Chapter Two

Introduction to Radiative Transfer

2.1 THE OPTICAL DEPTH: THE MOST FUNDAMENTAL QUANTITY IN RADIATIVE TRANSFER

It is possible to sit through an entire radiative transfer class and not develop an intuition for the subject at all, since it is a technical one that threatens to drown the student in formalism. Thus, my approach is not to derive the governing equation of radiative transfer first, but rather to introduce you to the most fundamental quantity one needs to appreciate in order to understand the subject: the optical depth, which we shall denote by τ. The optical depth quantifies if a medium is transparent or opaque, thereby informing us on the approximation that we may use to describe the passage of radiation. It offers us a vocabulary to describe why and how the size of a star or exoplanet varies across wavelength. It identifies the surface of an object from which photons emanate and travel to our telescopes. It allows us to define the transit radius. To develop any intuition for radiative transfer, one needs to thoroughly understand what the optical depth is.

If you stare at clouds in the sky, you will realize that there are large, transparent clouds; there are also small, opaque clouds. Such an observation already informs us that physical size alone is a poor indicator of whether an object is transparent or opaque to radiation, because we are missing two other key ingredients: how loosely or tightly the material is being packed (i.e., the number or mass density) and how absorbent the constituent atoms or molecules are (i.e., the cross section or opacity). Formally, the optical depth is defined as

$$\tau \equiv \int n\sigma \, dx. \tag{2.1}$$

It is the product of the number density (n), a macroscopic quantity, multiplied by the cross section (σ), a microscopic quantity, and integrated across the spatial extent of the object. Other equivalent definitions of the optical depth exist, as we shall see shortly.

When a photon travels through a medium, we may ask how far it will propagate before it is absorbed by an atom or molecule. The answer is that it travels a distance corresponding to an optical depth on the order of unity,

$$\tau \sim 1. \tag{2.2}$$

In Problem 2.8.2, you will learn that τ is roughly the number of interactions the photon has with the medium. The precise value of the optical depth per interaction depends on the detailed properties of the atmosphere, but it currently suffices to understand that it is $\tau \sim 1$. The photon would travel a long distance if the medium was tenuous or poorly absorbent—the formulation of the optical depth automatically takes these two possibilities into consideration. If we return to the thought experiment of clouds in the sky, we now have a vocabulary for describing them: transparent clouds have $\tau \ll 1$, while opaque ones have $\tau \gg 1$. In other words, photons traveling through transparent clouds do not suffer a single absorption[1] event on average, while those penetrating opaque clouds incur multiple absorptions and re-emissions. Another familiar example is when one walks through fog or a snowstorm: the visibility only extends out to a distance corresponding to $\tau \sim 1$.

Nuclear astrophysics teaches us that the Sun has a nuclear reactor residing at its core with temperatures reaching $\sim 10^6$–10^7 K. Applying Wien's law informs us that this corresponds to radiation in the X-ray range of wavelengths, invisible to the naked eye. Yet, the rays of the Sun reach our eyes as visible light. How do we resolve this paradox? It turns out that the core of the Sun resides at $\tau \gg 1$. Instead, it is the surface corresponding to $\tau \sim 1$, known as the *photosphere*, that our eyes see, at least in visible light. It is the surface of last absorption or scattering before the photons stream across space into our telescopes.

Another observed and interesting fact about the Sun is that it has different sizes, depending on whether one observes it at X-ray, ultraviolet, visible or radio wavelengths. This illustrates another important concept regarding the optical depth—it is wavelength dependent.[2] What we call the "radius" of the Sun is its surface corresponding to $\tau \sim 1$. Since the location of this surface depends on wavelength, $\tau \sim 1$ at different wavelengths corresponds to different radii. Thus, the optical depth provides a useful vocabulary for describing why and how the size of an object changes across wavelength.

Now, consider the intense light of a star impinging upon a close-in exoplanet (e.g., a hot Jupiter). How far does the starlight penetrate? Again, it corresponds to the surface where the optical depth is on the order of unity (at a specific wavelength), because such a concept applies symmetrically to both emission from an object and absorption by it. When an exoplanet transits its host star, the measured *transit radius* always picks out the chord, cutting across the limb of its atmosphere, corresponding to a chord optical depth on the order of unity (Figure 2.1).

Finally, we note that the *magnitude* of the optical depth determines how the radiation behaves as it passes through the medium. When $\tau \ll 1$, radiation is in the *freely streaming limit* and propagates at the speed of light. When $\tau \gg 1$, it is absorbed and re-emitted multiple times within a short distance, causing its

[1] The same arguments also apply to scattering events.

[2] As another example, we note that styrofoam is opaque to visible light, but it is transparent in the microwave.

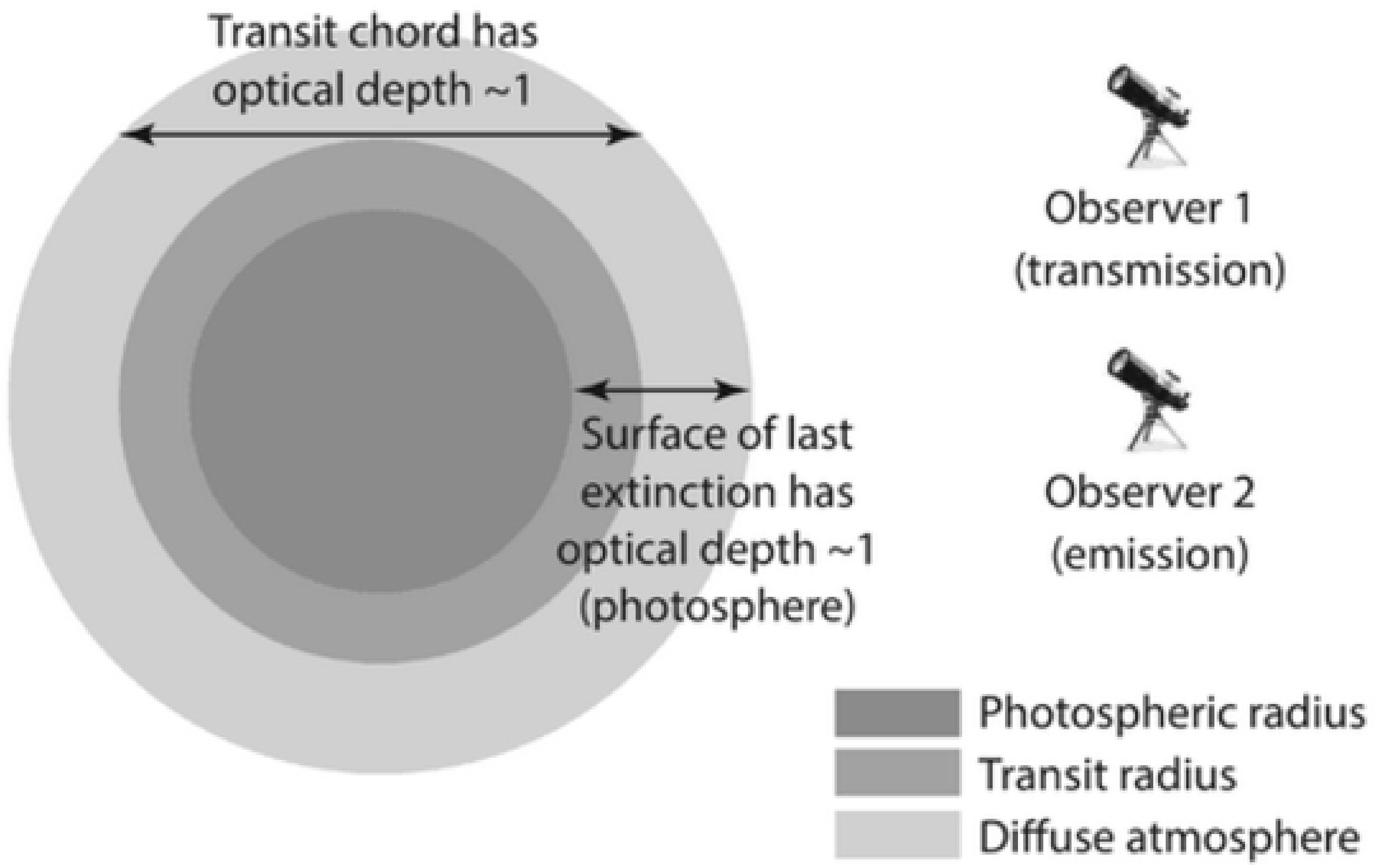

Figure 2.1: Schematic depicting the photospheric and transit radii. Observer 1 measures the transit radius of the exoplanet, which is the radius corresponding to a transit chord of optical depth ~ 1. Observer 2 measures photons emanating from the photosphere, which is the surface of last absorption or scattering. The optical depth between this surface and Observer 2 is ~ 1. The transit radius is generally larger than the photospheric radius [25], although not to the extent it has been exaggerated in this schematic.

overall passage through the medium to resemble diffusion. In Chapter 3, we will discuss one of the simplest implementations of radiative transfer known as the *two-stream approximation*, which is mostly designed to work when $\tau \lesssim 1$.

2.2 BASIC QUANTITIES IN RADIATIVE TRANSFER

2.2.1 Opacity, cross section and extinction coefficient

There are three ways of expressing the optical depth, because there is an equal number of ways to quantify how absorbent a medium is. We will use the term *extinction* to collectively describe both absorption and scattering.

The first measure of extinction is the cross section (σ), which we have already described. It may be visualized as the area of the "target" provided by an atom, molecule or particle—if a collidant passes within this target area, a collision will occur. The cross section per unit mass is the *opacity*, which we will generally denote by κ. A column of atmosphere has $\tilde{m}$ of mass per unit area—it is aptly termed the *column mass*. Thus, another definition of the optical depth, which

is more appropriate for atmospheric studies, is

$$\tau \equiv \int \kappa \ d\tilde{m}. \tag{2.3}$$

The relationship between the pressure and the column mass is essentially Newton's second law (cast in a per-unit-area form): $P = \tilde{m}g$, where g is the surface gravity of the exoplanet.

Yet another way of defining the optical depth comes from using the *extinction coefficient* ($\alpha_{\rm e}$).

$$\tau \equiv \int \alpha_{\rm e} \ dx. \tag{2.4}$$

Its reciprocal is the mean free path traveled by the photon ($l_{\rm mfp} = 1/\alpha_{\rm e} = 1/n\sigma$).

A source of confusion originates from the fact that some references refer to the cross section, opacity and extinction coefficient collectively as the "extinction coefficient,"[3] despite the fact that σ, κ and $\alpha_{\rm e}$ have physical units of cm^2, cm^2 g^{-1} and cm^{-1}, respectively. It is my hope that you will use these terms more judiciously.

2.2.2 The extinction efficiency

Consider a spherical particle with a radius of r. Its geometric cross section is πr^2. However, its actual cross section for extinction is generally larger than πr^2,

$$\sigma = Q_{\rm e} \pi r^2. \tag{2.5}$$

The factor $Q_{\rm e}$ is known as the *extinction efficiency* [191]. Its functional form and value depend on whether a particle is small or large. It may be separated into its absorption ($Q_{\rm a}$) and scattering ($Q_{\rm s}$) components: $Q_{\rm e} = Q_{\rm a} + Q_{\rm s}$.

The beauty about physics is that whether something is "small" or "large" is not a subjective judgment based on human emotion. For the interaction of radiation with matter, it is in comparison to the wavelength (λ). Specifically, a particle is small if $2\pi r/\lambda \ll 1$ and large if $2\pi r/\lambda \gg 1$. For example, a micron-sized particle is (radiatively) large when it interacts with radiation in the visible range of wavelengths, but it is small when the interaction is with photons in the far infrared.

To derive the exact functional form of $Q_{\rm e}$ requires a full-blown calculation of absorption and scattering by a given particle [214]. In practice, we, as atmospheric scientists, look up tables of $Q_{\rm e}(\lambda, r)$ and apply them to our models [52, 53, 136]. Roughly speaking, the extinction efficiency has a somewhat universal form,

$$Q_{\rm e} \sim \min\left\{ \left(\frac{2\pi r}{\lambda}\right)^4, 1 \right\}, \tag{2.6}$$

[3] For example, see Appendix 2 of the textbook by Goody & Yung [74].

with a transition from *Rayleigh scattering* ($2\pi r/\lambda \ll 1$) to scattering by large particles ($2\pi r/\lambda \gg 1$) occurring at a value of the normalized particle radius that depends on its composition. With the exception of resonant transitions that are specific to each material, the overall shape of the extinction efficiency is similar for different members of, for example, the same silicate family (e.g., the olivine group).

2.2.3 The single-scattering, geometric, spherical and Bond albedos

If we now distinguish between the absorption (σ_a) and scattering (σ_s) cross sections, then the *single-scattering albedo* is defined as

$$\omega_0 \equiv \frac{\sigma_s}{\sigma_s + \sigma_a}. \tag{2.7}$$

It quantifies the fraction of light reflected or scattered by a particle during a single scattering event. Another quantity needed to fully quantify scattering by a single particle is the *scattering asymmetry factor* (g_0), which describes how isotropic or anisotropic the scattering is; we will discuss g_0 in greater detail in Chapter 3. In practice, both $\omega_0(\lambda, r)$ and $g_0(\lambda, r)$ are tabulated quantities one obtains from detailed calculations [136]. If we have $g_0 = 0$, then the incident light is scattered equally in all directions (i.e., isotropically).[4] If we have $g_0 > 0$, then it is peaked in the forward direction; if it is $g_0 < 0$, then backscattering prevails. Even without engaging in first-principles calculations of g_0, it is interesting to note its asymptotic behavior [191],

$$g_0 \rightarrow \begin{cases} 0, & \frac{2\pi r}{\lambda} \ll 1, \\ 1, & \frac{2\pi r}{\lambda} \gg 1. \end{cases} \tag{2.8}$$

Evidently, small particles scatter isotropically, while large particles preferentially deflect radiation into the same direction from which it came. For example, a sub-micron-sized particle will scatter infrared radiation isotropically, but engage in the forward scattering of X-rays.

An atmosphere is composed of an enormous number of particles of different species, each with its own absorption and scattering properties. Collectively, they determine the fraction of incident light that is reflected out of the atmosphere at a given wavelength. As has been discussed by Seager [208], a historical and commonly used form of the albedo is the *geometric albedo* (A_g), which is the brightness of a planet or moon compared to an illuminated Lambertian disk[5] at a specific wavelength and full phase. This somewhat awkward definition means that it is possible to have $A_g > 1$. The geometric albedo is *not* the fraction of incident starlight scattered by an atmosphere.

[4] Strictly speaking, $g_0 = 0$ corresponds to a symmetric phase function. For example, Rayleigh scattering has $g_0 = 0$ but also exhibits a slight departure from isotropy. Conversely, the Henyey-Greenstein scattering phase function displays perfect isotropy when $g_0 = 0$.

[5] Isotropically scattering disk.

As the exoplanet orbits around its star, it shows different faces of its scattering surface to the astronomer, corresponding to different orbital *phases*. Integrating over all solid angles yields the *spherical albedo* [160, 204, 208, 237],

$$A_s = A_g \mathcal{P}_g, \tag{2.9}$$

with $\mathcal{P}_g$ being the *phase integral*. In other words, the geometric albedo is the spherical albedo at zero phase angle, which is also the moment of secondary eclipse or superior conjunction for an astronomer monitoring a transiting exoplanet. The relationship between the geometric and spherical albedos depends on the details of the scattering surface. For example, a Lambertian sphere (i.e., isotropic scattering) has $\mathcal{P}_g = 3/2$ [204], while Rayleigh scattering yields $\mathcal{P}_g = 4/3$ [208].

If the spherical albedo is in turn integrated over all wavelengths, we obtain the *Bond albedo* [160, 237],

$$A_B \equiv \frac{\int I_\star A_s \, d\lambda}{\int I_\star \, d\lambda}, \tag{2.10}$$

where $I_\star$ is the wavelength-dependent intensity of the star. Generally, $I_\star$ is not a Planck function. Essentially, the spherical albedo is integrated over all wavelengths, weighted by the stellar spectrum. The Bond albedo is a true measure of the incident energy that is scattered, and not absorbed, by the atmosphere.

In an astronomical context, the geometric albedo for a transiting exoplanet is directly obtained from measuring its secondary eclipse [28, 208],

$$A_g = \frac{F}{F_\star} \left(\frac{a}{R}\right)^2, \tag{2.11}$$

where $F/F_\star$ is the ratio of the flux of the exoplanet to its star *as observed at Earth*, a is the exoplanet-star separation and R is the radius of the exoplanet. In practice, this is done via photometry for a range of visible wavelengths, rather than at a single wavelength, which yields the wavelength-integrated geometric albedo. By assuming[6] the form of $\mathcal{P}_g$, one obtains the wavelength-integrated spherical albedo. If the range of wavelengths over which the secondary eclipse is measured encompasses the blackbody function of the star, then the wavelength-integrated spherical albedo is the Bond albedo. An important caveat is that the exoplanetary atmosphere is cool enough that its thermal emission is not contributing significantly in the same range of visible wavelengths [91]. This caveat may be violated for very hot exoplanets orbiting Sun-like stars or temperate exoplanets around cool red dwarfs.

In Chapter 4, we will see that there is a relationship between A_B, ω_0 and g_0 within the two-stream approximation of radiative transfer. Generally, this relationship is non-trivial to elucidate.

[6] At least, until astronomical techniques are advanced enough to measure the phase integral.

2.3 THE RADIATIVE TRANSFER EQUATION

We now derive the governing equation of radiation traveling through matter. Sophisticated derivations of the radiative transfer equation exist [74, 172], but I will opt for the shortest possible derivation. Namely, instead of working with length or spatial distance, we will directly work with the optical depth, since we are now convinced that it is the correct measure of transparency or opaqueness. Let the *intensity*[7] (energy per unit area, time, wavelength and solid angle subtended) of radiation be given by I. Its physical units are erg cm^{-3} s^{-1} sr^{-1}.

In general, the intensity impinges upon the surface of an object at an angle θ; we shall characterize this angle by its cosine, $\mu \equiv \cos\theta$. We choose the sign convention for μ to be such that reflection from this surface corresponds to $\mu > 0$, while penetration past the surface and into the object has $\mu < 0$. With this convection, the loss in intensity due to extinction is given by I/μ. This loss may be compensated by the fact that the object generally has a finite temperature and thus emits its own thermal emission ($-S/\mu$). The quantity S is aptly termed the *source function*. Thus, the change in intensity (ΔI), across the optical depth corresponding to the extent of the object ($\Delta\tau$), is

$$\Delta I = \frac{\Delta\tau}{\mu}\left(I - S\right). \tag{2.12}$$

If we take $\Delta\tau$ to be infinitesimally small, then we obtain the *radiative transfer equation*,

$$\mu\frac{\partial I}{\partial\tau} = I - S. \tag{2.13}$$

It looks deceivingly simple, because all of the complexity has been hidden within the source function. Besides thermal emission, S may include flux scattered from other directions into the intensity beam under consideration. It is often forgotten that the radiative transfer equation was first derived for stars by astrophysicists and later made its way into the atmospheric and climate sciences [191].

2.4 SIMPLE SOLUTIONS OF THE RADIATIVE TRANSFER EQUATION

2.4.1 Beer's law

The simplest solution of the radiative transfer equation comes from discarding the source function ($S = 0$), which yields

$$\int_{I_0}^{I} \frac{dI}{I} = \int_0^{\tau} \frac{d\tau}{\mu}. \tag{2.14}$$

[7] A potential source of confusion comes from the fact that atmospheric scientists sometimes prefer the terms *radiance* and *irradiance* when referring to the intensity and flux, respectively. I categorically avoid the use of these terms.

The integration is performed from the surface of the object, which we marked by $\tau = 0$, to some arbitrary depth or distance within it. The incident or initial intensity impinging upon the surface is given by I_0.

If we integrate the preceding expression, we obtain

$$I = I_0 e^{\tau/\mu}. \tag{2.15}$$

This is *Beer's law*, which tells us that the intensity passing through the object is being diminished exponentially. In this limit, $\tau = 1$ corresponds to one e-folding of the intensity, i.e., about 37% of the intensity beam is removed when an optical depth of unity is traversed. Note again that $\mu < 0$ in our convention.

In Chapter 4, we will derive a generalization of Beer's law that includes non-isotropic scattering. To a very good approximation, Beer's law describes how starlight is absorbed when it is incident upon an exoplanetary atmosphere. It is less appropriate for describing how the thermal emission of an exoplanet behaves, because in this case radiation has both incoming and outgoing components and the source function cannot be neglected.

2.4.2 Direct solution (pure absorption only)

Beyond Beer's law, the radiative transfer equation may be solved directly and analytically only in the limit of pure absorption. By "direct," we mean that one is solving for the intensity, rather than integrating the equation first over μ and solving for the angle-integrated quantities. To prove this point, we include coherent scattering in the radiative transfer equation and demonstrate that it is soluble only when scattering is absent. By "coherent," we mean that scattering does not alter the wavelength of a photon. The opposite limit is that of complete redistribution (or complete non-coherence), where the wavelength information is randomly distributed over some range [172]. Strictly speaking, spectral lines are poorly described by coherent scattering and better approximated by complete redistribution over the line profiles. However, if one is examining a wavelength interval or bin that is wide compared to the widths of individual lines, then coherent scattering is a decent approximation.

In the limit of coherent, isotropic scattering, the source function becomes [172],

$$S = \frac{\omega_0 J}{4\pi} + \left(1 - \omega_0\right) B, \tag{2.16}$$

where B is the blackbody or Planck function. The *total intensity* is

$$J \equiv \int_0^{4\pi} I \; d\Omega, \tag{2.17}$$

where $d\Omega = d\mu \; d\phi$ is the solid angle subtended and ϕ is the azimuthal angle. The limiting values of the source function appear to make sense: $S = B$ for pure absorption ($\omega_0 = 0$) and $S = J/4\pi$ (the mean intensity) for pure scattering ($\omega_0 = 1$).

With this choice of S, the radiative transfer equation becomes

$$\mu \frac{\partial I}{\partial \tau} = I - \frac{\omega_0 J}{4\pi} - (1-\omega_0)\, B. \tag{2.18}$$

Note that we have intentionally not labeled the intensity, source function and Planck function with a subscript, because our formulation applies regardless of whether these quantities are cast in per wavelength, per frequency or per wavenumber units. In practice, they may also apply to intervals or bins in wavelength, frequency or wavenumber. An argument for using frequency or wavenumber units is that spectral lines are more evenly distributed across them, rather than across wavelength.

Consider the transfer of radiation between two points labeled by 1 and 2. By integrating the equation between these points, we obtain

$$I_2 e^{-\tau_2/\mu} - I_1 e^{-\tau_1/\mu} = -\frac{1}{\mu} \int_{\tau_1}^{\tau_2} \left[\frac{\omega_0 J}{4\pi} + (1-\omega_0)\, B \right] e^{-\tau/\mu}\, d\tau. \tag{2.19}$$

Written in this form, it becomes apparent why one cannot proceed unless $\omega_0 = 0$, because one does not know the functional form of $J(\tau)$ a priori.

We now specialize to an atmosphere—we account for radiation traveling either upwards or downwards and subscript the quantities by $\uparrow$ and $\downarrow$, respectively.[8] We integrate equation (2.19) over all angles, while assuming that I_1, I_2 and B are constant. By writing the *fluxes* (energy per unit area, time and wavelength) as $F = \pi I$ (with the appropriate subscripts), we obtain

$$\begin{aligned} F_{\uparrow_1} &= F_{\uparrow_2} \mathcal{T} + \pi B \left(1 - \mathcal{T}\right), \\ F_{\downarrow_2} &= F_{\downarrow_1} \mathcal{T} + \pi B \left(1 - \mathcal{T}\right). \end{aligned} \tag{2.20}$$

The *transmission function* takes the form,

$$\mathcal{T} \equiv 2 \int_0^1 \mu e^{-\Delta\tau/\mu}\, d\mu, \tag{2.21}$$

where $\Delta\tau \equiv \tau_2 - \tau_1$ is the difference in optical depth between the two points and we demand $\tau_2 > \tau_1$ as a property of the coordinate system. The transmission function has a convenient range of values between 0 and 1. When $\mathcal{T} = 1$, no radiation is absorbed and blackbody emission is not produced. The incident flux simply passes right through the pair of layers—we have a completely transparent medium. When $\mathcal{T} = 0$, nothing passes through and the pair of layers radiates purely as a blackbody.

Writing the solutions in this form allows us to describe the transfer of radiation from Points 1 to 2 and vice versa, using either location as a starting point or boundary condition. In Chapter 3, we will see that formulating the problem as being radiative transfer between pairs of layers is the basis of the two-stream treatment.

[8] The directions of up and down correspond to $0 \le \theta \le \pi/2$ and $\pi/2 \le \theta \le \pi$, respectively.

2.5 A PRACTICAL CHECKLIST FOR RADIATIVE TRANSFER CALCULATIONS

Now that we have developed a basic intuition for radiative transfer and the various quantities associated with it, it is time to list all of the necessary ingredients required to perform a numerical calculation. They include:

- Specifying the appropriate form of the radiative transfer equation and any approximations taken, checking that the assumptions made are plausible and selecting the numerical method. In Chapter 3, we examine the *two-stream solutions* of the radiative transfer equation. Solutions of the radiative transfer equation are sometimes known as *forward models*, as they take a set of chemical abundances and opacities, compute forward, and translate them into a synthetic spectrum.

- Setting up a model atmosphere with a finite number of discrete layers, each with its own temperature, pressure and column mass, and calculating the fluxes passing through each layer. For example, equation (2.20) is a forward model in the limit of pure absorption and isothermal layers. To use it, we need the opacities as inputs, which allows us to compute the optical depths, transmission functions and fluxes associated with each layer. Knowledge of the fluxes allows the temperature of each layer to be calculated self-consistently.

- Deciding upon the chemical composition of the atmosphere being investigated (Chapter 7) and obtaining the opacities of the atoms, molecules and aerosols or condensates involved, typically from pre-computed tables (Chapter 5). Combining all of the opacities, weighted by the relative abundance of each species, across temperature, pressure and wavelength, yields the *opacity function* of the atmosphere.

- Numerically iterating to obtain a converged solution, since the opacities, optical depths and temperature are all interdependent quantities. The solution is judged to have converged once the temperature in each layer ceases to evolve, which in practice means that the change in temperature, between time steps, is less than a stated numerical threshold.

The converged solution yields the flux emerging from the top of the atmosphere, which is its synthetic spectrum. The temperatures in all of the layers collectively yield the temperature-pressure profile of the model atmosphere.

How do we relate the temperature and fluxes associated with each layer of the atmosphere? In the absence of atmospheric dynamics, one hopes to solve for *radiative equilibrium*, which occurs when the heat entering and exiting each layer vanishes. It is a statement of *local* energy conservation and originates from the first law of thermodynamics (Chapter 9),

$$\rho c_P \frac{DT}{Dt} = \frac{DP}{Dt} + \rho Q, \tag{2.22}$$

where ρ is the mass density, c_P is the specific heat at constant pressure and Q is the term associated with heating. If we focus just on radiative heating (and ignore convection and conduction), we have

$$\rho Q = -\nabla . \mathcal{F}_-, \tag{2.23}$$

where $\mathcal{F}_-$ is the wavelength-integrated net flux. The net flux of a layer is the *difference* between the flux entering and exiting it, which explains why I have labeled it with a minus sign. Geometrically, we can see that the preceding expression makes sense: when radiation emanates from the layer, the divergence of the net flux is positive, which leads to cooling ($Q < 0$). The opposite happens when radiation enters the layer. The minus sign is crucial in this respect—if it is missing, one gets the unphysical situation of runaway cooling or heating occurring.

If we ignore work done on the system ($\frac{DP}{Dt} = 0$) and atmospheric dynamics ($\frac{D}{Dt} = \frac{\partial}{\partial t}$), then we obtain the equation that is typically used to solve for radiative equilibrium in a one-dimensional model of an atmosphere,

$$\frac{\partial T}{\partial t} = -\frac{1}{\rho c_P} \frac{\partial \mathcal{F}_-}{\partial z}. \tag{2.24}$$

The computational iteration between the fluxes and temperatures, which depend on each other, is performed until the temperature stops changing in time,[9] at which point radiative equilibrium is formally obtained. Note that the spatial coordinate z increases with altitude.

Strictly speaking, if a calculation does not enforce radiative equilibrium, then we should not consider it to be a bona fide radiative transfer calculation. Radiative equilibrium is often confused with *global energy conservation*, which is the statement that the energy entering and exiting the atmosphere, as a whole, must equate. The latter is a necessary but insufficient condition for the former, a fact that is often unappreciated. We will cast this statement in precise mathematical terms in Chapter 4.

2.6 CLOUDS

The formation of aerosols or condensates in a gaseous environment is one of the outstanding puzzles of modern astrophysics and affects several of its branches, including that of exoplanetary atmospheres [162]. It is also a puzzle for climate scientists on Earth. In this textbook, I will use the terms *cloud, haze, aerosol* and *condensate* interchangeably, while being aware that there are subtle differences in their meaning related to their formation mechanisms. To compound matters, the usage of terminology differs between the various disciplines, so care must be

[9]In practice, one simply needs to get $\frac{\partial T}{\partial t}$ to reach a value that is below some numerical threshold (e.g., 10^{-6}). Obtaining a precise zero is impossible in such an iterative numerical calculation.

taken when reading the published literature. For example, hazes are typically associated with photochemical products,[10] while clouds follow the condensation curves set by thermodynamics. Forming clouds from first principles, out of the atmospheric gas, is a daunting theoretical challenge as it requires that we understand the following steps:

- Elucidating the details of *nucleation*, whereby the precursor seed particles form and onto which the further growth of the cloud particles may proceed [84]. A true theory should predict the shape, composition and size distribution of cloud particles formed.
- Calculating the gaseous and solid phases of the atmospheric chemistry self-consistently, meaning that as the gas gets depleted to form solids, its atomic and molecular abundances are adjusted accordingly.
- Treating the radiative transfer and opacities of the atmospheric gas and solid cloud particles self-consistently, including getting the details of absorption and scattering, associated with each component, correct. In other words, the opacities for both the gas and cloud are calculated from first principles (using quantum mechanics), across wavelength, temperature and pressure, rather than being described by ad hoc parameters.
- Performing this set of calculations self-consistently[11] against the dynamical background of the atmosphere. To be kept aloft, the aerosol needs to be dynamically supported, meaning that its terminal velocity, due to the action of gravity, is cancelled out by upwelling atmospheric flow. *Sedimentation* occurs when the condensed solids sink deeper into the atmosphere, out of reach of the photosphere.

It is difficult to *form* the cloud particles from first principles, but once they do it is hardly controversial to model their effects on the computed spectrum. For our purposes, the term *aerosol* is the most neutral one—it refers to any particle that may be described by its size, composition and index of refraction (or extinction efficiency).

To lowest order, the main effect of the aerosol is to introduce an extinction efficiency into the radiative transfer calculation. Crudely speaking, it may be approximated by this fitting formula [139],

$$Q_e = \frac{5}{Q_0 x^{-4} + x^{0.2}}, \tag{2.25}$$

where $x \equiv 2\pi r/\lambda$, r is the radius of the aerosol (which is assumed to be spherical) and λ is the wavelength. When $x \gg 1$, the extinction efficiency asymptotes to

[10]This is the accepted view in the planetary sciences, but the term "haze" in the Earth sciences is used as a measure of particle size.

[11]Rather than being parametrized by an *eddy mixing coefficient* (usually denoted by K_{zz}), which is essentially the use of a diffusion coefficient to crudely describe large-scale atmospheric circulation.

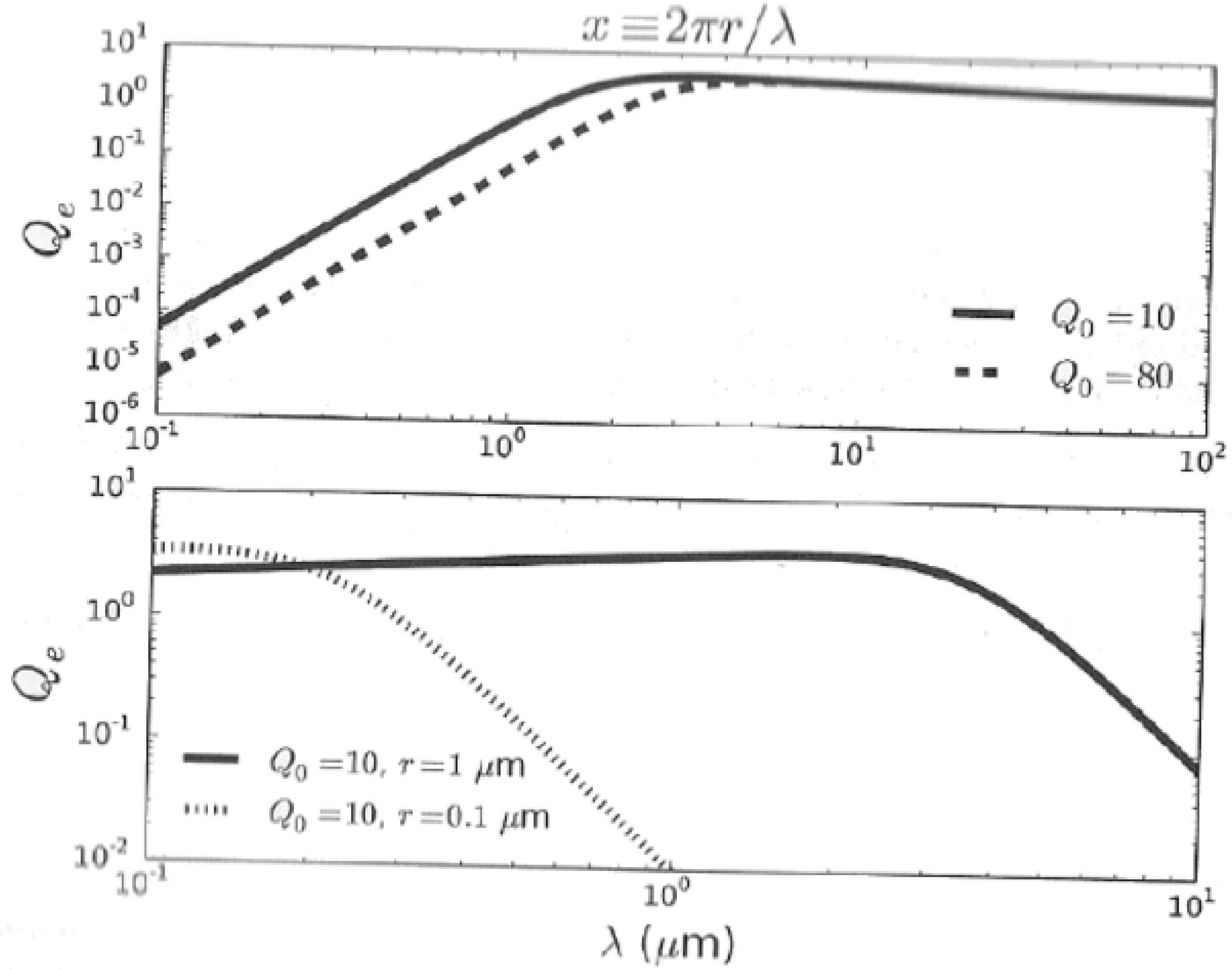

Figure 2.2: The extinction efficiency of spherical cloud particles versus the normalized radius (top panel) and wavelength (bottom panel). It is important to note that the curves in the top panel represent *all* particle radii. In other words, the extinction efficiency as a function of the normalized radius (x) is size-independent to lowest order. The different values of $Q_0 = 10$ and 80 mimic the effects of cloud composition, with the lower and higher values representing refractory and volatile species, respectively. The bottom panel shows the extinction efficiency associated with refractory cloud particles of two different radii: $r = 0.1$ and 1 μm.

a roughly constant value. When $x \ll 1$, it describes Rayleigh scattering by small particles. The dimensionless parameter Q_0 is a proxy for composition: refractory species (e.g., silicates) tend to have $Q_0 \approx 10$, while volatile ones have $Q_0 \approx 40$–80 (e.g., water).

Figure 2.2 shows the extinction efficiency associated with cloud particles of different sizes and compositions. The particle radius exerts a somewhat larger influence on the extinction efficiency than the composition. Generally, the influence of a cloud particle on the absorption and scattering of radiation is only felt at $\lambda \lesssim 2\pi r$. This insight generalizes straightforwardly to a cloud with a size distribution of particles.

To lowest order, these simple estimates allow us to glean some intuition for what clouds would do to the spectrum of an exoplanetary atmosphere. When the optical depth approaches unity, spectral features and lines become muted and "less sharp" for $\lambda \lesssim 2\pi r$ [23]. However, this effect may be negated by increasing the abundances of the atomic and molecular species responsible for these features and lines. In this way, one may understand how clouds introduce a degeneracy to the interpretation of chemical abundances in an atmosphere. To the next order, cloud particles are described by a single-scattering albedo and scattering asymmetry factor. A full-fledged radiative transfer calculation would take the wavelength- and layer-dependent nature of these quantities into account when implementing, for example, the two-stream solutions. More accurate calculations would employ *multi-stream* treatments [35].

2.7 ATMOSPHERIC RETRIEVAL

Traditionally, when one builds a model of an atmosphere, one starts with a governing equation—either of fluid dynamics, radiation, chemistry or some combination of them—and a set of assumptions and computes. Ideally, the calculation predicts some measurable quantity that may be confronted by the astronomical observations. A judgment may then be made on whether the model has successfully represented Nature in some way, but only against the preconceived notions we have about an atmosphere.

Atmospheric retrieval tries to shed itself of these preconceived notions by providing a framework in which one may invert a set of observables to obtain the properties of an atmosphere, such as its thermal structure or chemical abundances [18, 138, 139, 143, 144, 151, 152]. It has its origin in the atmospheric sciences, where inversion techniques were used to analyze remote sensing observations of Earth [199]. It has also been used in the planetary sciences to analyze the atmospheres of the Solar System bodies [109]. Typically, there is a combination of satellite and in-situ data to provide some ground truth to anchor the retrieval. As my colleagues in these fields of inquiry like to say, "We can simply let the data do the talking."

This faith in the data alone becomes less viable in the study of exoplanetary atmospheres, because we are essentially dealing with the remote sensing of unresolved point sources and in-situ measurements are out of the question. Furthermore, the data are often sparse and fragmented in the sense that one only sees incomplete facets of the exoplanetary atmosphere (e.g., its dayside but not its nightside, insufficient wavelength coverage) that do not allow for unique interpretations. Letting our interpretations be guided entirely by the data—for example, without enforcing radiative or chemical equilibrium—becomes foolhardy when the number of model parameters exceeds the number of data points. Our interpretation needs to be anchored by the laws of physics and chemistry to some degree. This is an ongoing debate.

In this section, we will discuss the basic ingredients and computational procedure that go into constructing and implementing an atmospheric retrieval model. We will then discuss the pitfalls involved in using these inversion techniques.

2.7.1 Basic ingredients for constructing a retrieval model

Earlier in the chapter, we already summarized the basic ingredients needed for radiative transfer. Here, we augment that list for atmospheric retrieval.

- The first ingredient one needs is a way of describing the temperature-pressure profile of the atmosphere using a set of parameters. One may either use an ad hoc fitting function [109, 151] or a simplified solution of the radiative transfer equation [18, 143, 144]. Chapter 4 discusses the latter.

- Next, one needs to specify the *identities* of the atoms and molecules being included in the retrieval (e.g., water, methane, carbon monoxide). The relative abundances of these atoms and molecules may be treated as being free parameters of the retrieval model. Alternatively, one may wish to enforce chemical equilibrium, where the molecular abundances are determined by the elemental abundances. When chemical equilibrium is enforced, the only free parameters present are the elemental abundances. Chapter 7 discusses simple models for computing the relative abundances of molecules in chemical equilibrium.

- For a fixed combination of parameter values for the temperature-pressure profile and chemical abundances, one may use a forward model to convert a set of opacities into a synthetic spectrum. In retrieval models, this conversion is performed once for each set of parameter values. There is typically no attempt to solve for radiative equilibrium, as the numerical iteration previously described is not performed.

Millions, if not billions, of combinations of these parameters are considered. Each synthetic spectrum is then compared against the observed one and a goodness of fit is computed for each model. In this manner, a range of best-fit models are found given the uncertainties on the data and the posterior distributions of the molecular abundances may be computed. In practice, a minimization algorithm is required to efficiently and thoroughly scan this large, multi-dimensional parameter space.

It is beyond the scope of the current edition of this textbook to cover the minimization algorithms used in atmospheric retrieval, but we will briefly mention them. The *optimal estimation method* exploits the property that if the prior distribution of an input quantity takes the form of a Gaussian function, then the expressions used for minimization are simplified [199]. If one desires a uniform prior, then the width of the Gaussian is taken to be very large to mimic this property [138]. *Monte Carlo methods* allow one to consider more general functional forms for the priors and have been implemented in various flavors [18, 143, 144, 151, 152].

2.7.2 Pitfalls

A retrieval model is only as good as the assumptions and techniques used. Here, we list some of the possible pitfalls.

2.7.2.1 Degeneracies

The spectral lines associated with different molecules inevitably overlap across wavelength, which naturally leads to degeneracies between their relative abundances as multiple combinations of relative abundances may lead to the same net absorption at a given wavelength. Aerosols or clouds introduce another form of degeneracy, as they act to diminish the strength of spectral lines, an effect that may be compensated by increasing the abundances of the molecules associated with these lines. Any good retrieval model will quantify these degeneracies in a formal and systematic way.

2.7.2.2 Being misled by your priors

The posterior distributions of the properties of the atmosphere inferred from the retrieval calculation are dependent on the *priors* assumed. As an example, the carbon-to-oxygen ratio (C/O) may be estimated using,

$$\mathrm{C/O} \approx \frac{\tilde{n}_{\mathrm{CO}} + \tilde{n}_{\mathrm{CH_4}} + \tilde{n}_{\mathrm{CO_2}}}{\tilde{n}_{\mathrm{CO}} + \tilde{n}_{\mathrm{H_2O}} + 2\tilde{n}_{\mathrm{CO_2}}}, \tag{2.26}$$

where $\tilde{n}$ generally denotes the *mixing ratio* of a molecule, which is its abundance normalized by that of molecular hydrogen (for a hydrogen-dominated atmosphere); self-explanatory subscripts indicate the specific molecule being referred to. If carbon monoxide (CO) is the dominant molecule, then we have $\mathrm{C/O} \approx 1$. If carbon dioxide (CO_2) is dominant, then we have $\mathrm{C/O} \approx 0.5$. If methane (CH_4) or water (H_2O) are dominant, then we have $\mathrm{C/O} \gg 1$ and $\mathrm{C/O} \ll 1$, respectively. In other words, even if we assume the prior distributions of these molecules to be uniform, we produce two artificial peaks in the prior distribution of C/O [143].

2.7.2.3 Local versus global energy conservation

In the ensemble of models being computed by a retrieval calculation, it is natural to discard those where the emergent thermal emission from the exoplanetary atmosphere exceeds the energy input from the star (assuming that the exoplanet produces negligible internal heat) [151]. This is simply a statement of global energy conservation—you cannot produce more than what you put in. What is less obvious is *local* energy conservation, which is radiative equilibrium—the net heating or cooling of each layer of the model atmosphere must be the same across layers (see Chapter 4) in the absence of atmospheric circulation.

2.7.2.4 Not everything is chemically possible

When an atmosphere is in chemical equilibrium, not every combination of the relative abundances is possible. As each layer of the model atmosphere has its own values of temperature and pressure, the mixing ratios of molecules take on specific values given the elemental abundances. One is not free to specify the mixing ratios as free parameters. When atmospheric dynamics is present and the dynamical mixing time scale is less than the chemical time scale, it is possible for the mixing ratios to depart from chemical equilibrium given the local values of temperature and pressure—but again, not every combination of mixing ratios is possible. If we invoke the quenching approximation (Chapter 6), then the mixing ratios originate from a deeper part of the atmosphere that is in chemical equilibrium given its local values of temperature and pressure. As another example, consider carbon dioxide in hot ($\gtrsim$ 1000 K), hydrogen-dominated atmospheres, which is never more abundant than carbon monoxide or water unless the metallicity is implausibly high [100].

2.7.2.5 Opacities: The devil is in the details

The task of calculating atmospheric opacities is a subtle and detail-oriented business. Across the ultraviolet to infrared range of wavelengths, one has to compute the shapes of millions to billions of lines—and this is for one molecule at a specific pairing of temperature and pressure. *Line-by-line calculations* require that one sample *all* of these lines with no approximation taken, which is a formidable task when the atmospheric temperatures attain $\sim$ 1000 K. Binning techniques such as the *k-distribution method* may be employed to speed up the calculation, but this comes at the cost of invoking a set of approximations (see Chapter 5). Hidden in all of these details is the issue of the pressure broadening of the far wings of spectral lines, which remains an unsolved physics problem. Furthermore, photometric observations of exoplanetary atmospheres record radiation integrated over a broad bandpass, implying that many combinations of molecules could produce the same opacity across this wavelength range. When you read published papers that omit these details, you should exercise one of the most precious instruments in your toolkit as a scientist: skepticism.

2.7.2.6 Numerical convergence

Running a retrieval model is a technically demanding task, as one needs to discretize space (via the temperature-pressure profile) and wavelength (the opacity function) and also consider a set of atoms and molecules. To complete the calculation within a reasonable amount of time, one often has to use a coarse numerical resolution across one of these axes. To determine if the resolution is sufficient requires a set of resolution tests to be performed—it is a thankless but necessary step. And as a rule of thumb, one should always adopt the *second-highest* resolution if the highest one attains convergence.[12]

2.8 PROBLEM SETS

2.8.1 The Sun

(a) The photosphere of the Sun has a temperature of about 5800 K. Using Wien's law, what is the peak wavelength of its blackbody function? What color of light does this peak correspond to?
(b) The *chromosphere* resides *above* the photosphere and has temperatures $\sim 10^4$ K. Why do we generally not see the chromosphere in visible light? At what wavelengths do we detect the chromosphere?

2.8.2 The random walk of a photon

Photons created at the center of the Sun must travel a distance of $R_\odot$ to escape from it, where $R_\odot \approx 700,000$ km is the solar radius. Each photon travels a distance of about the mean free path ($l_{\rm mfp}$) before being absorbed and re-emitted (or scattered).
(a) If the Sun were completely transparent, estimate the time it would take a photon to escape.
(b) Imagine if the Sun were not too opaque such that the center-to-surface optical depth (τ) were between 1 and 10. Convince yourself that the number of absorption/scattering events would be about $R_\odot/l_{\rm mfp}$. Show that the number of events would be $N \sim \tau$.
(c) Imagine if the Sun were very opaque such that the center-to-surface optical depth greatly exceeded unity. On average, the displacement of the photon would be zero. But the mean of the square of the displacement would be $N l_{\rm mfp}^2$. Thus, show that $N \sim \tau^2$.
(d) Consider the case of the opaque Sun. Taking the number density to be $n \sim 10^{24}$ cm^{-3} (pure hydrogen) and the cross section to be $\sigma \sim 10^{-24}$ cm^2 (Thomson scattering), estimate the value of $l_{\rm mfp}$. Estimate the time taken for a photon to leak out of the Sun from its center.
(e) What is a key weakness of this analysis? (Hint: it is *not* that the numbers are imprecise.)

2.8.3 The greenhouse effect on Earth

A naive estimate of the temperature on Earth would lead us to the conclusion that our atmosphere is warmer than expected. This greenhouse warming is caused by the variation in optical depth, across wavelength, of the Earth's atmosphere.

[12] A piece of advice I attribute to Scott Tremaine.

(a) On average, the Earth orbits the Sun at a distance of a. Let the effective temperature of the Sun be $T_\star$ and its radius be $R_\star$. Let the Bond albedo of the Earth's atmosphere be $A_{\rm B}$. If we take $A_{\rm B} = 0.3$, estimate the equilibrium temperature,

$$T_{\rm eq} = T_\star \left(\frac{R_\star}{2a}\right)^{1/2} (1 - A_{\rm B})^{1/4}, \tag{2.27}$$

at Earth. Convince yourself that the equilibrium temperature corresponds to the incident solar flux averaged over the entire surface of the Earth. Is $T_{\rm eq}$ less or greater than the freezing point of water?
(b) What is the optical depth of the Earth's atmosphere to sunlight?
(c) Sunlight absorbed by the Earth is reprocessed into infrared emission (such that the second law of thermodynamics is obeyed). What is the optical depth of Earth's atmosphere to the infrared emission?

2.8.4 The various forms of the Planck function

A common pitfall for novices of radiative transfer is the failure to properly elucidate the various forms of the Planck function. When written in per wavelength units, it is

$$B_\lambda = \frac{2hc^2}{\lambda^5} \left(e^{hc/\lambda k_{\rm B} T} - 1\right)^{-1}. \tag{2.28}$$

where h is the Planck constant, c is the speed of light, $k_{\rm B}$ is the Boltzmann constant and T is the temperature. (Note that we usually reserve ν for the kinematic viscosity, but in this problem we will use it to denote the frequency.)
(a) Let the Planck function in per frequency units be denoted by B_ν. By enforcing energy conservation,

$$\int B_\nu \, d\nu = \int B_\lambda \, d\lambda, \tag{2.29}$$

derive the functional form of B_ν.
(b) Next, let the Planck function in per wavenumber ($\tilde{\nu} \equiv 1/\lambda$) units be denoted by $B_{\tilde{\nu}}$ and derive its expression as well.
(c) Show that

$$\pi \int_0^\infty B_\lambda \, d\lambda = \pi \int_0^\infty B_\nu \, d\nu = \pi \int_0^\infty B_{\tilde{\nu}} \, d\tilde{\nu} = \sigma_{\rm SB} T^4. \tag{2.30}$$

What is the expression for $\sigma_{\rm SB}$? (Hint: you may require the use of Riemann zeta functions.) Hence, the Stefan-Boltzmann constant is not a fundamental constant, but is composed of other physical constants.

2.8.5 Direct solutions of the radiative transfer equation

(a) For the treatment of isothermal atmospheric layers discussed in Section 2.4.2, show that

$$\mathcal{T} = (1 - \Delta\tau) e^{-\Delta\tau} + (\Delta\tau)^2 \mathcal{E}_1, \tag{2.31}$$

where $\mathcal{E}_1(\Delta\tau)$ is the exponential integral of the first order, defined as [2, 7]

$$\mathcal{E}_i(\Delta\tau) \equiv \int_1^\infty x^{-i} e^{-x\Delta\tau} \, dx. \tag{2.32}$$

(b) Generally, the Planck function varies across optical depth. Even in this non-isothermal situation, it is possible to obtain a direct solution of the radiative transfer equation. Consider the following functional form for B,

$$B = \begin{cases} B_1 + B'(\tau - \tau_1), & \downarrow \text{ direction}, \\ B_2 + B'(\tau - \tau_2), & \uparrow \text{ direction}, \end{cases} \tag{2.33}$$

where

$$B' \equiv \frac{\partial B}{\partial \tau} = \frac{B_2 - B_1}{(\tau_2 - \tau_1)} \tag{2.34}$$

is the gradient of the Planck function across the medium bounded by Points 1 and 2. Quantities subscripted by 1 and 2 are evaluated at their respective locations. Mathematically, this functional form is the Taylor series expansion of B, about the point τ_1 or τ_2, truncated at the linear term. Show that this choice of B ensures that $B = B_1$ and $B = B_2$ when $\tau = \tau_1$ and $\tau = \tau_2$, respectively. Furthermore, show that the solutions of the radiative transfer are

$$\begin{aligned} F_{\uparrow_1} &= F_{\uparrow_2}\mathcal{T} + \pi B_2 (1 - \mathcal{T}) + \pi B' \left[\frac{2}{3}\left(1 - e^{-\Delta\tau}\right) - \Delta\tau\left(1 - \frac{\mathcal{T}}{3}\right)\right], \\ F_{\downarrow_2} &= F_{\downarrow_1}\mathcal{T} + \pi B_1 (1 - \mathcal{T}) - \pi B' \left[\frac{2}{3}\left(1 - e^{-\Delta\tau}\right) - \Delta\tau\left(1 - \frac{\mathcal{T}}{3}\right)\right], \end{aligned} \tag{2.35}$$

where we again have $\Delta\tau \equiv \tau_2 - \tau_1$.

2.8.6 The transit radius and the optical depth

Consider an exoplanet, with an atmosphere, transiting its star. Starlight cuts across the limb of the atmosphere along a chord. Let the spatial coordinate along this chord be x. Let the transit radius be R.
(a) Let x and R form the sides of a right-angled triangle with a hypotenuse given by $R + z$, where z is the vertical coordinate. By asserting that $z \ll R$, show that

$$z \approx \frac{x^2}{2R}. \tag{2.36}$$

(b) Assume an isothermal and hydrostatic atmosphere. Derive the number density (n) as a function of z and a reference value of the number density ($n_{\rm ref}$) corresponding to $z = 0$. By integrating along the transit chord, show that the chord optical depth is [62]

$$\tau = n_{\rm ref} \sigma \sqrt{2\pi H R}, \tag{2.37}$$

where H is the (isothermal) pressure scale height.
(c) Obtain an alternative and approximate derivation of the optical depth by asserting that $\tau = n_{\rm ref}\sigma l$ and the length scale l must be the mean of H and R. (Hint: think about the *type* of averaging involved.)
(d) Now, consider a non-isothermal atmosphere with the temperature being expressed as a Taylor series expansion truncated at the linear term [97],

$$T = T_{\rm ref} + \frac{\partial T}{\partial z} z, \tag{2.38}$$

where $T_{\rm ref}$ is a reference value of the temperature. Derive the number density profile, $n(z)$. What is the expression for the non-isothermal scale height? Finally, derive the expression for the chord optical depth.

Chapter Three

The Two-Stream Approximation of Radiative Transfer

3.1 WHAT IS THE TWO-STREAM APPROXIMATION?

The passage of radiation through an atmosphere generally occurs in many directions—photons are absorbed (and re-emitted) or scattered multiple times before they escape it. These interactions with matter may also alter their wavelengths. Thus, we are faced with a formidable problem spanning three dimensions and across a broad range of wavelengths or frequencies. To make any progress in understanding, we need to reduce the complexity of the problem down to a level suitable for pedagogy.

In Chapter 10, we will discuss the utility of the shallow water models for understanding exoplanetary atmospheres and point out that "shallow" has a subtle meaning in the context of fluid dynamics. Within the context of radiative transfer, it is more straightforward—atmospheres with extents much less than the radius of the exoplanet are considered shallow. In this case, the horizontal propagation of radiation may be neglected and it appears as if radiative transfer occurs only in two directions: upwards and downwards. This is the physical basis of the *two-stream approximation*. It is an acceptable approximation for Earth, Mars, Titan and Venus. It is decent even for calculations of hot Jupiters with model atmospheres that are still considerably thinner than their radii, because the extent of the model domain is set by the depth of penetration of stellar heating.

The mathematical basis of the two-stream approximation is a little subtler. The radiative transfer equation has two independent variables: optical depth and propagation angle. Rather than solve this partial differential equation, it is easier to first integrate over the outgoing (upwards) and incoming (downwards) hemispheres separately, before solving the resulting pair of ordinary differential equations that depend only on the optical depth. The problem with this approach is that one always ends up with one more dependent variable than the number of equations available, which renders the system under-determined. To close the system of equations requires that we make assumptions about the ratios of these dependent variables, which is the origin of the *Eddington coefficients*. The freedom to choose different values for a set of Eddington coefficients leads to an ambiguity in the types of *closures* used. The choice of closure is in turn related to fundamental issues of energy conservation.

It is worth mentioning that it is possible to perform separate sets of two-stream calculations for incident starlight and thermal emission from the exoplan-