. Note also that the phase lag of surface temperature relative to insolation differs from the mixed layer case. For the diffusion equation, the surface temperature lags the insolation by $\pi/4$ radians in the high frequency limit rather than $\pi/2$. The thinning of the active thermal layer keeps the surface temperature closer to instantaneous equilibrium than it would be in the fixed-depth case.

Apart from some exceptional circumstances, the thermal inertia of a solid surface has little effect on the seasonal cycle, though it can substantially moderate the diurnal cycle. This can be seen easily through the evaluation of tau_D in a few typical cases. First we consider the case of Antarctic or Arctic ice-covered regions. The flux coefficient based on a linear OLR fit in the temperature range 240K to 270K is $b = 2.16 W/(m^2 K)$. Using the heat capacity and thermal diffusivity for water ice, given in Table 8.1, we find $tau_D = 11 Earthdays$. At latitudes somewhat

away from the poles, the diurnal cycle of insolation becomes significant, particularly during the equinoxes. Since the time scale for the surface is shorter than that for the atmosphere, it would be more appropriate to use surface flux coefficients than OLR in analyzing the terrestrial diurnal cycle. As noted in Chapter 6, the turbulent heat transfer is strongly inhibited at night-time, when the boundary layer is statically stable. In this case, the flux coefficient is dominated by the radiative term $4\sigma T^3$ based on surface temperature. For temperatures around $255K$ this yields an even shorter response time $\tau = 4Earthdays$. In the midlatitudes and Tropics, the estimate differs only in the use of the slightly larger values of b appropriate to the warmer temperatures, and the somewhat different thermal properties of rock or soil, but the result remains that τ_D is on the order of a few days or less. For Mars, one may use $b = 4\sigma T^3$ based on $T = 200K$, given the thin atmosphere. This yields $tau_D = 15Earthdays$, which is still not sufficient to appreciably affect the seasonal cycle. It is only at the extremely cold temperatures of Titan that the response time of a solid ice surface becomes significantly longer (roughly 1300 Earth days), but even there the effect is of little interest, owing to the much longer response time of Titan's atmosphere. In sum, a solid surface can generally be considered to be in equilibrium for the purpose of computing temperature fluctuations on the seasonal time scale.

It should not be concluded from the above estimates that the thermal inertia of solid surfaces is sufficient to eliminate the diurnal cycle. The variation of insolation between noon and night-time is huge; On Earth, at a latitude where the Sun is overhead at noon, the amplitude of the variation is $1370W/m^2$, which leads to an undamped temperature fluctuation of 685K based on a flux coefficient $b = 2W/(m^2K)$. Even damped by a factor of 20, this amounts to a very considerable diurnal fluctuation. Similar considerations apply to the Martian diurnal cycle.

As a complement to the periodically forced solution, Figures 8.8 and 8.9 show the solutions for the diffusion equation in water ice which is initialized at a uniform temperature of 300K and allowed to cool without solar heating subject to a flux upper boundary condition. The heat loss from the surface was computed using an Earthlike OLR fit $OLR(T_s) = 48.461 + 1.5866(T_s - 180) + .0029663(T_s - 180)^2$ A quadratic fit was used so that the the fit would remain accurate over a large temperature range. Except for the high initial temperature, which turns out to be inconsequential, this problem can be thought of as representing the cooling of the Antarctic ice cap after permanent winter night closes in. Figure 8.8 illustrates the progressive penetration of the surface cooling into the depth of the ice; at time t, the cooling has penetrated to a depth on the order of $\sqrt{Dt}$, where D is the thermal diffusivity of the ice. Figure 8.9 shows that there is an extremely rapid initial cooling, owing to the thin layer of ice affected at short times. After a half day, the temperature has already fallen below freezing. Therafter, the temperature drop becomes slower, as the depth of the ice layer involved becomes greater. The reduction in OLR as temperature drop also contributes to the reduction in cooling rate. Nonetheless, after two months, the temperature has fallen to 190K, which is well below the 235K minimum temperature observed at the South Pole. Incorporation of the atmosphere's thermal inertia reduces the cooling rate somewhat, but does much increase the extremely cold temperature encountered at the end of the winter. Clearly, the Antarctic interior relies on heat transport from warmer latitudes to limit its winter temperature drop.

We conclude this section with a few remarks on the special effects of snow and ice (whether from water, CO_2 or some other substance) on the seasonal and diurnal cycle. Snow has a profound effect on the diurnal cycle, because of its very low thermal conductivity, which is nearly an order of magnitude lower than that of ice (see Table 8.1 for the case of water snow). The low thermal conductivity arises from the high proportion of the snow's volume which consists of air trapped in pores which are too small to allow the air to flow; since air itself has extremely low thermal conductivity, heat must primarily make its way through the contorted pathways of snow crystals in contact with each other. Other gases, trapped in snows made of other substance, have a

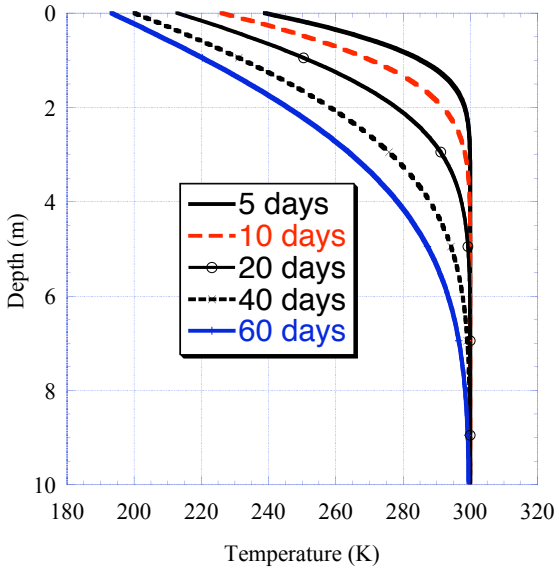

Figure 8.8: Temperature vs. depth at various times, for an ice layer subject to temperature-dependent heat loss at the surface. See text for specification of the heat loss rate.

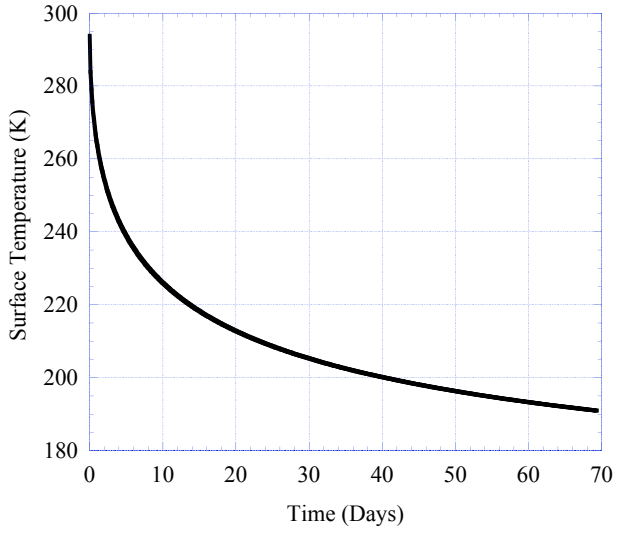

Figure 8.9: Time evolution of surface temperature for the solution shown in Figure 8.8

similar effect. The low conductivity dramatically reduces the characteristic response time of the surface, even for a snow layer of modest thickness. In the Antarctic case discussed above, τ_D drops to a mere 60 minutes for old snow, and 20 minutes for fresh snow. At night, the temperature of the snow surface plunges almost instantaneously to its equilibrium value. In the case of the Earth, the atmosphere has sufficient thermal inertia that it doesn't cool much at night, above the boundary layer. Given the suppression of turbulent flux in the stable nocturnal boundary layer, the night-time equilibrium temperature is maintained mainly by the downwelling infrared flux from the atmosphere, as discussed in Chapter 6. When the low level air temperature is 255K, the downwelling infrared flux is about $120W/m^2$, maintaining a snow surface temperature of 214K. On present Mars, the atmosphere cools down markedly at night, and in any event is too thin to provide much downwelling flux, so it is less obvious what limits the night-time temperature drop over the CO_2 snow fields that form in the winter hemisphere. One relevant consideration is that the flux coefficient b drops dramatically at very cold temperatures, leading to an increase of the relaxation time; when the surface temperature falls to 150K, τ_D increases to 23 hours even over snow. However, at such low temperatures the saturation vapor pressure of CO_2 is only 1.26mb, well below the ambient surface pressure. Hence, the night-time temperature minimum is likely to be governed by the latent heat release due to CO_2 condensation, which sets in at surface temperatures near 160K.

Snow cover on any planet can change rapidly in the course of the seasons, and on Earth, sea ice cover similarly expands and retreats. Since snow and ice have higher albedo than the surfaces they generally cover, this has an important feedback effect on the seasonal cycle. It enhances the winter-time cooling once ice or snow begin to accumulate, delays the springtime warming, but then accelerates the warming once ice or snow begin to retreat. The albedo feedback of snow is especially pronounced, since snow has a much higher albedo than ice. For water snow, for example, the albedo of fresh snow averaged over the solar spectrum can exceed .85, whereas a typical albedo for sea ice is on the order of .6. The high albedo of snow, like its low diffusivity, arises from its highly porous nature which offers many opportunities for light to encounter discontinuities in index of refraction, leading to scattering. It is a generic property of the snow of any weakly-absorbing substance. Note that the concept of "sea ice" is peculiar to planets with water oceans. On a planet with a liquid methane or CO_2 ocean "sea ice" would sink, and not have any chance to affect the surface albedo until the ocean were frozen to the bottom.

The presence of a solid phase on the surface of the planet also introduces a new form of thermal inertia, associated with the latent heat of phase change from the solid to liquid form. Where there is ice, whether it be in the form of sea ice or land glaciers, the surface temperature cannot rise above the triple point (the "melting point") until all the ice has been melted. The phenomenon is familiar from an experiment commonly performed in elementary school science classes, in which one tries to boil a pot of water containing ice cubes,and finds that the temperature doesn't start to rise above freezing until all the ice is gone. As an example, let us consider the melting of a 5m thick layer of ice with an albedo of .7. Based on the latent heat of melting, it takes $1.5 \cdot 10^9$ Joules to melt a square meter of this ice layer. In summertime at the pole, the ice absorbs $160W/m^2$ which would then take 110 days to melt the ice layer, even if all the absorbed solar energy is retained for melting, and none is lost by radiation to the atmosphere. Thus, a modest layer of ice can persist throughout a warm season, keeping the temperature from rising. One can clearly see this principle in operation in the summertime polar temperatures of the Earth, but it also operates in midlatitude areas subject to seasonal snow cover, in effect delaying the end of winter. The case of sublimation is somewhat similar, though subtler since there is no threshold temperature and the sublimated gas enters the atmosphere (with its stored energy), rather than flowing away in the form of rivers or ocean currents, as is the case for the liquid produced by melting. As discussed in Chapter 6, the latent heat flux due to sublimation (like evaporation) greatly increases the surface

energy loss for any given temperature. This reduces the warming of the surface required to balance the summertime increase in absorbed solar radiation. Unlike melting, it does not generally cap the surface temperature at some fixed value, but it does reduce the warming below what it would be without the presence of the sublimating ice or snow.

8.3.3 Summary of thermal inertia effects

The preceding discussion has revealed two limiting forms of behavior a planet can exhibit in the course of its seasonal cycle. A "waterworld," having high thermal inertia in the ocean-atmosphere system, responds primarily to the annual average insolation. Such a world will be coldest at the poles and warmest at the equator, unless the obliquity exceeds about 54^o, in which case the warmest climates will be found near the two poles. A "desertworld," having little thermal inertia in either the surface or the atmosphere, responds to the instantaneous insolation at each time of the year. The location of the highest temperature moves from some latitude North of the equator to the same latituded South of the equator, and back again, in the course of the year. For small obliquity, the poles are frigid throughout the year, and the hot spot executes modest excursions about the equator. For obliquities greater than about 18^o, the excursion goes all the way from pole to pole, assuming a uniform albedo. Geographical and temporal albedo variations alter this picture. Formation of permanent ice or snow cover near the poles will tend to keep the polar regions cold throughout the year; this effect is assisted by the thermal inertia implied by the latent heat required to melt or sublimate ice, which limits the summertime temperature increase. The Earth shows some characteristics of both limiting cases, with extreme continental climates and equable maritime climates. Thermal inertia sufficient to moderate the seasonal cycle can be provided either by a thick atmosphere or a well mixed liquid layer at the surface, which need only have a depth of some tens of meters. Heat storage provided by non-melting solid surfaces is almost never sufficient to have a significant affect on the seasonal cycle, though it can substantially moderate the diurnal cycle for planets with rotation periods on the order of a few Earth days or less.

8.4 Some elementary orbital mechanics

Sir Isaac Newton showed that the orbit of a single planet revolving about its star takes the form of an ellipse, with a focus of the ellipse at the center of mass of the system. Since stars are typically much more massive than their planets, the center of mass for most purposes is identical to the center of the star. The elliptical nature of orbits has an important effect on the seasonal cycle, since the planet is farther from its sun at some parts of the year than it is at others. This makes the solar "constant" L a function of time of year. On Earth, we don't notice this effect too much because our orbit is nearly circular. Nonetheless, the effect has an important influence on the long-term evolution of climate. On other planets, it can be even more important.

The distance of closest aproach of a planet to its star is called the *perihelion*, which we shall call r_p. The greatest distance is called the *aphelion*, which we shall call r_{ap}, The *semi-major axis* is then $a = (r_p + r_{ap})/2$. Let κ_1 be the angle made by the line between the star and the planet, defined so that $\kappa_1 = 0$ at the perihelion. Then, in polar coordinates, the equation of the elliptical orbit is

$$r = a\frac{1 - e^2}{1 + e\cos(\kappa_1)} \tag{8.25}$$

where e is the *eccentricity* of the orbit, which lies in the interval $[0, 1]$. $e = 0$ yields a circular orbit, while the ellipse becomes progressively more elongated as $e \to 1$. Specifically, perihelion is $(1-e)a$,

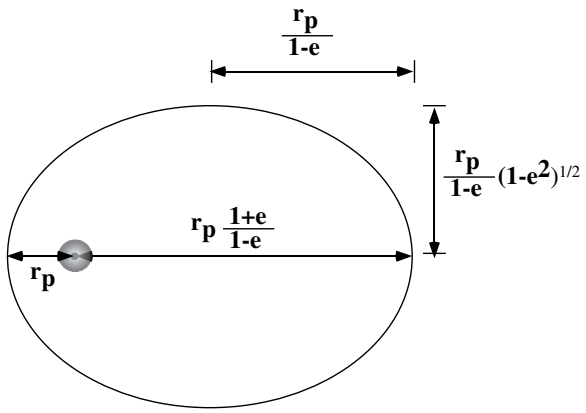

Figure 8.10: Geometry of an elliptical orbit with eccentricity $e = .66$

the aphelion is $(1 + e)a$ and the ratio of the distance at aphelion to the distance at perihelion is $(1 + e)/(1 - e)$. To get the semi-minor axis, we maximize $r(\kappa_1)sin(\kappa_1)$, yielding $a\sqrt{1 - e^2}$. Hence, the ratio of the minor to major axis is $\sqrt{1 - e^2}$. The geometry of the orbit is summarized in Figure 8.10.

Exercise 8.4.1 What eccentricity would yield an ellipse with a 3:1 axis ratio? Sketch such an ellipse, indicating the correct location of the Sun relative to the orbit.

The variation of the solar constant is then given by

$$L = \frac{1}{4\pi}\frac{I_o}{r^2} \tag{8.26}$$

where I_o is the power output of the star (e.g about $3.8x10^{26}$ Watts for the Sun at present). The annual variation in distance from the Sun leads to "distance seasons," which are synchronous between the hemispheres. This contrasts with the "obliquity seasons" (dominant for Earth and Mars) which are out of phase between the hemispheres, one hemisphere enjoying winter while the other suffers under the torrid heat of summer. In the limit of small eccentricity, the ratio of solar constant at aphelion to that at perihelion is $1 + 4e$. This represents a very considerable variation, even for modest eccentricity. For the present eccentricity of the Earth (.017), it amounts to 6.8%, or $93W/m^2$ difference in the solar constant between perihelion and aphelion. To turn this flux into a crude temperature estimate, we divide 4 to account for the averaging over the Earth's surface, and apply a typical terrestrial $OLR(T)$ slope of $2W/(m^2K)$, yielding a temperature difference of more than 11K between perihelion and aphelion. This represents the amplitude of the distance seasons. For Mars, with its present eccentricity of .093, the effect is even greater. The perihelion to aphelion flux variation is 37%, or $219W/m^2$. For Martian conditions, where the atmosphere has a weak greenhouse effect, this translates into an amplitude of 30K.

To determine the time dependence of r, we must know $\kappa_1(t)$. Because the orbit is no longer circular, the angular velocity is no longer constant; the planet moves faster when is close to the sun than when it is farther away. There is no analytic expression for the time variation of the orbital position. However, it can be easily computed by numerically solving a first order differential equation, which can be derived either from Kepler's equal-area law, or directly from angular momentum conservation. We shall take the latter route. Let $v_\perp$ be the component of

velocity perpendicular to the line joining the planet to its star. Then, by conservation of angular momemtum, $rv_\perp = J$ is independant of time. However, the angular velocity of the orbit is simply $v_\perp/r$, so the angle satisfies the equation

$$\frac{d\kappa_1}{dt} = \frac{J}{r^2} = \frac{J}{a^2}\frac{(1+e\cos(\kappa_1))^2}{(1-e^2)^2} \tag{8.27}$$

This equation shows that the angular velocity of the planet speeds up as it approaches perihelion, and slows down as it approaches aphelion. In consequence, the planet spends less time near the sun than it does at greater distances, and "distance summer" is shorter than "distance winter."

The average of the solar constant over the course of the year can be written

$$< L >= \frac{1}{4\pi}\frac{I_o}{a^2} < \frac{a^2}{r^2} >= L_a < \frac{a^2}{r^2} > \tag{8.28}$$

where angle brackets denote the average over the planet's year and L_a is the solar constant evaluated at a distance equal to the semi-major axis of the orbit. We can take advantage of the fact that the same $1/r^2$ factor appears in Eqn. 8.27 to relate the mean solar constant to the nondimensionalized duration of the planet's year. Specifically, integrating Eqn. 8.27 over one year and dividing by the length of the year yields

$$< L >= \frac{1}{\tau_y^*} L_a \tag{8.29}$$

where $\tau_y^* = \tau_y/(2\pi a^2/J)$, τ_y being the length of the year in dimensional terms. The quantity $2\pi a^2/J$ is the length of year for a circular orbit with radius a. Numerical integration of Eqn. 8.27 shows that the nondimensional year defined in this way decreases as the orbit becomes more eccentric. For $e = .1$, τ_y^* is .995, for $e = .25$, τ_y^* is .968, and $e = .5$, τ_y^* is .866.

Most planets have nearly circular orbits; leaving out Mercury and Pluto, the other planets have current eccentricities ranging from .007 to .093. Even Pluto, the most eccentric planet, only has a value of .244. Note that the difference between perihelion and aphelion distance is $O(e)$, whereas the ratio of major to minor axes deviates from unity by only $O(e^2)$. Hence, for small e, the orbit still looks like a circle, but with the Sun displaced from the circle's center by $O(e)$. For small e, Eqn. 8.27 can be solved approximately by a straightforward expansion in e. Substituting

$$\kappa_1(t) = \frac{J}{a^2}[t + eF(t) + e^2G(t)] \tag{8.30}$$

into the equation and matching like terms in e yields the solution

$$\kappa_1 = 2\pi t^* + 2e\cos 2\pi t^* + e^2[\pi t^* + \frac{5\pi}{2}\sin 4\pi t^*] + O(e^3) \tag{8.31}$$

where $t^* = tJ/(2\pi a^2)$. The first order term causes an $O(e)$ variation in the orbital angular velocity over the course of the year, but it this term by itself does not alter the length of the year. Taking into account the second order term, it may be inferred that the nondimensional length of the year is approximately $\tau_y^* = 1 - e^2/2$. In consequence, the annual mean insolation varies very little from what it would be for a circular orbit with radius equal to the semi-major axis. For $e = .1$, close to the present value for Mars, the eccentricity increases mean insolation by only .5%. For $e = .02$, similar to Earth at present, the increase is a meager .02%, or $.274 W/m^2$. Except in very unusual cases, orbital eccentricity affects the climate through the intermediary of the seasonal cycle, and not through any effect on the annual mean radiation budget.

The consequences of orbital eccentricity for a planet's climate derive from the way the distance seasons interact with the tilt seasons. Each of these types of seasons has a period of one planetary year, so the nature of the interaction is governed by the position in the orbit at which the Northern Hemisphere summer solstice occurs, measured relative to the position of the perihelion. This can be measured by an angle, called the *precession angle* or *precession phase*. We will define the phase such that when it is zero, the Northern Hemisphere solstice occurs at the perihelion. It is also common to define the phase as the angle between the perihelion and the Northern Hemisphere spring ("vernal") equinox. When the precession angle is zero, the distance seasons make the Northern Hemisphere seasonal cycle stronger, since "Northern tilt summer" happens when the planet is closest to the Sun and ""Northern tilt winter" happens when the planet is farthers from the Sun. Conversely, the Southern Hemisphere seasonal cycle is attenuated when the precession angle is zero. When the precession angle is 180^o, the situation is reversed between the hemispheres, with the Southern Hemisphere getting very hot summers and very cold winters, and the Northern Hemisphere experiencing more moderate seasons. When the precession angle is 90^o or 270^o, the solstices conditions are no longer modulated by the distance seasons, but instead the vernal equinox becomes warmer than the autumnal equinox, or *vice versa*.

Figure 8.11 illustrates the effect of eccentricity and precession on the seasonal cycle of insolation. These results were computed by numerically solving Eq. 8.27, and substituting $\kappa_1(t)$ into the flux distribution function given by Eq 8.10 and Eq 8.9, after shifting its phase to account for the precession angle. Given $\kappa_1(t)$, we also know $r(t)$. Using this, we multiply the flux factor by $(a/r(t))^2$ to account for the variations in orbital distance. This is the quantity plotted, at selected latitudes, in Figure 8.11. One multiplies this flux factor by the solar constant at a distance equal to the semi-major axis, in order to obtain the actual insolation in W/m^2. Using the symmetries of Eqn. 8.9, the results for precession angles of 180^o and 270^o can be obtained from those shown in Figure 8.11 by simply shifting the curves shown by a half year, and interchanging the two hemispheres, so these cases do not require separate discussion.

For both eccentricities, we see that the Northern Hemisphere extratropical seasonal cycle is made more extreme when the precession angle is 0^o, while that in the Southern Hemisphere is moderated. At the Equator, the two equinoxes have identical insolation, but the time of maximum equatorial insolation is shifted towards the Northern summer solstice, which is also the time of perihelion in this case. For the larger, Marslike, eccentricity ($e = .1$), the maximum equatorial insolation in fact occurs at the solstice. For the case of 90^o obliquity, the extratropical seasonal cycle has identical strength in both hemispheres, but the equinox conditions now differ from each other, the Autumnal equinox receiving less insolation than the Vernal (Spring) equinox. Also, the time of maximum and minimum extratropical insolation is also significantly displaced from what it would be for a circular orbit. The effect of orbital velocity variations on the seasonal cycle is just barely visible for the lower, Earthlike, eccentricity, but it is prominent for the higher eccentricity case. For 0^o precession, Summer is longer than Winter in the Southern hemisphere, while Winter is longer than Summer in the Northern hemisphere; for 90^o precession, there is a marked asymmetry between the rate of increase of insolation going into each season, and the rate of decrease coming out of it. For example, in the Northern Hemisphere, Summer sets in rapidly, but the transition to Winter takes a long time. In fact the Northern hemisphere, Southern hemisphere and equatorial insolation maxima are all bunched up within a period of about a quarter of a year, indicating that the distance seasons are beginning to dominate the tilt seasons even at this modest eccentricity. The effect of precession phase on the annual average insolation at each latitude is insignificant; for both the high and low eccentricity cases shown in Figure 8.11, changing the precession phase leaves the annual mean flux factor unchanged to at least four decimal places.

Note that the precession angle has a big effect on climate when the eccentricity is large, but

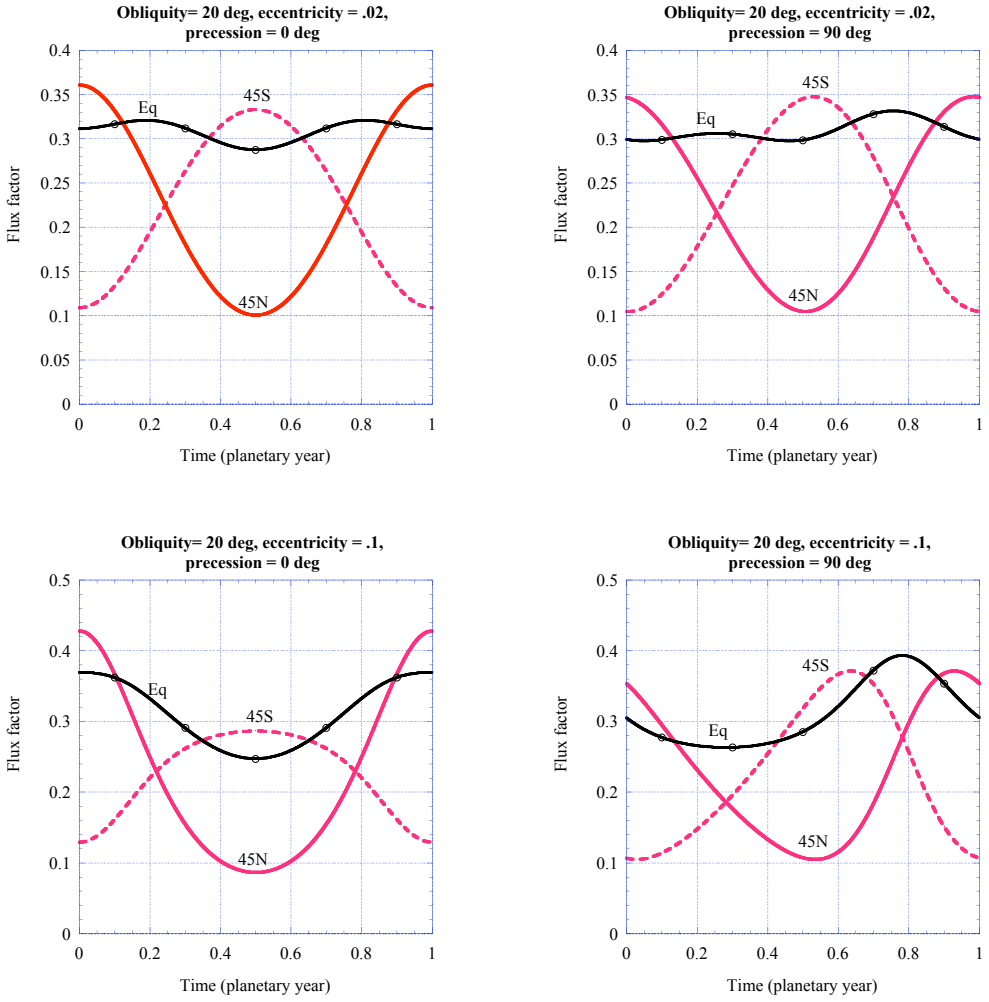

Figure 8.11: The seasonal cycle of solar flux factor for a planet with 20^o obliquity, at the Equator, 45N and 45S. To obtain the insolation at any given time of year, this flux factor is multiplied by the solar constant at the time of perihelion. Results are shown for an Earthlike eccentricity of .02 (top row), and a Marslike eccentricity of .1 (bottom row). The left column gives results for a precessional phase of zero degrees, while the right gives results for 90 degrees, both measured relative to the Northern Hemisphere summer solstice.

has no effect when the eccentricity is zero. The effects of precession angle and orbital eccentricity work in conjunction with each other, and cannot be disentangled.

At present, Earth's precession angle is close to 180^o, so that the Southern hemisphere is driven towards hotter summers and colder winters, while the Northern hemisphere is driven towards a weaker seasonal cycle. This pattern is not manifest in the observations (Figure 8.1) because the Northern Hemisphere has more land than the Southern Hemisphere, giving it a stronger seasonal cycle, owing to its lower thermal inertia. Relatively speaking, though, the Northern Hemisphere seasonal cycle is weaker than it would be if the precession angle were 90^o or 0^o. Coincidentally, the precession angle of Mars is also about 180^o at present, so that the Southern Hemisphere Martian winters are expected to be considerably colder than those in the North. Evidence that this indeed occurs, and its broader implications for Martian climate, will be taken up in Section 8.6.

The precession angles and orbital eccentricities of Earth and Mars have been different in the past, and will be different in the future. This has some extremely important implications for the evolution of climate, to which we now turn our attention.

8.5 Effect of long term variation of orbital parameters

The three *orbital parameters* that govern the seasonal and geographical distribution of insolation are the precession angle, obliquity, and eccentricity. All three change gradually on a scale of many thousands of years, owing basic laws of mechanics which apply to any planet in any solar system.

The evolution of the precession angle derives from a fairly elementary property of the mechanics of rigid-body rotation. The rotation axis of a rotating body subject to a net torque executes a rotation at constant rate about a second axis whose orientation is determined by the torque. The *precession rate* is determined by the magnitude of the torque and the angular momentum of the rotating body. The phenomenon of precession can be easily observed on a tabletop, by setting down a toy gyroscope with its axis inclined from the vertical. The top will precess, because there is a torque caused by the Earth's gravity and the force of the tabletop pushing up on the point of the top. For planets, the torque instead is provided by the slight deviations of the mass distribution from spherical symmetry. The equatorial bulge caused by rotation is a major player, but other asymmetries, including those due to the distribution of ice, and of major geographic features, are also of consequence.

Obliquity variations also stem from the basic properties of rigid-body rotation, but these variations arise from fluctuations in the torque on the planet, rather than the mean torque. The obliquity cycle is inextricably linked with the precessional cycle, which modulates the orientation of the aspherical planet with respect to the non-uniform gravitational field caused by the Sun, the planet's moon(s) (if sufficiently massive), and all the other planets.

Eccentricity evolves because the periodic elliptical orbit is a solution only of the two-body problem, consisting of a planet and its star in isolation. Although the gravity of the Sun greatly dominates that of the other planets in our Solar System (and most likely in other planetary systems as well) the relatively small tugs of the planets on each other causes eccentricity to change gradually. Early in the history of this subject, it was shown by Laplace and Lagrange that the semi-major axis remains very nearly constant in the course of such eccentricity changes. The results of the preceding section therefore imply that eccentricity cycles have only a weak effect on annual mean insolation, since the mean insolation changes little if the semi-major axis is held fixed, except for extremely non-circular orbits.

Tiny deviations of the stellar gravity field from the ideal $1/r^2$ law add up to significant effects on obliquity and eccentricity over sufficiently long periods of time. The fact that the Sun is not perfectly spherical enters the problem, and even general relativistic deviations from Newtonian gravity have major effects.

Eccentricity modulates the distance seasons, and precession determines whether they constructively or destructively interfere with the tilt seasons. Meanwhile, obliquity variations modulate the strength of the tilt seasons. The net result is a rich variety of rhythms and patterns in insolation, which may lead to dramatic cycles in the state of a planet's climate.

8.5.1 Milankovic cycles on Earth

Earth's precessional cycle is shown in Figure 8.12. The precession angle increases at a nearly constant rate, completing a cycle every 22,000 years. Though the variation in rate is not evident over any one cycle, the rate is not exactly constant, and therefore the phase drifts over the course of hundreds of thousands of years.

The precessional cycle is very rapid, and the precession angle has changed markedly even over historical times. Eight thousand years ago, when the first Sumerians poured into the valleys of the Tigris and Euphrates, the star we now call Polaris (the "Pole Star", in the tail of the Little Bear) was about 40^o of arc away from the star that the the North Polar axis then pointed to,and about which the constellations rotated at the time. The consequences of precession for change in seasonality are potentially highly consequential. In Figure 8.12, the July insolation at 65N is shown as a general indication of the magnitude of the seasonality effect; high northern July insolation in the precessional cycle goes with low January northern insolation, weak southern January (summer) insolation, and relatively strong southern July (winter) insolation. Ten thousand years ago, the Northern Hemisphere summer insolation was fully $40W/m^2$ greater than at present, and so the northern summers should have been considerably warmer than today, while the northern winters should have been considerably colder. The effect should show up especially over land, which is dark enough to absorb most of the solar radiation and has low enough thermal inertia to respond nearly instantaneously to seasonal changes. The climate system in its full glory is nonlinear and complex, so the response of climate to this change in seasonality could show up in any number of unexpected ways, and not simply as an enhancement of the Northern Hemisphere seasonal cycle over land.

The event which is most likely to be a recent manifestation of the precessional cycle is the "Climatic Optimum," covering the period of about 5000 to 7000 years ago (see Chapter 1). The term is most often used to refer to a period of generally warmer Eurasian temperatures. The "optimum" is sometimes said to be about 1-2K warmer than present, but it is difficult to get reliable estimates of global mean temperatures, or even annual means. What is certain is that some regions during some seasons were warmer than they were at recent pre-industrial times. At about the same time, the Sahara, which is now a torrid desert, experienced a period of greening, with currently dry riverbeds ("wadis") filled with water, and a teeming variety of animal life and flora not known at present. The greening of the Sahara is thought to be associated with atmospheric circulation systems known as "monsoons," forced to a greater extent by the enhanced heating of Northern Hemisphere subtropical land. A central question, though, is why the greening of the Sahara, and the Climatic Optimum occurred several thousand years after the precessional peak in Northern Hemisphere insolation. There are some indications that the warming may have *begun* as much as 10,000 years ago, but the question of the physics accounting for the time delay in response remains unsettled. Candidates for the necessary inertia in climate response include vegetation adaptation,

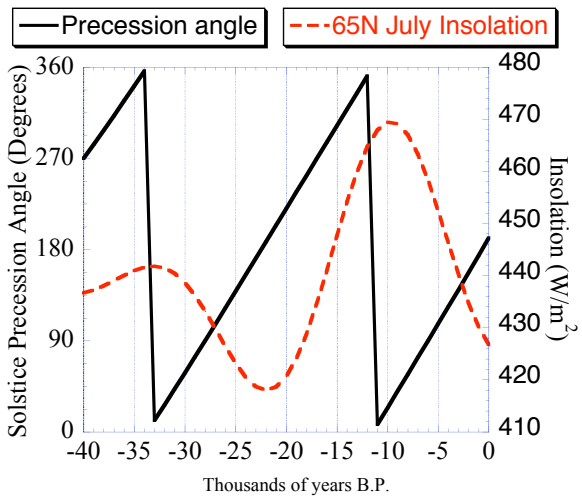

Figure 8.12: Evolution of precession angle relative to the Northern Hemisphere Summer Solstice, and the associated July insolation at 65N. Data taken from Berger and Loutre (1991).

land ice, and deep ocean heat storage.

Looking further back in time, the obliquity and eccentricity variations become significant, though of course, the precession cycle also continues to have a large effect. The Earth's obliquity and eccentricity cycle is shown in Figure 8.13. The amplitude of the obliquity cycle varies considerably over time, but it's dominant period is on the order of 40,000 years. The Earth's obliquity varies narrowly in a range from about 22^o to 24.5^o. At present, the Earth is in the middle of its obliquity range. Eccentricity varies on a longer time scale of approximately 100,000 years. However, in Figure 8.13 there are also hints of 400,000 year cycle of eccentricity, whose fingerprint consists of two high eccentricity cycles followed by two low eccentricity cycles. This visual impression is borne out by spectral analysis. Currently, the Earth is near the low end of its eccentricity range, though it has gotten quite close to zero during the past two million years. At the other extreme, Earth's eccentricity has gotten as high as .055, or more than half that of Mars.

The idea that ice ages are due to changes in Earth's orbital parameters is nearly as old as the discovery of ice ages themselves. The idea has gained currency, but it is nearly as hard to justify today on basic physical principles as it was when first proposed. The main reason for its acceptance is circumstantial, in that increasingly detailed data on the observed rhythm of the ice ages shows the unmistakable imprint of the calculated rhythm of the orbital forcing. James Croll first proposed in the 1870's that changes in the Earth's eccentricity led to ice ages, and his idea was refined a half century later by Milutin Milankovic, whose name is now generally attached to the theory. The centerpiece of Milankovic's idea is that ice ages require the accumulation of snow on land, and that

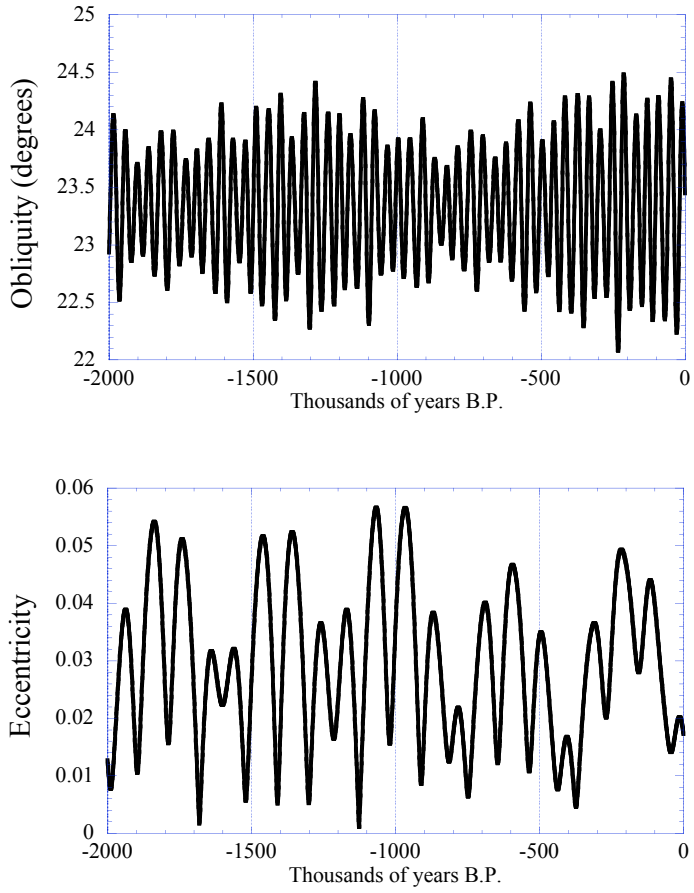

Figure 8.13: Evolution of the Earth's obliquity and eccentricity. Data taken from Berger and Loutre (1991).

this in turn is favored by mild summers (limiting melting of old snow and ice) and warmer, but still sub-freezing, winters (favoring snow accumulation, since warmer air contains more water). The gaping hole in Milankovic's theory is that it predicts that ice ages should follow the precessional cycle. In particular, the Northern Hemisphere and Southern Hemisphere should have ice ages in alternation every 10,000 years, with the severity of the ice ages modulated by the eccentricity cycle. This is not at all what is observed. Figure 8.14 shows the Antarctic temperature record for the past 400,000 years, together with eccentricity and the July insolation at 65N. Numerous other temperature proxies worldwide show that the Northern Hemisphere temperature, and global glacier ice volume, is nearly in phase with the Antarctic temperature record, so that the Antarctic temperature can be taken as an index of when the world is in an ice age. The dominant signal in the climate response is an approximately 100,000 year spacing in the major interglacial warm periods, and a similar spacing in the coldest glacial periods. Crudely speaking, each interglacial corresponds to a peak in eccentricity, and a time within which (during parts of the precessional cycle) the Northern Hemisphere seasonality is unusually strong. This is somewhat reminiscent of the Milankovic mechanism, but what filters out the high frequency precessional cycle? Why does the entire Earth fall into an ice age at the same time, rather than alternating between hemispheres? A closer examination of the 65N July insolation strongly suggests that major global deglaciations occur when the Northern Hemisphere seasonality is weak, suggesting that the Earth listens to the Northern hemisphere forcing more than the Southern, in deciding when to have an ice age. This probably has something to do with the fact that the Northern hemisphere has more land, and hence more seasonality, than the Summer, but the precise way this asymmetry influences global glaciation remains largely obscure.

The problem is not that the amplitude of radiative forcing associated with Milankovic cycles is small: it amounts to an enormous $100W/m^2$, with the amplitude determined by the eccentricity cycle. The problem is that the forcing occurs on the fast precessional time scale, whereas the climate response is predominately on a much slower 100,000 year time scale. One does not so much need an amplifier of Milankovic forcing, as a "rectifier," which is sensitive to the *amplitude* of the precessional variation, rather than to its mean. Recall that atmospheric CO_2 is observed to vary on the glacial-interglacial time scale. Certainly, this is a major piece of the puzzle, since the drop in CO_2 during glacial times is sufficient to account for a major portion of the cooling of the climate, particularly in the Southern Hemisphere (see Chapter 4). CO_2 is a globalizing effect, and (insofar as it is linked to the glacial-interglacial physical climate changes) an amplifying feedback. The circumstantial role of CO_2 in ice ages is also a reprise of an old idea. The 19th century physicist Tyndall, whose work on infrared spectroscopy is at the foundations of our current understanding of the greenhouse effect, was primarily interested in explaining the ice ages, and the association reappeared later in the work of Chamberlain. The mechanism of the CO_2 cycle not known, but almost certainly involves CO_2 storage in the deep ocean. The lack of a theory for the glacial-interglacial CO_2 cycle is the central impediment to a theory of the ice ages. The presence of ice does seem to be a prerequisite for a strong climate response to orbital forcing. Before the onset of permanent polar ice at the beginning of the Pleistocene, response to orbital forcing was weak (see Chapter 1). Besides CO_2, ocean circulations can potentially play a major role in globalizing and rectifying the Northern Hemisphere signal, through direct heat transport as well as indirect effects on CO_2. The answer to the mystery of the ice ages lies somewhere in the space: *ice,ocean,* CO_2, but how the system works its miracles to yield a 100,000 year cycle is still unknown.

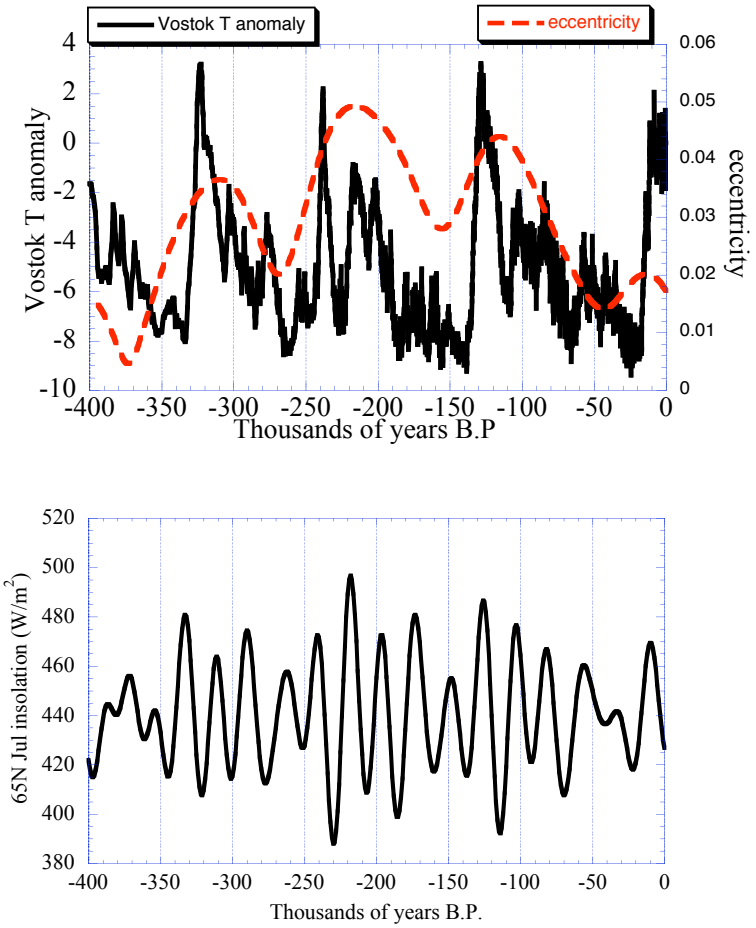

Figure 8.14: Comparison of Antarctic temperature reconstructed from Vostok ice core deuterium measurements, with the Earth's eccentricity cycle. The bottom panel shows the corresponding July insolation at 65N. Temperature is given as deviation from the mean modern value. Vostok temperature data was taken from Peteet *et al.* (2001).

8.5.2 Milankovic cycles on Mars

As expected from general mechanical considerations, Mars has Milankovic cycles analogous to those of Earth. Mars' cycles differ in some key respects, because of the lack of a massive moon, and because of the proximity of Jupiter.

As for Earth, the precession angle of Mars increases at a nearly constant rate. However, because Mars does not have a moon as massive as Earth's, the precession is dominated by Solar gravity, and is slower. The Mars precessional cycle has a period of approximately 50,000 Earth years. The current precession phase is 145^o, and will reach 180^o in about 5000 years.

The obliquity and eccentricity variations are shown in Figure 8.15. Obliquity has short term variations with amplitude on the order of 20^o. The period is not visible in the figure, but a finer scale examination of the data shows that the period is about 125,000 Earth years in recent times. The amplitude is markedly larger than that of Earth's obliquity cycle, but what is even more remarkable is that the obliquity drifts to values as large as 47^o over 10 million years. The extreme obliquity variations are directly linked to the absence of Earth's massive moon, which can be shown to provide a considerable damping effect on obliquity. This raises the intriguing possibility that a massive moon may be a necessary condition for a planet to avoid extreme climate fluctuations that could compromise its habitability. Calculations of the Earth's obliquity have also been carried out for tens of millions of years, and do not yield any greater variations than have been encountered in the past million years.

Mars is close to its maximum eccentricity at present, though it can get somewhat larger. The eccentricity of Mars undergoes quasiperiodic large amplitude cycles with a period on the order of 3 million years. In addition, there are short period, lower amplitude eccentricity variations with a period on the order of 100,000 years, rather similar to Earth's. In contrast, the very long period variations are not found in Earth's eccentricity.

Mars has no ocean, little thermal inertia, and a thin atmosphere that has a relatively modest effect on the planet's surface temperature. These features should lead to a different, and perhaps simpler, response to orbital forcing on Mars as compared to Earth. The predicted climate changes have been simulated in detail using comprehensive climate models, but we will confine ourselves here to some general remarks. The main effect of Martian Milankovic cycles is likely to be the redistribution of water deposits, in the form of either glaciers or permafrost. There are two aspects to this redistribution. On the short precessional time scale, the asymmetry between the Northern and Southern polar ice caps should reverse. For example, about 25000 years ago, the Southern hemisphere should have had milder summers and winters, while the Northern had cold winters and hot summers; the default reasoning would imply that at such times, the Southern ice cap should be large and be composed mainly of water ice, whereas the Northern ice cap becomes smaller and experiences massive seasonal CO_2 snow deposition. On the time scale of millions of years, the obliquity of Mars becomes much greater, leading potentially to a situation where water my migrate from poles that are seasonally very hot, and re-deposit in the tropics. At times of much lower obliquity, permafrost ice may migrate to both poles. The migration of water deposits and changes in patterns of deposition of CO_2 snow probably leaves some imprint on the surface geology of Mars, and the growth and decay of glaciers certainly does. These offer some prospects for reconstructing the consequences of Milankovic cycles on Mars. Even better information would be obtained through analyzing cores of the polar ice caps, much as is done in Antarctica and Greenland. It is a very exciting development that the technology for doing this robotically on Mars is already under development. With respect to Mars, we are more or less at the stage of Croll or Milankovic, who thought they found the key to Earth's ice ages. Data showed they were on the right track, but that the climate system is much more intricate than they imagined. Given

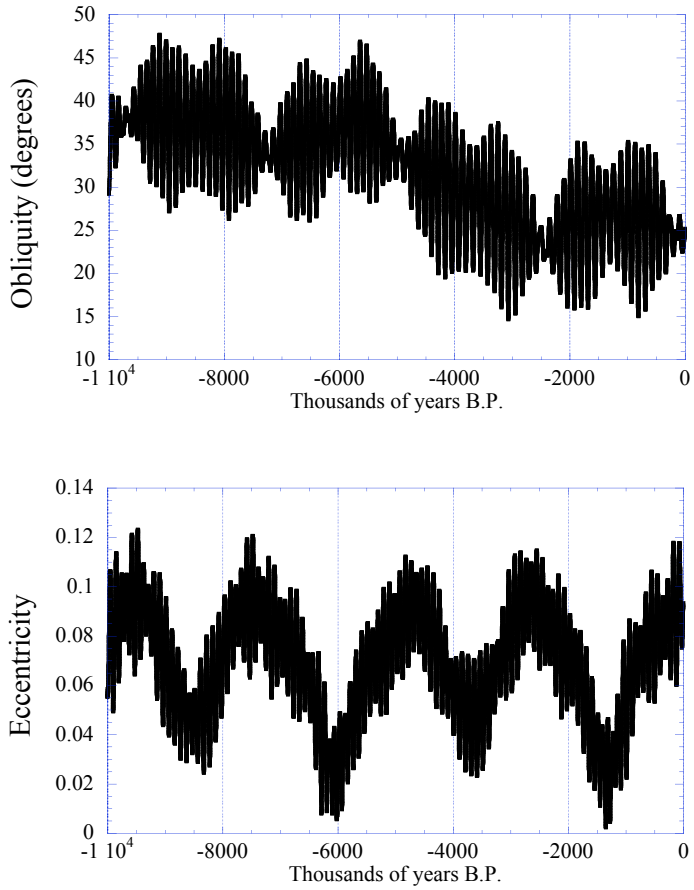

Figure 8.15: Evolution of Mars' obliquity and eccentricity. Data taken from Laskar *et al.* (2003).

that we do not yet have a satisfactory theory leading from orbital variations to climate response on Earth, one can look forward to many surprises, once data on the Martian climate response becomes available.

8.6 A palette of planetary seasonal cycles

In this section we apply the preceding ideas to a range of planets (including Earth under very different past climate regimes), to see how they play out. This section is still under development. The following provides an outline of topics to be covered. Each topic will be supplemented so far as possible by observations or (when observations are unavailable) GCM simulations, illustrating the behavior inferred from the basic reasoning.

8.6.1 Airless planets and moons

This includes a discussion of Earth's moon, as well as nearly airless exotic cases like Triton. Pluto is an interesting case, because of its high eccentricity; like Triton, it should have a seasonal nitrogen "micro-atmosphere."

8.6.2 Venus

Slow rotation implies no real distinction between seasonal and diurnal cycles. High thermal inertia, and efficient atmospheric heat transport. Therefore no significant geographical or seasonal/diurnal variation of temperature, except in the upper stratosphere.

8.6.3 Gas Giants

Issue of Solar vs. interior heating. Penetration depth of solar forcing. Estimates of thermal inertia of the upper atmospheres of gas giants, in comparison to their orbital periods. Possibilities for significant seasonal cycle on Uranus.

Extrasolar giant planets. "Roasters"

8.6.4 Mars, present and past

The current seasonal cycle on Mars. Asymmetries between the solstices; Southern winter is much colder than Northern winter. CO_2 condensation, and the seasonal cycle of surface pressure. Seasonal cycle of Ar as a diagnostic of polar condensation. The strong, deep diurnal cycle on Mars. Factors limiting the polar night temperature drop on Mars, and relevant observations.

A survey of GCM results on the redistribution of water ice in the course of Martian Milankovic cycles

Possible long term evolutions of Marslike planets. Revisit the toy model of the interaction between sublimation of a massive CO_2 glacier and the CO_2 greenhouse effect, with a seasonal cycle included this time.

Early Mars: Mars with a thick CO_2 atmosphere, but without major oceans. Estimates for various values of the obliquity. Can we have seasonal summertime melting? Can we form transient or permanent CO_2 glaciers?

8.6.5 Snowball Earth

A full glaciated Snowball Earth is more like Mars than like present Earth. Low thermal inertia of surface. High albedo implies weaker solar driving, and lower temperatures imply less role of water vapor and clouds. This section is based in part on my papers in Nature and JGR on the Snowball Earth, and datasets from those simulations will be provided for use with the Workbook.

8.6.6 Hothouse Earth

This section deals with climates like the Cretaceous, with no permanent polar ice. A particular emphasis is on the extreme seasonal cycle of polar continents, the moderate high latitude maritime seasonal cycle, and the difficulty of "getting rid of winter" in the interior of midlatitude continents.

8.6.7 Earth without a moon

This is an exercise in habitability and comparative planetology. If the Earth had no mssive moon, its obliquity would vary over a wider range, like that of Mars. What are the consequences for climate? For a waterworld, high obliquity actually leads to a fairly equable climate, but life on large continents becomes very problematic.

8.6.8 Titan

Remark on the obliquity of moons: one must be careful to distinguish the obliquity relative to the plane of the moon's orbit from the obliquity relative to the ecliptic. This is critical in cases like Titan, which are tide-locked to the parent planet.

Observations of Titan seasonal cycle. Expectations from thermal inertia and obliquity. Low rotation rate and thick atmosphere means that the atmosphere can transport heat very efficiently,and the very weak solar forcing means the atmosphere doesn't need to transport much heat to equalize the temperature. Upper atmospheric vs. surface solar heating. Implications of the methane and nitrogen "hydrological" cycles.

Chapter 9

Evolution of the atmosphere

This chapter goes over some of the basics of how atmospheres change over time. The primary attention is given to models of the long term CO_2 evolution in terms of silicate weathering, after the fashion of Walker and Kasting, and of Berner. Models of atmospheric mass loss (thermal escape, hydrodynamic escape, solar wind sputtering and ejection due to giant collisions) are also discussed. A very basic survey of some relevant atmospheric chemistry is also included. The focus here would be on the interplay of methane and oxygen in Earthlike conditions, the formation of tholin clouds on Titan and Early Earth, and the chemistry accounting for CO_2 stability on Mars. We will also talk a little bit about the methane/oxygen story and its possible role in the Early Earth climate.

9.1 The CO_2 weathering thermostat

$$[**Difeq for CO2 evolution] \tag{9.1}$$

$$\frac{W}{W_0} = (\frac{r}{r_0})^{\alpha}(\frac{p}{p_0})^{\beta}e^{-\frac{T-T_0}{T_U}} \tag{9.2}$$

where W is the weathering rate [**rate of removal of CO_2 from the atmosphere, in Moles per sq m per unit time] r is the runoff rate, p is the partial pressure of CO_2 in the atmosphere and T is the global mean temperature. W_0 is the weathering for the reference state with runoff r_0, carbon dioxide partial pressure p_0 and temperature T_0. α, β and T_U are empirically determined constants. The last of these represents the direct temperature sensitivity of the Urey reaction.

9.2 Methane and Oxygen

9.3 Escape of the atmosphere to space

9.4 Other topics in chemical evolution

Chapter 10

Consequences of heat transport

This chapter introduces the importance of meridional heat transport, and illustrates the qualitative effects of heat transport using diffusive energy balance models. Energy balance models are introduced not as a way of finding accurate solutions to climate problems, but as a way of probing general effects of heat transport in an exploratory fashion. It provides a motivation for the (hard) fluid dynamical issues to be taken up in Volume 2. The discussion of EBM's somewhat follows the development in my lecture notes from the WHOI GFD summer program. The limitations of diffusive EBM's are amply discussed.

Diagnosis of heat transports in the Earth atmosphere, based on ERBE data. Estimates of effective diffusivity.

10.1 Mechanisms of heat transport

Flux of dry static energy, and flux of latent heat. Expression for fluxes in terms of correlation with winds. General scaling arguments.

10.2 Formulation of energy balance models

The column-averaged heat budget. Representation of energy transport by a diffusion. Limitations of the diffusive approximation. The problem of determining diffusivity. The problem of representing latent heat transport. The problem of representing tropical heat transport. Problems of representing lapse rate, water vapor and cloud effects.

Alternate approach to formulation of a vertically integrated model: Form equations for vertically integrated entropy (amounts to diffusing potential temperature). In this case, heating terms show up as an entropy source and don't integrate out in terms of the T.O.A. budget.

Models based on diffusion of moist static energy.

Representing the hydrological cycle: Is diffusing moisture a good idea? Distinction between moisture as a radiative agent and moisture as a means of energy transport and source of precip.

Representing the tropics. Does diffusion of MSE represent the Hadley cell?

Representation of Top-of-Atmosphere and surface fluxes. Models which track surface and atmospheric temperature separately.

10.3 Equilibrium energy balance models

This section discusses general properties of the solutions of EBM's in the steady case.

10.4 The seasonal cycle revisited

Here we re-do some calculations of the seasonal cycle, this time incorporating a diffusive model of lateral heat transport.

10.5 Ice albedo feedback

This revisits the basic ice-albedo feedback problem in the context of energy balance models. First, the steady state problem is discussed. Then, the effect of the seasonal cycle on the bifurcation diagram is discussed.

The related problem of the CO_2 ice cap feedback on Mars is also discussed.