Chapter 3

Elementary models of radiation balance

3.1 Overview

Our objective is to understand the factors governing the climate of a planet. In this chapter we will be concerned with energy balance and planetary temperature. 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 vacuum 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. The wavenumber can be viewed as the frequency measured in alternate units, and so we will often refer to wavenumber and frequency interchangeably. Although mks units are preferred throughout this book, we follow spectroscopic convention and make an exception for wavenumber when dealing with infrared radiation, which will usually be measured in cm^{-1} since it yields comfortable and familiar ranges of numbers. Wavelengths themselves will sometimes be measured in μm (microns, or $10^{-6}m$). 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, just as the mass density of a three dimensional object describes the distribution of mass in three-dimensional space.

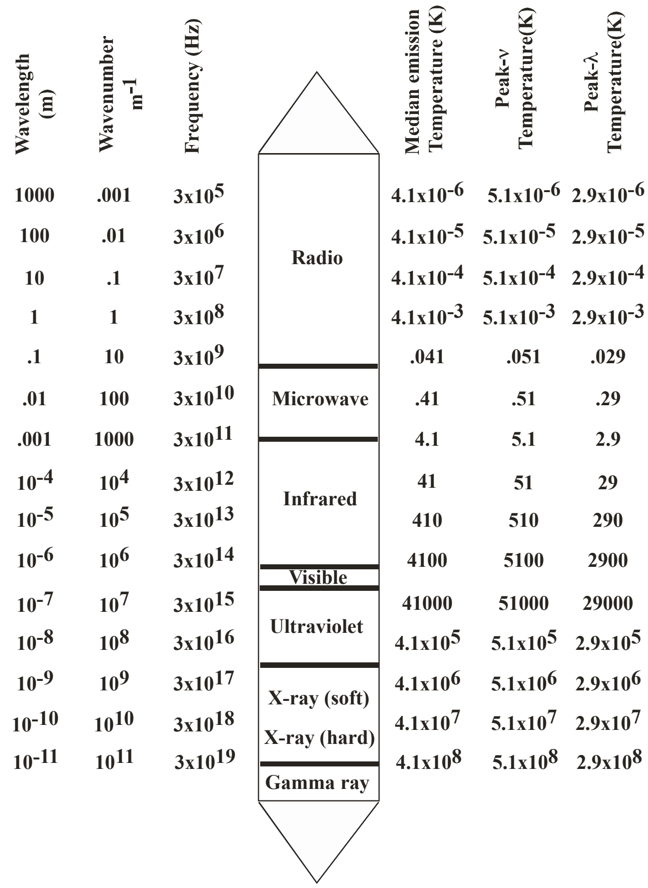

Figure 3.1: The electromagnetic spectrum. The Median Emission Temperature is the temperature of a blackbody for which half of the emitted power is below the given frequency (or equivalently, wavelength or wavenumber). The Peak-ν Temperature is the temperature of a blackbody for which the peak of the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

Before proceeding, we must pause and talk a bit about how the "size" of collections of directions are measured in three dimensions. For collections of directions on the plane, the measure of the "size" of the set of directions between two directions is just the angle between those directions. The angle is typically measured in radians; the measure of the angle in radians is the length of the arc of a unit circle whose opening angle is the angle we are measuring. The set of all angles in two dimensions is then 2π radians for example. A collection of directions in three dimensional space is called a *solid angle*. A solid angle can sweep out an object more complicated than a simple arc, but the "size" or measure of the solid angle can be defined through a generalization of the radian, known as the *steradian*. The measure in steradians 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 steradians, while a set of directions tracing out the entire sphere (i.e. all possible directions) has measure 4π steradians. If we choose some specific direction (e.g. the vertical) as a reference direction, then a direction in three dimensional space can be specified in terms of two angles, θ and ϕ, where *theta* is the angle between the reference direction and the direction we are specifying, and ϕ is the angle along a circle centered on the reference direction. These angles define a spherical polar coordinate system with the reference direction as axis; $0 \geq \theta \leq \pi$ and $0 \geq \phi \leq 2\pi$. In terms of the two direction angles, the differential of solid angle Ω is $d\Omega = \sin\theta d\theta\ d\phi = -(d\cos\theta)(d\phi)$. Generally, when writing the expression for $d\Omega$ in the latter form we drop the minus sign and just remember to flip the direction of integration to make the solid angle turn out positive. We recover the area of the unit sphere by integrating $d\Omega$ over $\cos\theta = -1$ to $\cos\theta = 1$ and $\phi = 0$ to $\phi = 2\pi$. A similar integration shows that the set of directions contained within a cone with vertex angle $\Delta\theta$ measured relative to the altitude of the cone has measure $2\pi(1-\cos\Delta\theta)$ steradians. A narrow cone with $\Delta\theta \ll 1$ has measure $\pi(\Delta\theta)^2$ steradians.

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 $\hat{n}$. The energy spectrum $\Sigma(\vec{r},\nu,\hat{n})$ at this point is defined such that 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 include, and $d\Omega$ measures the range of solid angles we wish to include.

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 $c\Sigma dA d\nu d\Omega$, 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. The flux spectrum defined in this way is usually called the *spectral irradiance*; integrated over all frequencies, it is called the *irradiance*.

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 $2000 Calories/day$. (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 $1 Calorie = 4184J$.

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 granularity 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 μm, 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 irradiance (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, $d\Omega = 2\pi d\cos\theta$ for the set of all rays making an angle θ relative to the normal to the patch. Integrating $B\cos\theta d\Omega$ 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 $\pi BdAd\nu$. 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 can be 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 $20814m^{-1}$, or a wavelength of 48 μm; 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

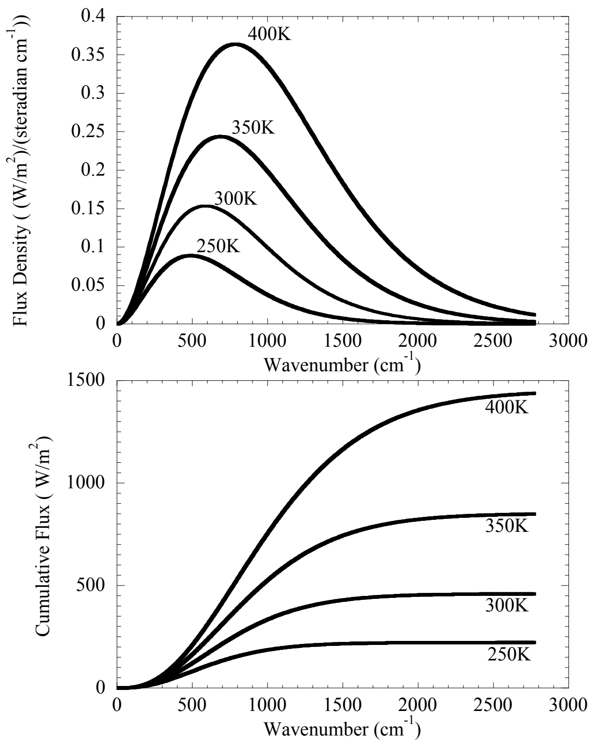

Figure 3.2: The spectrum of blackbody radiation for the various temperatures indicated on the curves. Upper Panel: The Planck density in wavenumber space. Lower Panel: The cumulative emission as a function of wavenumber. Note that the density has been transformed such that the density times dn is the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

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 $u = h\nu/kT = hc/\lambda kT$, 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. 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 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:

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

[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.

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 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:

[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 $6\zeta(4) = \pi^4/15$

- The only source of energy heating the planet is absorption of light from the planet's host star.
- The *planetary albedo*, or proportion of sunlight reflected by the planet as a whole including its atmosphere, 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 most other 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 $4\pi r_\circledast^2(\pi B(\nu, T_\circledast))$, where $r_\circledast$ is the radius of the star and $T_\circledast$ 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 $\pi B r_\circledast^2/r^2$, and the total flux is $\sigma T_\circledast^4 r_\circledast^2/r^2$. 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" , and will be denoted by $L_\circledast$, or simply L where there is no risk of confusion with latent heat. The solar (or stellar) "constant" depends on a planet's orbit, but the *luminosity* of the star is an intrinsic property of the star. The stellar luminosity is the net power output of a star, and if the star's emission can be represented as blackbody radiation, the luminosity is given by $\mathcal{L}_\circledast = 4\pi r_\circledast^2 T_\circledast^4$.

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 $L_{\circledast}$, the energy per unit time impinging on the planet's surface is $\pi a^2 L_{\circledast}$; the rate of energy absorption is $(1-\alpha)\pi a^2 L_{\circledast}$, 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

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

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 $L_{\circledast}$ in terms of the photospheric temperature, the result is

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

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 μm. 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 μm. An isothermal planet at the orbit of Mars would radiate with a median wavenumber of 550 cm^{-1}, corresponding to a wavelength of 18 μm.

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

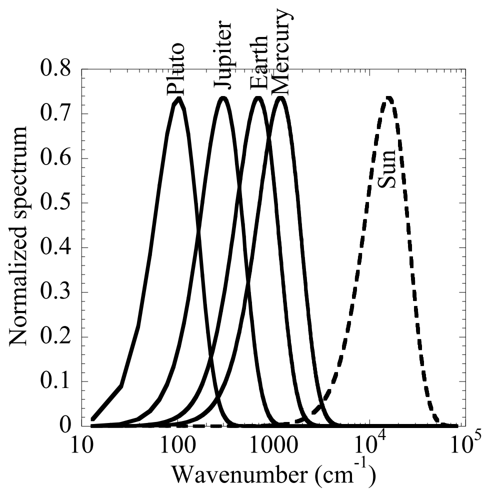

Figure 3.3: The Planck density of radiation emitted by the Sun and selected planets in radiative equilibrium with absorbed solar radiation (based on the observed shortwave albedo of the planets). The Planck densities are transformed to a logarithmic spectral coordinate, and all are normalized to unit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

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 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 $\frac{r}{r_\circledast}$ 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

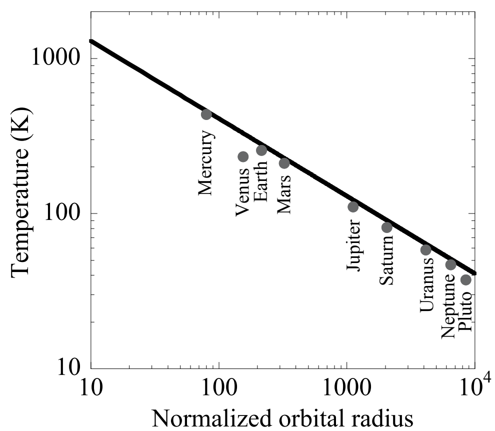

Figure 3.4: The equilibrium blackbody temperature of an isothermal spherical zero-albedo planet, as a function of distance from a Sun having a photospheric temperature of 5800K. The orbital distance is normalized by the radius of the Sun. Dots show the equilibrium blackbody temperature of the Solar System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

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 hierarchy 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 atmosphere, and assume that it is well mixed with uniform mass concentration q. This gas is transparent to solar radiation but interacts strongly enough with infrared that when a sufficient amount is mixed into a parcel of air, it turns that parcel into an ideal blackbody. Such a gas, which is fairly transparent to the incoming shortwave stellar radiation but which interacts

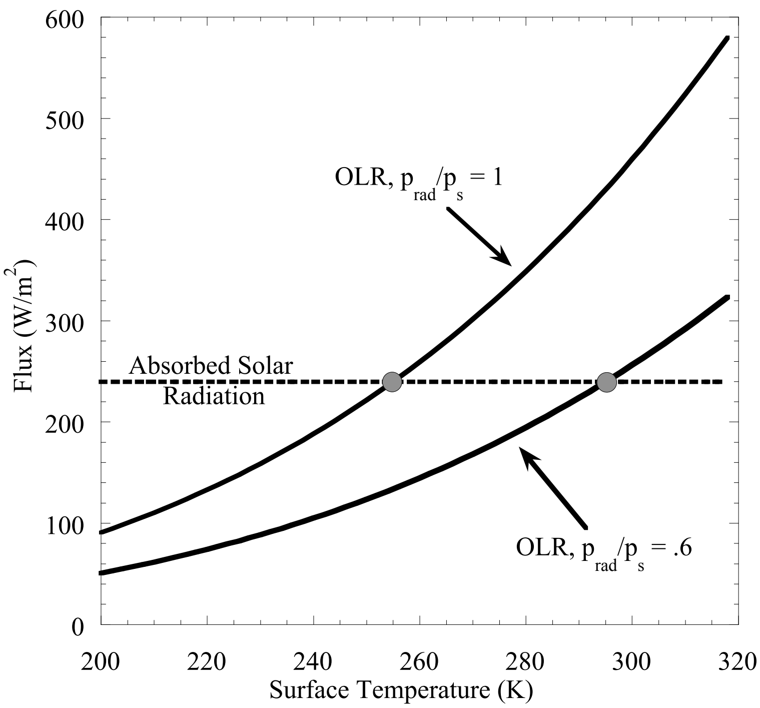

Figure 3.5: Determination of a planet's temperature by balancing absorbed solar energy against emitted longwave radiation. The horizontal line gives the absorbed solar energy per unit surface area, based on an albedo of .3 and a Solar constant of $1370W/m^2$. The OLR is given as a function of surface 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.

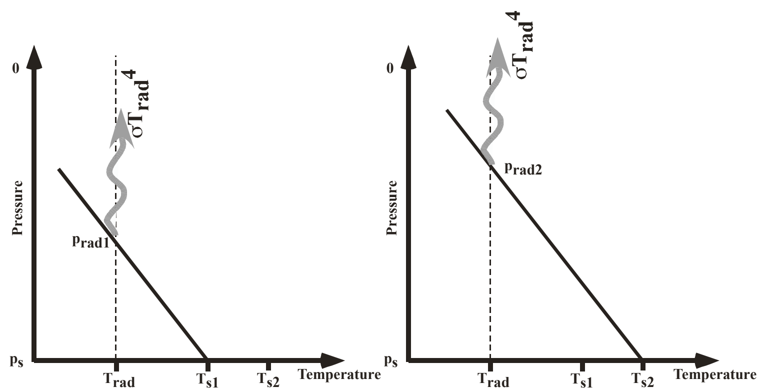

Figure 3.6: Sketch illustrating how the greenhouse effect increases the surface temperature. In equilibrium, the outgoing radiation must remain equal to the absorbed solar radiation, so T_{rad} stays constant. However, as more greenhouse gas is added to the atmosphere, p_{rad} is reduced, so one must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

strongly with the outgoing (generally infrared) emitted radiation is called a *greenhouse gas*, and the corresponding effect on planetary temperature is called the *greenhouse effect.* 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. The mass of greenhouse gas that must be mixed into a column of atmosphere with base of 1 m^2 in order to make that column act begin to act like a blackbody is characterized by the *absorption coefficient* κ, whose units are m^2/kg. Here we'll assume κ to be independent of frequency, temperature and pressure, though for real greenhouse gases, κ depends on all of these. Since the mass of greenhouse gas in a column of thickness Δp in pressure coordinates is $q\Delta p/g$, then the definition of κ implies that the slab acts like a blackbody when $\kappa q\Delta p/g > 1$. When $\kappa q p_s/g < 1$ then the entire mass of the atmosphere is not sufficient to act like a blackbody and the atmosphere is said to be *optically thin.* For optically thin atmospheres, infrared radiation can escape from the surface directly to space, and is only mildly attenuated by atmospheric absorption. When $\kappa q p_s/g \gg 1$, the atmosphere is said to be *optically thick.*

If the atmosphere is optically thick, we can slice the atmosphere up into a stack of slabs with thickness Δp_1 such that $\kappa q\Delta p_1/g = 1$. Each of these slabs radiates like an ideal blackbody with temperature approximately equal to the mean temperature of the slab. Recall, however, that another fundamental property of blackbodies is that they are perfect *absorbers* (though if they are only blackbodies in the infrared, they will only be perfect absorbers in the infrared). Hence *infrared radiation escapes to space only from the topmost slab.* The OLR will be determined by the temperature of this slab alone, and will be insensitive to the temperature of lower portions of the atmosphere. The pressure at the bottom of the topmost slab is Δp_1. We can thus identify Δp_1 as the characteristic pressure level from which radiation escapes to space, which therefore will be called p_{rad} in subsequent discussions. The radiation escaping to space – the OLR– will then be approximately $\sigma T(p_{rad})^4$. Because temperature decreases with altitude on the adiabat the OLR 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. All other things being equal 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. The key insight to be taken from this discussion is that *the greenhouse effect only works to the extent that the atmosphere is colder at the radiating level than it is at the ground.*

For real greenhouse gases, the absorption coefficient varies greatly with frequency. Such gases act on the OLR by making the atmosphere very optically thick at some frequencies, less optically thick at others, and perhaps even optically thin at still other frequencies. In portions of the spectrum where the atmosphere is more optically thick, the emission to space originates in higher (and generally colder) parts of the atmosphere. In reality,then, 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}). Some of the implications of a frequency-dependent absorption coefficient are explored in Problem **??**, and the subject will be taken up at great length in Chapter 4.

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, 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 is treated 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 $238W/m^2$, corresponding to a radiating temperature of $255K$. This is much less than the observed surface temperature of $298K$, which would radiate at a rate of $446W/m^2$ 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 prevalence 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

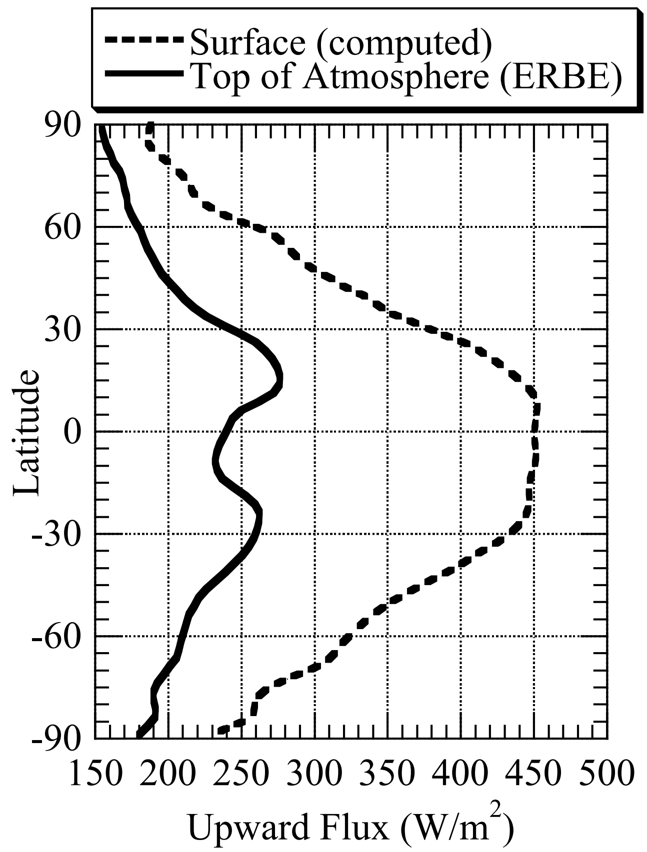

Figure 3.7: The Earth's observed zonal-mean OLR for January, 1986. The observations were taken by satellite instruments during the Earth Radiation Budget Experiment (ERBE), and are averaged along latitude circles. The figure also shows the radiation that would be emitted to space by the surface (σT_s^4)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

condensed substances absorb infrared as well as water does. Liquid methane (important on Titan) and CO2 ice (important on present and early Mars) are comparatively poor infrared absorbers. 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 atmospheric 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 and neglect of variations in c_p, 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

	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.

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 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. 2.33, 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. 2.27 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 phenomenon, 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 what 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. Indirect inferences based on the rate of ocean heat uptake indicate 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 [3]. 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.

[3] At the time of writing, top-of-atmosphere satellite measurements are not sufficiently accurate to permit direct observation of this imbalance

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 $(1-\alpha)L_{\circledast}$; the energy balance determining the ground temperature is then $\sigma T^4 = (1-\alpha)L_{\circledast}$, 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 consideration. 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. The term *insolation* will be used to refer to an incident solar flux of this type, sometimes with additional qualifiers as in "surface insolation" to distinguish the flux reaching the ground from that incident at the top of the atmosphere. For an isothermal planet $S = \frac{1}{4}L_{\circledast}$, while at the opposite extreme $S = L_{\circledast}$ for the instantaneous response at the *subsolar point* – the point on the planet at which the sun is directly overhead. 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

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.

variations will be given in Chapter 7.

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.

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 7, 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

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

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 $L_\circledast$. In this plot, we have taken $OLR = \sigma T^4$, which assumes no greenhouse effect [4]. 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 $L_\circledast$. When $L_\circledast$ is sufficiently small (as in the case $L_\circledast = 1516 W/m^2$ 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 $L_\circledast$ is sufficiently large (as in the case $L_\circledast = 2865 W/m^2$ 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 $L_\circledast$, 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 $L_\circledast$. 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 $L_\circledast$ is increased smoothly from low values, the temperature of 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 $L_\circledast$ is further increased. As $L_\circledast$ 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 equilibrium 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

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

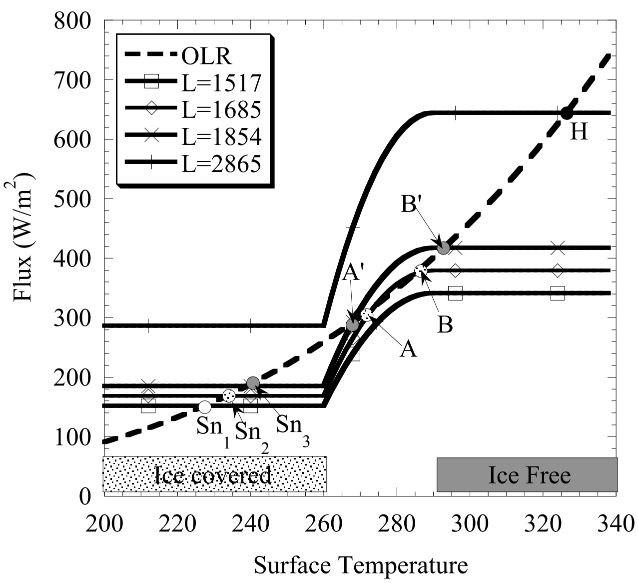

Figure 3.8: Graphical determination of the possible equilibrium states of a planet whose albedo depends on temperature in accordance with Eq. 3.9. The OLR is computed assuming the atmosphere has no greenhouse effect, and the albedo parameters are $\alpha_o = .1, \alpha_i = .6, T_i = 260K$ and $T_o =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

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

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

where $G = \frac{1}{4}(1-\alpha(T_s))L_\circledast - OLR(T_s)$. 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 continuos 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.

Moreover, if the net flux $G(T)$ is continuos 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 continuously by some parameter like $L_{\circledast}$, then new solutions are born in the form of a single marginal state which, upon further change of $L_{\circledast}$ 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

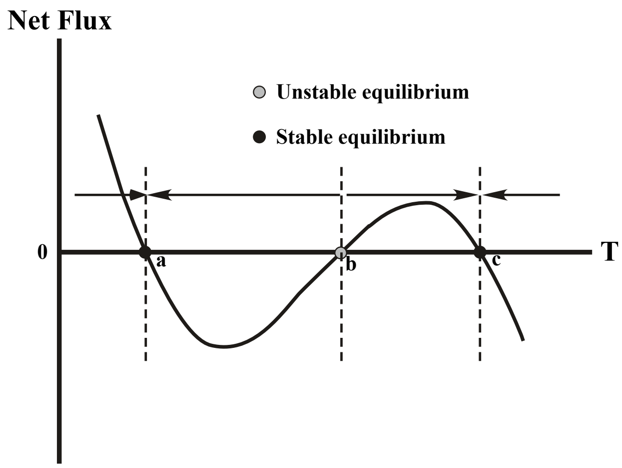

Figure 3.9: Sketch illustrating stable vs. unstable equilibrium temperatures.

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 $L_\circledast$ 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 $L_\circledast = 960W/m^2$, 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 $L_\circledast$ 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 $L_\circledast$ 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

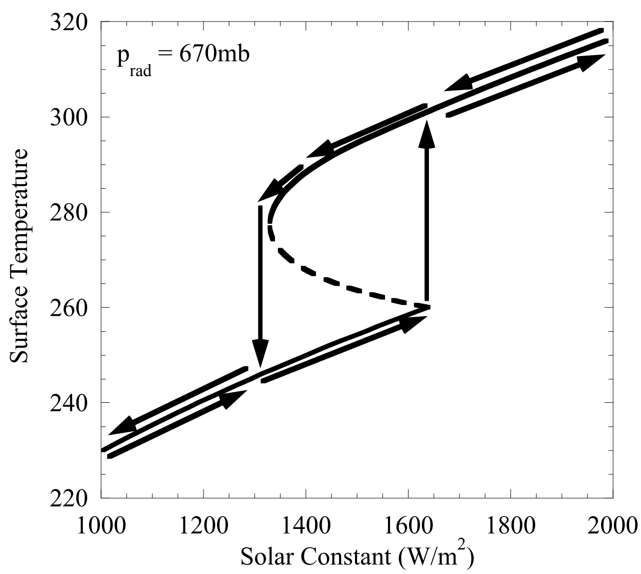

Figure 3.10: Hysteresis diagram obtained by varying $L_\circledast$ with p_{rad}/p_s fixed at .67. Arrows indicate path followed by the system as $L_\circledast$ is first increased, then decreased. The unstable solution branch is indicated by a dashed curve.

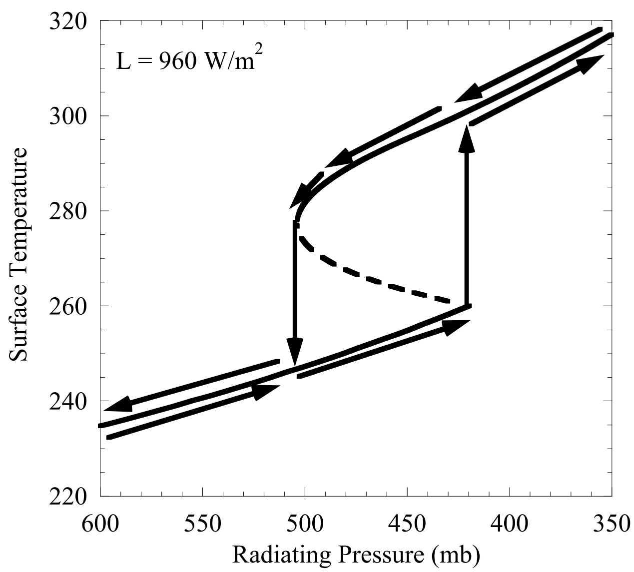

Figure 3.11: As in Fig. 3.10, but varying p_{rad} with $L_{\circledast} = 960W/m^2$.

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 8. Fig. 3.11 shows how the equilibria depend on p_{rad}, with $L_{\circledast}$ 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.

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 $L_{\circledast} = 1000W/m^2$, where the system is inevitably in a Snowball state with $T = 230K$. Let's now gradually increase $L_{\circledast}$. When $L_{\circledast}$ 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 $L_{\circledast}$ 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 $L_{\circledast}$ 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 $L_\circledast$ decreases, so that when we find ourselves once more at $L_\circledast = 1500 W/m^2$ 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 $L_\circledast$ is decreased further, the warm branch disappears at $L_\circledast = 1330 W/m^2$ 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 $L_\circledast$ is reduced further. The trajectory of the system as $L_\circledast$ 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 $L_\circledast$ 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 $L_\circledast = 960 W/m^2$, 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, $L_\circledast \approx 1170 W/m^2$, 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, $L_\circledast \approx 1290 W/m^2$ 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 8.

Exercise 3.4.2 Assuming an ice albedo of .6, how high does $L_\circledast$ 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 $L_\circledast(t) = L_{\circledast p} \cdot (.7 +$

$(t/22.975) + (t/14.563)^2)$, where t is the age of the Sun in billions of years ($t = 4.6$ being the current age) and $L_{\circledast p}$ 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 continuos 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 Chapter 8. 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 hierarchy 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 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 down 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

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

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 unambiguously 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 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 absorptivity 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 justification 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

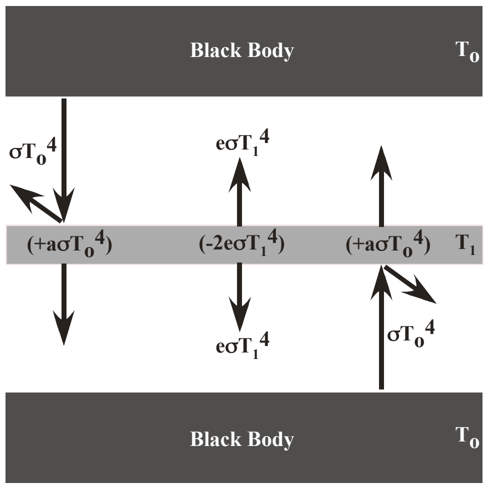

Figure 3.12: Sketch illustrating thought experiment for demonstrating Kirchoff's Law in the mean over all wavenumbers. In the annotations on the sketch, $a = \bar{a}(T_o)$ and $e = \bar{e}(T_1)$.

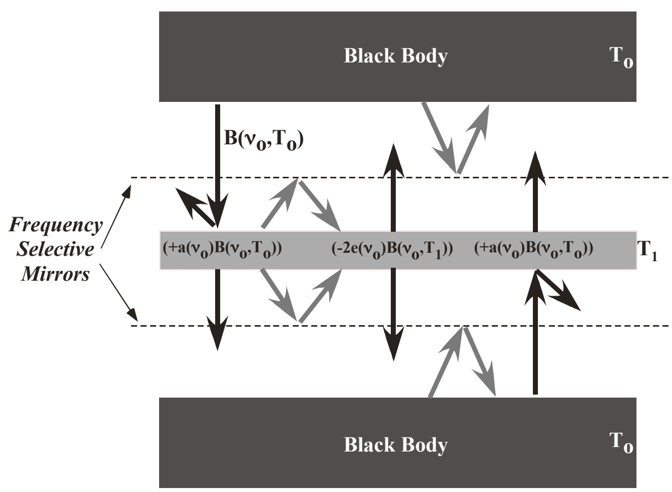

Figure 3.13: Sketch illustrating thought experiment for demonstrating Kirchoff's Law for a narrow band of radiation near frequency ν_0. The thin dashed lines represent ideal frequency-selective mirrors, which pass frequencies close to ν_0, but reflect all others without loss.

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 derivation 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. $\frac{1}{4}L_{\circledast}$ for the global mean or $L_{\circledast}$ 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 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

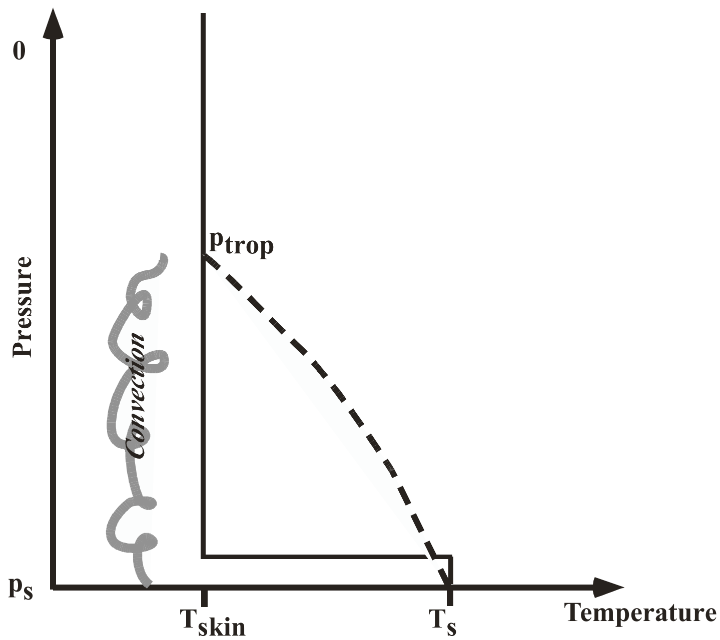

Figure 3.14: The unstable pure radiative equilibrium for an optically thin atmosphere (solid line) and the result of adjustment to the adiabat by convection (dashed line). The adjustment of the temperature profile leaves the surface temperature unchanged in this case, because the atmosphere is optically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

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 solar flux 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

$$2e_{ir}\sigma T^4 = e_{ir}OLR + a_{sw}S \tag{3.26}$$

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 case of present Mars, the fast thermal response of the atmosphere and surface 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 summer 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 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

	Skin temperature
Venus	213.
Earth	214.
Mars (255K sfc)	214.
Jupiter	106.
Titan	72.

Table 3.3: Computed skin temperatures of selected planets.

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 atmospheric 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, temperature will decrease 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; as will be discussed in Chapter 4, radiative effects can cause the temperature to continue decreasing with height above the top of the convectively mixed layer. Yet a further complication is that, in midlatitudes, large scale winds associated with storms are probably more important than convection in carrying out the stirring which establishes 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 $5mb$ surface pressure puts the tropopause at $2.4mb$, 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 detail in Chapter 4, 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.

3.7 For Further Reading

For more information on electromagnetic waves and electromagnetic radiation, I recommend:

- Jackson JD 1998: *Classical Electrodynamics.* Wiley
- Feynman RP, Leighton RB and Sands M 2005: *The Feynman Lectures on Physics, Vol 2.* Addison Wesley.

An engaging and accessible intellectual history of the quantum theory can be found in

- Pais A 1991: *Niels Bohr's Times.* Oxford University Press

For a derivation of the Planck distribution, see Chapter 1 of

- Rybicki GB and Lightman AP 2004: /it Radiative Processes in Astrophysics. Wiley-VCH.

The reader seeking a comprehensive introduction to nonrelativistic quantum theory will find it in Volume 1 of

- Messiah A 1999: *Quantum Mechanics*. Dover

In the Dover edition, this book is a bargain, and repays a lifetime of close study.