Principles of Planetary Climate

R. T. Pierrehumbert

March 9, 2008

Preface

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

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

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

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

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

ii

Contents

Preface			i	
Co	onten	ts		1
1	The	$\operatorname{Big}\mathbf{Q}$	uestions	3
	1.1	Overvie	ew	3
	1.2	Close t	o home	3
	1.3 Into deepest time: Faint Young Sun and habitability of the Earth		epest time: Faint Young Sun and habitability of the Earth	8
	1.4	Goldilo	cks in space: Earth, Mars and Venus	15
	1.5	Other S	Solar System planets and satellites	20
	1.6	Farther	afield: Extrasolar planets	22
	1.7	Digress	ion: About climate proxies	27
		1.7.1	Overview of proxy data	27
		1.7.2	Isotopic proxies	29
		1.7.3	Hydrogen and Oxygen isotopes in sea water and marine sediments	34
		1.7.4	Forams to the rescue	36
	1.8	The Pr	oterozoic climate revisited:Snowball Earth	38
	1.9	The ho	thouse/icehouse dichotomy	45
		1.9.1	The past 70 million years	45
		1.9.2	Hothouse and icehouse climates over the Phanerozoic	50
	1.10	Pleisto	cene Glacial-Interglacial cycles	53
		1.10.1	The Marine Sediment Record	53
		1.10.2	Ice core records	55
	1.11	Holocer	ne climate variation	57
	1.12	Back to	b home: Global Warming	59
	1.13	The fat	e of the Earth, the lifetime of biospheres	67

	1.14	For Further Reading	68	
2	The	rmodynamics in a Nutshell	71	
	2.1	A few observations	71	
	2.2	Dry thermodynamics of an ideal gas	75	
		2.2.1 The equation of state for an ideal gas	75	
		2.2.2 Specific heat and conservation of energy	78	
		2.2.3 Entropy, reversibility and Potential temperature; The Second Law	79	
	2.3	Static stability of inhomogeneous mixtures	82	
	2.4	The hydrostatic relation	83	
	2.5	Thermodynamics of phase change	85	
	2.6	The moist adiabat	90	
	2.7	For Further Reading	95	
2	Flor	nontany models of radiation balance	07	
J	2 1	Energy halance and temperature	97	
	3.1 3.2	Blackbody radiation	91	
	0.2 2.2	Blackbody radiation		
	3.5 3.4		116	
	0.1	3.4.1 Faint Young Sun Snowball Earth and Hysteresis	191	
		3.4.2 Climate sensitivity, radiative forcing and feedback	121	
	35	Partially absorbing atmospheres	121	
	3.0 3.6	Partially absorbing atmospheres		
	3.0 3.7	For Further Boading	133	
	5.7	For Further Reading	197	
4	Rad	liative transfer in temperature-stratified atmospheres	139	
	4.1	Overview	139	
	4.2	Basic Formulation of Plane Parallel Radiative Transfer	140	
		4.2.1 Optical thickness and the Schwarzschild equations	140	
		4.2.2 Some special solutions of the Two-Stream equations	145	
	4.3	The Grey Gas Model	150	
		4.3.1 OLR and back-radiation for an optically thin grey atmosphere \ldots .	151	
		4.3.2 Radiative properties of an all-troposphere dry atmosphere	152	
		4.3.3 A first look at the runaway greenhouse	156	
		4.3.4 Pure radiative equilibrium for a grey gas atmosphere	160	

		4.3.5 Effect of atmospheric solar absorption on pure radiative equilibrium	163
	4.4	Real gas radiation: Basic principles	167
		4.4.1 Overview: OLR through thick and thin	167
		4.4.2 The absorption spectrum of real gases	170
		4.4.3 I walk the line	177
		4.4.4 Behavior of the band-averaged transmission function	182
		4.4.5 Dealing with multiple greenhouse gases	188
		4.4.6 A homebrew radiation model	191
		4.4.7 Spectroscopic properties of selected greenhouse gases	193
		4.4.8 Collisional continuum absorption	204
		4.4.9 Condensed substances: Clouds	209
	4.5	Real gas OLR for all-troposphere atmospheres $\ldots \ldots \ldots$	211
		4.5.1 CO_2 and dry air $\ldots \ldots \ldots$	212
		4.5.2 Pure CO_2 atmospheres: Present and Early Mars, and Venus $\ldots \ldots \ldots \ldots \ldots$	214
		4.5.3 Water vapor feedback	217
		4.5.4 Greenhouse effect of CO_2 vs CH_4	225
	4.6	section:OLR.CO2vsCH4	225
	4.7	Another look at the runaway greenhouse	228
	4.8	Pure radiative equilibrium for real gas atmospheres	230
	4.9	Tropopause height for real gas atmospheres	240
	4.10	The lesson learned	243
	4.11	For Further Reading	243
5	Scat	ttering 2	247
0	5.1	Overview	247
	5.2	Basic concepts	248
	5.3	3 Scattering by molecules: Rayleigh scattering	
	5.4	Scattering by particles	263
	5.5	The two-stream equations with scattering	268
	5.6	Some basic solutions	270
	5.7	Numerical solution of the two-stream equations	277
	5.8	Water and ice clouds	277
	5.9	Things that go bump in the night: Infrared-scattering with gaseous absorption	282
	5.10	Effects of atmospheric solar absorption	282

CONTENTS	
CONTENIS	

	5.11	For Further Reading	282
6	The	e Surface Energy Balance	283
	6.1	Formulation of the surface flux problem	283
	6.2	Radiative exchange	285
		6.2.1 Shortwave radiation	285
		6.2.2 The behavior of the longwave back-radiation $\ldots \ldots \ldots \ldots \ldots \ldots$	285
		6.2.3 Radiatively driven ground-air temperature difference	288
	6.3	Basic models of turbulent exchange	291
		6.3.1 Sensible heat flux	292
		6.3.2 Latent heat flux	293
	6.4	Joint effect of the fluxes on surface conditions	297
	6.5	Monin-Obukhov theory	299
	6.6	Mass balance and melting	299
	6.7	Precipitation-temperature relations	299
	6.8	Simple models of sea ice in equilibrium	300
	6.9	For Further Reading	300
	Radiative-convective models		
7	Rad	liative-convective models	301
7	Rad 7.1	liative-convective models Dry grey-gas atmospheres	301 301
7	Rad 7.1 7.2	liative-convective models Dry grey-gas atmospheres	301 301 301
7	Rad 7.1 7.2 7.3	liative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres	 301 301 301 302
7	Rad 7.1 7.2 7.3 7.4	liative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes	 301 301 301 302 302
7	Rad 7.1 7.2 7.3 7.4 7.5	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse	 301 301 301 302 302 302
7	Rad 7.1 7.2 7.3 7.4 7.5 7.6	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine	 301 301 302 302 302 302 302
7	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height	 301 301 302 302 302 302 303
7	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited	 301 301 302 302 302 302 303 303 303
7	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited Mars, present, past and future	 301 301 302 302 302 302 303 303 303
7	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 7.10	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited Mars, present, past and future Titan	 301 301 302 302 302 303 303 303 303 303
7	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 7.10 7.11	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited Mars, present, past and future Gas Giants	 301 301 302 302 302 303 303 303 304
8	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 7.10 7.11 Var	Hiative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited Mars, present, past and future Titan Gas Giants iation of temperature with season and latitude	 301 301 302 302 302 303 303 303 304
8	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 7.10 7.11 Var 8.1	liative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited Mars, present, past and future Titan Gas Giants iation of temperature with season and latitude A few observations of the Earth	 301 301 302 302 302 303 303 303 304 305
8	Rad 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 7.10 7.11 Var 8.1 8.2	liative-convective models Dry grey-gas atmospheres Surface vs. Top of Atmosphere Budgets: Who controls the surface temperature? Real gas atmospheres Sensitivity of climate to CO_2 changes Methane-dominated greenhouse The atmosphere as a heat engine Effect of the diurnal cycle on tropopause height The runaway greenhouse revisited Mars, present, past and future Titan Gas Giants Station of temperature with season and latitude A few observations of the Earth Distribution of incident solar radiation	 301 301 302 302 302 303 303 303 304 305 306

		8.3.1 Thermal inertia for a mixed-layer ocean	316
		8.3.2 Thermal inertia of a solid surface	320
		8.3.3 Summary of thermal inertia effects	326
	8.4	Some elementary orbital mechanics	326
	8.5	Effect of long term variation of orbital parameters	331
		8.5.1 Milankovic cycles on Earth	332
		8.5.2 Milankovic cycles on Mars	337
	8.6	A palette of planetary seasonal cycles	339
		8.6.1 Airless planets and moons	339
		8.6.2 Venus	339
		8.6.3 Gas Giants	339
		8.6.4 Mars, present and past	339
		8.6.5 Snowball Earth	340
		8.6.6 Hothouse Earth	340
		8.6.7 Earth without a moon	340
		8.6.8 Titan	340
	8.7	For Further Reading	340
0	Fuel	lution of the atmosphere	2/1
9	0.1	The CO methoding the master	941 941
	9.1	The CO_2 weathering thermostat \ldots	041 941
	9.2	Escape of the streambars to more	041 949
	9.5	Other tenies in chamical matrices	042 040
	9.4	Examples in chemical evolution	042 040
	9.5	For Further Reading	342
10	Con	sequences of heat transport 3	343
	10.1	Mechanisms of heat transport	343
	10.2	Formulation of energy balance models	343
	10.3	Equilibrium energy balance models	344
	10.4	WTG models of the tropics	344
	40 5	The seasonal evel revisited	344
	10.5		011
	10.5 10.6	Ice albedo feedback	344

CONTENTS

Chapter 1

The Big Questions

1.1 Overview

This chapter will survey a few of the major questions raised by observed features of present and past Earth and planetary climate. Some of these questions have been answered to one extent or another, but many remain largely unresolved. This will not be a comprehensive synopsis of Earth and planetary climate evolution; we will be content to point out a few striking facts about climate that demand a physical explanation. Then, in subsequent chapters, we'll develop the physics necessary to think about these problems. Although we hope not to be too Earth-centric in this book, in the present chapter we will perforce talk at greater length about Earth's climate than about those of other planets, because so much more is known about Earth's past climate than is known about the past climates of other planets. A careful study of Earth history suggests generalities that may apply to other planets, and also raises interesting questions about how things might have happened differently elsewhere, and it is with this goal in mind that we begin our journey.

1.2 Close to home

When the young Carl Linneaus set off on his journey of botanical discovery to Lappland in 1732, he left on foot from his home in Uppsala. He didn't wait until he reached his destination to start making observations, but found interesting things to think about all along the way, even in the plant life at his doorstep. So it is with climate as well.

To the discerning and sufficiently curious observer, a glance out the window, a walk through the woods or town, a short sail on the ocean, all raise profound questions about the physics of climate. Even without a thermometer, we have a perception of "heat" or "temperature" by examining the physical and chemical transitions of the matter around us. In the summertime, ice cream will melt when left out in the sun, but steel cooking pots don't. Trees and grass do not spontaneously burst into flame every afternoon, and a glass of water left outdoors in the summer does not boil. Away from the tropical regions, it often gets cold enough for water to freeze in the wintertime, but hardly ever cold enough for alcohol to freeze. What is it that heats the Earth? Is it really the Sun, as seems intuitive from the perception of warmth on a sunny day? In that case, what keeps the Earth from just accumulating more and more energy from the Sun each day, heating up until it melts? For that matter, why don't temperatures plummet to frigid wintery values every night when the Sun goes down? Similarly, what limits how cold it gets during the winter?

With the aid of a thermometer, such questions can be expressed quantitatively. The first, and still most familiar, kinds of thermometers were based on on particular reproducible and measurable effect of temperature on matter – the expansion of matter as it heats up. Because living things are composed largely of liquid water, the states of water provide a natural reference on which to build a temperature scale. The *Celsius* temperature scale divides the range of temperature between the freezing point of pure water and the boiling point at sea level into 100 equal steps, with zero being at the freezing point and 100C at the boiling point ¹.

Through observations of fire and forge, even the ancients were aware that conditions could be much hotter than the range of temperatures experienced in the normal course of climate. However, they could have had no real awareness of how much *colder* things could get. That had to await the theoretical insights provided by the development of thermodynamics in the nineteenth century, followed by the invention of the refrigeration by Carl von Linde not long afterwards. By the close of the century, temperatures low enough to liquify air had been achieved. This was still not as low as temperatures could go. The theoretical and experimental developments of the nineteenth century consolidated earlier speculations that there is an *absolute zero* of temperature, at which random molecular motions cease and the volume of an ideal gas would collapse to zero; no temperature could go below this absolute zero. On the Celsius scale, absolute zero occurs at -273.15C. Most of thermodynamics and radiation physics can be expressed more cleanly if temperatures are given relative to absolute zero, which led to the formulation of the Kelvin temperature scale, which shifts the zero of the scale while keeping the size of the degrees the same as on the Celsius scale. On the Kelvin scale, absolute zero is at zero degrees, the freezing point of water is at 273.15K, and the sea-level boiling point of water is at 373.15K. Viewed on the Kelvin scale, the temperature range of Earth's climate seems quite impressively narrow. It amounts to approximately a $\pm 10\%$ variation about a typical temperature of 285K. A 20% variation in the Earth's temperature (as viewed on the Kelvin scale) would be quite catastrophic for life as we know it. This remark can be encapsulated in a saying: "Physics may work in degrees Kelvin, but Earth life works in degrees Celsius,"

There is more to climate than temperature. Climate is also characterized by the amount and distribution of precipitation (rainfall and snowfall), as well as patterns of atmospheric winds and oceanic currents. However, temperature will do for starters. In this book we will discuss temperature at considerable length, and venture to a somewhat lesser degree into the factors governing the amount of precipitation. We will not say much about wind patterns, though some of their effects on the temperature distribution will be discussed in Chapter 10.

If you live outside the tropical zone, you will come to wonder why it is hotter in summer than in winter, and why the summer/winter temperature range has the value that it does (e.g. 30C in Chicago) and why the variation is generally lower over the oceans (e.g. 7C in the middle of the Pacific Ocean, at the same latitude as Chicago). If you communicate with friends living in the Arctic or Antarctic regions, and other friends living near the Equator, you will begin to wonder why, on average, it is warmer near the Equator than in the polar regions, and why the temperature difference has the value it does (e.g. 40C difference between the annual average around the Equator vs. the annual average at the North Pole). The physics underlying the seasonal cycle and the pole

¹The scale is named for the Swedish astronomer Anders Celsius, who originally formulated a similar temperature scale in 1742. Celsius' scale was reversed relative to the modern one, putting 100 at the freezing point and zero at the boiling point. The Celsius scale is sometimes called *centigrade*, but Celsius is considered to be the preferred term. The official definition of the temperature scale is now based on standards that are more precise and unambiguous than the freezing and boiling point of water.

1.2. CLOSE TO HOME

to equator temperature gradient is discussed in Chapters 8 and 10. If you climb a mountain (or even observe the snow-capped peaks of a mountain from the valley floor on a hot summer day), or if you go up in a hot-air balloon, or fly in an airliner which informs you of the outdoor temperature – you will notice that the air gets colder as one goes higher in altitude? Why should this be? This turns out to be a general feature of planetary atmospheres, and the basic physics underlying the phenomenon is discussed in Chapter 2.

The air that surrounds us is itself a matter of interest. We know that it is there because it has a temperature, exerts pressure, and because it is necessary that we breathe it in order to remain alive. But what is the air made of, and why does it have the composition it does? We can see water condense out of the air, but why don't other components condense in the course of natural weather and climate variations? How much air is there? And has it always been there with its present composition, or has it changed over time? If so, how much and how quickly?

We know that our planet journeys through the hard vacuum of outer space, clothed in a thin blanket of air – our atmosphere. It is natural to wonder how our atmosphere affects the Earth's climate. The airless moon shares the same orbit of the Earth, at the same distance from the Sun, so one can look to the Moon to get an idea of what the Earth's climate would be like if it had no atmosphere. We know the moon is airless because a reasonably thick atmosphere would bend the light rays from the Sun and stars, just as objects appear displaced when viewed through the surface of a swimming pool. But how to measure its temperature?

Of course, one could go there with a thermometer (and this did eventually happen) but people became curious about Lunar conditions long before it seemed likely that anybody would ever get there. Dante Alighieri himself, in the *Paridiso* written between 1308 and 1321, devoted fully one hundred cantos to a learned discussion between himself and Beatrice concerning the source of Lunar light and the solidity of the Lunar surface. By the mid nineteenth century, science had progressed to the point that the questions could be formulated more sharply, and the means for an answer had begun to emerge. With the discovery of infrared light by Sir William Herschel in 1800, astronomy opened a new window into the properties of planets and stars. Over the coming decades, it gradually became clear that all bodies emit radiation according to their temperature. This is known as *blackbody radiation* and will discussed in detail in Chapter 3. Infrared light from the Moon was detected by Charles Piazzi Smyth in 1856, and the first attempt to use it to estimate temperature was by the Fourth Earle of Rosse in 1870. The instruments available at the time were not up to the task. In 1878, Langlev invented the bolometer, which made good observations of Lunar infrared possible. However, while Langley made the first accurate observations of Lunar infrared, theory was not quite up to the task of interpreting the observations. These issues were largely sorted out by 1913, though Langley gave up on his earlier estimates rather reluctantly. By 1913 it was pretty clear that the daytime temperature of the Moon at the point where the Sun is directly overhead is well in excess of 373K (the sea-level temperature of boiling water on Earth). Night-time temperatures were harder to determine accurately, since the infrared emission from cold objects is weak; however it was clear that temperatures at night dropped by well over 140Krelative to the daytime peak. Pettit and Nicholson observed the temperature of the Moon during the Lunar eclipse of 1927, using the Mt. Wilson telescope. They found something even more remarkable: over the span of the few hours of the eclipse, the Lunar temperature fell from 342Kat the point of observation to 175K. Modern measurements show the daily average temperature at the Lunar equator to be around 220K, while the mean temperature at 85N latitude is 130K

It appears that without an atmosphere or ocean, the Earth would be subject to extreme swings of temperature between day and night. The Moon's "day" is 28 Earth days, since it always shows the same face to the Earth; on that basis, one could imagine that the day/night extremes were due to a longer night offering more time to cool down, but the rapid cooling during an eclipse gives the lie to this idea. Given the rapid cooling of an airless body at night, it is likely that the Earth's summer/winter temperature difference would be far more extreme in the absence of an atmosphere. Further, a comparison of the pole to equator gradient in daily mean temperature with that on Earth suggests that the atmosphere significantly moderates this gradient, too. What is it about the atmosphere or ocean that damps down day/night or summer/winter swings in temperature? This subject will be taken up in Chapter 8, where we'll also learn why summer is warmer than winter and why the poles are on the average colder than the Equator. How does an atmosphere or ocean moderate the temperature difference between pole and equator? We'll learn something of that in Chapter 10.

At its hottest the Moon gets much hotter than Earth, and at its coldest it gets much colder. But how does the Moon's mean temperature stack up against that of Earth? The 220K mean equatorial temperature of the Moon is very much colder than the observed mean tropical temperature on Earth, which is on the order of 300K. If the Earth's mean temperature were as low as that of the Moon, the oceans would be solidly frozen over. The cold mean temperature of the Moon does not come about because the Moon reflects more sunlight than the Earth; the Moon looks silvery but measurements show that it actually reflects *less* than Earth. Why is the Earth, on average, so much warmer than the Moon? Does this have something to do with our atmosphere, or is it the case that Earth is warmed by some internal heat source that the Moon lacks?

The search for the first stirrings of an answer to this problem takes us back to 1827, when Fourier published his seminal treatise on the temperature of the Earth. Fourier could not have known anything about the temperature of the Moon, but he did know a great deal about heat transfer – having in fact largely invented the subject. Using his new theory of heat conduction in solids, Fourier analyzed data on the rate at which average temperature increases as one descends deeper below the Earth's surface ; he also analyzed the attenuation of day/night or summer/winter temperature fluctuations with depth. (Fourier's solution for the latter problem will be derived in Chapter 8.) Based on these analyses, Fourier concluded that the flow of heat outward from the interior of the Earth was utterly insignificant in comparison to the heat received from the Sun. We'll see shortly that this situation applies to other rocky planets as well: dry rock is a good insulator, and doesn't let internal heat out very easily.

If the Earth is continually absorbing solar energy, it must also have some way of getting rid of it. Otherwise the energy would have accumulated over the past eons, leading to a molten, incandescent uninhabitable planet (see Problem ??) – which is manifestly not the case. Fourier seems to have known that there was little or no matter in the space through which planets plied their orbits, and so he posited that planets lose heat almost exclusively through emission of infrared radiation (called "dark heat" at the time). ² He also knew that the rate of emission of "dark heat" increased with temperature, which provided a means for an equilibrium temperature to be achieved: a planet would simply heat up until it radiated infrared energy at the same rate as it received energy from the Sun. Finally, Fourier refers to experiments showing that something in the atmosphere emits infrared radiation downward toward the ground, and seems to have been aware also of the fact that something in the atmosphere absorbs infrared. Based on these somewhat sketchy observations, Fourier inferred that the Earth's atmosphere retards the emission of infrared to space, allowing it to be warmer than it would be if it were airless.

Fourier's treatise made it clear the the thermal emission of infrared light was not just useful for astronomical observations – it was in fact part and parcel of the operation of planetary climate.

²Fourier also refers to the importance of heating from what he calls the "temperature of space." It is unclear whether he thought there was some substance in space that could conduct heat to the atmosphere, or whether he was referring to some invisible radiation which pervades space. His inferences regarding the importance of this factor were erroneous – the only real error in an otherwise remarkable paper.

1.2. CLOSE TO HOME

At Fourier's time the state of understanding of infrared radiation emission was not sufficiently developed as to allow him to complete the calculation he set up. Nonetheless, he correctly formulated the problem of terrestrial temperature as one of achieving a balance between the rate at which solar radiation is absorbed and the rate at which infrared is emitted. With this great insight, the modern era of study of planetary temperature had begun. Fleshing out the "details," however, required major advances in several areas of fundamental physics. The basic principles of planetary energy balance, and of the manner in which an atmosphere increases planetary temperature, are introduced in Chapter 3 and elaborated on in the earlier parts of Chapter 4.

One of the many details that needed to be settled was the question of which components of the atmosphere affected the transmission of infrared radiation. In 1859 Tyndall found that the dominant components of the Earth's atmosphere – nitrogen and oxygen – are very nearly transparent to infrared radiation. He found instead that it was two relatively minor constituents - water vapor and CO_2 - which accounted for most of the infrared absorption and emission by Earth's air. Gases of this sort, which let solar energy through virtually unimpeded but strongly retard the outward loss of infrared radiation, are known as "greenhouse gases." Their warming effect on the lower portions of a planet's atmosphere, and on its surface (if it has one) is called the "greenhouse effect." The term was not coined by Fourier, and in some ways is misleading, since real greenhouses do not work by blocking infrared emission. However, the glass or plastic enclosure of a real greenhouse does warm the interior by reducing heat loss to the environment while allowing solar heating, and in that sense - viewed as a broader metaphor for the implications of energy balance – the analogy is apt. Besides CO_2 and water vapor, we now know of a number of additional greenhouse gases, including CH_4 (methane), which may have played a very important role on the Early Earth, and plays some role even today. In fact, it turns out that in some very dense atmospheres such as that of Titan, even nitrogen can become a greenhouse gas. What determines whether a molecule is or is not a good greenhouse gas, and how do we characterize the effects of individual gases, and thus the influence of atmospheric composition on climate? These questions will be taken up in the latter half of Chapter 4.

In thinking about the effect of greenhouse gases on climate, it is important to distinguish between *long-lived greenhouse gases* which are removed slowly from the atmosphere on a time scale of thousands of years or more, and *short-lived greenhouse gases* which are removed on a time scale of weeks to years by condensation or rapid chemical reactions. The short-lived greenhouse gases act primarily as a feedback mechanism. Their concentration adjusts rapidly to other changes in the climate, serving to amplify or offset climate changes caused by other factors – including changes due to long-lived greenhouse gases. Long-lived greenhouse gases can also participate in feedbacks, but only on time scales longer than their typical atmospheric adjustment time. Whether a greenhouse gas is long-lived or short-lived depends on environmental conditions. On the Earth, CO_2 is a long-lived greenhouse gase but water vapor is a short-lived greenhouse gas; however, on Mars, which gets cold enough for CO_2 to condense, that gas can be considered short-lived.

Greenhouse gases are largely invisible, but the atmosphere also holds a readily visible component that exerts a profound influence over our planet's energy balance – the clouds. Clouds on Earth are composed of suspended droplets of condensed water, in the form of liquid or ice. Clouds, like water vapor, act as a short-lived greenhouse gas affecting the rate at which infrared can escape to space. The infrared opacity of clouds is used routinely in weather satellites, since this property makes cloud patterns visible from space even on the night side of the Earth. However, clouds affect the other side of the energy balance as well, because cloud particles quite effectively reflect sunlight back to space. The two competing effects of clouds are individually large, but partly offset each other, so that small errors in one or the other term lead to large errors in the net effect of clouds on climate. Moreover, the effect of clouds on the energy budget depends on all the intricacies of the physics that determine things like particle size and how much condensed water remains in suspension. For this reason, clouds pose a very severe challenge to the understanding of climate. This is the case not just for Earth, but for virtually any planet with an atmosphere. The physics underlying the effects of clouds on both sides of the radiation balance will be discussed in Chapters 4 and 5

1.3 Into deepest time: Faint Young Sun and habitability of the Earth

The Solar system was not always as we see it today. It formed from a nebula of material collapsing under the influence of its own gravitation, and once the nebula began to collapse, things happened very quickly. The initial stage of formation of the Solar system was complete by about 4.6 billion years ago. By this time, the Sun had begun producing energy by thermonuclear fusion; the formation of the outer gas giant planets and their icy moons by condensation, and the formation of the inner planets by collision of smaller rocky planetesimals, were essentially complete. The last major event in the formation of the Earth was collision with a Mars-sized body 4.5 billion years ago, which formed the Moon and may have melted the Earth's primitive crust in the process. All these collisions left behind a great deal of heat that had to be gotten rid of before the crust could stabilize. To determine how long it takes to get rid of this heat, we must learn about the mechanisms by which planets lose energy, and about how the rate of energy loss depends on temperature and atmospheric composition; this will happen in Chapters 3 and 4. It turns out that a planet loses energy almost exclusively by radiation of infrared light to space. While the precise rate of loss depends on the nature of the atmosphere, all estimates show that the surface of the Earth quickly cools to 2000K, at which point molten rock solidifies; in the absence of an atmosphere, this process takes a thousand years or less, while with a thick atmosphere it could take as long as two or three million years.

Once a solid crust forms, the flow of heat from the interior of the Earth to the surface is sharply curtailed, because heat diffuses very slowly through solid rock. In this situation, supply of heat from the interior becomes insignificant in comparison with the energy received from the Sun, and the Earth has settled into a state where the climate is determined by much the same processes that determine today's climate: a competition between the rate at which energy is received from the Sun against the rate at which energy is lost to space by radiation of infrared light. This is very likely to have been the case by 4.4 billion years ago, if not earlier. There are no actual rocks as old as this, but there are individual zircon crystals embedded in the Jack Hills formation of Western Australia which are 4.4 billion years old. Zircons of a similar age are also found within the 3.7 billion year old crustal rocks of the neighboring Narryer Gneiss Complex. These crystals provide indisputable evidence for the existence of at least some continental crust of a sort very like that we see today; they also provide convincing though less certain evidence of the existence of liquid water in contact with the early continental crust. The existence of liquid water does not in itself put much constraint on temperature, since water can be maintained in a liquid state even at temperatures in excess of 500 degrees Kelvin, provided the pressure exerted by water vapor in the atmosphere is high enough. The thermodynamics needed to address this issue will be introduced in Chapter 2. Certain aspects of the chemical composition of the zircons, however, suggest that they interacted with near-surface water having a temperature of 100C or less. By 4.4 billion years ago, it would appear, the Earth was no longer a molten volcanic inferno.

The precise nature of the climate evolution between 4.5 and 3.8 billion years ago is obscure at present. Depending on the composition of the atmosphere, the surface temperature could have

1.3. INTO DEEPEST TIME: FAINT YOUNG SUN AND HABITABILITY OF THE EARTH 9

been as high as 200C or low enough to cause the ocean (if any) to freeze over completely, and the climate could well have swung wildly between the two extremes. In addition, the dates of Lunar craters indicate that the Earth very likely underwent a period of heavy bombardment by interplanetary debris between 4.1 and 3.8 billion years ago; it is generally supposed that this *late heavy bombardment* affected the rest of the inner Solar system as well, though that is far from certain. The energy brought in by impacts during this period could easily have been sufficient to bring surface temperatures episodically to values well in excess of 100C, sterilizing any nascent ecosystems. Life, if any, may have waged and won a battle for survival in deep ocean refugia.

By 3.8 billion years ago, the veil begins to lift. This is the age of the oldest intact rocks, found today in the Isua Greenstone Belt of Greenland. The appearance of these rocks marks the end of the *Hadean eon*, and the dawn of the *Archaean eon*. Remnants of 3.7 billion year old shales in the Isua formation show the unmistakable signs of deposition of sediments in open water. More intriguingly, these shales are rich in organic carbon, and this carbon preserves a chemical signature generally associated with microbial activity – life. The Barberton formation of South Africa and the Warrawoona formation of Australia, both about 3.5 billion years old, contain layered carbonate sedimentary structures known as *stromatolites*, which in later times are known to be laid down by microbial mats. This is not an unambiguous sign of life, since inorganic processes can also produce stromatolite-like features. Be that as it may, the early stromatolites certainly require ponds of open water evaporating into air. The Barberton and Warrawoona formations also contain microscopic features that are suggestive of bacterial fossils, though not unambiguously so.

The record of surface conditions during the subsequent billion years is hardly continuous, but preserved rocks dating to this period very commonly show a sedimentary character of a type most easily explained by deposition in an open, unfrozen ocean. The first truly unmistakable microbial fossils date to 2.6 billion years ago, where they are found in the Campbell formation of Cape Province, South Africa, and argue for open water conditions having a moderate temperature. At about this time, we bid farewell to the Archean eon, and enter the *Proterozoic eon*, which extends to the appearance of animal life 544 million years ago. Certain fine-grained silica based sedimentary rocks known as *cherts* preserve information about past temperatures, as well as a wealth of fossils. Very ancient cherts contain no unambiguous microbial fossils, but certain aspects of their chemical composition point to temperatures as high as 70C at 3.5 billion years ago, declining to 60C at 2 billion years ago, and declining further to 30C at 1 billion years ago. Well-preserved ancient cherts are rare, however, so this data by no means implies that temperatures were uniformly warm on the young Earth. It only indicates that the Earth attained high surface temperatures at least part of the time; there is ample room to hide lengthy cold periods within the gaps in the chert record, as we shall soon see.

The earliest geological indication of the presence of glaciers on Earth occurs in the upper part of the Pongola formation of South Africa, and dates to 2.9 billion years ago. The evidence consists of glacial sedimentary deposits called *diamictites*, material of a sort usually transported by floating ice, and even glacier-scratched rocks. This does not mean that there were no earlier glaciations, but in light of the chert record and widespread occurence of marine sedimentary rock it seems fairly certain that the Earth did not spend the bulk of its earlier history locked in a deep-freeze. Still, the Pongola glaciation seems to mark the beginning of Earth's long flirtation with ice. The Makganyene glaciation begining around 2.3 billion years ago, recorded in rocks of the Transvaal group of Southern Africa, was a big one, and may well have been global. We know this because a record of the Earth's magnetic field is preserved in the rocks, and this can be used to infer the latitude at which the rocks were located when the glacial deposits were laid down. This *paleomagnetic data* shows that there was ice within 12 degrees of the Equator, strongly suggestive of a global glaciation. The first unambiguous bacterial microfossils (found in the Campbell group of South Africa) date to 2.6 billion years ago, shortly before the Makganyene glaciation. While earlier fossil and geochemical evidence is very strongly suggestive of life, the Campbell group fossils are the ocular proof that biology was well underway. These fossils mark a watershed in another important way, in that they are identifiable as cyanobacteria – the type of organisms that produce oxygen by photosynthesis. The issue of when cyanobacteria evolved is hotly debated, with some lines of indirect evidence putting their appearance early in the Archaean and others dating their onset to the time of the Campbell Group microfossils. Be that as it may, the appearance of these fossils speaks for a fairly benign environment, with open water and temperatures no more than about 40C. After the Makganyene glaciation, microbial fossils become quite abundant. The two billion year old Gunflint Chert of Canada is one of many such marine sedimentary formations in which cyanobacterial microfossils are preserved.

So far, no glaciations have been reported in the period between two billion years ago and 800 million years ago, though there are abundant sedimentary rocks dating to this time. The record is far from continuous and the lack of glaciations in this period may be an artifact of preservation, but the evidence certainly indicates that icy climates were not dominant at this time, and were probably quite rare. The long hiatus in ice is terminated by the massive – and possibly global – Snowball Earth glaciations of the Neoproterozoic, about 700 million years ago. Thereafter, the climate alternated between fairly lengthy periods when the Earth was ice-free or nearly so, and periods when there was at least some ice in polar regions. The ice never again, however, reached the nearly global proportions it attained during the Neoproterozoic suggesting that the Earth passed some new threshold of climate stability in the Neoproterozoic. What might that be? This is one of the central questions of climate science.

Our overall picture of Earth history is that liquid water and moderate temperatures appeared at least episodically very shortly after the Moon-forming collision, and that the next three billion years had widespread open water with temperatures probably not exceeding 70C and generally much less. These conditions were punctuated by occasional glaciations, only a very few of which may have been global in extent. It was an environment that could support the evolution and survival of life, including (by 2.6 billion years ago, if not before) photosynthetic life requiring moderate temperature conditions. Let's keep this picture of relative stability in mind as we go on to discuss long-term changes in the atmosphere and the Sun – the two principle ingredients determining the Earth's climate, or indeed that of any planet.

There are many processes at play that cause the composition of a planet's atmosphere to evolve over time. In the earliest times, bombardments can help supply atmosphere-forming volatiles such as water, nitrogen and carbon dioxide. Equally, however, sufficiently energetic bombardments can cause loss of atmosphere through literally splashing it into orbit. On a volcanically active planet with a hot interior, such as the Earth or Venus, or the younger Mars, new atmosphere is continually being supplied by outgassing of volatile substances from the interior of the planet. The heat needed to keep the interior of the planet churning so it can recycle minerals formed at the surface and cook out volatile gases in the hot interior is supplied by leftover heat from formation of the planet and by radioactive decay. How long this process can continue before the planet freezes out becomes tectonically inactive depends on the size of the planet and the stuff it is made of; the nature of the gases which come out of volcanoes and other types of vents depends on the chemistry of the planet. For example, the early segregation of iron in the Earth's core made it harder to bind up oxygen in minerals, and therefore resulted in fairly oxidized gases like carbon dioxide (CO_2) , and sulfur dioxide (SO_2) being released in preference to gases like methane (CH_4) and hydrogen (H_2) - though some of the latter two do nonetheless escape. The interior Earth also outgasses water vapor (H_2O) , which is cooked out of hydrated minerals; the volume of the oceans appears to have

1.3. INTO DEEPEST TIME: FAINT YOUNG SUN AND HABITABILITY OF THE EARTH11

been in a steady state for a long time, though, indicating that the rate of release is balanced by the rate of formation and subduction of new hydrated minerals at the surface. Nitrogen (N_2) is fundamentally different from other current and past constituents of the Earth's atmosphere as it doesn't readily get incorporated into the minerals that form the bulk of Earth's crust and interior. Unlike, say, CO_2 , nitrogen does not cycle through the Earth's interior. The bulk of the Earth's N_2 is in its atmosphere and stays there, where it has probably been for almost all of our planet's history. This is likely to be the case as well for any other rocky planet made of stuff similar to the Earth – Iron, oxygen (mostly bound up in minerals), silicon, magnesium and sulfur.

While atmosphere is being supplied by outgassing from the interior, other processes cause material to be lost from the atmosphere. Parts of a planet's atmosphere extend far out from the surface, where hot, fast-moving molecules can reach escape velocity and escape to space. Besides escape from random molecular motions, the solar-heated tenuous outer atmosphere can sustain fluid flows which cause atmospheric mass to fountain systematically into the void. In addition, the solar wind can literally blow way the outer portions of an atmosphere; the extent to which this happens is affected by the intensity of the planet's magnetic field, which shields the atmosphere from the solar wind. As outer parts of the atmosphere are eroded, new gases from lower altitudes well up to replace the lost material, sustaining the gradual loss of atmosphere. All three mass loss processes preferentially remove lighter molecules, either because lighter molecules move more swiftly for a given temperature, or because the outer atmosphere is enriched in gases having a lower molecular weight. For a given density, a smaller planet has lower surface gravity, and so binds its atmosphere less tightly; in consequence, escape of atmosphere to space proceeds more rapidly on a small planet. Impacts by large, swift bodies can impart sufficient energy to part of the atmosphere to blast it into space. This mechanism of atmosphere loss does not discriminate as to molecular weight, but as with the other mechanisms, it is easier for a small planet to lose atmosphere this way. Overall, the Moon or Mars is more prone to lose atmosphere than more massive bodies such as the Earth or Venus, to say nothing of Jupiter or Saturn. For Earth and Venus, escape to space is significant only for H_2 and He, and of these the latter is important only as an indicator of planetary history rather than as a physically or chemically active component of the atmosphere. Saturn's satellite Titan is an interesting case, as it maintains a mostly N_2 atmosphere more massive than that of Earth (per unit surface area) despite having a surface gravity lower than that of the Moon. The very cold temperature of Titan helps it retain its atmosphere, but it is nontheless likely that the persistence of the atmosphere requires a substantial rate of resupply from the interior of the planet.

Some components of the atmosphere can also be lost through chemical reactions with rocks at the Earth's surface. A particularly important example of this is the *Urey reaction*, which removes CO_2 from the atmosphere. When CO_2 dissolves in water, it forms a weak acid (carbonic acid), which reacts with silicate minerals (e.g. $CaSiO_3$) to form carbonate minerals (e.g. $CaCO_3$, or "limestone"). The reaction takes place only in the presence of liquid water, so if a planet becomes so hot that liquid rain never reaches the surface, or if it somehow loses its water altogether, then CO_2 outgassed from the interior of the planet will accumulate in the atmosphere until the interior source is depleted or the rate of supply is balanced by loss to space. On Earth, all of the CO_2 presently in the atmosphere could be removed by the Urey reaction within 5000 years, forming a layer of limestone a mere 5 millimeters thick; if all the carbon stored in ocean water were to outgas as CO_2 and react to form limestone, the process would take a half million years and form a layer a half meter thick.

Life itself, once it appears, has a profound effect on atmospheric composition. While little methane escapes directly from the Earth's interior, bacteria known as *methanogens* can synthesize it from H_2 and CO_2 or from organic material produced by other organisms. Methanogens may

well have dominated the ecosystems of the Earth's first two billion years, potentially allowing a methane-dominated atmosphere to build up. The advent of life also had a profound effect on nitrogen cycling. The bonds holding together N_2 are so strong that in the abiotic world only rare energetic events such as lightning strikes can form nitrogen compounds. In fact, though nitrogen is an essential ingredient of all living material, higher forms of life – including all plants and animals - are incapable of performing the chemical magic that makes N_2 available to organisms. This trick is accomplished by nitrogen-fixing bacteria, which can efficiently transform atmospheric N_2 into ammonium (NH_4) , in turn transformed into nitrate (NO_3) which can be used by other organisms in the synthesis of living matter. Other bacteria, in oxygen-starved conditions, can make a living combining the oxygen in nitrate with carbon, returning N_2 to the atmosphere in the process. It was only in 1910, with the invention of the Haber Process for turning atmospheric N_2 into ammonia (NH_3) that humans caught up with the bacteria. This innovation has become essential to the human population, as the demands of industrial-style agriculture have far outstripped the ability of the natural bacterial ecosystem to supply nitrate (which is not to deny that other forms of agriculture might be able to live within the means provided by our bacterial friends). Nitrogenfixing bacteria are still way ahead of industry in terms of chemical sophistication, though, since the Haber Process requires molecular hydrogen (made from fossil methane), an iron catalyst, temperatures exceeding 400C, and operates at a pressure over two hundred times that of air at the Earth's surface; Nitrogen-fixing bacteria can do the same trick, in contrast, in ambient temperature/pressure conditions and using materials found readily in their immediate environment.

In the absence of oxygen-producing photosynthetic life, only minuscule amounts of oxygen would be present in the atmosphere, since only a trickle could be produced through the breakdown of H_2O by exposure to sunlight. That there was very little oxygen in the early atmosphere (under 0.2%, compared to 20% today) is confirmed by the widespread presence of striking rock structures known as *banded iron formations* until 2.4 billion years ago. Banded iron formations can be laid down only when iron is very soluble in the ocean and can be transported long distances. This requires low oxygen, since in the presence of oxygen iron forms compounds that are not very soluble in water. Additional evidence stemming from the chemical composition of certain sulfurcontaining minerals indicates that during at least some periods earlier than 2.6 billion years ago the atmospheric oxygen content might in fact have been orders of magnitude lower than 0.2%.

Note that the appearance of oxygen-producing photosynthesis is not synonymous with the rise of oxygen in the atmosphere. For oxygen to accumulate, a sufficient proportion of the organic matter must be buried before it is oxidized by other bacteria, taking the synthesized oxygen right back out of the atmosphere. Further, if the Earth had accumulated a great stock of available organic matter in the ocean during its anoxic phase, this backlog would have to be worked off before oxygen could begin to accumulate to any significant degree. Be that as it may, banded iron formations begin to falter somewhat after the time of the Campbell Group cyanobacterial microfossils, and disapear entirely by around 2 billion years ago. By this time, oxygen may have made up at least 3% of the atmosphere. Once oxygen made its appearance in the atmosphere at significant concentration, it changed all the rules of atmospheric chemistry, since it is so powerfully reactive. In particular, it made it much harder for CH_4 and H_2 to accumulate in the atmosphere, since the former oxidizes readily to CO_2 and the later to H_2O . The rise of oxygen may also have fostered another great biological innovation – the eukaryotic cell, which has a complex internal organization with specialized structures, including a nucleus within which genetic material is segregated. We are made of eukaryotic cells, as are all animals and plants. Eukaryotic cells make their first unambiguous appearance in the fossil record about 1.5 billion years ago, in the Roper Group shales of Australia, though more indirect evidence suggests that eukaryotic life may have evolved much earlier. However the answer to that issue may shake out, it is certain that eukaryotic life - even of the single-celled variety – did not proliferate and diversify until much later in the Proterozoic.

1.3. INTO DEEPEST TIME: FAINT YOUNG SUN AND HABITABILITY OF THE EARTH13

There was a sporadic reappearince of banded iron formations at the time of the Neoproterozoic Snowball events (600-700 million years ago), but by 600 million years ago oxygen was approaching its present concentration and banded iron formations disappeared for good (at least so far). This second pulse of oxygenation made the rise of multicellular animals possible, which happened in a great flurry of biological innovation known as the *Cambrian Explosion*, occupying a remarkably short span of years around 543 million years ago. It is even possible that the rise of animals helped to stabilize atmospheric oxygen levels, by providing a reliable means of transporting organic carbon to the sea floor where it can be buried and preserved. Simple multicellular oceandwelling plants appeared much earlier, as might be expected from the fact that photosynthetic organisms are not dependent on oxygen.

All of this atmospheric evolution takes place against a backdrop of a gradually brightening Sun. The energy produced within a star leaves the star almost exclusively in the form of electromagnetic radiation – loosely speaking, light of all wavelengths. The net power output is called the *luminosity*, and is measured in *Watts* (a measure of energy per unit time), just as if the star were a light bulb. Stars like the Sun get their energy by fusing hydrogen into helium, and as time goes on the proportion of helium in the Sun increases, thus increasing the mean molecular weight of the Sun. This in turn means that the core of the Sun needs to contract and heat up in order to maintain the pressure required to balance gravity. The increased density and temperature increases the rate of fusion more than the reduced availability of hydrogen reduces it, so the rate of production of energy – and hence the Solar luminosity – *increases* with time. The resulting evolution of luminosity over time rests on fundamental aspects of solar physics that are not seriously in question, and does not depend greatly on the finer points of solar modeling. The results of the standard solar evolution model can be fit by the expression

$$I_{\odot}(t) = \frac{I_{\odot}(t_{\odot})}{1 + \frac{2}{5}(1 - t/t_{\odot})}$$
(1.1)

where $I_{\odot}(t)$ is the luminosity of the Sun at time t and t_{\odot} is the current age of the Sun, usually taken as 4.6 billion years. This formula was originally proposed to describe the younger Sun, but it continues to be reasonably accurate out to times 4 billion years in the future as well.

It follows from Eq. 1.1 that 4 billion years ago the Earth received solar energy at only 75% of the rate it does today. All other things being equal – atmospheric composition in particular – that means the Earth would have been colder than it is today. How much colder? We will learn how to do this calculation in the simplest way in Chapter 3, and add sophistication to the calculation in Chapter 4. It turns out that if the atmospheric composition were the same as today's atmosphere throughout Earth's history, then Earth should have been completely frozen over 4 billion years ago, and given that ice reflects sunlight so well, it should still be completely frozen over today. However, we know that incidents of global ice coverage are rare in Earth history, if indeed they happened at all; we know with rather more certainty that the Earth is not solidly frozen over today. This contradiction is generally known as the Faint Young Sun Paradox, though of course, like most paradoxes it is not paradoxical at all once one understands what is going on. Calling it a "paradox" is just a way of starkly bringing home the fact that to account for the basic facts of Earth's climate history, the atmosphere must indeed have undergone massive changes - changes of a sort that could substantially affect climate. How much would we need to increase CO_2 or CH_4 in order to make up for the faintness of the young Sun? This is another one of the big questions. It will be answered in Chapter 4. A related Big Question is the extent to which CH_4 (or some other long-lived greenhouse gas) substituted for CO_2 in maintaining the Earth's warmth during the Archaean. There is scattered evidence from mineral composition of fossil soils that during some periods of the Archaean the CO_2 concentration might not have been high enough to offset the Faint Young Sun. This has led some researchers to jump to the conclusion that CH_4 played the

decisive role throughout the Archaean, but such a viewpoint rests on exceedingly shaky evidence. It matters a great deal whether CO_2 or CH_4 did the trick, since the long term control of the two gases is governed by very different processes, with very different time scales. The matter, at present, must be considered unresolved.

A corrollary to the above resolution of the Faint Young Sun Paradox is that any atmosphere that would be sufficient to keep the Early Earth unfrozen would make it uninhabitably hot with the present Solar output. The atmosphere must have somehow changed in lock step with the brightening Sun, in precisely such a manner as to keep the Earth in a habitable temperature range – one where liquid water exists at or near the surface, but where the water never gets hotter than about 100C – or indeed, even hotter than about 50C if cyanobacteria are to survive. It defies belief that the required co-evolution of atmosphere with solar output could maintain the required temperature range purely by coincidence, so it seems likely that some sort of temperature-regulating mechanism is in operation. In Chapter 9 we'll see how the Urey reaction can participate in a geochemical thermostat for the Earth and similar planets. The Snowball episodes represent temporary, and evidently rare, breakdowns of the temperature-regulating mechanism. Whatever the regulating mechanism may be, it must be sufficiently fail-safe to allow for recovery from global or near-global glaciations.

Given what we laid out earlier concerning the myriad processes churning out turmoil in atmospheric composition, the reader should quite rightly feel it a bit silly to make the stipulation "all other things being equal" in the statement of the Faint Young Sun Paradox. Indeed, even if the solar output were constant over time, any number of the aspects of atmospheric evolution we discussed earlier would have been sufficient to freeze or fry the Earth, so the gradual increase in solar output has no special claim on our attention in this regard. To be fair, when the "paradox" was first laid out, much less was known about the ways the Earth's atmosphere evolved, and with much less certainty. From the standpoint of current science, however, the traditional framing of the Faint Young Sun Paradox leaves a lot to be desired. It would be more satisfactory to refer to the Habitability Problem, which could be stated as follows: *How can the temperature of a planet be maintained in the habitable range for billions of years, in the face of geological, geochemical and biological turmoil in atmospheric greenhouse gas composition, and in the face of gradual increase in solar output?*. This is indeed one of the grandest of questions. The material discussed in Chapter 9 provides a plausible solution, but the book is far from closed on the Habitability Problem.

The history of the Faint Young Sun problem reveals something magnificent and deeply inspiring about the nature of discovery in Earth and planetary climate. The basic physics underpinning the energy source of stars was worked out by Hans Bethe by 1939, and the existence of benign conditions on the Early Earth was known (or at least inferred) even earlier. Yet, it was not until 1972 that Carl Sagan and George Mullen put the two bits together and inferred that there was indeed a big problem, requiring a profound answer ³. This insight sparked a revolution in thinking about planetary climate, that was in its own way as earthshaking as the discovery of DNA was in its own domain. This history highlights a lovely thing about our subject: The most profound new phenomena are often discovered by putting together a few bits of basic physics and chemistry in creative new ways. For the most part, new ideas come from playing with simple models, not from enormous incomprehensible computer simulations that take huge teams to put together. The entire goal of this book is to teach students to think the way Carl Sagan and other innovators did, and to provide the tools needed to build the simple models needed to turn a bright idea into real science.

³Sagan and Mullen proposed accumulation of atmospheric ammonia as a resolution to the paradox, but later work on atmospheric chemistry showed that sunlight destroys ammonia too rapidly to allow it to build up to the required concentrations. Attention later shifted to CO_2 and CH_4 .

1.4. GOLDILOCKS IN SPACE: EARTH, MARS AND VENUS

The problem of maintaining long-term climate stability is not just an issue for Earth. Planets that could support some form of life are naturally of special interest, and while other forms of life than those we know of might prefer quite different conditions than prevail on Earth, the long-term maintainence of climate in a fairly narrow "habitable" range clearly would make it easier for life to evolve and persist. Any potentially life-bearing planet in any solar system anywhere must negotiate a way to maintain long-term habitability in the face of the gradual increase of the brightness of its sun and gradual or not-so-gradual changes in the composition of its atmosphere. Naturally, such considerations apply to the evolution of the climates of other planets in our very own solar system. Venus and Mars did not manage to maintain their habitability, or perhaps were never in a habitable state. How close did we come to the same fate?

1.4 Goldilocks in space: Earth, Mars and Venus

Until well into the 1960's, science fiction stories about Venus generally portraved it as a steamy jungle planet, but one where intrepid explorers could perhaps survive unprotected on the surface. The idea of a jungle and breathable air was of course unfounded speculation, but the general picture of the climate was not wholly without merit. After all, the dense reflective cloud deck of Venus was readily observable – it is what makes Venus so bright as the "evening star" – and the reflection of sunlight could easily make up for the fact that Venus is closer to the Sun than is Earth. In fact in Chapter 3 we'll see that the reflectivity *more than* makes up for the proximity. In the late 1950's the picture of Venus as a habitable world began to unravel. Recall that the temperature of Earth's Moon was determined by examining infrared radiation from that body. Viewed in the infrared spectrum Venus appeared quite cool, but in the microwave ("radar") spectrum it was far too bright. In fact, seen in the microwave, Venus radiated like a body with a temperature well in excess of 600K (327C). A popular hypothesis at the time was that the anomalous microwave emission arose in the upper portions of the Venusian atmosphere. Another view held that the microwave emission came from the surface of the planet, and that the atmosphere was transparent to microwaves but relatively opaque to infrared. The latter idea suffered from the lack of a plausible mechanism to make the surface of Venus so hot. Then, in 1960 the young Carl Sagan proposed that Venus has a very thick atmosphere rich in greenhouse gases, which would heat up the surface to the required temperature. Little was known about the mass of the atmosphere or its composition at the time, but Sagan developed simple models of the greenhouse effect of a thick atmosphere, which showed that the trick could be accomplished with an atmosphere consisting of mostly carbon dioxide with some water vapor mixed in, having a total mass three or four times that of the Earth's atmosphere. Sagan even recognized that since the planet was too hot for water vapor at the hypothesized concentration to condense and reach the surface as liquid, the Urey reaction (which removes carbon dioxide from the atmosphere and turns it into limestone) could not take place. This would make it easy for carbon dioxide to accumulate in the atmosphere, though even Sagan did not envision just how far this would go.

A series of interplanetary probe missions over the next two decades – four US Mariner missions, two US Pioneer missions and sixteen Soviet Venera missions including eight Venera missions that returned data successfully from the surface – refined the estimates of surface temperature and substantially revised the conception of the atmospheric mass and composition. By the late 1970's, it was known that the surface temperature was nearly uniform at 737K. The atmosphere was found to be much more massive than originally thought, in fact sufficiently massive to raise the surface pressure to 92 times that of Earth's atmosphere. And, it was found that the atmosphere consisted almost entirely of carbon dioxide, with only traces of water vapor remaining. The thick clouds that give Venus its high reflectivity were found to be made not of water, but of droplets of sulfur dioxide and concentrated sulfuric acid. It took the better part of another decade before the challenges of dealing with the effect of such an exotic atmosphere on climate were fully mastered and a fully satisfactory account of the high surface temperature could be given. Still, the initial exploration of the problem was carried out with simple models very like the ones we'll introduce in the first half of Chapter 4. The discovery of the true nature of Venus climate is another illustration of the tenet that big ideas grow from little models.

Mars yielded up its climatic secrets somewhat earlier, because it was not hidden behind the thick atmospheric veil that complicated observation of Venus. Mars was observed in the infrared using the Mt. Wilson telescope during the opposition (time of closest approach) of 1926 and 1927. Like all infrared observations, the interpretation was complicated by interference from the Earth's atmosphere. By 1947, the understanding of the effect of atmospheres on the emission of infrared light had progressed to the point that the mass and composition of the Martian atmosphere could be estimated. Based on these measurements, Kuiper estimated that Mars had an atmosphere that was almost entirely CO_2 , with a sufficient mass that the surface pressure on Mars would be only .03% of the surface pressure on Earth. This turns out to be an underestimate of the true mass by about a factor of twenty, but even so, the picture of Mars as a nearly airless planet was not far wrong. By way of comparison, infrared observations of Venus interpreted in the 1940's using similar techniques suggested that the surface pressure of Venus was a fifth that of Earth an underestimate by a factor of nearly 500. The Mt. Wilson infrared observations of Mars also indicated that the atmosphere was almost entirely CO_2 , with almost no water – so little, in fact, that water vapor wouldn't condense until temperatures fell below -60C, and at those temperatures it would condense into frost, not liquid. Infrared observations showed further that the visible polar ice caps of Mars are most probably made of water ice rather than frozen CO_2 . Temperature estimates based on the Mt. Wilson infrared observations were less informative. Nicholson and Pettit, in the same paper in which they discuss Lunar temperatures, noted a very large day/night cycle of Martian infrared, indicating extreme diurnal constrasts unlike those found on Earth. They concluded that this was due to the lack of water vapor in the Martian atmosphere, but we shall encounter the true reason in Chapter 8. Writing in 1947, Adel reported quantitative estimates of Martian surface temperature ranging from as low as 236K to as high as 318K (or even higher if the surface was assumed to emit infrared inefficiently, as does granite). Some of the variation in reported temperatures may have been due to the fact that these were not whole-disk observations, insofar as Mars exhibits extreme temperature contrasts. The higher end of the estimates based on Mt. Wilson turned out to be far greater than the actual maximum ground temperature encountered anywhere on the planet.

Given the thin atmosphere, what does theory lead us to expect about the Martian surface temperature? By 1947, it was a simple exercise to compute the expected temperature of airless bodies like the Moon, using arguments like those we'll discuss near the beginning of Chapter 3. Planets with atmosphere present more of a problem. The atmosphere of Mars is thin compared to that of Earth, but how thin does an atmosphere have to be (particularly if it's pure CO_2) in order to have a minimal effect on planetary temperature? We'll learn how to answer that question in the latter portions of Chapter 3. Based on similar reasoning, Kuiper, writing in 1947, inferred correctly that the atmosphere would have only minor effect on T, in particular allowing severe night-time temperature drops. By the 1940's Mars was already looking inhospitable – a mostly cold, dry nearly airless body where (it was still hoped) conceivably lichens might eke out a living but certainly not Thuvia, Maid of Mars.

Ground-based and theoretical estimates of the Martian climate improved gradually over the next decade, but the real breakthrough came with the Mariner flyby of 1965 and the two Viking orbiters and landers of 1976. Spaceborne infrared observations gave the first detailed picture of geographic variations of the ground and air temperature, and Viking provided in situ air temperature measurements of the ground. These observations consolidated the picture of Mars as a planet which (as we find it now) has more in common with the airless moon than with Earth. Even hopes of lichens were dashed, though it is too soon to give up on bacterial life, especially in view of the innovative chemical entrepreneurship shown by non-photosynthetic bacteria on Earth. In discussing Martian temperatures, one must take care to distinguish the air temperature from the temperature of the ground itself, as the two differ considerably. At high noon in the tropics, the ground can indeed briefly get as warm 300K, though temperature drop by 100K or more at night. The air temperature also shows an extreme day/night cycle, but the peak daytime air temperatures are far less than the peak ground temperature; at the Viking Lander 1 site, in the tropics at 22N latitude, the daytime air temperature never exceeds 260K, and plummets to under 200K at night. The reason the air temperatures are so much lower than the ground temperatures will become clear in Chapters 4 and 6.

There are considerable seasonal and latitudinal variations in daytime temperature, and the Southern Hemisphere polar summer is notably warmer than the Northern Hemisphere polar summer. Night-time temperatures are comparatively uniform both geographically and seasonally; the Mars surface cools so fast that that once it is dark, it evidently doesn't matter much whether it has been dark for a few hours or a hundred days. The Southern Hemisphere winter pole does get notably colder than the rest of the planet, dropping to as low as 160K. The Viking landers also provided the first clear picture of surface pressure variations on Mars. They showed that the Martian surface pressure varies from a high of about 1% of Earth's surface pressure to a low of about 0.6%, with the lowest values occurring in Southern Hemisphere winter. Since surface pressure gives a measure of the mass of air in the atmosphere (2), the large variation of pressure indicates that a considerable portion of Mars' CO_2 atmosphere snows out over the North pole in Northern winter and the South Pole in Southern winter, only to sublimate back into the atmosphere as spring approaches. A theory for the seasonal cycle of Martian temperature and pressure will be developed in Chapter 8.

So, Venus, like the porridge tasted by Goldilocks is too hot. Mars is too cold, and Earth is just right. One could quite reasonably object that this is a view prejudiced by our own status as a form of terrestrial life, and that conditions "too hot" by our standards could well be "just right" for somebody else. However, it appears that there's nobody home on either Venus or Mars (not even a microbial somebody), so if conditions there are indeed "just right" for somebody, it must not be very easy for such a somebody to evolve, given that it didn't happen in the past four billion years.

Presumably, Venus started out with a composition rather similar to Earth. What went wrong? Why did it keep most of it's CO_2 in its atmosphere, whereas most of Earth's CO_2 got bound up in carbonate rocks? Where did its water go? The answer came in 1967 with the theory of the *runaway greenhouse*, formulated first by M. Kombayashi and shortly thereafter independantly rediscovered by Andrew Ingersoll of Caltech. This theory puts together two simple bits of physics, the first being that water vapor content of a saturated atmosphere increases exponentially with temperature (Chapter 2), and the second being that water vapor is a greenhouse gas (Chapter 4). When the two are put together, it is found that a planet which receives sufficient solar radiation can get into a runaway cycle where the planet warms in response to absorbed sunlight, which causes more water vapor to enter the atmosphere, which causes more greenhouse effect, which leads to further warming in an unstable feedback loop that doesn't end until the entire ocean is evaporated into the atmosphere. At that point, the water vapor in the upper atmosphere breaks down into hydrogen and oxygen under the influence of high energy solar radiation, and the hydrogen escapes to space while the oxygen reacts with rocks. Without liquid water, the Urey reaction which turns CO_2 into limestone can't take place, so all the outgassed CO_2 stays in the atmosphere. The runaway greenhouse theory (explored in Chapter 4) gives rather precise predictions of the circumstances under which a runawy can occur, and explains why Earth did not undergo a runaway despite the fact that it has a water ocean. The work of Kombayashi and Ingersoll is another example of the general idea that big ideas come from simple models. Their reasoning was based on simple radiation models of the sort developed in the first half of Chapter 4. This work also illustrates another general principle of planetary climate: profound results can be obtained by combining a few bits of very basic physics in a novel way.

Radar mapping from the Magellan orbiter of 1990 revealed another remarkable fact about Venus: Unlike the Earth, with long-lived continents and a gradually subducted sea floor, Venus has a young-looking uncratered surface, suggesting that the crust may have been engulfed and resurfaced as recently as 500 million years ago. This has important implications for planetary habitability. Evidently, the formation of a planetary crust, as at end of Hadean, is not end of the peril from fires in the deep. For habitability, the crust has to be relatively stable, engulfed slowly in subduction zones (as is the Earth's sea floor) rather than being subject to episodic catastrophic volcanism as seems to have been the case on Venus. However, if the crust is engulfed *too slowly*, then limestone that forms by the Urey reaction is only sluggishly recycled, leading to a drawdown or atmospheric CO_2 and (under some circumstances) a very cold planet; since photosynthesis uses CO_2 as a feedstock, low CO_2 impedes habitability even if a planet is in an orbit where it doesn't need the greenhouse effect of atmospheric CO_2 in order to stay warm. The question of when a planet has *plate tectonics* like Earth or when it has episodic catastrophic resurfacing like Venus, is another one of the great questions of planetary science, though one we will not take up at any great length in this book.

Mars may be impoverished in atmosphere compared to Venus, but it has something Venus lacks: a geological record of the distant past preserved in its crust. In fact, the ancient features on the surface of Mars are far better preserved than is the case on Earth. Mars appears to have lost most of its atmosphere quite early on, leading to a near-halt in the rate of erosion of surface features. The first high resolution data of Martian surface features, returned by the Mariner mission, revealed a startling fact. Evidently, Mars was not always the dry, frigid planet cloaked in a tenuous atmosphere that we see today. The Mariner photographs revealed dry river-like channels for which the only likely explanation is flowing surface water, which would be impossible under the conditions of the present Martian climate. A more recent image of this type of feature is shown in Figure 1.1. The rate of cratering of a planet goes down with time, so the features can be dated by counting superposed craters; many of the major river-like features date to very early in Mars history, perhaps 4 billion years ago. This led to the concept of a "warm, wet Early Mars." But how could Mars have been so much warmer than it is at present, at a time when the Sun was so much fainter. Mars presents an even more extreme version of the Faint Young Sun paradox than does the Earth. It will probably come as no surprise that it was Carl Sagan who first pointed out the implications of Martian dry river networks. It is still somewhat disputed whether the surface features really demand that Early Mars be warm and wet, but adopting the warm-wet view, the resolution of the Faint Young Sun problem, as was the case for Earth, lies in the supposition that the Early Mars atmosphere had a substantially stronger greenhouse effect. What kind of atmosphere could warm Mars to the point that liquid water could flow long distances at the surface? That question is taken up in Chapters 4 and 5.

If Mars started out with such a dense atmosphere, where did it go? Some possible answers are suggested in Chapter 9. Modern high-resolution images suggest other forms of massive climate change on Mars. In particular, there are tropical landform features suggesting that at some time in the past, glaciers formed in the Martian tropics, whereas virtually all the ice is today sequestered

1.4. GOLDILOCKS IN SPACE: EARTH, MARS AND VENUS



Figure 1.1: Nanedi Vallis on Mars, observed by the Mars Orbiter Camera on the Mars Global Surveyor Mission. The image covers an area 9.8km wide and 18.5km tall.

at the cold polar regions. The tropical glacier landforms suggest that at some time in the past, the equatorial regions of Mars were colder than the poles? How could that be? The answer is provided in Chapter 8.

The fact that Earth maintained habitable conditions while Venus succumbed to a runaway greenhouse and got to hot while Mars lost its atmosphere and got too cold raises the question of just how narrowly Earth escaped the fate of Mars and Venus. How much could Earth's orbital distance be changed before it turned into Mars or Venus, and how would the answer to this question change if Earth were more massive (making it easier to hold onto atmosphere) or less massive (making it easier to lose atmosphere? If Mars were as large as Earth, would it still be habitable today? What if Venus were as small as Mars? Perhaps if the orbits of Mars and Venus were exchanged, our solar system would have three habitable planets, instead of just one.

The range of orbital distances for which a planet retains Earthlike habitability over billions of years is know as the *habitable zone*. Determing the habitable zone, and how it is affected by planetary size and composition as well as the properties of the parent star, is one of the central problems of planetary climate.

1.5 Other Solar System planets and satellites

For the gas giant planets – Jupiter and Saturn – many of the most striking questions that arise are fluid dynamical in nature. These questions include the origin of the banded multiple-jet structure of the atmospheric flow, and the dynamics of long-lived atmospheric vortices, most famously Jupiter's Great Red Spot. We shall have little to say about such fluid dynamical questions in this book. However, the thermal structure o the atmosphere provides an essential underpinning for any dynamical inquiry. Moreover, the thermal structure determines the rate of heat loss from the planet, and therefore plays a crucial role in the long-term evolution of the gas giants. The thermal structure also affects atmospheric chemistry and the nature of the colorful clouds that allow us to visualize the spectacular fluid patterns on these planets.

The gas giants present an interesting contrast to rocky terrestrial-type planets, because of the lack of a solid surface. Instead of solar radiation having the possibility of penetrating to the ground and being absorbed there, thus heating the atmosphere from below, solar radiation on the gas giants is deposited continuously throughout the upper portion of the atmosphere as the solar beam propagates downward and attenuates. Further, unlike the case of the rocky planets where heat flux from the interior is an insignificant player in climate, the fluid nature of the gas giants allows considerable heat flux from the interior to escape to space. For both Jupiter and Saturn, this heat flux is comparable to the flux of energy received from the Sun. One of our objectives in subsequent chapters will be to learn how the distinct nature of atmospheric driving on the gas giants affects the thermal structure. The gas giants also offer an interesting opportunity to test ideas about how climate is affected by atmospheric composition. These planets are mostly made of H_2 , with a lesser amount of He and trace amounts of a range of other substances, including ammonia (NH_3) , methane (CH_4) and water, the latter three of which exist in both gaseous and condensed forms. The composition affects the thermodynamics of the atmosphere, as well as the optical properties for both infrared and visible light.

Uranus and Neptune are like the gas giants in that they have no distinct solid surface at any depth that could significantly affect the atmosphere. However, they are usually classified separately as *ice giants* because they contain a much higher proportion of ice-forming substances such as water, ammonia and methane. The composition of the outer portions of the atmospheres can be determined by spectral observations, and contain a high proportion of hydrogen and helium. The overall density of the planets, however, constrains them to be composed primarily of an *ice* mantle. In the case of Uranus, the ice mantle must make up between 9.3 and 13.4 Earth masses worth of the total mass of the planet, which is 14.5 Earth masses. Similar proportions apply to Neptune. The commonly used term "ice mantle" is somewhat misleading, since the substance is actually a hot, slushy mixture that would be more aptly described as a water-ammonia ocean. Whatever term is used, the very thermal structure that determines the nature of the transition between the ice mantle and the more gaseous outer atmosphere engages all the same issues of atmospheric energy balance as one encounters on other planets. A novel feature of Uranus is its axial tilt. It's axis of rotation is very nearly perpendicular to the plane of the orbit. That means that in the Uranian Northern Hemisphere summer, the North pole is pointing directly at the Sun and the entire Southern Hemisphere is in darkness. By way of contrast, the Earth's axis is only tilted by 23.4° at present. The high axial tilt potentially gives Uranus an extreme seasonal cycle, though it will take a long time to observe it since Uranus' year lasts 84 Earth years. The effect of axial tilt on seasonal cycles of planets are discussed in general terms in Chapter 8. The very low solar radiation received at the distant orbits of Uranus and Neptune leads to extremely cold outer atmospheres, particularly in the case of Neptune. These planets provide an opportunity to examine the novel features of an atmosphere driven by an exceedingly weak trickle of solar energy, supplemented by an equally feeble trickle of heat from the interior. Despite the weak thermal driving, Neptune has by far the strongest winds in the Solar system, as well as a variety of interesting meteorological features. We will not say much about planetary winds, but as in the case of the gas giants, a good understanding of the thermal structure is a prerequisite for any attack on the meteorology.

The gas and ice giants also challenge our notion of *habitable zones* – orbits where a planet has some region where there are Earthlike temperatures allowing for liquid water. The gas and ice giants have no distinct surface, but there is some depth on each of them where the temperature is Earthlike and liquid water can exist. The atmosphere also have plenty of chemical feedstocks for organic molecules, including ammonia and methane. The pressures are no greater than those seen at the bottom of the Earth's ocean. One may have some prejudice in favor of surfaces for life to live on, but it must be recalled that on Earth life first arose in the oceans and indeed stayed there for many billions of years before venturing onto land. The gas and ice giants could just as well be thought of as being "all ocean" rather than "all atmosphere" so it is far from clear that they are inhospitable, at least for chemosynthetic forms of life that don't need much sunlight. Our thinking about habitable zones is overly prejudiced toward life that carries out its existence on a rocky surface.

From the standpoint of planetary climate, one of the most interesting Solar System bodies is not a planet at all, but a satellite. Titan, which orbits Saturn, is a fairly large icy body with a radius of that is 76% of that of Mars. Because it is composed of ice rather than rock, the surface gravity is low: $1.35 m/s^2$, which is actually lower than the Moon's surface gravity, though the Moon is smaller than Titan. What makes Titan interesting, however is its dense atmosphere. The atmosphere of Titan consists mainly of nitrogen, with a surface pressure about 1.5 times that of Earth. What is even more interesting is that the lower portion of the atmosphere is about 30% methane. At the cold temperatures of Titan (about 95K) methane can rain out, and participates in a hydrological cycle analogous to that of water on Earth – but operating at a much colder temperature. In subsequent chapters we will develop the physics to examine the similarities and differences between the role of methane on Titan vs. water on Earth. Titan's atmosphere is also a seething organic chemical factory, with complex long-chain hydrocarbon hazes being manufactured from methane in the upper atmosphere. These hazes absorb solar radiation, shade the surface, and are a key player in Titan's climate. Such organic hazes were first discovered on Titan, but there are speculations that similar hazes could have been present in methane-dominated atmospheres of the Early Earth.

A major question about Titan is why it has an atmosphere left at all. Given the low gravity, the N_2 atmosphere would be expected to escape fairly quickly (we'll have a look at the relevant physics in Chapter 9). Moreover, the chemical reactions in the atmosphere should gradually convert all the methane into a tarry sludge sequestered at the surface. In some way or another, the atmosphere of Titan must be dynamically maintained by recycling of chemicals deposted on the surface, and by outgassing of N_2 (probably in the form of ammonia) and CH_4 from the interior. Precisely how this happens is one of the Big Questions of Titan.

Even icy moons without an appreciable atmosphere can manifest features of considerable interest. Jupiter's moon Europa is a case in point. This satellite has a water ice crust between 10 and 50 km thick, but beneath the crust there lies a liquid water ocean. Europa shows an intriguing range of crustal features, including some that suggest melt-through of the ocean. Of course, the existence of the ocean has attracted attention as possible habitat for life. The icy moons challenge the epistemological boundaries of planetary science. At the cold temperatures of Europa's surface, as on Titan, water ice is basically a rock, just as sand can be considered an "ice" of SiO_2 on Earth. The ice-rock forms minerals with ammonia and methane and other compounds, and when warm enough the ice-minerals can flow or melt and lead to cryovolcanism. When studying the crust and interior of Europa or Titan or other icy moons are we doing geology or oceanography or glaciology? Whatever one wants to call it, these moons are, as has been said, "always icy, never dull."

1.6 Farther afield: Extrasolar planets

Up until 1988, the Solar system was the only field of play for students of planetary climate, and it provided the only example against which theories of planetary formation could be tested. Revolutionary improvements in detection methods led to the first confirmed detection of a planet orbiting another star in that year, and instrumentation for planetary detection has continued to improve by leaps and bounds. As of the time of writing, over 228 planets orbiting stars within 200 parsecs (652.3 light years) of Earth have been detected, and the rate of detection of new planets is if anything accelerating. Certainly, much of the excitement surrounding the new zoology of planets has revolved around the prospect for detecting a planet that is habitable for life as we know it – or have known it to be in the past few billion years of Earth history. Perfectly aside from the habitability question, though, the rich variety of new planets discovered offers the student of planetary climate stimulus for thinking well outside the box of how the climates of known Solar System planets operate and have evolved over time.

Planets have been detected in orbit about a variety of different kinds of stars, so it is necessary to learn something of how stars are characterized. At the most basic level, stars are classified according to their luminosity (i.e. their net power output) and the temperature of the star at the surface from which the starlight escapes to space. The luminosity is determined by measuring the brightness of the star as seen from Earth's position, and then correcting for the distance between the star and the Earth. For relatively nearby stars, the distance can be measured directly by looking at the tiny shift in angular position of the star as seen from opposite ends of the Earth's orbit, but for stars farther than 500 parsecs (1630.8 light years), more indirect inferences are required. The stellar effective temperature is determined by measuring the spectrum of the starlight – how the brightness changes when the star is observed through different filters. Hotter stars have colors towards the blue end of the spectrum, while cooler stars tend toward the red. Hot stars emit more energy per unit surface area, but a reddish cool star can still have very high luminosity if it is very large, since it then has more surface area that is emitting. The energy sustaining the emission of starlight comes from the fusion of lighter elements into heavier elements. Since hydrogen is by far the most common element in the Universe that can participate in thermonuclear fusion, the overwhelming majority of stars ignite by fusing hydrogen into helium. These stars are known as *Main Sequence* stars, and stellar structure models predict that there is a distinct relation between the luminosity and emission temperature for stars that get their energy in this way. The stellar structure theories imply that the position of a star on the Main Sequence is determined primarily by its mass, with more massive stars being both hotter and more luminous. Moreover, stars spend most of their lifetimes on the main sequence, so that a scatter plot of luminosity vs. temperature of stars – a *Hertzsprung- Russell diagram* shows most stars to be clustered along the main-sequence curve. Indeed, the Main Sequence was discovered by plotting catalogs of stars in this way long before the energy source of stars was discovered.

Stars do not evolve *along* the main sequence. Rather, they enter it at a certain position when fusion ignites, remain near the same point for a certain amount of time while gradually brightening, and then leave the main sequence for a comparatively short afterlife as a brighter star with a more rapidly evolving spectrum. What happens when a star leaves the Main Sequence depends on the star's mass. The Sun will spend a billion years or so as a Red Giant, before collapsing into a gradually fading white dwarf. Main Sequence stars are thought to be the best candidates for hosts of habitable planets, since they provide relatively long-term stable stellar environments. Once a star leaves the main sequence, the climates of any unfortunate planets the star may have had will have been radically disrupted, if indeed the planets continue to exist at all; there will be relatively little time for any new form of life to establish itself. Nonetheless, planetary systems do exist off the main sequence, and these planets, too have their points of interest. The first confirmed detection of a planet orbiting a Main Sequence star did not occur until 1995.

Just as the luminosity of the Sun increases over time, the luminosity of other Main Sequence stars increases during their time on the Main Sequence. However, the lifetime of a star on the Main Sequence varies greatly with the mass of the star. The mass of the star determines the amount of nuclear fuel available to sustain the star's life on the main sequence, while the luminosity give the rate at which this fuel is consumed. The Main-Sequence lifetime of a star with mass M and luminosity I, scaled to values for the Sun, is estimated by

$$\tau = \tau_{Sun} \frac{M}{M_{Sun}} \frac{I_{Sun}}{I} \tag{1.2}$$

Main sequence stars have a power law mass-luminosity relationship $I \propto M^{3.5}$, so on the Main Sequence the lifetime scales with luminosity according to the law

$$\tau = \tau_{Sun} \left(\frac{I_{Sun}}{I}\right)^{1.29} \tag{1.3}$$

Bright, hot, blue massive stars thus have a much shorter lifetime on the Main Sequence than dim, cool reddish dwarf stars. The Sun has a Main Sequence lifetime of about 10 billion years, and is nearly halfway through it's time on the Main Sequence.

Figure 1.2 shows the Hertzsprung-Russell diagram for stars that are known to host one or more planets. Astronomers designate the colar (equivalently surface temperature) of stars according to *spectral classes* given by the letters O, B,A,F,G,K,M extending from hottest to coldest, with numbers appended to indicate subdivisions within a spectral class. Our Sun is a class G2 Main Sequence star. The diagram shown in the Figure represents a tiny subset of the many millions of stars that have been catalogued, and none of these stars (so far) have spectral classes hotter than F. The stars cluster along the Main Sequence simply because there are more stars in general along



Figure 1.2: Scatter plot of luminosity vs. effective surface temperature for stars about which at least one orbiting planet has been detected as of 2007. Luminosity is given as multiples of luminosity of the Sun. The colder stars are redder while the hotter stars are more blue in color. The letters at the top indicate the standard spectral classification of stars in this temperature range, and the dashed line approximately locates the Main Sequence. The Sun is a class G2 Main Sequence star.

the Main Sequence than elsewhere. There is also a selection bias due to the technologies available at present for detecting planets, which work better for some kinds of stars than for others. Thus, the gap in detections between K and M0 stars may be an artifact of detection bias rather than a reflection of some basic feature of planetary formation.

There is a rich supply of planets orbiting stars between spectral classes G0 and K, as well as a good handful of planets around bright red giant stars off the main sequence. The cluster of detections around M5 dim red-dwarf stars are particularly interesting, as many of these turn out to represent the most Earthlike planets to date – again a detection bias, because it is easier to detect low-mass Earthlike planets around low mass stars using present technology. These M-dwarf stars are very dim, so a planet has to be in a very close orbit about its star in order to be as warm as Earth. In compensation, these systems have very long lifetimes compared to the Sun and other brighter stars. According to Eq. 1.3, an M5 dwarf spends about 100 times as long on the main sequence as the Sun does, and so this kind of star will brighten only very slowly over time. Such a star provides a very stable climate to its planet, and requires much less adjustment of atmospheric conditions than the Earth had to accomplish to resolve the Faint Young Sun problem. In contrast, an F0 star would spend under a billion years on the Main Sequence, and if the history of life on Earth is anything to go by, life around an F0 star would be snuffed out at the prokaryotic stage. before it could even begin to think of making oxygen. Aside from affecting the lifetime, the spectral class of the star affects the degree of absorption of stellar radiation by whatever atmosphere the star's planet may have, and this too provides a lot of novel things to think about when pondering the climates of extrasolar planets.

1.6. FARTHER AFIELD: EXTRASOLAR PLANETS

That takes care of the stars, but what is known about the extrasolar planets themselves? Here, too, there is a detection bias, since it is much easier to detect massive planets comparable in mass to Jupiter than it is to detect more Earthlike planets. Most planets detected to date are very massive planets which, according to theories of planetary formation, are likely to be gasgiants like Jupiter or Saturn, or ice giants like Neptune or Saturn. The full variety of planetary climates offered by the various combinations of planetary mass, orbital characteristics and stellar characteristics is hard to convey by looking at just a few graphs. We'll give a small sampling of this variety below. In the course of the exercises in this book the student will have ample opportunities to explore the wider universe of exoplanets.

One of the key determinants of planetary cliamte is the rate at which the planet receives energy from its star. This is a function of the luminosity of the star and the distance of the planet from the star, which varies to some extent in the course of the planet's year (as discussed in Chapter 8). Planets that receive more stellar energy flux than the Earth will tend to be hotter, all other things being equal, whereas planets that receive less will tend to be colder. The left panel of Fig. 1.3 shows the mass of the planets discovered so far and plotted against the stellar flux heating the planet at the time of the planet's closest approach to its star. The masses are measured relative to the mass of Jupiter and the fluxes are measured relative to Earth's solar heating. One Earth mass is 0.00315 times the mass of Jupiter. On this diagram, a planet with a relative flux of unity would have would have an Earthlike temperature if its atmosphere were like Earth's. We see that a great many planets with a mass one tenth Jupiter mass or greater have been discovered; these are all likely to be gas giants or ice giants made mostly of hydrogen and helium. Though these are in some ways like Jupiters, their climate has no real analog in the Solar system, since most of them are in orbits where the planet receives at least as much stellar energy as the Earth receives from the Sun. These are all "hot Jupiters," and represent climates very different from anything in the Solar System. Could Jupiters receiving Earthlike stellar radiation be habitable for life? That is certainly a Big Question, requiring understanding of the climate of such planets. Even more exotic are the giant planets receiving vastly more stellar flux than the Earth - up to one thousand times as much, in fact. These are "roasters" – very hot gaseous planets in close orbits about their host stars. Planetary formation theory gave no inkling that such things should exist, and indeed the existence of roasters poses real challenges for the theory.

Another way that the new extrasolar planets offer novel climates is in the nature of their orbits. Solar System planets, except for Pluto, have fairly circular orbits. However, most exoplanets have highly elongated orbits with a considerable difference between the distance of closest approach to the star (the *periastron*) and the distance of farthest remover from the star (the *apastron*). The range of orbital elongation is shown in the right panel of Fig. 1.3. Since the stellar energy goes down like the square of the distance from the star, the planets with highly elongated orbits will have novel seasonal cycles unlike any encountered in the Solar System. They would tend to heat up to a great degree at periastron and cool down, perhaps freezing over any ocean, at apastron. Could such a planet be habitable? The answer depends a lot on the thermal response time of the planet's atmosphere and ocean, which could average out the orbital extremes. The relevant physics is discussed in Chapter 8.

The orbital period, or "year", of the extrasolar planets also varies widely, as shown in the left panel of Fig. 1.4. Planets have been discovered with a wide range of orbital periods, ranging from as little as two Earth days to as much as 6000 Earth days. Planets in close-orbits with short orbital periods are likely to be *tide locked* and always present the same face to the star, just as the Moon always presents the same face to the Earth. The Super Earth's to be discussed shortly are mostly in this category. Tide-locked planets offer novel possibilities for planetary climate. The night-side could get very cold, and if the planet has an ocean, it might freeze over completely.



Figure 1.3: Left panel: Scatter plot of mass of extrasolar planets in units of Jupiter masses vs. the flux of stellar energy impinging on the planet at the time of closest approach (the periastron). Right panel: Scatter plot of the ratio of farthest (apastron) to closest (periastron) distance of the planet from its star in the course of the planets orbit vs. the stellar energy at the time of closest approach. The stellar energy flux is given as a multiple of the corresponding flux for the Earth.

The day side would be hot, but there could be a habitable zone near the ice margin. Further, the transport of moisture and heat from the day side to the night side poses interesting questions; the answers are important, since such transports in large measure will determine the nature of the planet's climate.

Low mass planets are of particular interest because according to theories of planetary formation they have the best chance to have a rocky composition similar to that of Earth, Mars or Venus. Relatively few planets with a mass of 10 Earth masses (0.03 Jupiter masses) or less have been discovered, but there has been recent progress in this area. The planets discovered so far in this range are all considerably more massive than Earth, and are therefore called *Super Earths*. The left panel of Fig. 1.3 shows that a handful of Super Earths have been discovered with stellar fluxes ranging from .25 that of Earth (yielding a too-cold planet) to 1000 times that of Earth (yielding a planet far too hot. The closest to being "just right" is Gliese 581c, with a flux of about three times that of Earth. Would such a planet be habitable, or would it turn into a Venus? The physics needed to answer such questions will be developed in Chapter 4.

The right panel of Fig. 1.4 gives an indication of the masses of planets that have been discovered about stars having various temperatures. The stellar temperature is of interest to climate since it determines the spectrum – the redness or blueness – of the starlight, which in turn affects the absorption of starlight by various atmospheric constituents. Hotter stars also put out more of the energic ultraviolet radiation, which can have a profound effect on atmospheric chemistry. We see that a few low mass Super Earths have been found around class G or K stars, but examination of the associated planetary data reveals that these planets are in close orbits and receive several hundred times as much stellar energy as the Earth receives from the Sun. They will unquestionably be extremely hot places, and unlikely to be able to sustain an atmosphere or liquid water. There is, however, a small cluster of Super Earths orbiting very cool stars with temperatures under 3700K. These stars are Main Sequence "M-dwarfs." They are very red, very small, and very dim, but by virtue of their dimness planets in close-orbits can still have a chance to be habitable. Moreover, dim stars like M-dwarfs have a long lifetime, and therefore provide a stable environment for their planets. Gliese 581 is such a system, but it appears that the two most Earthlike planets still miss being habitable – one likely to turn into a Venus if it has an



Figure 1.4: Left panel: Scatter plot of orbital period (in Earth days) vs. the flux of stellar energy impinging on the planet at the time of closest approach (the periastron). Right panel: Scatter plot of the mass of the planets (in units of Jupiter masses) vs. the effective radiating temperature of the host star.

ocean and the other likely to freeze into a Snowball. Part of the reason for interest in M-dwarfs comes purely from detection technology. It is comparatively easy to detect low-mass planets in close orbit around a low-mass star, and at the same time low-mass dim stars give a planet in a close orbit a chance to be habitable. As time goes on, it is likely that a habitable world around an M-dwarf will be discovered. As detection technology improves, the possibility for discovering Earthlike habitable planets will expand to other spectral classes of stars.

Naturally, one is at this point intensely curious as to the composition of these planets, whether they have water, and what their atmospheres (if any) are made of. Some information can be inferred from planetary formation theory and theories of atmospheric evolution, but there is as yet no great ability to determine atmospheric composition from observations. That will change in the next decade or two, as satellite-borne instruments come online which will be able to determine the spectrum of emission and reflection from extrasolar planets. The anticipated instruments will only return a single spectrum averaged over the whole visible surface of the planet, but a great deal can nonetheless be inferred about atmospheric composition from such data. Learning how to make the most of this single-pixel planetary astronomy, in which the Earth would appear as a pale blue dot (in Carl Sagan's words) opens a whole range of Big Questions. Much of the same radiative transfer used in calculating planetary climate bears on the interpretation of planetary spectrum as well.

1.7 Digression: About climate proxies

1.7.1 Overview of proxy data

Instrumental records of climate – that is, records of measurements of temperature and other quantities by scientific instruments – date back at most a few hundred years. The first accurate thermometer was invented in 1654 by Ferdinando II de'Medici, and two hundred years passed before anything like a global network of reliable temperature measuring stations began to become available. Written historical records of such events as frost dates, encounters with sea ice and depictions of mountain glacier length provide some information about the climate of the past few millennia, but for the most part one must rely on *climate proxies* for information on what climate was doing a century or more ago. A climate proxy is any measurable thing preserved in the geological record of Earth or other planets, from which some aspect of climate can be inferred; to be useful, a climate proxy must come with a *chronology*, that is, some means of telling what period the proxy dates to. There is a vast and ever-improving array of climate proxies. We have encountered a few already. For example, the existence of river networks on the surface of Mars tells us that at some time in the past the surface of Mars must have been warm enough to support a liquid (probably water) flowing for long distances along the surface; in this case, the chronology at the time of writing comes from counting the number of craters superimposed on the features. Similarly, the existence of marine sedimentary deposits and stromatolites during the Archean provides compelling evidence that the Earth was warm enough to support open ocean water through much of its early history.

Some of the more intuitive kinds of proxies derive from plants and animals that live on land, since physiology places certain constraints on the conditions in which various organisms can grow or thrive. The presence of cold-blooded animals like crocodiles is a sure sign that winters cannot have been much below freezing for extended periods of time. Some kind of plants require tropical conditions to survive, while others require cooler conditions; even the shape of leaves provides information about temperature. A great deal of this evidence is preserved in the fossil record. Where there are trees, the width of tree rings provides a record of annual variations in the temperature of the growing season (though it is sometimes hard to distinguish temperature from rainfall effects). Land-based proxies are only available after the time at which a fairly diverse ecosystem established itself on land. While the first primitive plants colonized land as early as 470 million years ago (in the *Ordovician period*), and the first plants with stem structures able to conduct water appeared perhaps 430 million years ago (the *Silurian period*, a diverse land ecosystem did not really get underway until the *Devonian period*, which began about 416 million years ago. Once plants colonized land, animals in search of a free lunch followed not long afterwards.

One of the richest sources of information about past climates of the Earth comes from material preserved in sea floor sediments. Sediment is laid down in layers like the pages of a book, in which the history of Earth's climate can be read. The Earth is a dynamic planet, and sea floor is constantly being re-created at mid-ocean ridges, and likewise pulled down into the Earth's interior for recycling at subduction zones. For this reason, the deep-ocean marine sediment record is mostly limited to the past 100 million years, and is rather sketchy for the first half of that period. The oldest remaining deep-sea floor is about 180 million years old, and there is precious little of that. Near-shore deposits on continental shelves, on the other hand, can be uplifted and preserved for hundreds of millions, or even billions, of years. Many of the key marine deposits that tell us about the climate of the Neoproterozoic (about 600 million years ago) are now high and dry in Namibia, while others are found in Arctic Canada. These various deposits have a lot of individual personality and are known to geologists by names such as the Acasta Gneiss, the Warrawoona Formation, the Akademikerbreen or the Isua Greenstone Belt; we've met many of these already in the preceding sections.

Numerous aspects of the sedimentary record have been used to infer past conditions, and the ingenuity of paleoclimatologists is adding to the list all the time. Some sedimentary proxies involve the physical structure of the sediments, and are independent of chemistry or biology. *Diamictites* are a class of sediments known to be carried by land glaciers discharging into the ocean. They are a sure sign of very cold conditions, since it is only in such cases that a glacier can survive to sea level. *Ice-rafted debris* (*IRD*) is coarse material that can only be carried offshore by hitching a ride on icebergs or sea-ice. *Dropstones* – individual stones the size of pebbles and larger which fall with enough force to deform sediment layers – are a particularly striking form of ice-rafted debris. Inasmuch as stones do not float, they present very convincing evidence for ice. Continental dust

1.7. DIGRESSION: ABOUT CLIMATE PROXIES

in sediments provides an indication of the strength of wind, since larger particles require stronger winds to loft and transport them; the mineral composition of the dust can often mark where the dust came from, and hence the direction of the wind. Surveying the species of fossil algae found in sediments can provide information about the temperature of the layers in which the algae grew, since some organisms require colder temperatures while other require warmer temperatures.

A great deal of information can be gleaned from the chemical composition of the ocean, and of the microscopic creatures which dwell within it. Some chemical proxies do not rely on biology, and others make use of organisms mainly as nearly passive recorders of the composition of the ocean. Still other proxies are more intimately tied to biology, through the effect of temperature on the rate at which an organisms makes use of one element in preference to another. For example, the ratio of Magnesium (Mq) to Calcium (Ca) in corals and in the shells of certain micro-organisms is dependent on the temperature at which the organisms grew. The Strontium-Calcium substitution has been used similarly. Organic molecular proxies represent an important new source of data; it has been discovered that certain kinds of micro-organisms produce long-chain organic molecules with a somewhat variable chain-length which depends on the temperature at which the organism grows. When these molecules are robust enough to be preserved in the sedimentary record, the ratio of chain-lengths can be used to infer past temperatures. Alkenone and Tex-86 proxies fall into this class. They are useful only over time spans for which organisms producing the molecules exist, and can be presumed to have similar biochemistry to modern analogue organisms. Both alkenones and Tex-86 have been used to probe climates lying many tens of millions of years in the past, as well as more recent climates of the past few hundred thousand years.

If the chemical composition of a sedimentary sample has been altered by interaction with ocean water at some time after the sediment was first formed, then the information the sediment provides about past conditions is severely compromised. Post-depositional alteration is known as *diagenesis*, and it is a problem that plagues interpretation of geochemical sedimentary proxies. Paleoceanographers have had to become very clever sedimentary detectives in order to check for the nefarious effects of diagenesis, particularly when dealing with samples more than a few million years old. In some cases the existence of diagenesis has gone undetected for a decade or more, as will be discussed later in connection with the problem of hothouse climates.

1.7.2 Isotopic proxies

The chemical properties of an element are primarily determined by the number of protons in the nucleus (the *atomic number*, which also determines the configuration of the electron cloud. Nucleii also contain neutrons, and atoms having the same atomic number can appear in forms with different numbers of neutrons. These differing forms are known as *isotopes* of the element. Isotopic proxies have proved to be a versatile source of information about past climates. Some isotopes are unstable, and decay into other elements; these can be used as "clocks," to determine when things happened. The original, and most famous, such application is radiocarbon dating, which makes use of the decay of the form of carbon having a molecular weight of 14 (carbon-14, or ^{14}C . Stable isotopes do not decay, and instead provide a tracer of past chemical reactions in which the substance participated. For elements heavier than Helium the stable isotopic composition of a planet is determined primarily by the synthesis of the elements in the supernova explosion which gave birth to the material eventually incorporated into the Solar system. The isotopic ratio can in some cases be further changed by the process of planetary formation. For example, oxygen has three stable isotopes: ${}^{18}O$, ${}^{17}O$ and ${}^{16}O$, the latter of which is by far the most common. About 1 in 500 atoms of oxygen on Earth are ${}^{18}O$, which is nearly the same ratio as found in the Sun and which presumably represents the composition of the primordial Solar nebula. ^{17}O is less
commonly used as a proxy because it is so much rarer than ${}^{18}O$ (only 1 in 2500 oxygen atoms on Earth), but variations in its composition are readily detectable with modern measurements, and have important specialized uses. Hydrogen has two stable forms: Deuterium (D) which has a proton and a neutron, and normal hydrogen (H) which has only a proton. About one hydrogen molecule in 6500 occurring in Earth's ocean water is D. This differs from the 1:1700 ratio in the outer Sun, because Deuterium is destroyed by nuclear fusion in the Sun and in the early Solar nebula; curiously, the composition of the outer Sun is nearly the same as Jupiter. The stable isotopes of carbon are ${}^{13}C$ and ${}^{12}C$, of which the latter is by far the more common. About 1 in 100 carbon atoms on Earth are ${}^{13}C$. The carbon isotopic system attracts particular attention because carbon is the basis of organic chemistry – the stuff of life as we know it. Stable isotopes of S, N, Ar and many other elements have proved useful as proxies. In fact, it is hard to think of any measureable isotopic ratio that hasn't proved useful as a proxy for some aspect of climate or atmospheric composition.

The utility of isotopes as climate proxies derives from what they tell us about processes that sort them out into different reservoirs. Isotopes of an element have *nearly* the same chemical and physical behavior, but not *exactly* the same behavior. The subtle differences come about in part because, at a given temperature, the molecules of the lighter isotopes have a higher typical velocity than those of the heavier isotopes, since all molecules have the same mean kinetic energy at any given temperature. Among other things, this means that lighter isotopic forms of a substance evaporate more readily than heavier ones, so that the vapor phase of a substance is typically depleted in heavy isotopic forms relative to the condensed phase with which it is in contact. We will find this effect particularly useful in the interpretation of water isotopes, but a very similar process applies to the gradual "evaporation" of a planet's atmosphere into outer space, and can be used to constrain the proportion of atmosphere that has been lost in such a fashion.

The rate at which a given element or a molecule containing that element undergoes chemical reactions also depends on the isotope involved. Other things being equal, heavier isotopes would tend to react more slowly, because they have lower speeds and therefore lower collision rates with other reactants than their lighter, nimbler cousins. However, there are more subtle effects involved that can either enhance or retard reaction rates. Specifically, the difference in molecular weight between isotopes can affect the characteristic vibration frequencies of molecule bonds, and this can in turn affect the probability that colliding reactants will stick together. Very often, the degree of preference for one isotopic form over another – the degree of fractionation occurring between reactant pool and reaction product – depends on the temperature at which the reaction is taking place. When this is so, the isotopic composition of the product provides a useful paleo-thermometer. In Section 1.3 we stated that "certain aspects" of the chemical composition of zircons constrained the temperature of Hadean water with which the zircons were in contact; with this cryptic phrase, we were in fact referring to the ratio of ${}^{16}O$ to ${}^{18}O$ in the Hadean zircons which, like all silicate minerals, contain oxygen (their chemical formula is $ZrSiO_4$). Similarly, it was the oxygen isotopic ratio of cherts (later confirmed by silicon isotopic ratios) that was used to constrain Archaean temperatures in Section 1.3. A problem with all such paleothermometers is that fractionation is relative to the isotopic composition of the reactant pool, so that one needs to know the likely composition of the reactant pool in order to infer temperatures from the reaction product (e.g. the cherts or zircons), which are preserved long after the pool of reactants have been dissipated.

Biochemistry also has isotopic preferences. Photosynthetic Earth life (even of the sort that doesn't produce oxygen) prefers the lighter forms of carbon, so that organic material of photosynthetic origin is enriched in ${}^{12}C$ and depleted in ${}^{13}C$ relative to the inorganic carbon left behind. A record of the isotopic composition of inorganic carbon is preserved in carbonate minerals (e.g. $CaCO_3$) precipitated out and deposited on the ocean floor, with or without the

help of shell-forming organisms. Organic material is directly preserved in sedimentary rocks such as shales. For example, the organic material in the Isua shales is presumed to be of biological origin because its carbon is isotopically light – enriched in ${}^{12}C$ relative to the average composition of carbon on Earth. Respiration – eating organic matter and combining it with oxygen to release energy – does not fractionate carbon to any significant extent, so the isotopic signature imprinted by photosynthesis is carried over to those of us creatures in the non-photosynthetic organic realm.

The simplest kind of isotopic fractionation to characterize is equilibrium fractionation. To keep things concrete, we'll illustrate the concept using oxygen isotopes. Consider two physically or chemically distinct substances, each of which contains oxygen atoms. For example, the two substances could consist of water in its vapor phase and water in its liquid phase. Alternately, the two substances might be different chemical compounds containing oxygen, such as calcium carbonate ($CaCO_3$) and water (H_2O), or silica (SiO_2) and water; the latter pair leads to the chert-based paleothermometer mentioned earlier. Each of the two substances will have some initial ratio of ¹⁸O to ¹⁶O. Now imagine bringing the two substances into contact, whereupon the two substances exchange heavier and lighter forms of oxygen, changing the initial ratios. After a very long time, the isotopic ratios in each substance will reach equilibrium and stop changing. At that point, we can define the equilibrium fractionation factor $f_{1,2}$ via the relation

$$r_1 = f_{1,2} r_2 \tag{1.4}$$

where r_1 is the ratio of ¹⁸O to ¹⁶O in the first substance after equilibrium has been reached, and r_2 is the corresponding ratio for the second substance. The fractionation factors differ for different pairs of substances, but typically (though not invariably) exhibit the following characteristics:

- The fractionation factor is typically quite close to unity
- The fractionation factor deviates most from unity at low temperatures, and approaches unity as temperature increases
- The deviation of the fractionation factor from unity increases as the contrast between the masses of the two isototopes increases (e.g. more fractionation for ${}^{18}O$ vs ${}^{16}O$ than for ${}^{17}O$ vs ${}^{16}O$

For example, the oxygen in silica (SiO_2) is isotopically heavy in comparison with the oxygen in the water with which it is in equilibrium; the fractionation factor is 1.036 at 20*C* but falls to 1.018 at 100*C*. Similarly, oxygen in liquid water is isotopically heavy in comparison to that in the water vapor with which it is in equilbrium; in this case, the fractionation factor is 1.01 at 20*C* and falls to 1.005 at 100*C*. It is the temperature dependence of the fractionation that makes it possible to use isotopic ratios as paleothermometers. Some kinds of fractionation appear to operate in nearly the same way whether the reactions happen within organisms or inorganically. This appears to be the case for carbonate precipitation, which fractionates in much the same way regardless of whether it happens inorganically or in shell-forming organisms. Other forms of fractionation, such as the fractionation of carbon isotopes in photosynthesis, are more inherently biologically mediated, though even in such cases the fractionation factors tend to be similar across broad classes of organisms sharing the same biochemical pathways.

Isotopic composition is usually described using δ notation, which is defined as follows. Let r_A be the ratio of the number of molecules of isotope A in a sample to the number of molecules of the dominant isotope. Typically, r_A will be a rather small number. Next let $r_{A,S}$ be the isotopic ratio for a standard reference sample. Isotopic composition is invariably reported relative to a standard because the analytical instruments currently in use cannot measure the absolute composition to

very high accuracy, but they can measure the difference relative to a standard very accurately; the standard is an actual physical substance, natural or manufactured, which can be put into the analytical instrument and serve as a basis for comparison. The choice of standard is a matter of convention, and there are various standards typically used in different contexts. For example, the standard for oxygen and hydrogen isotopes in ice or water is usually taken to be VSMOW, which stands for Vienna Standard Mean Ocean Water. The ratio of ¹⁸O to ¹⁶O in VSMOW is 1/498.7, and of D to H is 1/6420; this approximates the mean present composition of water in the ocean.

Once one agrees upon a standard, the isotopic composition of a sample can be described in terms of the quantity

$$\delta A \equiv \frac{r_A - r_{A,S}}{r_A,S} \tag{1.5}$$

Thus, negative values of δ indicate that the sample is depleted in isotope A relative to the standard, whereas positive values indicate that the sample is enriched. The δ value is usually expressed as a *per mil*, or parts-per-thousand, value. For example, a δ value of .001 would usually be expressed as 1 *per mil* or 1‰. A difference of 1‰is often equivalent to a miniscule variation in isotopic concentration, requiring high analytical precision to measure. For example, for the case of $\delta^{18}O$, a 1‰difference is equivalent to changing the ratio of ¹⁸O molecules to ¹⁶O molecules by .001 $\cdot \frac{1}{498.7}$, or a hair over 2 parts per million. Deuterium (D) is even tougher. For deuterium, a difference of 1‰amounts to a change in the D to H ratio of only 0.16 parts per million, though the challenge is offset by the fact that fractionation in the D/H system is considerably stronger than fractionation in the ¹⁸O/¹⁶O system owing to the lesser relative mass contrast in the latter case.

Carbonate minerals (e.g. $CaCO_3$ or $MgCO_3$) are important recorders of the oxygen and carbon isotopic composition of past environments. For carbonates, the isotopic composition is usually reported relative to the *PDB* standard, named after a mineral powder made from a naturally occurring fossil carbonate (the "Peedee Belemnite") The physical powder standard no longer exists, so it has been supplanted by a synthetic equivalent, known as *VPDB*. It is important to keep the standards in mind when interpreting the isotopic literature. Oxygen isotopes occur in both carbonate and water, and so can be reported relative to either the *VSMOW* or *VPDB* standard. The conversion between the two is given by

$$\delta^{18}O(VSMOW) = 1.03091\delta^{18}O(VPDB) + 30.91 \tag{1.6}$$

A useful rule of thumb to keep in mind when interpreting oxygen isotopes in marine carbonates is that a carbonate reading zero relative to VPDB would be in equilibrium with water having zero $\delta^{18}O(VSMOW)$, if the carbonate formed at a temperature of around 18C. This means that a carbonate $\delta^{18}O(PDB)$ of "around" zero goes along with the waters in which they formed having $\delta^{18}O(VSMOW)$ of "around" zero. Later, we'll clarify just what we mean by "around," and how that depends on temperature.

Having introduced the δ notation and the VPDB standard, we are now in a position to get more quantitative about the things to be learned from the stable isotopes of carbon preserved in carbonates. The quantity of interest is $\delta^{13}C$, reported relative to the VPDB reference. Carbon dioxide is cooked out of carbonates in the interior of the Earth, and outgases from volcances, subduction zones and the mid-ocean ridge with $\delta^{13}C \approx -6\%$. In a steady state, the flux of carbon in this carbon dioxide is balanced by the burial of carbon in the form of inorganic carbonates (e.g. $CaCO_3$ and organic carbon (schematically CH_2O). In the long run, most of this burial is in sea-floor sediments, since whatever forms on land tends to eventually get washed into the ocean. Note that even carbonate precipitated biologically in the form of shells of organisms is considered inorganic material, and acts pretty much (though not exactly) like inorganically precipitated carbonate. There are some kinds of recently evolved plants that use photosynthetic pathways that

1.7. DIGRESSION: ABOUT CLIMATE PROXIES

fractionate carbon very little, but for the most part, photosynthetically produced organic carbon has $\delta^{13}C$ values that are about 25‰ lower than that of the carbon dioxide reservoir from which the photosynthetic organisms make their substance. For example, CO_2 in the atmosphere today has $\delta^{13}C \approx -8\%$, while land plants have $\delta^{13}C$ values running from -32‰ to -25‰ and marine organic carbon has $\delta^{13}C \approx -25\%$. Fossil fuels, which are made from ancient land plants (in the case of coal) or marine organisms (in the case of oil) have $\delta^{13}C \approx -25\%$.

If f_{org} is the fraction of carbon which is buried in the form of organic carbon, δ_o is the $\delta^{13}C$ of carbon outgassed from the Earth's interior, δ_{org} is the $\delta^{13}C$ of the organic carbon buried and δ_{carb} is the $\delta^{13}C$ of the carbonate carbon buried, then mass balance implies

$$\delta_o = f_{org} \delta_{org} + (1 - f_{org}) \delta_{carb} \tag{1.7}$$

If data from both organic and carbonate sediments are available, this formula can be used directly to infer f_{org} . It is instructive, however, to make use of the intrinsic fractionation between inorganic carbon and photosynthetically produced organic carbon to infer the isotopic compositions of the two burial fluxes as a function of f_{org} . Note that because photosynthetic fraction is relative to the composition of the inorganic carbon pool, it is incorrect to assume that $\delta_{org} \approx -25\%$. It could be considerably heavier, if the inorganic pool has very positive $\delta^{13}C$. Let's consider two limiting cases. Suppose that $f_{org} \approx 1$, so that nearly all of the carbon outgassed from the Earth's interior is intercepted by photosynthesis and buried as organic carbon. In that case, mass balance requires that $\delta_{org} \approx \delta_o \approx -6\%$. This can only happen if the inorganic carbon pool consists of isotopically heavy carbon with $\delta^{13}C \approx (-6+25)\% = 19\%$. Since the carbonate that precipitates from an inorganic carbon pool tends to be isotopically heavier than the pool itself, the trickle of carbonate precipitated in this situation will have $\delta^{13}C$ in excess of 19%. The precise value depends on aspects of ocean chemistry we will not pursue here. In the opposite limit, when $f_{org} \approx 0$ and there is little organic carbon burial, then $\delta_{carb} \approx \delta_o \approx -6\%$. The inorganic carbon pool this carbonate precipitates from is somewhat isotopically light compared to the carbonate itself, and the organic carbon fractionates relative to that value, yielding $\delta_{org} < (-6 - 25)\% = -31\%$. The typical situation over the past two billion years has been for δ_{carb} to be somewhat positive, between 2‰ and 5‰, while δ_{org} hovers around -22‰. There are periods, however, when carbon isotopes undergo considerable excursions from the typical situation. These carbon isotope excursions provide an important window into big events in the carbon cycle.

The above picture applies only when the carbon cycle is in a steady state. When any part of the carbon cycle is significantly out of equilibrium – for example, when one is building up a new pool of stored organic carbon in land plants and soils – the simple input-output isotopic calculation no longer works. One must then do a detailed accounting of the flows of carbon between the various reservoirs involved, and the attendent isotopic fractionations. This can be a very intricate process, especially since the fractionations involved are generally temperature dependent. There are other important aspects of the isotopic carbon cycle we have swept under the rug, such as the important information that can be gained from vertical gradients of $\delta^{13}C$ in carbonates.

There is another biological process that can leave a distinct mark on the carbon isotopic record, namely *methanogenesis*. When there is oxygen around, organic matter generally gets decomposed into CO_2 by respiration. In anaerobic environments, methanogens get the goodies instead, and turn organic matter into CH_4 . This is a multi-step process, each step of which fractionates carbon. The result is CH_4 which is much lighter isotopically than the organic feedstock from which it was produced. Biologically produced methane today has $\delta^{13}C$ values on the order of -50‰. When the atmosphere-ocean system is rich in oxygen, as has been the case for the past half billion years, methanogenesis plays a very minor role in the carbon cycle and usually leaves little imprint in the isotopic record. A possible exception to this general rule may occur as a result of gradual accumulation of large amounts of methane in the form of exotic ices called *clathrates*, which can form in ocean floor sediments and under arctic permafrost laters. If some event occurs which suddenly releases this stored methane into the atmosphere or ocean, the isotopically light methane quickly oxidizes into CO_2 which works its way into the carbonates. The net result is a negative carbonate carbon isotope excursion, the magnitude of which tells us something about the quantity of methane released relative to the net carbon in the ocean-atmosphere system.

At present, the arsenal of climate proxies is much more limited for other planets than it is for Earth. Biologically mediated proxies are obviously not in the cards for planets that seem to have no biology, but many of the abiological proxies used on Earth would be equally useful on other planets; cherts on Mars would provide much the same kind of information as cherts have provided on Earth. The use of these chemical proxies is limited primarily by the weight and power consumption of analytical instruments needed to carry out some of the analyses that are typically done on Earth materials. The same constraint applies to many of the means of determining chronology based on radioactive decay. Such techniques can be applied to other planets with a preserved geological record (notably Mars), but must await sample return missions. Meanwhile, a considerable amount has been accomplished by remote-sensing from planetary orbiters, and from low-power instrumentation on landers. The landforms of Mars have been imaged in great detail, and constitute a proxy for past climates with regard to occurence of water and glacial activity. The minerology of the surface can be determined by a range of remote-sensing techniques, so a fair amount is known about the occurrence of clay minerals (a signature of water and weathering) in the ancient crust of Mars, and other minerals such as the iron compound hematite also tell us something about the aqueous environment of Early Mars. On all planets, a study of the isotopic composition of atmospheric gases (possible by *in situ* and remote spectroscopic means) provides valuable information on the source of the atmosphere and how much has been lost over time, insofar as lighter isotopes escape to space more readily than heavy isotopes. With sample return missions and improved robotic exploration, the future promises a rich expansion in planetary proxy studies, not least the prospect of drilling the Martian polar glaciers to see what climate mysteries they record.

1.7.3 Hydrogen and Oxygen isotopes in sea water and marine sediments

We will turn our attention now to the isotopes of hydrogen and oxygen contained in water and in sediments precipitated from the water column. We'll learn what the concentration of these isotopes tells us about the volume of glacier ice and the temperature of various parts of the ocean. "Normal" water is $H_2^{16}O$, but other isotopes of hydrogen or oxygen can substitute for the most prevalent isotopes, leading to various forms of heavy water, notably $HD^{16}O$ and $H_2^{18}O$.

At any given temperature light molecules, on average, travel with higher velocity than heavier molecules. This implies that during evaporation, the lighter isotopic forms of water evaporate more readily than the heavier forms, leading the vapor to be enriched in light water and depleted in heavy water relative to the liquid reservoir. Furthermore, when water vapor condenses into liquid or ice, the heavier forms condense more readily because the slower moving molecules can more easily stick together without bouncing off each other. In consequence, the rainfall is enriched in heavy forms relative to the vapor in the air, while the vapor left behind in the air is further enriched in the light forms and further depleted in the heavy forms. This is a form of distillation, very similar to the process by which one makes brandy from wine (or moonshine from fermented corn mash). Alcohol is more volatile than water, so the vapor in contact with heated wine is enriched in alcohol relative to the liquid; if part of the vapor cools and condenses, the water condenses out preferentially, leaving a potent essence at the end of the still if the remaining vapor is

1.7. DIGRESSION: ABOUT CLIMATE PROXIES



Figure 1.5: Sketch showing how the growth of ice sheets on land affects the isotopic composition of ocean water. The water vapor which evaporates from the ocean is enriched in lighter forms of water, and becomes more isotopically light as the heavy forms of water preferentially rain or snow out before the remainder is deposited on the glacier. This process systematically transfers isotopically light water to the glacier, leaving the ocean isotopically heavy.

then condensed separately.

The way the distillation process affects the isotopic composition of sea water is sketched in Fig. 1.5. Let's suppose that the ocean starts with $\delta^{18}O$ of zero relative to VSMOW. Water will evaporate into the air until the air becomes saturated with water vapor (a concept that will be made precise in Chapter 2). Since heavy molecules evaporate less readily than light molecules, the water vapor will be depleted in ${}^{18}O$ relative to the ocean – in other words, it will be *isotopically light.* More precisely, the ratio of ${}^{18}O$ to ${}^{16}O$ for the water vapor in the air will be less than the ratio of the original ocean water, leading to a negative $\delta^{18}O$ for the vapor. The amount of depletion depends weakly on temperature. At 273K, the water vapor $\delta^{18}O$ is shifted by -11.7 % relative to the ocean. At 290K the shift is -10.1% and at 350K the shift is -6.0%. This reduction in isotopic contrast between reservoirs as temperature increases is typical of almost all isotopic fractionation problems. Because the amount of water stored in the form of water vapor in the atmosphere is utterly dwarfed by the amount of water in the ocean, the selective removal of light isotopes makes the ocean only very, very slightly isotopically heavy. But what if we removed the atmosphere's water vapor, sequestered it in a glacier on land, and repeated the process many times over until a substantial fraction of the ocean water had been transformed into an isotopically light glacier? In that case, the systematic removal of large volumes of isotopically light water from the ocean would leave the ocean water isotopically heavy by a significant amount – it would have a significantly positive $\delta^{18}O$. Thus, the degree of to which the ocean is enriched in isotopically heavy forms of water tells us how much ice has built up on land. As ice volume becomes greater, the $\delta^{18}O$ (or similarly, δD) of the remaining ocean water becomes more positive. As an example, let's suppose we build an ice sheet by removing 200*m* depth of water from an ocean with a mean depth of 4km, assuming the glacier to be built from vapor with $\delta^{18}O = -10\%$. If δ_i is the $\delta^{18}O$ of the ice and δ_o is that of the ocean water, then conservation of molecules implies that $200\delta_i + (4000 - 200)\delta_o = 0$, if the ocean started with $\delta^{18}O = 0$. From this we conclude $\delta_o = .526\%$.

In fact, for the reasons sketched in Fig. 1.5, the water that eventually snows out to form glaciers is much more isotopically light than the -10% value one might expect from just looking at the vapor in equilibrium with ocean water. The initially evaporated water vapor may have $\delta^{18}O =$ -10%, but on the way to the cold polar regions, some of that water will rain out back into the ocean, and the condensed water is isotopically heavy relative to the vapor, since heavy species condense more readily. That means that each time some atmospheric water vapor is lost to rainfall or snowfall back into the ocean, the vapor left behind becomes lighter. The precise extent of the additional lightening by the time the snow eventually falls out on a glacier depends on the amount of water lost on the way, which is in turn a function of the temperature difference between where the water was picked up from the ocean and where it was dropped on the glacier. Over the past 100,000 years, the $\delta^{18}O$ of the snow falling on Greenland has varied from -42 ‰in the coldest times to values around -35% today. Antarctic ice is somewhat isotopically lighter than Greenland ice, and has $\delta^{18}O$ values ranging from -40% to -55% depending on location and age of the ice. Because of the additional fractionation on the way to the pole, the formation of the glacier would leave the ocean much more enriched in heavy isotopes than our previous estimate suggested. For glaciers having isotopic compositions comparable to the present ones, removing 200m of ocean to build glaciers would leave the ocean enriched by about +2%, rather than a mere .526\%. The preceding discussion also shows that in order to translate the δ value of the ocean into an ice volume, one needs some estimate of the isotopic composition of the glaciers being formed. For the present glaciers, this can be determined by drilling into the ice, but for past ice that no longer exists one must rely on modelling.

So, if we could go back in time and grab a bucket of sea water, measuring its isotopic composition would tell us the volume of ice on the Earth, and this would tell us much about how cold the planet was. Wouldn't it be awfully nice if there were some way to do that?

1.7.4 Forams to the rescue

As it happens, Nature has provided a handy way of determining the isotopic composition of past ocean waters, via the good works of single-celled shelly amoeba-like organisms known as foramanifera, nicknamed forams (see Fig. 1.6). These creatures build distinctive calcium carbonate $(CaCO_3)$ shells which record the state of the water in which they grew. Because the shells have such diverse and unmistakable shapes, it is easy to recognize and select out the species which live at the the depth level one wishes to investigate. The two principle types of forams are *benthic* which live near the sea floor, and *planktonic* which require light and live near the ocean surface. The shells of both types wind up in tidy layers in the marine sediments, allowing one to read the state of the ocean in both depth and time, from a single sediment core. Benthic forams appear in the fossil record as early as 525 million years ago, but their use as paleoclimate indicators has been primarily restricted to the period over which significant portions of seafloor have survived – approximately the past 70 million years.

1.7. DIGRESSION: ABOUT CLIMATE PROXIES



Figure 1.6: The shell of a departed benchic foram (*Cibicidoides Robertsonianus*. The specimen is about one half millimeter in diameter.

Forams are not just passive recorders of the isotopic composition of ocean waters, however. As is the case for any chemically distinct pair of reservoirs in contact, the oxygen in foram carbonate is systematically fractionated relative to the isotopic composition of sea water. There is some dispute as to the extent to which this fractionation can be thought of as equilibrium fractionation, but the fractionation factors behave much like inorganic equilibrium fractionation factors and do not differ greatly amongst species. Carbonate prefers the heavier forms of oxygen, and at a temperature of 18C, the ${}^{18}O$ to ${}^{16}O$ of precipitated carbonate is greater than the ratio for the water in which it precipitates by a factor of about 1.03. In other words, carbonate is about 30% enriched in ¹⁸O compared to the water with which it is in equilibrium. As is typically the case for equilibrium fractionation, the degree of enrichment increases as temperature decreases. The change in fractionation with temperature is usually expressed as a *paleotemperature equation*. Many paleotemperature equations have been given, based on laboratory-cultured organisms, on field observations of recently living forams, and on laboratory measurements of inorganic carbonate precipitation. All give similar results, though the differences are important if one is interested in high accuracy. A general feel for the numbers is adequately given by the following paleotemperature equation, which applies to the benthic foram Uvigerina.

$$T = 17.97 - 4.0 \cdot \left(\delta_c(VPDB) - \delta_w(VSMOW)\right) \tag{1.8}$$

where T is the temperature in degrees C at which the foram grew, $\delta_c(VPDB)$ is the measured $\delta^{18}O$ of the foram carbonate reported relative to VPDB, and $\delta_w(VSMOW)$ is the $\delta^{18}O$ of the water in which the foram grew, reported relative to VSMOW. Both δ values in the above equation are to be expressed in permil units.

The temperature dependence of foram fractionation is a two-edged sword. On the one hand, the temperature dependence allows forams to be used as paleothermometers. On the other hand, the temperature dependence means it is hard to disentangle ice-volume effects from temperature effects. According to Eq. 1.8, a $\delta^{18}O$ variation in foram carbonates of 2‰ could represent an temperature change of 8K where the forams grew, or instead a change of 2‰ in the water in which the forams grew – corresponding to an ice volume change equivalent to roughly 200m of sea level. The use of benthic forams mitigates this ambiguity to some extent, since the deep ocean temperature is much more uniform than surface temperature. This is so because the deep ocean is filled with waters that are created at the coldest parts of the surface ocean, typically located near the poles. When the climate is in a state having ice at one or both poles, this temperature hovers around the freezing point of sea water, whence the benthic oxygen isotopes primarily reflect ice volume rather than temperature – though the temperature effect is still by no means negligible. A 2K variation in deep ocean temperature, which is not implausible even in icy conditions, leads to about a 0.5‰ variation in the $\delta^{18}O$ of carbonates, which if attributed instead to sea water composition would translate into an ice volume variation equivalent to about 50m of sea level. At the other extreme, when the climate is in a state without ice at either pole, the isotopic composition of ocean water itself can be considered nearly fixed, and the benthic foram isotopes provide an indication of the polar temperature, regardless of where the sediment core is actually drilled. In this case, a 2‰ increase in benthic foram $\delta^{18}O$ would indicate an 8K warming of polar temperature.

Benthic forams thus provide a valuable overall indicator of the state of the climate, giving an indication of polar minimum temperature in ice-free climates, and ice volume in icy climates. Since greater ice volume generally goes with a colder climate, and both high ice volume and low temperature make the $\delta^{18}O$ of foram carbonate more positive, high benthic foram $\delta^{18}O$ indicates a cold climate while low benthic foram $\delta^{18}O$ goes along with a relatively warm climate. In icefree climates, planktonic forams can be used to estimate surface temperature, but in icy climates the need to subtract out the ice volume effect makes it hard to get accurate surface temperature estimates by this means.

Forams also preserve other chemical signatures that are useful in reconstructing the past state of the climate system. Notably, they provide a record of the $\delta^{13}C$ of inorganic carbon of the ocean. Abiologically precipitated sea floor carbonates, or carbonates not associated with individual microfossils, also do this, but the additional depth information available by using benthic vs. planktonic forams provides valuable information about the state of the oceanic carbon cycle. The use of fractionation as a paleothermometer is not limited to isotopes. Notably, magnesium (Mg)substitutes for calcium (Ca) in foram shell carbonates, to an extent that depends on temperature; hence the Mg to Ca ratio in foram shells can be used as a paleothermometer ⁴. As with oxygen isotopes, the fractionation is relative to the composition of sea water, but the Mg to Ca ratio of ocean water is not affected by formation of glaciers, and hence evolves relatively slowly over geologic time. For this reason, magnesium-calcium paleothermometry has become a crucial tool for probing the past few million years of climate history, and the tool is being extended farther backward as understanding of the longer term evolution of the oceanic Mg to Ca ratio improves.

1.8 The Proterozoic climate revisited: Snowball Earth

With a few more tools at our disposal, we once more pick up the thread of Earth's climate history, starting with a more detailed look at certain aspects of the Proterozoic – the geological eon that extends from 2.5 billion years ago to 543 million years ago, and is subdivided into the Paleo- Meso- and Neo- Proterozoic, in order of decreasing age. From now on, we will have increasing need to refer to the various subdivisions of the geological time scale by name, so these are summarized in Fig. 1.7.

⁴Magnesium-calcium paleothermometry can also be used with cores drilled into corals. The growth zone of corals has a distinct depth preference, which has proved particularly useful in estimating ocean surface temperature



Figure 1.7: The geological time scale. Numbers on the scales represent millions of years before present. The entire span of time before the Phanerozoic is also known as the Precambrian.

So far we have not said much about continents and continental drift. Continents consist of light material that has floated to the outer portions of the solid Earth and been incorporated into the crust. This material is too buoyant to be subducted into the interior of the planet to any significant degree, so once continental material had completely segregated, it remained at the surface as a kind of scum resting atop the churning cauldron of Earth's interior fluid motions. The continental material is constantly being pushed around, broken up and re-arranged in a process known as *continental drift*. The Earth is the only planet in the Solar System that has continents in this sense, but it is likely that extrasolar rocky planets of a sufficient size to retain the heat that drives internal motions, and having a surface temperature similar to Earth, would also exhibit the dichotomy of drifting continents vs. areas of rapidly recycled mantle material (analogous to Earth's sea floors). It is at present unresolved whether a water ocean is necessary to maintaining this state of affairs. That is indeed one of the Big Questions of planetary science, but it is not one which we will take up in this book.

Continents are important to climate for three main reasons: They are a platform upon which polar glaciers can form; They are the primary sites of the silicate weathering reaction that governs atmospheric CO_2 , and the amount of weathering is strongly affected by the continental configuration; They affect the geometry of ocean basins, and hence the ability of oceans to transport heat from one latitude to another. In addition, they provide distinct habitat for novel forms of life, though they were not colonized to any great degree (if at all) until land plants evolved in the late Ordovician and early Silurian.

There is not enough preserved continental crust to get a clear idea of the distribution of the continents until the very end of the Proterozoic. Geophysical modelling and indirect geochemical evidence has lead to a prevailing belief that the total volume of continental material was similar to the present volume throughout the Proterozoic, but this is not a very well settled area of geophysics. By the close of the Proterozoic, however, the picture of the distribution of continents begins to clarify; we will begin showing maps of paleogeography when we come to discuss that time.

The Proterozoic was the Age of Microbes. Indeed, in terms of the functioning of the biogeochemical cycles needed to sustain life, we are all still basically the guests of the prokaryotes, but up until the very end of the Proterozoic single-celled organisms had the world to themselves, with the more complex eukaryotic form of microbial life making its appearance roughly midway through the Proterozoic. The Proterozoic was above all a time of adjustment of biosphere and climate to the massive changes wrought by oxygenation of the atmosphere and ocean. It is a matter of current debate as to whether the oxygenation began in this eon because oxygenic photosynthesis only evolved at this time, or because other factors kept photosynthetic oxygen from accumulating earlier; be that as it may, the oxygenation occurred in fits and starts throughout the Proterozoic, and atmospheric oxygen levels probably reached values comparable to modern ones by the close of the eon. One effect of oxygenation we have already noted is that it is likely to have reduced the importance of CH_4 as a greenhouse gas, with CO_2 (aided by water vapor) becoming increasingly dominant. More speculatively, oxygenation could have changed the abundance of other greenhouse gases that could conceivably have been significant in the anoxic atmosphere, such as N_2O and SO_2 . The nature of this greenhouse turnover, and the extent to which it constituted a habitability crisis for our planet, is another of the Big Questions. In Chapter 4 we'll learn how to evaluate the relative importance of the various greenhouse gases. Oxygenation would have also sharply limited the possibility for H_2 to accumulate in the atmosphere, which would have consequences for methanogenic ecosystems that used H_2 and CO_2 as a feedstock.

Oxygenic photosynthesis turns CO_2 into O_2 and organic carbon, represented schematically as CH_2O . In order for the oxygen to accumulate, the organic carbon produced by photosynthesis must be buried before it can be re-oxidized by other bacteria, which would just take the free oxygen back out of the system and turn it back into CO_2 . For this reason, the evolution of oxygen is intimately tied in with the carbon cycle. Since organic carbon is isotopically light compared to the carbon in CO_2 outgassed from the interior of the Earth, the long term evolution of $\delta^{13}C$ in carbonates gives us a window into the carbon cycle, and a good overview of what is going on in the Proterozoic. As noted earlier, when the carbon cycle is approximately in a steady state, the $\delta^{13}C$ of carbonate is driven more positive as the proportion of organic carbon burial to total carbon burial increases. One must be cautious about this interpretation, however, since the carbon cycle is thrown wildly out of equilibrium in the course of the Snowball episodes that constitute the most dramatic features of Proterozoic climate.

With regard to oxygenation, however, organic carbon burial is not the whole story. Various compounds of sulfur also play a key role in the cycling of oxygen through the Earth system; there is evidence that the role of the sulfur cycle is much more prominent in the Proterozoic than during later times. Bacteria are clever, and have ways of oxidizing organic matter that don't use up free O_2 . In a process called *sulfate reduction*, certain bacteria can react the sulfate ion (SO_4^{--}) with organic carbon to produce bicarbonate (HCO_3^{--}) and the stinky rotten-egg gas hydrogen sulfide (H_2S) , which reacts with iron oxides and a little bit of free oxygen to produce water and the mineral pyrite (FeS_2) . If the pyrite is then buried without being oxidized further, the net process turns organic matter into mineral carbonate while leaving much of the O_2 liberated by oxygenic photosynthesis in the atmosphere/ocean system. Precisely how much is left in the atmosphere and ocean depends on where the oxygen in sulfate and iron oxides comes from, but the net result is that pyrite burial liberates oxygen in a way that doesn't show up in the carbon isotope record.

The participation of sulfur in a variety of reactions involving oxygen makes the stable isotopes of sulfur $-{}^{32}S$, ${}^{33}S$, ${}^{34}S$ and ${}^{36}S$, of which the first is dominant – a key source of information about past behavior of oxygen. By comparing the degree of fractionation for the three minor isotopes, one can get additional information. One form of fractionation is *mass-independent fractionation* (MIF), which is nearly the same for all three minor isotopes. This kind of fractionation is believed to be produced only in photochemical reactions involving high energy ultraviolet light, and photochemical models of the atmosphere indicate that mass-independent sulfur fractionation can't be preserved in the sedimentary record unless the atmospheric oxygen concentration is extremely low – 10^{-5} of the present concentration or less, according to current estimates. It is the sulfur MIF proxy that tells us that Archaean oxygen are nearly zero; other oxygen proxies only require that Archaean oxygen be below 1% of present atmospheric concentration 5 .

The more conventional mass-dependent sulfur fractionation is mediated by sulfate-reducing bacteria. The fractionation nearly disappears when sulfate concentration in ocean water is low, so a strong mass-dependent sulfur fractionation indicates both high sulfate concentration and strong productivity of sulfate-reducing bacteria. An increase in sulfate concentration, in turn, is generally taken as indicative of a rise in atmospheric oxygen, since that permits more oxidation of pyrite into sulfate on land. Once oxygen builds up to the point that at least near-shore bottom waters become oxygenated, a host of additional bacterially-mediated sulfur reactions, called *disproportionation reactions* become possible, and these provide additional means of producing sedimentary sulfides (e.g. pyrite) that are isotopically light in sulfur. The interpretation of mass-dependent sulfur isotope fractionation is an exceedingly complex subject, which is likely to remain in a considerable state of flux for some time to come.

The proxy record shows a great deal of activity towards the beginning of the Proterozoic (during the *Paleoproterozoic*), and also towards the end of the Proterozoic (in the later parts of the

 $^{^{5}}$ The atmospheric chemistry models upon which this interpretation of the sulfur MIF is based rest on somewhat shaky assumptions, however.

Neoproterozoic). In between lies a billion-year period that has sometimes been called "the most boring period in Earth history." During this Big Yawn, which stretched from about 1.8 billion years ago to 800 million years ago, carbonate $\delta^{13}C$ held steady near 0%, indicating steady organic carbon burial. Mass dependent sulfur isotope fractionation implies oxygen levels of around 10% modern concentrations during this period. There is no indication of any significant glaciation. Eukaryotes arose towards the beginning of the Big Yawn, but appeared to have little effect on biogeochemical cycling or climate evolution – unless perhaps they were somehow the cause of the long period of climate stability. It would not be surprising if closer study eventually revealed more features of interest in this period, but at this point we'll turn our attention to the more manifestly dramatic doings at the beginning and end of the Proterozoic.

All lines of evidence point to a Great Oxidation Event at the beginning of the Proterozoic. Preservation of mass-independent sulfur fractionation in sediments ceases abruptly between 2.45 and 2.3 billion years ago, and is never seen again throughout the rest of Earth history. A hiatus in banded iron formations begins at about this time. It is estimated that the oxygen content of the atmosphere soared to values well in excess of 1% of the present level but crashed back to lower levels afterwards, as witnessed by a transient reappearance of banded iron formations; the peak value in the event is at present not well constrained. A subsequent increase in mass-dependent sulfur isotope fractionation preserved in the sediments is indicative of an increase in sulfate concentration in the ocean, most likely associated with an increase of oxidation of pyrite on land. Based on this evidence and disappearance of banded iron formations, it is estimated that oxygen levels in the atmosphere recovered to somewhere around 10% of present atmospheric concentrations around 1.7 billion years ago, and stayed there until 700 million years ago when there was a further oxygenation event.

The Paleoproterozoic is characterized by wild swings in the $\delta^{13}C$ of carbonates. Around the time of the Great Oxygenation event, $\delta^{13}C$ has a major positive excursion, reaching values as high as 10 ‰ before eventually subsiding to the lower values characteristic of the middle Proterozoic. This indicates a major transition in the carbon cycle, most likely an increase in the proportion of organic carbon burial. It is very suggestive of a take-off of oxygenic photosynthesis, but whether the cause is evolutionary, ecological or a matter of factors that allow better burial of carbon is a matter of dispute. There are several major glaciations within the Paleoproterozoic, of which one – the Makganyene alluded to earlier – was a Snowball event in which ice reached tropical latitudes. The Makganyene Snowball occurred within the interval between 2.32 and 2.22 billion years ago, and fine-scale examination of arguably synchronous glacial deposits in the Duitschland formation (Transvaal, S. Africa) indicates extreme carbon isotope excursions associated with major Paleoproterozoic glaciations: Carbonate $\delta^{13}C$ was around 5 ‰ before the glaciation, then dropped to zero or even negative values as the glaciation progressed, recovering slowly afterwards. We'll be able to probe similar features in more detail in connection with the Neoproterozoic Snowballs. The Paleoproterozoic presents us with a puzzle whose pieces include oxygen, the effect of oxygen on greenhouse gases, the carbon cycle, and glaciation. Figuring out how these pieces fit together is one of the Big Questions.

The Neoproterozoic has many features in common with the Paleoproterozoic. The extreme carbonate carbon isotope excursions which had been dormant for so long resume in the Neoproterozoic. There are several major glaciations during the Neoproterozoic, and two of these were Snowball events in which ice reached tropical latitudes. The more recent of the two Snowballs is the Marinoan event, which occurred about 640 million years ago; the older is the Sturtian, centered on 710 million years ago. Neoproterozoic Snowball-related geological formations exhibit a distinctive sequence of events. The scene starts with high carbonate $\delta^{13}C$, up to 5‰, which is in fact higher than the modern value and indicative of a greater proportion of organic carbon burial than is the case at present. Then, the $\delta^{13}C$ drops, falling to zero or even negative values. At some point in this decline, one sees diamictites and other glacial deposits. The $\delta^{13}C$ continues to drop, and above the glacial deposits one finds *cap carbonates* – very unusual carbonate features that are believed to require very high deposition rates from waters highly supersaturated in carbonate. In the carbonates overlying the glacial deposit, the $\delta^{13}C$ becomes negative, typically aroung -5‰ which is about the value for abiotic carbon outgassing from the Earth's interior. The $\delta^{13}C$ gradually recovers to positive values over a long (but somewhat unconstrained) period of time afterwards. A particularly clear depiction of this sequence of events is given in the review by Hoffman and Schrag listed in the Further Readings for this chapter.

However, not all major carbon isotope excursions are associated with Snowball events. In fact, the greatest carbon isotope excursion in Earth history – the *Shuram excursion* – sets in gradually after a conventional glaciation which is thought to reach only to midlatitudes (the *Gaskiers* glaciation). The Shuram excursion brings the carbonate $\delta^{13}C$ all the way down to -12 ‰. There is no known process that could bring the carbonate $\delta^{13}C$ so far below the mantle outgassing value if the carbon cycle is in equilibrium. Indeed, the $\delta^{13}C$ is so implausibly low in the Shuram that it was long thought to be an artifact of diagenetic alteration. The Shuram is an enigmatic event – indeed one of the Big Questions. Current thinking has it that the Shuram is associated with a transient reorganization of the carbon cycle, in which a large isotopically light pool of suspended organic carbon in the ocean is oxidized and deposited as carbonate.

In fact, a lot is going on with oxygen across the Neoproterozoic, though it is a bit hard to determine what, where and when. What is certain is that oxygen must have been high – even near present levels – right down to the ocean bottom by the end of the Neoproterozoic, since bottom-dwelling animals appear in the fossil record by this time, and it is unquestionable that such creatures require a great deal of oxygen. At the other side of the Neoproterozoic, around 700 million years ago, there is further evidence of oxygenation, in that a sharp rise in mass dependent fractionation in sedimentary sulfides indicates the expression of sulfur disproportionation reactions, which indicates an oxygenation of at least some of the bottom waters. Another important clue as to what is going on is the reappearance of banded iron formations in connection with the Neoproterozoic Snowball events, suggesting that the ocean once more went anoxic, most plausibly as a result of global ice cover shutting down photosynthesis.

A very Big Question is why all this excitement suddenly resumed after nearly a billion years of stasis.

The Snowball events of the early and late Proterozoic are some of the most dramatic events of Earth history. We have used the term "Snowball" to refer to any glaciation where there is evidence of glaciation at tropical latitudes, but it is a matter of considerable debate whether the oceans were indeed nearly completely frozen over all the way to the equator during these events. Sometimes the term *Hard Snowball* is used to refer specifically to a state with near-total ice cover. A Big Question of climate physics is whether it is possible to cool down the planet enough to yield land-based ice sheets discharging into the tropics, without also freezing over the tropical ocean completely. This requres an understanding of ice-albedo feedback, which will be developed at several places throughout the book. However, it also involves ocean heat transports (which are good at melting ice) and glacier dynamics, which for the most part are subjects that will be left for another time and place.

The Snowball phenomenon is pregnant with Big Questions, the most obvious of which are: How do you get in? And how do you get out? And if your planet does succumb to a global Snowball, how long does it take to get out again? Is it a matter of centuries, millions of years or billions of years? On Earth, the upper limit set by the geological record for the duration of Neoproterozoic snowballs is about 20 million years, and the duration could well have been shorter. However, without a clear understanding of the nature of the event, it is hard to determine whether we just got lucky or whether the event could have lasted much longer.

Most theories for the entry into a Snowball involve the drawdown of whatever greenhouse gas had previously been maintaining the planet's warmth – usually CO_2 and CH_4 in some combination. Various hypotheses include methane destruction by oxygen, weathering enhancement triggered by catastrophic methane release from sediments, weathering enhancement due to continental configuration or production of weatherable rock by massive volcanic eruptions, and (more speculatively) drawdown of CO_2 through runaway photosynthesis and oxygenation. Whatever the mechanism, a key requirement is that the mechanism be compatible with the observed reduction in $\delta^{13}C$ before the onset of glaciation. Not all of the relevant biogeochemistry will be treated in this book, but in order to evaluate the hypotheses, it is certainly necessary that one have the tools to assess how low any given greenhouse gas has to go in order to trigger a global glaciation. These tools will be provided in subsequent chapters.

Assuming for the moment that the cooling process caused a Hard Snowball, the next question is how to deglaciate the planet. We'll see in Chapter 3 that one would have to wait a billion years or more to exit from a Snowball if the exit were due to increase in solar output alone. Based on rather simple reasoning of the sort that will be covered in the remainder of this book, Kirschvink proposed that once the Earth freezes over, the weathering of silicate to carbonate (which requires liquid water washing over weatherable rocks) ceases, so that CO_2 outgassed from the Earth's interior accumulates in the atmosphere until it reaches concentrations sufficient to cause a deglaciation. This is another illustration of the principle that *Big Ideas come from Simple Models*. A Big Question (treated in subsequent chapters) is: how high does CO_2 have to go in order to trigger deglaciation of a globally ice-covered planet? Much hangs on the answer.

Both cap carbonates and the persistent negative carbon isotope excursion following the Snowball events are consistent with a massive buildup of CO_2 in the atmosphere during the frozenover period. Once the planet gets warm enough to deglaciate, the powerful precipitation in the ensuing hothouse world would wash great quantities of land carbonates into the ocean, where they would precipitate to form cap carbonates. Further, if photosynthesis nearly shuts off during the glaciated phase, the inorganic carbon that accumulates in the ocean-atmosphere reservoir would have $\delta^{13}C$ comparable to the mantle outgassing value of about -6%. As this reservoir is gradually transformed by silicate weathering into carbonate sediments, the isotopically light carbon works its way into the carbonates. If the reservoir is big enough at the termination of the snowball, it can keep the sedimentary carbonate $\delta^{13}C$ light even in the face of a resumption of photosynthesis. When interpreting the carbon isotope excursions in the course of a Snowball, it is essential to keep in mind that the carbon cycle is likely to be far out of equilibrium in the course of these events. A full understanding of the connection between the carbon isotope evolution and the sequence of events surrounding the Snowball requires a detailed accounting of flows of carbon between the many different carbon reservoirs in the Earth system – land carbonate rocks, marine carbonate sediments, atmospheric carbon dioxide, various species of dissolved inorganic carbon, and organic carbon.

Assuming that the exit from a Snowball state does indeed proceed from accumulation of a great deal of CO_2 in the atmosphere, several Big Questions arise in connection with the post-glacial climate. Is there a risk of triggering a runaway greenhouse? If not, how hot does the climate get in the tropics and polar regions? Would it be hot enough to sterilize the planet to any great degree? How long is the post-snowball recovery? In other words, how long does it take for weathering processes to draw the CO_2 back down to more normal levels?

1.9. THE HOTHOUSE/ICEHOUSE DICHOTOMY

Other Big Questions include: Why were there no Snowball event during the Big Yawn period of the middlle Proterozoic? Why did Snowballl events cease at the beginning of the Phanerozoic? Could they happen again, or is this particular threat behind us?

But let us not forget that another Big Question is whether there is a climate state with significant amounts of open water in the Tropics, which is nevertheless consistent with the full range of geological data accompanying the Marinoan and Sturtian events. One could pose the same question for the Makganyene event, but there is less data to constrain the answer. The early and late Proterozoic manifests a lot of climate "weirdness" that is not seen elsewhere in Earth history, and such striking signatures would seem to call for an equally dramatic cause, rather than just a minor variation on the theme of ice ages. The global Snowball seems to fit the bill, but it remains to be seen whether other explanations are possible. The climate physics developed throughout this book will give the reader the underpinning needed to assess hypotheses as they develop, and even perhaps to formulate new ones. fits the bill

Regardless of whether a true global Snowball glaciation ever happened on Earth, the Snowball certainly represents a state a water-covered planet could fall into if the right stellar and atmospheric conditions are encountered. Once a planet falls into a Snowball state, it is likely to stay there for a long time, and the consequences for existing life and evolution of new forms of life are profound. As such, Snowball states are a potential habitability crisis that extrasolar planets need to avoid or surmount. It is therefore worth understanding, in general terms, the physics of entry into, exit from, and duration of Snowball states, as well as the nature of the climate at various stages of the sequence and the effect of the sequence on life. This constitutes another Big Question, about which will have much to say.

1.9 The hothouse/icehouse dichotomy

The present climate has ice at both poles, and the ice volume has fluctuated episodically between the present amount and a considerably larger extent for the past two million years (about which more anon). However, the present icy climate is not at all typical of Earth history. A careful study of the climate evolution over the past seventy million years illustrates a transition between climate states archetypical of a theme that has been played over with variations during the 543 million years since the close of the Proterozoic. With this latest eon, known as the *Phanerozoic*, we complete the repertoire of the major divisions of geologic time, as summarized in Fig. 1.7. Though the very first preserved multicellular organisms appear at the close of the Proterozoic, the Phanerozoic is the eon in which multicellular organisms of a generally modern form become abundant and diversify, first in the ocean with colonization of land coming towards the middle of the eon. Though the Phanerozoic was not subjected to extreme variations of climate and atmospheric composition rivaling the Snowball or oxygenation transitions of earlier eons, the events of the Phanerozoic are by no means inconsequential.

1.9.1 The past 70 million years

Figure 1.8 shows the paleogeography at the end of the Cretaceous, 65 million years ago. The continent of Antarctica has approached the South Pole, and will continue to drift over the next 40 million years or so until it is more nearly centered on the pole. There is open water at the North Pole in the late Cretaceous, and the open Arctic Ocean continues throughout the subsequent time through the present. The modern continents of North and South America, Eurasia, and Africa



Figure 1.8: Position of the continents at the end of the Cretaceous, 65 million years ago. Continents are shown on a Mollweide map projection, with the North Pole at the top and the South Pole at the bottom. Light areas are continents while dark grey areas are ocean.

are still early in their separation, leaving a narrow Atlantic ocean and a very broad Pacific. The continents will continue drift apart as they approach their modern configuration; the Atlantic widens steadily throughout the span of time under discussion.

The record of benthic foram $\delta^{18}O$ in Fig. 1.9 provides a good overview of the climate evolution for the past 70 million years. Towards the beginning of the period, $\delta^{18}O$ is considerably lower than it is in the modern ocean. It reaches a minimum value of -0.1‰ around 51 million years ago. Independent geological evidence shows that there was no significant amount of ice sequestered in land glaciers until 36 million years ago, so the variations in $\delta^{18}O$ before this time can be interpreted as polar ocean temperature changes. Melting the present ice in Antarctica and Greenland would leave the ocean with a $\delta^{18}O$ of around -.7‰ relative to VSMOW. Plugging this into the paleotemperature equation, we estimate a high-latitude ocean temperature of 15.5*C*. This estimate applies to the coldest seasonal temperature attained either in the Arctic ocean or in the waters surrounding Antarctica. This is well above the freezing point of sea water, and so we conclude that the oceans were ice-free year round, even in the Arctic and Antarctic regions which are very cold in the modern climate. We'll refer to this kind of climate state as a *hothouse climate*. The late Cretaceous polar temperatures were about 4C cooler than those at the peak warmth, but still warm enough to guarantee ice-free conditions.

Other lines of evidence also support warm Northern high-latitude conditions and abovefreezing winter conditions. In the early twenty-first century, the first useful deep-time Arctic marine cores were recovered, and TeX-86 proxies applied to these cores indicated Arctic ocean up to 22C during the time of the spike marked *PETM* in Fig. 1.9, with temperatures in the range of 17-18*C* before and after. Fossil vegetation from Arctic land also supports temperatures in this range. Moreover, the abundant evidence that lemurs and crocodiles were able to survive in high Northern latitudes points toward mild winters, since these creatures cannot survive sub-freezing temperatures for any significant length of time.

While evidence for warm, ice-free polar conditions in the Eocene and late Cretaceous is unambigous, the nature of the tropical climate is somewhat problematic. Up until the year 2001, most



Figure 1.9: A composite record of benthic foram $\delta^{18}O$ (vs. PDB) over the past 70 million years, based on the average of several marine sediment cores. Data is from Zachos *et al.* 2001.

paleoceanographers would have said, based on planktonic for $1^{8}O$ that the Cretaceous tropical sea surface was no more than two or three degrees warmer than present, and the Eocene tropical sea surface was no warmer than today, and might even have been cooler. This posed the paradox of the "low gradient" climate – how to warm up the planet enough to prevent polar ice, without frying the tropics. In 2001, it was discovered that most of the evidence for a cool tropics was spurious, having been affected by diagetic alteration of sediments. The surviving non-altered data indicated a warmer tropics, but there was precious little data left after the diagenetically contaminated data was discarded. Gradually, new sediments and new proxies have come to the fore, and the story continues to develop. Tex-86 proxies and Mq/Ca proxies now indicate tropical temperatures of up to 34C or even 37C in places, in the Eocene warm interval. Tropical surface temperatures were two or three degrees cooler in the late Cretaceous. This still gives a considerable reduction in the pole-equator temperature gradient as compared to the modern climate, but the problem is not as severe as it was. It is quite certain that the tropical sea surface temperatures could not have been as high as 40C, since the planktonic forams that are seen in the sedimentary record could not have survived at such high temperatures. A striking feature of the data is that tropical surface temperatures seem to remain fairly constant throughout the Eocene, though polar temperatures (as indicated by the benthic forams) are decrease towards the Oligocene.

Returning to Fig. 1.9 we see that following the peak Eocene warmth of 51 million years ago, the climate commenced a long slide toward the icehouse climate characterizing the latter part of the record. Between the peak warmth and the beginning of the Oligocene 34 million years ago, the minimum polar temperature dropped by 8C as indicated by benthic $\delta^{18}O$. At this point, small ephemeral ice sheets began to form on Antarctica, culminating in a more substantial glaciation of Antarctica that lasted until 26 million years ago, somewhat before the beginning of the Miocene. The Oligocene glaciation is visible as a pronounced ditch in the $\delta^{18}O$. This first attempt at glaciating Antarctica didn't last however, since the climate recovered and returned to a period of generally cold Antarctic conditions with sea ice but with land ice sheets having volume below 50% of the present volume. This situation lasted until 15 million years ago, when the slide towards icehouse conditions resumed. Antarctic ice sheets grew again, but the Northern hemisphere land glaciation had not yet been initiated by this time. The first abundant evidence of sea ice, based on ice-rafted debris in polar sediment cores, appears at 14 million years ago. The increase in $\delta^{18}O$ in the next several million years is due to a combination of continued cooling and Antarctic ice sheet growth, culminating in the initiation of the major Northern Hemisphere ice sheets around 6.5 million years ago, as we enter the Pliocene period. The oxygen isotopes begin to show substantial fluctuations at this time, which grow in amplitude as one enters the Pleistocene. These fluctuations are due to waxing and waning of ice sheets, predominantly in the Northern Hemisphere – in other words, the coming and going of "ice ages." The nature of the fluctuations will be examined more closely in Section 1.10.

What accounts for the nature of the hothouse climate state, and for the subsequent hothouse/icehouse transition? This is another of the Big Questions. There is no support in astrophysics for solar variability of a magnitude and type that could explain the transition, so attention has settled primarily on long term fluctuations in the greenhouse gas content of the atmosphere. In an oxygenated atmosphere like that of the Phanerozoic, CO_2 is the only known long-lived greenhouse gas that can build up to concentrations sufficient to cause climate changes of a magnitude comparable to those seen over the Phanerozoic; to get fluctuations of the requisite magnitude requires amplification of the direct CO_2 effect by water vapor feedbacks, and cloud feedbacks can also substantially modify the response. Another thing that makes CO_2 a good suspect for the role of primary agent in Phanerozoic climate evolution is the fact that it is a central participant in all aspects of the inorganic and organic carbon cycle, which offers many possible mechanisms whereby CO_2 could evolve over the long term. There are a great many unresolved issues regarding the CO_2 theory of Phanerozoic climate evolution, but a central part of testing the theory is to understand the way CO_2 and water vapor act in concert to determine the temperature of a planet; the necessary estimates will be given in Chapter 4, with further refinements to be given in Chapter 7. One needs to understand how much CO_2 it would take to account for Eocene warmth, before one can decide whether there are plausible geochemical mechanisms that could lead to the required concentration.

The greatest impediment to testing the CO_2 theory of the hothouse/icehouse dichotomy is the difficulty of estimateing past CO_2 levels. There are various geochemical and fossil proxies that can be brought to bear on the problem. For example, algae preferentially take up ^{13}C at a rate that depends somewhat on the CO_2 concentration in the water in which they grow. Carbon isotopes in fossil soil carbonate, also preserves information about past CO_2 , as does the density of pores (stomata) in fossil leaves. All estimates known to date are subject to considerable uncertainty. Nonetheless there is support for the idea that CO_2 concentrations around 70 million years ago could have been 6-10 times modern pre-industrial values. The evidence points to a general decline of CO_2 since that time, but there are also some indications that CO_2 may have already attained quite low levels during some periods well before the Pliocene icehouse climate set in. This is a rapidly evolving subject, however, and nothing definitive can be said at present. What is certain is that there are known geochemical mechanisms associated with the Urey reaction and silicate weathering, which have the potential for causing changes in atmospheric CO_2 of the required magnitude, and on a time scale consistent with the observations. These mechanisms will be discussed in Chapter 9. As with any climate problem, on Earth or elsewhere, uncertainties regarding cloud feedbacks complicate the problem of testing theories of climate response. It is not out of the question that part of the answer lies in modulation of cloud albedo by, say, biologically produced sulfur compounds that seed cloud formation. Further, if data should ultimately support the low-gradient picture of the Cretaceous and Eocene hothouse climates, some mechanism will be needed to keep the tropics from overheating while the poles are warmed by elevation of CO_2 . This, too, may involve clouds, or it may involve changes in ocean circulation. It has even been suggested, with considerable physical support, that increases in hurricane intensity in a warmer world could

1.9. THE HOTHOUSE/ICEHOUSE DICHOTOMY

provoke ocean circulation changes of the sort required to provoke a low-gradient climate.

Figure 1.9 exhibits a dramatic climate event of considerable importance. The spike in $\delta^{18}O$ at the Paleocene/Eocene boundary (marked "PETM" in the figure, for Paleocene/Eocene Thermal Maximum) is not a glitch in the data. It represents a real, abrupt and massive transient warm event. The spike looks small in comparison to the range of isotopic variation over the past 70 million years, but in fact it represents the planet accomplishing two million years worth of warming in a warm spike that (on closer examination) sets in within under 10,000 years and has a duration of around 200,000 years. This isotopic excursion corresponds to a global warming of about 4C; other proxy records show that the warming had similar magnitude in the Arctic and at the Equator, and that is extended to the deep ocean. This climate event triggered a mass extinction of benthic species, probably due to a combination of warming, oxygen depletion, and ocean acidification. An important clue as to the cause of the warming is that the record of $\delta^{13}C$ from the same core (not shown) exhibits a major negative excursion at the same time, going from values of about +1.2% down to about zero at the bottom of the excursion. This indicates a catastrophic release of large quantities of isotopically light carbon into the climate system, which presumably increased the atmospheric greenhouse effect and led to warming. One possibility is that the release came in the form of methane from destabilized clathrate ices in the ocean sediments; another is that the isotopically light carbon came from oxidation of suddenly exposed organic carbon pools on land, releasing large quantities of CO_2 . Based on analysis of the carbon isotope record, it has been estimated that approximately 4200 gigatonnes of carbon were released into the ocean-atmosphere system, if the release were from organic matter. This compares with 700 gigatonnes of carbon in the form of CO_2 in the modern atmosphere. This would considerably enhance the atmospheric greenhouse effect, though the effect would largely wear off after a thousand years or so, over which time about 80% of the released carbon works its way into the ocean. It is far from clear that one can account for the observed magnitude and duration of *PETM* warming with the amount of carbon one has at ones' disposal. This is one of the Big Questions. We shall not answer it in this book, but the reader will be provided with the tools needed to assess the warming caused by various amounts of CO_2 or methane, and also (in Chapter 9) a bit of insight about the partitioning of carbon between atmosphere and ocean. These are tools one must have at hand in order to evaluate any theory of the PETM.

The Cretaceous is closed by the impact of a large asteroid or comet (known generically as a bolide). This is known as the KT impact event (for "Cretaceous/Tertiary," Tertiary being an obsolete term for the period following the Cretaceous). This event has little or no expression in the isotope record shown in Fig. 1.9, and is instead identified by global presence of a layer of the element iridium. The impact crater has also been identified, which allows an estimate of the energy of the impactor. The KT impactor had effects of extreme consequence despite the lack of an expression in the oxygen isotope record. Notably, it was the dinosaur killer – though many other species went extinct at the same time. Examination of the carbon isotope record shows also that the ocean carbon cycle remained highly perturbed for millions of years. There are many Big Questions associated with the consequences of a bolide impact. What is the mechanism by which the impact causes extinction? Is it direct blast and heat, or some longer-lasting change in the climate? It has been estimated that the impactor released about $5 \cdot 10^{23}$ Joules of energy. How does this energy compare to other energy sources in the climate system, and what effects should it have on the atmosphere and ocean? What are the broader effects of a bolide impact on climate, and how long do they last? Does the impact cause a warming (perhaps through release of greenhouse gases) or a cooling (through lofting of a dust and soot cloud)? Some of the climate questions will be taken up in Chapter 4.

The KT impact event is the archetype of impact events, which have been episodically impor-



Figure 1.10: Known occurrences of land glaciation during the Phanerozoic. Tall shaded bars indicate periods of major glaciation in which ice reaches to latitudes of 50N or 50S, while short shaded bars indicate periods of more minor glaciation.

tant throughout Earth history. Similarly the Earth has experienced many other mass extinctions besides that at the KT boundary, not all of which are clearly associated with a bolide impact. All mass extinctions lead to Big Questions.

1.9.2 Hothouse and icehouse climates over the Phanerozoic

The Cretaceous hothouse climate and the Pleistocene icehouse climate represent opposite extremes of the Earth's typical climate state of the past half billion years. Going back further in time, the Snowball Earth represents an ultra-extreme on the cold end; going further afield in space, the runaway greenhouse represents an ultra-extreme on the hot end, though one that evidently never occurred on Earth. The Earth has experienced many individual hothouse and icehouse episodes in the past half billion years. For times earlier than 70 million years ago one cannot rely on $\delta^{18}O$ records from sea-floor sediments to provide a cross-check on ice volume. Instead, one must look for evidence of glaciation or temperate polar climate preserved as glacial features or fossil animal and plant occurrences on land and in near-coastal environments. This record is far less complete and well-preserved. An accurate long term sea level record would make it possible to estimate ice volume in the distant past, but recovering and interpreting sea level has proved difficult. It is certainly possible to detect the existence of major glaciations, but good estimates of ice volume are not available for pre-Mesozoic glaciations. The periods between major glaciations – tentatively labeled as hothouse climates below – could well have undetected episodes of polar ice embedded within them. The known episodes of major and minor land glaciation are summarized in Fig. 1.10.

The Cretaceous hothouse conditions extended back throughout the entire Mesozoic, and well into the late Permian. There is evidence that some of these periods, notably the early Jurassic, may have been even warmer than the Eocene. To find another era of glaciation to rival that of the Pliocene and Pleistocene icehouse, one needs to go back to late Carboniferous and early Permian. The sixty million year period centered on the Permo-Carboniferous boundary 300 million years ago

1.9. THE HOTHOUSE/ICEHOUSE DICHOTOMY

was a time of very extensive glaciation, reaching to lower latitudes even than the Pleistocene ice ages, though not attaining Snowball proportions. This period is a crucial one for the CO_2 theory of Phanerozoic climate, since it is a time when there is quite strong evidence for low CO_2 . The earlier Phanerozoic exhibits comparatively minor glaciations at the end of the Devonian, in the mid-Silurian and for a brief period in the mid-Cambrian, but so far as the Phanerozoic goes, the Permo-Carboniferous glaciation is the big one to explain.

The changing paleogeography, shown in Fig. 1.11, is likely to have influenced climate. In particular, it is bound to be easier for a glacier to accumulate on land if there is land at or near one or both of the poles. Certainly there is land at the South Pole during the Carboniferous glaciation and the current glaciation that began in the mid-Cenozoic. However, this is clearly not the whole story, as there was plenty of land at the South Pole already 400 million years ago, but the Carboniferious glaciation didn't set in until nearly a hundred million years later. Likewise, Antarctica was already near the South Pole during the Cretaceous, but Antarctic glaciation only took off in the mid-Cenozoic. Most likely, fluctuations in CO_2 – probably itself affected by continental configuration – play a crucial role in the timing of glaciations.

A general theme in the evolution of paleogeography is the assembly and breakup of *supercontinents*. From 500 million years ago to 400 million years ago one can see the Southern supercontinent Gondwana near the South Pole, though there are a few leftover bits of land that are not part of Gondwana. By 300 million years ago, Gondwana has merged with these bits to form the global supercontinent Pangea, which then breaks up into the present continents over the course of the rest of the Phanerozoic. The interiors of supercontinents are isolated from the moderating effects of the oceans on climate, and so could be expected to experience harsh seasonal swings in temperature. Do we expect supercontinent interiors to be deserts or steaming, moist fern forests? This is another of the Big Questions.

Is there a clear dominance of hothouse or icehouse conditions over the past half billion years? The record of the past hundred million years certainly supports the notion that the largely ice-free hothouse is the preferred state of the Earth's climate, but going further back in time it is harder to say whether the apparent dominance of hothouse conditions is an artifact of poor preservation of the polar deposits where glaciers are most likely to have occurred. Some of the episodes we think of as hothouse climates could well have had significant amounts of ice.

In any event, the delineation of the circumstances which favor icehouse or hothouse climates, and the factors governing the transition between the two, constitutes one of the Big Questions of climate science. It seems likely that if the hothouse/icehouse transition of the past 70 million years can be understood, similar mechanisms could be applied to the rest of the Phanerozoic. Variations on the theme would include a greater range of different continental configurations – notably the breakup of supercontinents – as well as biological innovations such as the colonization of land and the evolution of deepwater carbonate shell-forming micro-organisms, both of which can affect the global carbon cycle.

During the Phanerozoic, life on Earth went through a number of mass extinctions rivaling or exceeding the end-Cretaceous event. The biggest mass extinction of all occurred at the end of the Permian, wiping out 96% of all marine species, 70% of all land vertebrates, and a large fraction of all land plants and invertebrates. It is the only mass extinction that included insects to any great extent. There is no clear evidence for a bolide impact at this time, though it remains possible that an impact occurred but failed to leave a trace in the fossil record. In any event, the cause of the end-Permian mass extinction ranks as one of the Big Questions.



Figure 1.11: Evolution of geography over the Phanerozoic. Map projection and shading as for Fig. 1.8. The numbers give the time in millions of years ago.

1.10 Pleistocene Glacial-Interglacial cycles

Now we'll take a closer look at what's been going on in the past five million years. The earlier portion of this time is known as the *Pliocene* epoch, and the latter portion, beginning around 1.8 million years ago, is the *Pleistocene*. The choice of the Pliocene-Pleistocene boundary is based on an obsolete notion of the time when Northern Hemisphere ice sheets attained nearly their modern extent. At the level of geological periods, the Neogene-Quaternary more closely approximates this distinction. The Quaternary period extends all the way to the present, but the Pleistocene is terminated somewhat arbitrarily at the end of the last major ice sheet retreat around 10,000 years ago. The remainder of time up to the present is the *Holocene* epoch. This is a very human-centric division, since the time we are experiencing now is a fairly ordinary Pleistocene-type interglacial period – or at least it was up until the commencement of the industrial revolution. A more rational division would extend the Pleistocene up to about the year 1700, whereafter we would enter something with a name like "Anthropocene" (for reasons to be discussed in Section 1.12).

The Pliocene and Pleistocene are a time of establishment of the great Northern Hemisphere ice sheets. During this time, the ice sheets settled into a rhythm of expansion and retreat – the rhythm of the coming and going of *ice ages*. The notion of an "ice age" is distinct from that of the "icehouse climate state" introduced earlier. The latter term refers to a span of time (usually several million years) during which there is permanent ice at one or both poles. Within the time embraced by an icehouse climate state, ice volume is not constant, but fluctuates episodically with variations in ice volume that can be on the order of a factor of two (judging from the Pleistocene). Individual episodes of large ice volume within an icehouse climate are referred to as *ice ages*, with the warmer periods in between referred to as *interglacials*, though the ice doesn't come close to disappearing completely. In the Pleistocene, the fluctuation in ice volume is dominated by changes in Northern Hemisphere ice sheets, but as ice ages come and go, the entire globe becomes colder and warmer.

1.10.1 The Marine Sediment Record

Figure 1.12 shows the $\delta^{18}O$ of benthic forams in a tropical Pacific core over the past 4 million years. The short-period fluctuation represents fluctuations in both ice volume and benthic temperature, and in addition there is a downward trend in temperature and increasing trend in ice volume over the earlier two million years of the period. By 2 million years ago, the fluctuations have settled into a fairly regular pattern, with a dominant period of about 40,000 years. The period may be crudely estimated by counting peaks in the isotope record. This says that major ice advances in the early Pleistocene occur roughly every 40,000 years. About 800,000 years ago, there is a major transition in which the amplitude of the glacial-interglacial cycle becomes markedly larger, and the periodicity lengthens to about 100,000 years. During this period, an asymmetry between glaciation and deglaciation becomes readily apparent: the climate cools and ice builds up over long periods of time, but deglaciation occurs rather precipitously.

The periodicity of the ice ages, and the reason for the transition to a dominant 100,000 year cycle later in the Pleistocene is another of the Big Questions of climate science. We will learn in Chapter 8 that the periodicities are almost certainly connected with the quasiperiodic variations of the Earth's orbital characteristics – namely the tilt of its rotation axis and the departure of the orbit from circularity. The cycles are known as *Milankovic cycles*, after the scientist who first formulated a detailed theory connecting orbital parameters with the coming and going of ice ages. Milankovic's theory was largely ignored for decades, because not enough was known about the



Figure 1.12: Benthic foram oxygen isotopes from the Ocean Drilling Program Core ODP 849 (see Mix *et al* in the Further Readings for this chapter). Values are reported relative to the PDB standard. The core is located in the tropical Pacific, but the benthic data is representative of the global climate state. Note that the vertical axis has been reversed, so that upward excursions represent warmer and less icy times.

pattern of ice ages to give the theory a fair test. It was only revived in the 1970's, when data of the sort given in Figure 1.12 first became available. Even today, the means by which cycles in orbital characteristics are expressed in the climate record are far from clear, and remain a subject of active research.

This is another case in which the lesson learned from Earth carries over to other planets, as the orbital parameters of Mars also show Milankovic cycles. Compared to Earth, however, the study of how these cycles are expressed in the paleoclimate record of Mars is in its very infancy.

1.10.2 Ice core records

The ice at the base of the Antarctic ice sheet is about a million years old, so one can also retrieve a record of past climate by drilling cores into the ice. Many aspects of climate are recorded in the ice, but the ones that will concern us here are stable isotopes of water (δD and $\delta^{18}O$), recorded in the ice itself, and composition of past air preserved in bubbles within the ice. The stable isotopes are essentially recorders of temperature; the ice becomes more isotopically light as the polar temperature becomes colder, since, more of the heavy isotopes have been distilled out in that case.

The upper panel of Fig. 1.13 shows the δD time series from the Vostok and Epica ice cores in the Antarctic. The Vostok data is systematically below the Epica data because Vostok is further inland, higher and colder, but otherwise tracks the Epica data. This record only covers the period within which the glacial-interglacial cycles have already settled into a 100,000 year cycle; older ice is too distorted to yield a useful record. This record confirms the 100,000 year cycle seen in the marine cores, and also confirms the asymmetry between slow buildup of glaciers and rapid deglaciation. It also shows that there is a strong Antarctic warming and cooling that occurs in association with the glacial/interglacial cycles.

The CO_2 data is shown in the lower panel of Fig. 1.13. In the course of the glacial/interglacial cycles, CO_2 fluctuates between 180 parts per million (by count of molecules) and 300 parts per million. Moreover, the fluctuation is very nearly synchronous with the warming and cooling. The correlation between CO_2 and temperature does not determine which causes which (or whether they mutually reinforce each other), but since CO_2 is a greenhouse gas, it is certain that the rise in CO_2 warms the planet and reinforces the termination of ice ages, whereas conversely the fall of CO_2 enhances cooling and reinforces the onset of glaciation. The origin of the glacial/interglacial CO_2 cycle is another of the Big Questions. Our understanding of glacial/interglacial cycles cannot be complete without resolving this question.

Greenland ice also records past climate, as seen in Fig. 1.14. The base of the Greenland ice cap is not as old as the Antarctic ice, so one can only go back in time about 100,000 years here. By way of compensation, though, the rate of snow accumulation in Greenland is much higher than in Antarctica, so one can see the past with much higher time resolution.

The Greenland record records a sharp deglaciation leading to the modern era, consistently with the Antarctic record. However, we can see in Greenland that there are many high frequency temperature fluctuations embedded within the glacial period –especially the period from about 60,000 years ago to 10,000 years ago. These don't have a strict periodicity, but have a time scale on the order of a thousand years, and hence are collectively referred to as *millennial variability*. The expressions of millennial variability seen in Greenland isotopes are called *Dansgaard-Oeschger events* after their discoverers. There are many other climate proxies that reflect millennial variability of the sort seen in Greenland isotopes, and these often represent precipitous switches in the



Figure 1.13: Data from the Vostok and Epica Antarctic ice cores. The upper panel shows the variation in deuterium depletion of the ice, which is a proxy for temperature. Higher (less negative) values indicate warmer conditions. The lower panel gives the variation in the CO_2 concentration in air bubbles trapped in the ice.



Figure 1.14: Oxygen isotope data from the GISP-2 Greenland ice core. Larger (less negative) values correspond to warmer temperatures.

climate state – referred to as *abrupt climate change*.. One of the most striking of these abrupt change events occurs as the planet is coming out of the most recent ice age – the *Last Glacial Maximum* (or *LGM*). The spike marked "B" in the figure is the *Bolling* warm period, and represents a recovery of the climate to full interglacial warmth. However, in the wake of the Bolling the climate abruptly reverted to full glacial temperatures, in an event known as the *Younger Dryas* (marked "YD" in the figure). The Younger Dryas is believed to have been triggered by a massive draining of a glacial lake into the ocean, but generally speaking the mechanisms of both the Younger Dryas and of millennial variability remain as Big Questions that are yet to be resolved. This is an especially important question because the flat, quiescent period at the end of the Greenland record – the *Holocene* period we have lived in for all of the history of civilization – represents an abrupt cessation of high amplitude millennial variability. What accounts for the uncommonly stable climate of the Holocene, and what would it take to break this situation? This is a Big Question with considerable import indeed.

1.11 Holocene climate variation

The climate variations of the Holocene have been (at least so far) more subtle than the massive variations we have discussed previously. In part, this is simply because the Holocene is a short period of time, and there hasn't been enough time for something really dramatic to happen, given the relatively slow pace of many of the geological processes that modify climate. But the limited span of the Holocene is not the whole story. The Holocene has not witnessed extreme millennial scale variability of the sort seen in the preceding glacial time. Indeed, one of the Big Questions concerning the Holocene is the physical basis for the relatively stable Holocene climate. In particular, given the massive assault on climate by industrial society (see Section 1.12) it becomes all the more pressing to understand what it would take to break the equable and steady Holocene climate enjoyed during the rise of civilization.

Still, the Holocene has not been without its points of interest so far as climate variations

go, the more so because there were human civilizations around to witness and be affected by these variations. A key driver of Holocene climate change is the *precessional cycle*, to be discussed in Chapter 8. It is the same precessional cycle that plays a role in the rhythm of the Pleistocene ice ages. The Earth's spin axis precesses like a top, so that the way the "tilt seasons" line up with the "distance seasons" associated with the varying distance of the Earth from the Sun goes through a cycle lasting about 22,000 years. A quarter cycle is only 5500 years, which brings these cycles within the span of recorded history. At present, the Earth is farthest from the sun when the Northern Hemisphere points towards from the Sun (i.e. during Northern summer), and is closest to the Sun when the Northern Hemisphere points away (i.e. during Northern winter). This gives us relatively warm winters and relatively cool summers. 11000 years ago the situation was reversed, and Northern Hemisphere summers received considerably more sunlight than they do today, particularly at high latitudes. This should have made polar regions considerably warmer than today, but for reasons that are only partly understood, the time of warmest summers was delayed several thousand years, to a time called rather tendentiously the *Climatic Optimum* about 7000 years ago. It is not entirely clear what the climate was "optimal" for, and this is in any event a Northern-centric view as Southern hemisphere continents if anything experienced a weak cooling at this time (as one would expect from the nature of the precessional cycle). The alternate term Altithermal is preferred, as being less value-laden. In any event, at this time the Northern polar regions were getting up to 10% more solar radiation in summer than they are today, leading to warmer summers. The warming is expected to be greatest over land, since oceans average out the warm summers and cold winters. Tree ring records indicate that high-latitude land masses were between 2C and 4C warmer in the summer at the time of the Altithermal. These estimates are corroborated by an expansion of the tree line to higher altitude land in the European Arctic. Cold-tolerant trees are primarily sensitive to growing-season temperatures, and are little effected by a moderate decrease in winter temperatures; in fact, more severe winters can in some cases be favorable to tree growth, since cold winters interrupt the life cycle of various insect pests. While the Altithermal is certainly connected with the precessional cycle, the magnitude of the warming, the cause of the delay in warming (partly associated with leftover glaciers from the last ice age), and the role of ocean circulations and vegetation changes in the altithermal, are all active subjects of research.

The precessional cycle has also had a profound effect on the distribution of precipitation in low latitudes. The Sahara is a desert today, but the dry river features known as *wadis* were flowing with water six thousand years ago, at which time the desert was a savannah grassland. This wet period commenced about 14,500 years ago and the Sahara abruptly reverted to desert about 4700 years ago. There are also intriguing indications that the time of initiation of present tropical mountain glaciers follows a precessional cycle. For example, the Peruvian Andean glaciers of the Southern Hemisphere date back to the last ice age, while the Kilimanjaro ice fields were laid down during the African Humid Period about 10,000 years ago and Himalayan glaciers of the northern subtropics tend to be even younger. The connection between the seasonal cycle of solar radiation and the tropical precipitation distribution involves atmospheric circulations – monsoons and the Hadley circulation – that cannot be treated without a full understanding of atmospheric fluid dynamics. We will therefore have only limited opportunities to pursue the precessional precipitation cycles in the course of this book, though the treatment of the precessional cycle in solar forcing will provide the student with the necessary background for further study.

The Little Ice Age is another Holocene climate fluctuation of considerable interest. This term refers to a period of generally cool Northern Hemisphere extratropical land temperatures extending from approximately 1500 to 1800. Tree ring estimates suggest the Northern Hemisphere mean temperature dropped by something over 0.5C between the year 1400 and 1600. The cooling is corroborated by records of advances of mountain glaciers, sailors' observations of sea ice, and

agricultural records. The Little Ice Age is too short and too recent to have anything to do with the precessional cycle, and while it is possible that fluctuations in the ocean circulation could have produced the cooling, the prime candidate for an explanation of the Little Ice age is a temporary slight dimming of the Sun. Sunspot observations do indicate a cessation in the normal solar sunspot cycle – called the *Maunder Minimum*– at about the time of the Little Ice Age. However, most estimates of the associate solar output change are far too small to yield a significant cooling. Various mechanisms are under investigation which could amplify the response to the small solar fluctuation, but in the grand scheme of things the Little Ice Age is a rather subtle event, and accordingly hard to understand, particularly in terms of simple models.

To put the Holocene in perspective, it is salutory to note that the "abrupt" PETM event discussed in Section 1.9.1 lasted nearly 200,000 years, and set in over a period of around 10,000 years – as long as the full length of the Holocene. There really hasn't been much time for things to happen in the Holocene, and the time span of Figure 1.9 no doubt contains many 10,000 year periods as quiescent as the Holocene has been up until recently. Short as the Holocene is, we will see next that human activities have been able to cause some dramatic changes in the composition of the atmosphere and consequently in the Earth's climate. One wonders what it would take to trigger a hyperthermal event such as the PETM, which is over in the wink of an eye by the standards of Figure 1.9, but has a duration twenty times as long as the span of human civilization to date.

1.12 Back to home: Global Warming

We have seen that CO_2 has been a major factor in determining climate througout Earth's history, and that life, in turn, has greatly shaped the carbon cycle. Life is in the midst of disrupting the carbon cycle once more, but this time it's technological life that has provided the necessary innovations. Over billions of years, a great deal of organic carbon has been sequestered in the Earth's crust without oxidizing. Most of this carbon is in very dilute forms which cannot easily be tapped to provide economically useful amounts of energy. However, a very small portion winds up in nearly pure forms that are moreover chemically altered in a fashion that makes them especially convenient as fuels. These are the *fossil fuels* – coal formed from land plants and oil formed in marine environments. Natural gas can be produced by thermal alteration of either coal or oil. Fossil fuels represent concentrated solar energy stored in the form of organic carbon, which has been accumulating over hundreds of millions of years. This pool of readily oxidizable carbon exists precisely because it is in geological formations that have kept it apart from oxygen over the ages. It is only the evolution of technical civilization that is making it possible to dig up and oxidize hundreds of millions of years worth of stored fossil carbon within a few centuries.

In the year 2005, over 8 gigatonnes $(8 \cdot 10^{12} kg)$ of carbon were released by fossil fuel burning, and annual emissions continue to grow rapidly. There are several ways to see that this is a very big number – a major upset to the natural carbon cycle. First, the pre-industrial atmosphere contained about 600 gigatonnes of carbon, so the 2005 annual emission is fully 1.3% of the undisturbed atmospheric content. If the same amount were released into the atmosphere each year, it would take only 75 years to double the atmospheric CO_2 content, provided all the released CO_2 stayed in the atmosphere. Alternately, one could compare the fossil fuel emissions to the volcanic outgassing which in the long term balances silicate weathering and sustains the carbon cycle. Precise estimates of volcanic outgassing are hard to come by, but generally are on the order of 0.1 gigatonnes of carbon per year or less. Thus, fossil fuel carbon emissions are eighty times larger than background volcanic outgassing. In fact, the very largest carbon flux number involved in the whole carbon cycle is the net CO_2 carbon fixed into organic carbon each year by worldwide photosynthesis, and fossil fuel emissions even look impressive when compared to this number. Based on satellite chlorophyll observations, it has been estimated that photosynthesis fixes 100 gigatonnes of carbon each year, about half on land and half in the oceans. The year 2005 fossil fuel emissions were fully 8% of this number. In other words, worldwide photosynthetic productivity would have to increase by 8% to take up the fossil fuel CO_2 and 100% of that carbon would have to be buried as organic matter without being recycled by respiration. That, of course, would be a completely absurd situation, as virtually all of the photosynthetically fixed carbon is quickly respired back into the atmosphere, largely by bacteria who have had several billion years to become proficient at making use of organic carbon wherever they find it. As an example, land photosynthesis fixes about 50 gigatonnes of carbon each year, but the flux of organic carbon to the oceans in all the world's rivers is a mere 0.4 gigatonnes per year (one twentieth of fossil fuel carbon emissions). And there is no evidence that much of the remainder of the photosynthetically fixed carbon is remaining on land as soil organic carbon. To say that humans have become a force of geological proportions vastly understates the case, for by this measure human influences on the carbon cycle overwhelmingly dominate the natural sources.

The result of all our busy digging and burning has been a steady increase in atmospheric CO_2 . Figure 1.15 shows the time series of atmospheric CO_2 concentration since 1750. CO_2 has a very atmospheric lifetime, so it is well-mixed. In consequence, one finds nearly the same CO_2 concentration whereever one measure it, so long as the measurement is not in the immediate vicinity of major sources or sinks. The part of the record since 1950 comes from direct analyses of air samples at the Mauna Loa observatory, whereas the earlier part of the record comes from air trapped in bubbles in the ice of the Siple Dome site, Antarctica, but the two records match up well where they meet. At the dawn of the industrial era CO_2 concentrations have exceeded 380ppmv - fully 35% above the pre-industrial value. Most of the increase has happened since the mid-twentieth century, and the rate of increase seems to be accelerating along with population and economic growth.

Not all of the carbon released by fossil fuel burning has remained in the atmosphere. Estimates based on careful historical inventories suggest that only about half of the total carbon released to date remaines in the atmosphere as carbon dioxide. Most of the remainder has slowly infiltrated the ocean, with a lesser amount having been taken up by the terrestrial ecosystem (net of deforestation). In fact, it has been shown that the rate at which the ocean can take up the excess CO_2 is limited by the mixing between the upper ocean and the deep ocean. This is a slow process, and if all fossil fuel burning were to suddenly cease, it would take in excess of 600 years for 80% of the excess CO_2 to be taken out of the atmosphere. The remainder would stay in the atmosphere for millennia longer, owing to certain chemical processes (discussed briefly in Chapter 9) which limit the ability of the ocean to take up CO_2 . The slow net removal rate of CO_2 allows fossil fuel emissions to accumulate in the atmosphere. Another consequence of the long lifetime of CO_2 in the atmosphere is that the climatic effects of elevated CO_2 will persist for centuries to millennia, even after any (much to be hoped-for) dramatic restriction of fossil-fuel burning. Allowing for uptake by the ocean, there are enough fossil fuel reserves- primarily in the form of coal – to ultimately increase the atmospheric CO_2 concentration to at least six times the pre-industrial value. The number could go much higher if the ocean sink were to become less efficient, or if land ecosystems were to turn around and become a CO_2 source rather than a sink.

This all leads us to a series of very Big Questions: if the rise in CO_2 is allowed to continue to a doubling of the pre-industrial value, how much will the Earth warm? How will the warming



Figure 1.15: Annual mean CO_2 concentration from 1750 to the time of writing. The earlier part of the record is from air trapped in bubbles in the Siple Dome Antarctic ice core. The more recent part of the record is from instrumental measurements at the Mauna Loa observatory. The units are molecules of CO_2 per million molecules of air.

be distributed? How much will sea level rise as a result of melting land ice and thermal expansion of ocean water? What will happen to precipitation patterns? How will all of this affect human societies and natural ecosystems? The basic physics needed to treat these questions is identical to what is used to account for the influence of CO_2 and other long lived greenhouse gases on past climates. The problem in this instance has more immediacy as many generations our descendents will be living with the consequences of our fossil fuel emissions in the next several decades. In order to understand what kind of planet we are leaving these descendents, there is a demand for greater detail in the understanding of the climate changes to be wrought by these rapid increases in atmospheric CO_2 .

Interest in the effect of CO_2 changes on climate long predates the kind of data shown in Figure 1.15 which showed that CO_2 was on the increase, and in fact predates the realization that human activities really could cause CO_2 to increase appreciably. Likewise, global warming was a concern long before it was confirmed that the Earth really was warming in response to increases of CO_2 . These things were all anticipated theoretically a century or more before global warming burst onto the scene as an issue of political consequence, and the driving force was basic curiosity about the physics governing planetary temperature. It's a line of inquiry that extends right back to Fourier's pathbreaking inquiry into how an atmosphere affects the energy budget of a planet, and hence its temperature. The discovery of global warming is a great triumph of two centuries of developments in fundamental physics and chemistry. It is not a matter of people having noticed that both CO_2 and temperature were going up, and concluding that the first must be somehow causeing the second. Both the rise of CO_2 as a consequence of fossil fuel burning, and the consequent rise in temperature as a response to the Earth's perturbed energy balance, were anticipated long before either was observed.

After Fourier, the tale resumes with Tyndall, whose work on the infrared absorption of CO_2 and water vapor was mentioned near the beginning of this chapter. Tyndall was interested in these gases because of the questions raised by Fourier regarding the factors governing planetary temperature. He was also interested in the recently-discovered phenomenon of the ice-ages, and with several contemporaries thought perhaps ice ages could arise from a reduction in CO_2 . In that, he was partly right; the Pleistocene ice ages are cold partly because of the glacial interglacial CO_2 cycle, even though the ultimate pacemaker of the ice ages is the rhythm of Earth's orbital parameters. Tyndall died, however, before he ever had the chance to translate his measurements into a computation of the Earth's temperature. That task was left to the Swedish physical chemist Svante Arrhenius, who in 1896 performed the first self-consistent calculation of the Earth's temperature incorporating the greenhouse effect of water vapor and CO_2 . Interestingly, Tyndall's measurements were not sufficient to provide the information about weak absorption over long path lengths, so for the absorption data he needed he turned to Langley's observations of infrared emitted by the Moon. It was a felicitous re-use of data intended originally for determination of the Moon's temperature, and indeed was a more correct use of the data than Langley was able to accomplish. This shows the benefit of curiosity-driven science: measurements taken to satisfy curiosity about lunar temperature wound up being instrumental in permitting an evaluation of the effect of the Earth's atmosphere on the Earth's temperature. Astronomers initiated the study of infrared as an observational technique, but the radiative transfer work stimulated by their needs soon provided the crucial tool needed to understand planetary climate. Arrhenius not only estimated the Earth's then-current temperature, but also estimated how much it would warm if the amount of CO_2 in the atmosphere were to double. Using clever scaling analyses from Langley's data, he was able to do this without a firm knowledge of just what the atmosphere's CO_2 content actually was. Not long afterwards, he realized that industrial burning of coal was dumping CO_2 into the atmosphere, and could eventually bring about a doubling; he described this process as "evaporating our coal mines into the atmosphere." At then-current rates of consumption, it appeared that a doubling would take up to a millennium, and Arrhenius would no doubt have been surprised to know that his own great-grandchildren could well live to witness the doubling. This takes our story to about 1900. What happened then?

A long hiatus. Two unfortunate turns of events held back the study of global warming for many decades. The first was a highly touted experimental study published in 1900 by the prominent physicist Angstrom, which purported to show that the radiative effects of CO_2 are "saturated," i.e. that the gas already absorbs as much as it can at the atmosphere's then-present concentration, so increases would have no effect. A concomitant and closely associated (and equally wrong) idea was that the strong absorption of water vapor would completely swamp any effect CO_2 might have. The experiment turned out to be wrong, but such was Angstom's reputation and such was the resistance to the idea that humans could change climate that it was decades before anybody checked the result. Moreover, it turns out that even if Angstrom had been right, it would not have negated the greenhouse effect; this misunderstanding hinged on the poorly developed understanding of radiative transfer in a temperature stratified atmosphere. The "grey gases" we will study in the first half of Chapter 4 are "saturated" in the sense of Angstrom, but nonetheless permit a very strong greenhouse effect. The second barrier to progress was the belief that the huge carbon content of the ocean would buffer the atmosphere, overwhelming anything human industry could have thrown at it. The carbonate chemistry needed to defeat this idea could have been done at the time of Arrhenius, and indeed requires nothing more than is taught routinely in high-school chemistry courses today. However, the logjam was not broken until Revelle and Suess took up the challenge in 1957, and showed that the upper ocean could not take up nearly as much CO_2 as had previously been thought. No doubt, if it hadn't been for Angstrom's insistence on saturation, this chemistry might have been worked out much sooner. Close on the heels of Revelle and Suess's result came the urgency of actually measuring atmospheric CO_2 , which was initiated by Charles Keeling, whose work culminated in the Mauna Loa data shown in Fig. 1.15. The techniques for recovering past CO_2 from air bubbles trapped in ice was not to be developed until

the 1970's, so Keeling had to wait a decade or so before it was clear that CO_2 was really rising, and a bit more time after that before there was a clear idea of just how high CO_2 already was relative to the pre-industrial value. This work, together with developments in infrared radiative transfer stimulated by astronomical observation and military interest in infrared target detection lead to new breakthroughs in the formulation of radiative transfer. The work culminated in 1967 with the calculation by Manabe and Wetherald of the Earth's temperature using modern radiative physics. They were also able to calculate the warming due to a doubling of CO_2 , allowing for expected changes in water vapor content as the planet warmed. This was not the end of the story, which indeed continues today, since there was much to be done in terms of embedding the radiative transfer in a fully consistent computation incorporating the fluid dynamics of the atmosphere – a *general circulation model*. It was, however, the beginning of the modern chapter of the study of global warming. With the publication of the Charney report by the US National Academy of Sciences in 1979, global warming began to be perceived as a real threat. The powers that be were slow to awaken to the magnitude of the problem, and several more years were to pass before the creation of the Intergovernmental Panel on Climate Change in 1988, which initiated regular, comprehensive surveys of the state of the science surrounding global warming. At the time of writing, the world still awaits substantive action to curb fossil fuel emissions.

All aspects of the essential chemistry, radiative physics and thermodynamics underlying the prediction of human-caused global warming have been verified in numerous laboratory experiments or observations of the Earth and other planets. Other aspects of the effect of increasing greenhouse gases rely on complex collective behavior of the interacting parts of the climate system; this includes behavior of clouds and water vapor, sea ice and snow, and redistribution of heat by atmospheric winds and ocean currents. Such things are impossible to test in laboratory experiments. To some extent, aspects of our theories of the collective behavior have been tested against the seasonal cycle of Earth, interannual variability, and past climates, as well as attempts to simulate other planetary climates. The ultimate test of the theory, though, is to verify it against the uncontrolled and inadvertent experiment we are conducting on Earth's own climate. Can we see the predicted warming in data? This is not an easy task. For one thing, the atmospheric CO_2 increase is only a small part of the way towards doubling, and the climate has not even fully adjusted to the effect of this amount of extra radiative forcing: oceans take time to warm up, and delay the effect for many years (for reasons to be discussed in Chapter 8. Thus, so far the signal of the human imprint on climate is fairly small. Set against that is a fair amount of noise complicating the detection of the signal. Climate, even unperturbed by human influence, is not steady from year to year, but is subject to a certain amount of natural variability. This can be due to volcanic eruptions and subtle variations in the brightness of the Sun. There are also various natural cycles in the ocean-atmosphere system that cause the planet to be a bit warmer or colder from one year to the next. Chief among these is the El Nino phenomenon of the tropical Pacific. During El Nino years, the coupled dynamics of the tropical ocean and atmosphere causes warm water to spread throughout the Pacific, leading to a warming of mean surface temperatures both in the tropics and further afield. La Nina years represent a bunching up of the warm water, and an accentuated upwelling of cold water, leading to cold years. The two phases alternate erratically, with a typical time scale of three to five years.

The fact that the signal is hard to detect does not mean that global warming is of little consequence. The difficulty arises precisely because we are trying to detect the signal before it becomes so overwhelmingly large as to be obvious. Given the long lifetime of CO_2 in the atmosphere, it would be highly desirable to keep the signal from ever getting that large, as if it ever does it will take many centuries to subside. Let's now take a look at some of the data, and see if there are any signs that the theoretically anticipated warming is really taking place.

Figure 1.16 shows a times series of estimated global mean surface temperature, based on recorded temperatures measured with thermometers. There is a lot of arduous statistics and data archaeology behind this simple little curve. Particularly for the data going into the early part of the curve, there has been a need to standardize measurements to allow for the various different ways of taking a temperature reading. Most of the oceanic measurements, for example, were taken by commercial or military ships of one sort or another, and some of the entries in ships' logs record things like "bait tank temperature" or "engine inlet temperature." There has also been a need to screen out stations that have been strongly affected by local land use changes such as urbanization, and to avoid spurious trends due to changes in the spatial distribution of temperature measurements (e.g. fewer Antarctic readings once Antarctic whaling essentially ceased). To help correct for biases in individual classes of temperature measurements, the long-term trends are presented as *anomalies* relative to the average of a station's long-term average standardized to a fixed base period (e.g. 1951-1980 for the data in Fig. 1.16).

There is little temperature trend between 1880 and 1920, but between 1920 and 2005 the temperature has risen by nearly 1C. The rise hasn't been steady and uninterrupted, however. It takes the form of an early rise between 1920 and 1940, followed by a 30 year period when temperatures remained fairly flat, whereafter the temperature rise resumes and has continued to the present. Given that CO_2 has been rising at an ever-increasing rate over the industrial period, why was the warming interrupted between 1940 and 1970? The answer lies largely in another effect of human activities on climate. Burning of fossil fuels, and especially coal, releases sulfur compounds into the atmosphere which form tiny highly reflective droplets known as *sulfate aerosols*. By 1995, the effect was finally quantified with sufficient accuracy to permit reasonable estimates of the effect, and it began to appear that most of the evolution of climate of the twentieth and twentyfirst centuries could be accounted for by a combination of rising greenhouse gases (mainly CO_2) due to human activity, with an offsetting cooling effect of sulfate aerosols. The reason small particles are so good at scattering light back to space is discussed in Chapter 5, where the optical properties of sulfate aerosols will be discussed in detail. By the year 2000, the greenhouse warming signal had unquestionably risen above both the noise of natural variability and the offsetting effect of aerosol cooling. Sulfur is an active element in many actual and hypothetical planetary atmospheres, and so the study of sulfate aerosols on Earth informs other planetary problems, including the clouds of Venus and Venus-like extrasolar planets.

The Earth's emissions in the microwave spectrum have been monitored continuously by satellite-borne instruments since 1979, and these observations make it possible in principle to obtain reconstructions of atmospheric temperature trends which are independent of the somewhat inhomogeneous surface station network. Processing the microwave data acccurately enough to obtain reliable temperature trends proved very difficult, and there were many false steps along the way. Nonetheless, the main problems were resolved by early in the twenty-first century. The microwave temperature retrievals give the temperature of the atmosphere averaged over fairly deep layers, in constrast to the surface stations which measure near-surface air temperature. The left panel of Figure 1.17 shows the satellite retrieval of temperature in layer of the atmosphere known as the lower troposphere – extending from sea level to roughly 5 km in altitude. The satellite record tracks the GISS surface station record very closely, with the exception of the very strong 1997 El Nino, during which the satellite indicates that the lower tropospheric layer warmed considerably more than the near-surface air. Both satellite and GISS records reproduce the cooling caused by the El Chicon eruption (which overwhelmed the 1982 El Nino) and the Pinatubo eruption (which accentuated the La Nina cooling following the 1991 El Nino, leading to a very cold year). The substantial agreement between the satellite and surface station record proves beyond doubt that the warming observed in recent times is not an artifact of any supposed inadequacies of the surface station record.



Figure 1.16: Global average annual mean surface temperature since 1870, estimated from surface temperature observations. Data source is the NASA GISS surface station analysis. The temperature is given as an anomaly relative to the mean temperature for the years 1951 to 1980. To turn these into actual global mean temperatures in degrees Celsius, add 14C to the anomaly.



Figure 1.17: Atmospheric temperature time series derived from analysis of microwave satellite data. Left panel: Mean temperature anomaly for the layer of the atmosphere below about 5km, compared with the GISS instrumental record. Temperature is given as an anomaly relative to the same base value as used in the GISS instrumental record. Right panel: Temperature anomaly for the lower stratosphere (layer from about 15 km to 25 km). In both figures, the El Chichon and Pinatubo volcanic eruptions are marked. In the left panel, major El Nino events are indicated by upward open-shafted arrows.
The situation looks quite different higher up in the atmosphere. The right panel of Fig. 1.17 shows the temperature trend in a portion of the atmosphere called the *lower stratosphere*, extending from about 15 to 25 km in altitude. Here, volcanic eruptions produce a pronounced warming, as opposed to the cooling seen at lower layers. This suggests that volcanic aerosols heat the upper atmosphere by absorbing sunlight. The pattern of upper level cooling and lower level warming produced by high altitude solar-absorbing layers will be discussed in Chapter 4, and the reflective effect of aerosols will be brought into the picture in Chapter 5. Leaving out the warm spikes associated with volcanic eruptions, the lower stratosphere appears to have undergone a pronounced cooling over the span of the satellite record. Is stratospheric cooling compatible with CO_2 -induced warming in the lower troposphere? This is a Big Question that is resolved in Chapter 4. Ozone destruction also cools the stratosphere, since ozone absorbs sunlight. That portion of the cooling should go away as ozone recovers as a consequence of the Montreal Protocol banning ozone-destroying chlorofluorocarbons.

The Big Question of how much the Earth will warm upon a doubling or quadrupling of CO_2 , and how fast it will do so, engages a number of associated questions. Insofar as water vapor is itself a powerful greenhouse gas, any tendency for water vapor content to increase with temperature will amplify the warming caused by CO_2 . This is known as water vapor feedback. This feedback is now considered to be on quite secure ground, but the study of the behavior of water vapor in the atmosphere offers many challenges, and is a problem of considerable subtlety. In subsequent chapters, we'll provide the underpinnings needed for a study of this host of questions. Clouds present an entirely greater order of difficulty, as they warm the planet through their effect on outgoing infrared radiation, but cool the planet through their reflection of solar radiation. The net effect depends on the complex processes determining cloud height, cloud distribution, cloud particle size and cloud water or ice content. The infrared effects of clouds will be discussed in Chapter 4 and the reflective effects of clouds on sunlight will be discussed in Chapter 5. Uncertainties about the behavior of clouds are the main reason we do not know precisely how much warmer the planet will ultimately get if we double the CO_2 concentration. Typical predictions of equilibrium global average warming for a doubling of CO_2 range from a low of around 2C to a high of around 6C, with some potential for even greater warming with a low (but presently unquantifiable) probability. Because of other uncertainties in the system (particularly the magnitude of the aerosol effect and especially the indirect aerosol effect on cloud brightness) simulations with a range of different cloud behaviors can all match the historical climate record so far, but nonetheless yield widely different forecasts for the future. There is no analysis at present that excludes the possibility of the higher end of the forecast range, for which the effects would likely be catastrophic. There are other feedbacks in the climate system that complicate the forecast. These include feedbacks from melting snow and ice, and from the dynamics of glaciers on land. They also include changes in vegetation, and changes in the ocean circulation which can affect the delay due to burial of heat in the deep ocean.

Global warming – perhaps more aptly called "global climate disruption" – is an event of geological proportions, but one which is caused by human activities. The natural range of CO_2 for the past 800,000 years, and almost certainly for the entire two million years of the Pleistocene, has been 180 to 280 molecules per million. Owing to human activities, the CO_2 concentration is already far above the top of the natural range that has prevailed for the entire lifetime of the human species, and without action will become much higher still. The human species and the natural ecosystems we share the Earth with have adapted over the Pliocene and Pleistocene to glacial-interglacial cycles, but a world with doubled CO_2 will subject them in the course of two centuries or less to a temperature jump to levels far warmer than the top of the range to which societies and organisms have adapted. Even if climate sensitivity is at the low end of the predicted range and if human societies hold the line at a doubling of CO_2 , the resulting 2C warming represents a substantial

climate change; it takes a great deal to change the mean temperature of the entire globe, and a 2C global mean increase is a summary statistic that masks much higher regional changes and potentially quite massive effects on sea ice, glaciers and ecosystems. If climate sensitivity turns out to be at the high end, the warming could be 4C or more, and if that is compounded by an increase to four times pre-industrial CO_2 the global mean increase could reach 8C. That is twice the degree of warming in the PETM, and though the PETM looks abrupt, it is very likely to have set in on a longer time scale than it would take human industrial society to burn the remaining reserves of fossil fuels. If this is allowed to happen, it will take thousands of years for the climate to recover to a normal state. Could global warming disrupt the natural glacial-interglacial cycle? What would the consequences of that be? Those are indeed Big Questions.

As seen by paleoclimatologists ten million years in the future, whatever species they may be, the present era of catastrophic release of fossil fuel carbon will appear as an enigmatic event which will have a name of its own, much as paleoclimatologists and paleobiologists refer today to the PETM or the K-T boundary event. The fossil carbon release event will show up in ^{13}C proxies of the carbon cycle, in dissolution of ocean carbonates through acidification of the ocean, through mass extinctions arising from rapid warming, and through the moraine record left by retreating mountain glaciers and land-based ice sheets. As an event, it is unlikely to permanently destroy the habitability of our planet, any more than did the K-T event or the PETM. Still, a hundred generations or more of our descendents will be condemned to live in a planetary climate far different from that which nurtured humanity, and in the company of a greatly impoverished biodiversity. Biodiversity does recover over the course of millions of years, but that is a very long time to wait, if indeed there are any of our species left around at the time to do the waiting. Extinction may not be precisely forever, but it is close enough.

1.13 The fate of the Earth, the lifetime of biospheres

Even if a planet enters a habitable phase at some stage in its life, it will not remain habitable forever; various kinds of crises can bring its habitability to an abrupt or gradual end. This brings us to the Big Question of lifetime of biospheres; the answer has implications for how likely it is that complex or intelligent life will have had time to evolve elsewhere in the Universe.

Certainly, the Earth's habitability will end when the Sun leaves the main sequence and expands into a Red Giant. Perhaps some of the outer planets or their satellites will enter a brief habitable phase at that time, but it will not be long lasting. That particular crisis is about four billion years in Earth's future, but other habitability crises are likely to set in long before then. In particular, as the Sun continues to brighten, at some point the brightness will outstrip the ability of the silicate weathering process to compensate by drawing down CO_2 . At that point the Earth would succumb to a runaway greenhouse, become lethally hot, and eventually lose its water to space. When will that happen? That is a Big Question, and some current estimates put the remaining natual lifetime of Earth's biosphere at as little as a half billion years. Given that it took four billion years of Earth History before intelligent life emerged, that makes our existence look like quite a close call. Even before the runaway stage, silicate weathering will draw down CO_2 to the point where most forms of photosynthesis will no longer be able to operate. Can more efficient forms of photosynthesis fill in the gap? That's a Big Question as well, but one of a primarily biological nature that we will not attempt to answer.

As the Sun's luminosity increases, Earth may become uninhabitable, but other planets in the Solar System may become more hospitable; in any event they will go through interesting transformations. Mars will warm up, but given that it has little or no active tectonics to generate a new atmosphere, it is unlikely to become Earthlike unless some artificial means is found to give it a more substantial atmosphere. Could Europa melt and become a waterworld? What will happen to Titan as the Sun gets brighter?

Alternately the end could come by ice rather than fire. Earth's life and climate are ultimately maintained by a brew consisting of solar energy and the CO_2 outgassing from the interior. The CO_2 has a warming effect of its own, which can be modified by organisms that intercept it and transform it into oxygen, methane, or other compounds. If the tectonic release of CO_2 ceases, as it will once the Earth exhausts its interior heat sources, all that will come to an end. There will be nothing to offset silicate weathering, and CO_2 will draw down until the Earth turns into a snowball – unless the runaway greenhouse from a brightening Sun gets us first.

This class of questions naturally generalizes to the question of how the time scales that limit the biosphere's lifetime would be different for planetary systems around other stars. We have already mentioned that hotter stars have a shorter life on the main sequence, while cooler stars last longer; the former will have planets with short-lived biospheres compared to the latter. The question of how long the silicate weathering thermostat can cope with changing stellar luminosity, and how long outgassing can sustain the climate, is far subtler, and will have interesting dependences on the planet's size, composition and orbit. There could well be other chemical cycles other than silicate weathering and CO_2 outgassing, which could provide climate regulation; the search for such possibilities is still in its infancy. There could also be novel habitability crises, associated with long term evolution of planetary systems with highly eccentric orbits or systems perturbed by binary star companions.

All this climate catastrophe presupposes no intervention by the inhabitants. In fact, there are quite realistic possibilities for technologically adept inhabitants to stave off the catastrophe at least until their star leaves the Main Sequence. A runaway greenhouse could be prevented by simply reducing the effective stellar brightness, through orbital sunshades or injection of reflecting aerosols into the upper atmosphere. Indeed, such *geoengineering* fixes have been proposed to offset the global warming effect of anthropogenic CO_2 increases. They are a rather desperate and alarming prospect as a solution to global warming, since they offset a climate forcing lasting a thousand years or more with a fix requiring more or less annual maintainence if catastrophe is not to strike; far better to keep CO_2 from getting dangerously high in the first place. However, if the alternative a half billion years out is a runaway greenhouse, the risk of maintaining sunshades will no doubt seem quite acceptable. The loss of CO_2 outgassing as Earth's tectonic cycle ceases also has a relatively easy technical fix. Inhabitants could use a small portion of the energy received from the star to cook CO_2 back out of carbonates, in a process nearly identical to that by which cement is manufactured. Given the slow rate of silicate weathering, only modest quantities of carbonate would have to be processed. All this can be done, but it would appear to require long term planning and intelligent intervention. A good understanding of the principles of planetary climate will be needed by any beings contemplating such interventions. This book, we hope, will be a good place to start.

1.14 For Further Reading

Conditions during the earliest period of Earth's history, the time required for the crust to form and cool, and the time evolution of heat flux from the interior of the Earth to the surface are discussed in:

• Sleep NH, Zahnle K and Neuhoff PS 2001: Initiation of clement surface conditions on the

1.14. FOR FURTHER READING

earliest Earth, Proc Nat Acad Sci, 98, 3666-3672.

• **TurcottePaper AUTHOR,FF YEAR: Title, Journal, VOL, nn-nn.

The long-term evolution of the brightness of the Sun and similar stars is discussed in:

- Gough, DO 1981: Solar interior structure and luminosity variations, Solar Physics, 74, 21-34.
- Sackmann I-J, Boothroyd AI, and Kraemer KE 1993: Our Sun. III. Present and Future, Astrophysical Journal, 418, 457-468.

A good review of the history of oxygen on Earth and the proxy methods used to infer this history can be found in

• Canfield DE 2005: The Early History of Atmospheric Oxygen: Homage to Robert M. Garrels Annu Rev Earth Planet Sci, **33**, 1-36. doi:10.1146/annurev.earth.33.092203.122711

Evolution of Phanerozoic climate, occurrence of glaciations, and evolution of CO_2 content of the atmosphere are discussed in

- Crowley TJ and Berner RA 2001: *CO*₂ and climate change, *Science*, **292**, 870-872. DOI: 10.1126/science.1061664
- Zachos J et al. 2001: Trends, Rhythms and Aberrations in Global Climate 65 Ma to Present, Science, 292, 686-693. DOI: 10.1126/science.1059412

Veizer's long term fossil ${}^{18}O$ tropical temperature record, discussed in Crowley and Berrner (2001), is not generally considered reliable.

CHAPTER 1. THE BIG QUESTIONS

Chapter 2

Thermodynamics in a Nutshell

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

2.1 A few observations

The temperature profile in Figure 2.1, measured in the Earth's tropics introduces most of the features that are of interest in the study of general planetary atmospheres. It was obtained by releasing an instrumented balloon (radiosonde) which floats upwards from the ground, and sends back data on temperature and pressure as it rises. Pressure goes down monotonically with height, so the lower pressures represent greater altitudes. The units of pressure used in the figure are *millibars* (*mb*). One *bar* is very nearly the mean sea-level pressure on Earth, and there are 1000 *mb* in a *bar*.

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

However, for various reasons one might nevertheless want to know at what altitude a given pressure level lies. By altitude tracking of the balloon, or using the methods to be described in Section 2.2, the height of the measurement can be obtained in terms of the pressure. The right



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

panel of Figure 2.1 shows the relation between altitude and pressure for the sounding shown in Figure 2.1. One can see that the height is very nearly linearly related to the log of the pressure. This is the reason it is often convenient to plot quantities vs. pressure on a log plot. If p_o is representative of the largest pressure of interest, then $-\ln(p/p_o)$ is a nice height-like coordinate, since it is positive and increases with height.

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

The sounding we have shown is typical. In fact, a similar pattern is encountered in the atmospheres of many other planets, as indicated in Figure 2.2 for Venus, Mars, Jupiter and Titan. In common with the Earth case, the lower portions of these atmospheres exhibits a sharp decrease of temperature with height, which gives way to a region of more gently decreasing, or even increasing, temperature at higher altitudes. In the case of Venus, it is striking that measurements taken with two completely different techniques, probing different locations of the atmosphere at different times of day and separated by a decade, nonetheless agree very well in the region of overlap of the measurements. This attests both to the accuracy of the measurement techniques, and the lack of what we would generally call "weather" on Venus, at least insofar as it is reflected in temperature variability. In the case of Mars, the temperature decrease shows most clearly in the summer afternoon when the surface is still warm from the Sun. As nignt approaches the upper level temperature decrease is still notable, but the lower atmosphere cools rapidly leading to a low level *inversion*, or region of temperature increase. In the Martian polar winter, the whole atmosphere cools markedly, and is much more isothermal than in the other cases.

The temperature decrease with height in the Earth's atmosphere has long been known from experience of mountain weather. It became a target of quantitative investigation not long after the invention of the thermometer, and was early recognized as a challenge to those seeking an understanding of the atmosphere. It was one of the central pre-occupations of the mountaineer and scientist Horace Bénédict de Saussure (1740-1799). In the quest for an explanation, many false steps were taken, even by greats such as Fourier, before the correct answer was unveiled. As will be shown in the remainder of this chapter, some simple ideas based on thermodynamics and vertical mixing provide at least the core of an explanation for the temperature decrease with height. Towards the end of Chapter 3 we will introduce a theory of tropopause height that captures the essence of the problem; the theory of tropopause height will be revisited with increasing sophistication at various points in Chapters 4 and 7. Nonetheless, some serious gaps remain in the state of understanding of the rate of decrease of temperature with height, and of the geographical distribution of tropopause height. In Chapters 3 and 4 we will see that the energy budget of a planet is crucially affected by the vertical structure of temperature; therefore, a thorough understanding of this feature is central to any theory of planetary climate.



Figure 2.2: The vertical profile of temperature for a portion of the atmosphere of Jupiter (upper left panel), for the atmosphere of Venus (upper right panel), and for the atmosphere of Mars (lower left panel) and that of Titan (lower right panel). The Venus Magellan and Mars data derive from observations of radio transmission through the atmosphere, taken by the Magellan (late 1980's) and Global Surveyor orbiters, respectively. The information on the lower portion of the Venus atmosphere comes from one of the four 1978 Pioneer Venus probes (the others show a similar pattern). The Jupiter data derives from *in situ* deceleration measurements of the Galileo probe. The full Mars profile dataset reveals considerable seasonal and geographical variation. The profiles shown here were taken in the Southern Hemisphere by Mars Global Surveyor. The warmest one is in the late afternoon of 1998 in the Summer subtropics while the next warmest is at night-time under otherwise similar conditions. The coldest Mars sounding shown in the plot is from the winter South Polar region. 100 $Pascal = 1 \ mb$

2.2 Dry thermodynamics of an ideal gas

2.2.1 The equation of state for an ideal gas

The three thermodynamic variables with which we will mainly be concerned are: temperature (denoted by T), pressure (denoted by p) and density (denoted by ρ). Temperature is proportional to the average amount of kinetic energy per molecule in the molecules making up the gas. We will always measure temperature in degrees Kelvin, which are the same as degrees Celsius (or Centigrade), except offset so that absolute zero – the temperature at which molecular motion ceases - occurs at zero Kelvin. In Celsius degrees, absolute zero occurs at about -273.15C. Pressure is defined as the force per unit area exerted on a surface in contact with the gas, in the direction perpendicular to the surface 1 . It is independent of the orientation of the surface, and can be defined at a given location by making the surface increasingly small. In the mks units we employ throughout this book, pressure is measured in Pascals (Pa); 1 Pascal is 1 Newton of force per square meter of area, or equivalently 1 $kq/(ms^2)$. For historical reasons, atmospheric pressures are often measured in "bars" or "millibars." One bar, or equivalently 1000 millibars (mb) is approximately the mean sea-level pressure of the Earth's current atmosphere. We will often lapse into using mb as units of pressure, because the unit sounds comfortable to atmospheric scientists. For calculations, though, it is important to convert millibars to Pascals. This is easy, because 1 mb = 100 Pa. Hence, we should all learn to say "Hectopascal" in place of "millibar." It may take some time. When pressures are quoted in millibars or bars, one must make sure to convert them to Pascals before using the values in any thermodynamic calculations.

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

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

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

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

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

In mks units, $R^* = 8314.5 Pa \cdot m^3 kg^{-1}K^{-1} = 8314.5(m/s)^2 K^{-1}$ We can also define a gas constant $R = R^*/M$ particular to the gas in question. For example, dry Earth air has a mean molecular weight of 28.97, so $R_{dryair} = 287 \ (m/s)^2 K^{-1}$, in mks units.

¹Pressure can equivalently be defined as the amount of momentum per unit area per unit time which passes *in* both directions through a small hoop placed in the gas. This definition is equivalent to the exerted-force definition because when a molecule with velocity v and mass m bounces elastically off a surface, the momentum change is 2mv, but only half of the molecules are moving toward the surface at any given time. The momentum flux definition, in contrast, counts molecules going through the hoop in both directions

If μ is measured in kilograms, then $1/\mu$ is the number of protons needed to make up a kilogram. This large number is known as a *Mole*, and is commonly used as a unit of measurement of numbers of molecules, just as one commonly counts eggs by the dozen. For any substance, a quantity of that substance whose mass in kilograms is equal to the molecular weight of the substance will contain one Mole of molecules. For example, 2 kg of H₂ is a Mole of Hydrogen molecules, while 32 kg of the most common form of O_2 is a Mole of molecular Oxygen. If n were measured in $Moles/m^3$ instead of molecules per m^3 , then density would be $\rho = n \cdot M$. One can also define the gram-mole (or mole for short), which is the number of protons needed to make a gram; this number is known as Avogadro's number, and is approximately $6.022 \cdot 10^{23}$.

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

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

$$p_A = kn_A T, p_B = kn_B T \tag{2.3}$$

or equivalently,

$$p_A = R_A \rho_A T, p_B = R_B \rho_B T \tag{2.4}$$

where $R_A = R^*/M_A$ and $R_B = R^*/M_B$. The same temperature appears in both equations, since thermodynamic equilibrium dictates that all components of the system have the same temperature. The ratio of partial pressures of any two components of a gas is a convenient way to describe the composition of the gas. From Eq. 2.3, $p_A/p_B = n_A/n_B$, so the ratio of partial pressure of A to that of B is also the ratio of number of molecules of A to the number of molecules of B. This ratio is called the *molar mixing ratio*. When we refer to a mixing ratio without qualification, we will generally mean the molar mixing ratio. Alternately, one can describe the composition in terms of the ratio of partial pressure of one component to total pressure of the gas $(p_A/(p_A + p_B))$ in the two-component example). Summing the two partial pressure equations in Eq. 2.3, we see that this is also the ratio of number of molecules of A to total number of molecules; hence we will use the term *molar concentration* for this ratio ². If η_A is the molar mixing ratio of A to B, then the molar concentration is $\eta_A/(1 + \eta_A)$, from which we see that for the molar concentration and molar mixing ratio are nearly the same for substances which are very dilute (i.e. $\eta_A \ll 1$).

Exercise 2.2.1 Show that a mixture of gases with molar concentrations $\eta_A = n_A/(n_A + n_B)$ and $\eta_B = n_B/(n_A + n_B)$ behaves like a perfect gas with mean molecular weight $M = \eta_A M_A + \eta_B M_B$.

²The term *volumetric* mixing ratio or concentration is often used interchangeably with the term *molar*, as in "ppmv" for "parts per million by volume." The reason for this nomenclature is that the volume occupied by a given quantity of gas at a fixed temperature and pressure is proportional to the number of molecules of the gas contained in that quantity. To see this, write n = N/V, where N is the number of molecules and V is the volume they occupy. Then, the ideal gas law can be written in the alternate form V = (kT/p)N. Hence the ratio of standardized volumes is equal to the molar mixing ratio, and so forth. Abbreviations like "ppmv" for molar mixing ratios are common and convenient, because the "v" can unambiguously remind us that we are talking about a volumetric (i.e. molar) mixing ratio or concentration, whereas in an abbreviation like "ppmm" one is left wondering whether the second "m" means "mass" or "molar."

(i.e. derive the expression relating total pressure $p_A + p_B$ to total density $\rho_A + \rho_B$ and identify the effective gas constant). Compute the mean molecular weight of dry Earth air. (Dry Earth air consists primarily of 78.084% N_2 , 20.947% O_2 , and .934% Ar, by count of molecules.)

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

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

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

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

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

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

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

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

sea level (1000mb)? At the top of Qomolangma (a.k.a. "Mt. Everest," about 300mb)? Does the temperature of the air (within reasonable limits) affect your answer much?

2.2.2 Specific heat and conservation of energy

Conservation of energy is one of the three great pillars upon which the edifice of thermodynamics rests. When expressed in terms of changes in the state of matter, it is known as the First Law of Thermodynamics. When a gas expands or contracts, it does work by pushing against the environment as its boundaries move. Since pressure is force per unit area, and work is force times distance, the work done in the course of an expansion of volume dV is pdV. This is the amount of energy that must be added to the parcel of gas to allow the increase in volume to take place. For atmospheric purposes, it is more convenient to do write all thermodynamic relations on a per unit mass basis. Dividing V by the mass contained in the volume yields ρ^{-1} , whence the work per unit mass of the substance while holding volume fixed changes the energy stored in the various motions of the molecules by an amount $c_v dT$, where c_v is a proportionality factor known as the specific heat at constant volume. For example it takes about 720 Joules of energy to raise the temperature of 1kg of air by 1K while holding the volume fixed. For ideal gases, the specific heat can depend on temperature, though the dependence is typically weak. For non-ideal gases, specific heat can depend on pressure as well.

Exercise 2.2.3 There are 20 students and one professor in a well-insulated classroom measuring 20 meters by 20 meters by 3 meters. Each person in the classroom puts out energy at a rate of 100 Watts (1 Watt = 1 Joule/second). The classroom is dark, except for a computer and LCD projector which together consume power at a rate of 200 Watts. The classroom is filled with air at a pressure of 1000mb (no extra charge). The room is sealed so no air can enter or leave, and has an initial temperature of 290K. How much does the temperature of the classroom rise during the course of a 1 hour lecture?

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

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

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

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

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

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

2.2.3 Entropy, reversibility and Potential temperature; The Second Law

One cannot use Eqn 2.8 to define a "heat content" Q of a state (p, T) relative to a reference state (p_o, T_o) , because the amount of heat needed to go from one state to another depends on the path in pressure-temperature space taken to get there; the right hand side of Eqn 2.8 is not an exact differential. However, it can be made into an exact differential by dividing the equation by T and using the perfect gas law as follows:

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

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

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

The Second Law endows the Universe with an arrow of time. If one watches a movie of a closed system and sees that the system starts with large fluctuations of temperature (low entropy) and proceeds to a state of uniform temperature (high entropy), one knows that time is running



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

forward. If one sees a thermally homogeneous object spontaneously generate large temperature inhomogeneities, then one knows that the movie is being run backwards. Note that the Second Law applies only to closed systems. The entropy of a subcomponent can decrease, if it exchanges energy with the outside world and increases the entropy of the rest of the Universe. This is how a refrigerator works.

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

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

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

To understand why the presence of cold air above warm air in the sounding of Figure 2.1 does not succumb immediately to instability, we need only look at the corresponding profile of potential temperature, shown in Figure 2.3. This figure shows that potential temperature increases monotonically with height. This profile tells us that the air aloft is cold, but that if it were pushed down to lower altitudes, compression would warm it to the point that it is warmer than the surrounding air, and thus being positively buoyant, will tend to float back up to its original level

rather than continuing its descent. We see also where the stratosphere gets its name: potential temperature increases very strongly with height there, so air parcels are very resistant to vertical displacement. This part of the atmosphere is therefore strongly stratified.

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

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

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

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

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

Exercise 2.2.4 (a) A commercial jet airliner cruises at an altitude of 300mb. The air outside has a temperature of 240K. To enable the passengers to breathe, the ambient air is compressed to

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

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

a cabin pressure of 1000mb. What would the cabin temperature be if the air were compressed adiabatically? How do you think airlines deal with this problem? (b) Discuss whether the lower portion of the Venus temperature profile shown in Figure 2.2 is on the dry CO_2 adiabat. Do the same for the Summer afternoon Mars sounding. (c) Assume that the Jupiter sounding is on a dry adiabat, and estimate the value of R/c_p for the atmosphere. Based on your result, what is the dominant constituent of the Jovian atmosphere likely to be? What other gas might be mixed with the dominant one?

2.3 Static stability of inhomogeneous mixtures

An atmosphere is statically unstable if an air parcel displaced from its original position tends to continue rising or sinking instead of returning to its original position. Such a state will tend to mix itself until it becomes stable. For a well-mixed atmosphere, the potential temperature profile tells the whole story about static stability, since, according to the ideal gas law, the density of an air parcel with potential temperature θ_0 will be $\rho_0 = p_1/(R\theta_0 \cdot (p_1/p_0)^{R/c_p})$ upon being elevated to an altitude with pressure $p_1 < p_0$. The ambient density there is $\rho_1 = p_1/(R\theta_1 \cdot (p_1/p_0)^{R/c_p})$. The displaced parcel will be negatively buoyant and return toward its original position if $\rho_0 > \rho_1$, which is true if and only if $\theta_0 < \theta_1$, i.e. if the potential temperature increases with height. For an inhomogeneous atmosphere, this is no longer the case, since the gas constant R depends on the mean molecular weight of the mixture, which varies from place to place. As an example, we may consider an atmosphere which has uniform θ , but which consists of pure N_2 for $p > p_0$ and pure CO_2 for $p < p_0$. In this case, the difference in density between a lifted N_2 parcel and that of the

2.4. THE HYDROSTATIC RELATION

surrounding CO_2 is

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

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

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

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

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

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

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

2.4 The hydrostatic relation

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

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

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

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

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

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

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

An important consequence of the hydrostatic relation is that it enables us to determine the total mass of an atmosphere through measurements of pressure taken at the surface alone. Integrating Eqn 2.15 from the ground $(p = p_s)$ to space (p = 0) yields the relation

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

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

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

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

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

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

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

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

2.5. THERMODYNAMICS OF PHASE CHANGE

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

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

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

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

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

2.5 Thermodynamics of phase change

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

Condensible substances play a central role in the atmospheres of many planets and satellites. On Earth, it is water that condenses, both into liquid water and ice. On Mars, CO_2 condenses into dry ice in clouds and in the form of frost at the surface. On Jupiter and Saturn, not only water but ammonia (NH_3) and a number of other substances condense. The thick clouds of Venus are composed of condensed sulfuric acid. On Titan it is methane, and on Neptune's moon Triton nitrogen itself condenses. Table 2.1 lists the latent heats for the liquid-vapor (evaporation), liquidsolid (fusion) and solid-vapor (sublimation) phase transitions are given for a number of common constituents of planetary atmospheres. Water has an unusually large latent heat; the condensation of 1 kg of water vapor into ice releases nearly five times as much energy as the condensation of 1 kg of carbon dioxide gas into dry ice. This is why the relatively small amount of water vapor in Earth's present atmosphere can nonetheless have a great effect on atmospheric structure and dynamics. Ammonia also has an unusually large latent heat, though not so much so as water. In both cases, the anomalous latent heat arises from the considerable energy needed to break hydrogen bonds in the condensed phase. Like most thermodynamic properties, latent heat varies somewhat with temperature. For example, the latent heat of vaporization of water is $2.5 \cdot 10^6 J/kg$ at 0C, but only $2.26 \cdot 10^6 J/kg$ at 100C. For precise calculations, the variation of latent heat must be taken into account, but nonetheless for many purposes it will be sufficient to assume latent heat to be constant over fairly broad temperature ranges.

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

Typically, the solid phase is more dense than the liquid phase, but water again is exceptional. Water ice floats on liquid water, whereas carbon dioxide ice would sink in an ocean of liquid carbon dioxide, and methane ice would sink in a methane lake on Titan. This has profound consequences for the climates of planets with a water ocean such as Earth has, since ice formed in winter remains near the surface where it can be more readily melted when summer arrives.

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

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

All gases are condensible at low enough temperatures or high enough pressures. On Earth (in the present climate) CO_2 is not a condensible substance, but on Mars it is. The ability of a gas to condense is characterized by the *saturation vapor pressure*, p_{sat} of that gas, which may be a function of any number of thermodynamic variables. When the partial pressure p_A of gas A is below $p_{sat,A}$, more of the gas can be added, raising the partial pressure, without causing condensation. However, once the partial pressure reaches $p_{sat,A}$, any further addition of A will condense out. The state $p_A = p_{sat,A}$ is referred to as "saturated" with regard to substance A. Each condensed state (e.g. liquid or solid) will have its own distinct saturation vapor pressure. Rather remarkably, for a mixture of perfect gases, the saturation vapor pressure of each component is independent of the presence of the other gases. Water vapor mixed with 1000 mb worth of dry air at a temperature of 300K will condense when it reaches a partial pressure of 38mb; a box of pure

water vapor at 300K condenses at precisely the same 38mb. If a substance "A" has partial pressure p_A that is below the saturation vapor pressure, it is said to be "subsaturated," or "unsaturated." The degree of subsaturation is measured by the saturation ratio $p_a/p_{sat,A}$, which is often stated as a percent. Applied to water vapor, this ratio is called the relative humidity, and one often speaks of the relative humidity of other substances, e.g. "methane relative humidity" instead of saturation ratios. Note that the relative humidity is also equal to the mixing ratio of the substance A in a given mixture to the mixing ratio the air would have if the substance were saturated. This is different from the ratio of specific humidity to saturation specific humidity, or the ratio of molar concentration molar concentration except when the mixing ratio is small.

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

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

where ρ_v is the density of the less condensed phase, ρ_c is the density of the more condensed phase, and L is the latent heat associated with the transformation to the more condensed phase. For vapor to liquid or solid transitions, $\rho_c \gg \rho_v$, enabling one to ignore the second term in the denominator of Eqn 2.21. Further, upon substituting for density from the perfect gas law, one obtains the simplified form

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

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

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

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

Figure 2.4 summarizes the features of a typical phase diagram. Over ranges of a few bars of pressure, the solid-liquid boundary can be considered nearly vertical. In fact the exact form of the Clausius-Clapeyron relation (Eq. 2.21) tells us why the boundary is nearly vertical and how



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



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

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

Below the triple point temperature, the favored transition is gas/solid, and so the appropriate latent heat to use in the Clausius-Clapevron relation is the latent heat of sublimation. Above the triple point, the favored transition is gas/liquid, whence one should use the latent heat of vaporization. The triple point (T, p) provides a convenient base for use with the simplified Clausius-Clapeyron solution in Eqn. 2.23, or indeed for a numerical integration of the relation with variable L. Results for water vapor are shown in Figure 2.5. These results were computed using the constant L approximation for sublimation and vaporization, but in fact a plot of the empirical results on a logarithmic plot of this type would not be distinguishable from the curves shown. The more exact result does differ from the constant L idealization by a few percent, which can be important in some applications. Be that as it may, the figure reveals the extreme sensitivity of vapor pressure to temperature. The vapor pressure ranges from about .1 Pascals at 200K (the tropical tropause temperature) to 35mb at a typical tropical surface temperature of 300K, rising further to 100mb at 320K. Over this span of temperatures, water ranges from a trace gas to a major constituent; at temperatures much above 320K, it rapidly becomes the dominant constituent of the atmosphere. Note also that the distinction between the ice and liquid phase transitions has a marked effect on the vapor pressure. Because the latent heat of sublimation is larger than the latent heat of vaporization, the vapor pressure over ice is lower than the vapor pressure over liquid would be, at subfreezing temperatures. At 200K, the ratio is nearly a factor of three.

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

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

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

2.6 The moist adiabat

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

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

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

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

where R is the gas constant for the substance making up the atmosphere. Without loss of generality, we may suppose that T_o is taken to be the surface temperature, so that $p_{sat}(T_o)$ is the surface

pressure p_s . Since the logarithm is negative, the temperature decreases with altitude (recalling that lower pressure corresponds to higher altitude). Further, the factor multiplying the logarithm is the ratio of the surface temperature to the characteristic temperature L/R. Since the characteristic temperature is large, the prefactor is small, and as a result the temperature of saturated adiabat for a one-component atmosphere varies very little over a great range of pressures. For example, in the case of the CO_2 vapor-ice transition, an atmospheric surface pressure of 7mb (similar to that of present Mars) would be in equilibrium with a surface dry-ice glacier at a temperature of 149K; at .07mb – one one-hundredth of the surface pressure – the temperature on the saturated adiabat would only fall to 122K.

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

Unless there is a reservoir of condensate at the surface to maintain saturation, it would be rare for an atmosphere to be saturated all the way to the ground. Suppose now that a onecomponent atmosphere has warm enough surface temperature that the surface pressure is lower than the saturation vapor pressure computed at the surface temperature. In this case, when a parcel is lifted by convection, its temperature will follow the dry or noncondensing adiabat, until the temperature falls so much that the gas becomes saturated. The level at which this occurs is called the lifted condensation level. Above the lifted condensation level, ascent causes condensation and the parcel follows the saturated adiabat. Since the temperature curve along the saturated adiabat falls with altitude so much less steeply than the dry adiabat, it is very easy for the two curves to intersect provided the surface temperature is not exceedingly large. An example for present Summer Martian conditions (specifically, like the warmest sounding in 2.2) is shown in Figure 2.6. A comparison with the Martian profiles in Figure 2.2 indicates that something interesting is going on in the Martian atmosphere. For the warm sounding, whose surface temperature is close to 255K, the entire atmosphere aloft is considerably warmer than the adiabat, and the temperature nowhere comes close to the condensation threshold, though the very lowest portion of the observed atmosphere, below the 200 Pa level, does follow the dry adiabat quite closely. Clearly, something we haven't taken into account is warming up the atmosphere. A likely candidate for the missing piece is the absorption of solar energy by CO_2 and dust.

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

As a final step up on the ladder of generality, let's consider a mixture of a condensible substance with a substance that doesn't condense under the range of temperatures encountered in the atmosphere under consideration. This might be a mixture of condensible methane on Titan with non-condensible nitrogen, or condensible carbon dioxide on Early Mars with noncondensible nitrogen, or water vapor on Earth with a non-condensible mixture of oxygen and nitrogen. Whatever the substance, we distinguish the properties of the condensible substance with the subscript "c," and those of the non-condensible substance by the subscript "a" (for "air"). We now need to do the energy budget for a parcel of the mixture, assuming that it has been cooled down enough for the condensible substance to reach saturation, so that any further cooling results in formation of enough condensate (with concomitant release of latent heat) to keep the system from becoming supersaturated. We further introduce the assumption that essentially all of the condensate is immediately removed from the system, so that the heat storage in whatever mass of condensate is left in suspension may be neglected. This is a reasonable approximation



Figure 2.6: The adiabatic profile for a pure CO_2 atmosphere with a surface pressure of 450 Pa (4.5mb) and a surface temperature of 255K. The conditions are similar to those encountered on the warmer portions of present-day Mars.

for water or ice clouds on Earth, but even in that case the slight effect of the mass of retained condensate on buoyancy can be significant in some circumstances. In other planetary atmospheres the effect of retained condensate could be of greater importance. The temperature profile obtained by assuming condensate is removed from the system is called a *pseudoadiabat*, because the process is not truly reversible. One cannot return to the original saturated state, because the condensate is lost. At the opposite extreme, if all condensate is retained, it can be re-evaporated when the parcel is compressed, allowing for true reversibility.

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

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

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

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

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

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

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

that

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

and

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

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

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

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

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

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

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

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

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

Figure 2.7 shows a family of solutions to Eqn 2.30, for the case of water vapor in Earth air. When the surface temperature is 250K, there is so little moisture in the atmosphere that the profile looks like the dry adiabat right to the ground. As temperature is increased, a region of weak gradients appears near the ground, representing the effect of latent heat on temperature. This layer gets progressively deeper as temperature increases and the moisture content of the atmosphere increases. When the surface temperature is 350K, so much moisture has entered the atmosphere that the surface pressure has actually increased to over 1300mb. Moreover, the moisture-dominated

94



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

region extends all the way to 10 Pa (.1mb), and even at 100 Pa (1mb) the atmosphere is 10% water by volume. Thus, for moderate surface temperatures, there is little water high up in the atmosphere. When the surface temperature approaches or exceeds 350K, though, the "cold trap" is lost, and a great deal of water is found aloft, where it is exposed to the destructive ultraviolet light of the sun and the possibility of thermal escape to space. In subsequent chapters, it will be seen that this phenomenon plays a major role in the life cycle of planets, and probably accounts for the present hot, dry state of Venus.

2.7 For Further Reading

CHAPTER 2. THERMODYNAMICS IN A NUTSHELL

Chapter 3

Elementary models of radiation balance

3.1 Energy balance and temperature

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

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

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

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

3.2 Blackbody radiation

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

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

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

If a field of radiation consists of a mixture of different frequencies and directions, the mixture is characterized by a *spectrum*, which is a function describing the proportions of each type of radiation making up the blend. A spectrum is a *density* describing the amount of electromagnetic energy contained in a unit volume of the space (3D position, frequency, direction) needed to characterize the radiation, just as the mass density of a three dimensional object describes the distribution of mass in three-dimensional space.

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



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.

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 Pwhich 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; $0 \ge \theta \le \pi$ and $0 \ge \phi \le 2\pi$. 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},ν,\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 $c\Sigma dAd\nu d\Omega$, which defines the flux spectrum $c\Sigma$. In mks units, the flux spectrum has units of $(Watts/m^2)/(Hz \cdot steradian)$, where the Hertz (Hz) is the unit of frequency, equal to one cycle per second. 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 $1 \text{ Newton} \cdot \text{meter/sec.}$ A Watt (*W*) is 1J/sec. A typical resting human in not-too-cold weather requires about 2000Calories/day. (A *Calorie* is the amount of energy needed to increase the temperature of 1Kg of pure water by 1K.) Convert this to a power consumption in *W*, using the fact that 1Calorie = 4184J.

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

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

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

$$B(\nu,T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1}$$
(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° 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 6240GigaHz, corresponding to a wavenumber of $20814m^{-1}$, or a wavelength of 48microns; this is in the far infrared range.

In terms of u, the Planck function can be rewritten

$$B(\nu,T) = \frac{2k^3T^3}{h^2c^2} \frac{u^3}{e^u - 1}$$
(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 $\nu \approx (2.821k/h)T \approx$ $58.78 \cdot 10^9T$. 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 $9.9 \cdot 10^{-9}$ of the peak emission. The microwave emission from a portion of the Earth's atmosphere with temperature 250K (having peak emission in the infrared) is readily detectable by satellites, whereas the emission of visible light is not.

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



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.

If one transforms to wavelength space, however,

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

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

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

$$F_{cum}(\nu,T) = \int_0^\nu \pi B(\nu',T)d\nu' = \int_\infty^\lambda \pi B_\lambda(\lambda',T)d\lambda'$$
(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 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 Bang¹. 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:

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

where ${}^2 \sigma = 2\pi^5 k^4/(15c^2h^3) \approx 5.67 \cdot 10^{-8} Wm^{-2} K^{-4}$. 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

 $^{^{1}}$ 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.

²The 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 $6\zeta(4) = \pi^4/15$

3.3. RADIATION BALANCE OF PLANETS

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:

- The only source of energy heating the planet is absorption of light from the planet's host star.
- The *albedo*, or proportion of sunlight reflected, is spatially uniform.

- The planet is spherical, and has a distinct solid or liquid surface which radiates like a perfect blackbody.
- The planet's temperature is uniform over its entire surface.
- The planet's atmosphere is perfectly transparent to the electromagnetic energy emitted by the surface.

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

Light leaving the upper layers of the Sun and 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 parallelbeam 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 58,000,000 km, the Sun (whose radius is 695,000 km) occupies an angular width in the sky of only about 2.695,000/58,000,000 radians, or 1.4° .

The solar ³ flux impinging on the planet is also reduced, as compared to the solar flux leaving the photosphere of the star. The total energy per unit frequency leaving the star is $4\pi r_{\odot}^2(\pi B(\nu, T_{\odot}))$, where r_{\odot} is the radius of the star and T_{\odot} is the temperature of its photosphere. At a distance r from the star, the energy has spread uniformly over a sphere whose surface area is $4\pi r^2$; hence at this distance, the energy flux per unit frequency is $\pi B r_{\odot}^2/r^2$, and the total flux is $\sigma T_{\odot}^4 r_{\odot}^2/r^2$. The latter is the flux seen by a planet at orbital distance r, in the form of a beam of parallel rays. It is known as the solar "constant", and will be denoted by L_{\odot} , 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 $I_{\odot} = 4\pi r_{\odot}^2 T_{\odot}^4$.

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

106

 $^{^{3}}$ We'll often use the adjective "solar" to refer to properties of starlight in general, and not just that from our own Sun.

3.3. RADIATION BALANCE OF PLANETS

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

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

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

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

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

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



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.

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 $\frac{r}{r_{\odot}}$ as low as 5. Such planets can have equilibrium blackbody temperatures as much as a third that of the photosphere of the parent star. For these planets, the distinction between the behavior of incoming and outgoing radiation is less sharp.

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



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.

the surface.

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

We will now consider an idealized thought experiment which illustrates the essence of the way an atmosphere affects OLR. Suppose that the atmosphere has a temperature profile T(p) which decreases with altitude, according to the dry or moist adiabat. Let p_s be the surface pressure, and suppose that the ground is strongly thermally coupled to the atmosphere by turbulent heat exchanges, so that the ground temperature cannot deviate much from that of the immediately overlying air. Thus, $T_s = T(p_s)$. If the atmosphere were transparent to infrared, as is very nearly the case for nitrogen or oxygen, the OLR would be σT_s^4 . Now, let's stir an additional gas into the portion of the atmosphere between the ground and a pressure $p_{rad} < p_s$, and suppose that the gas is transparent to solar radiation, but interacts so strongly with infrared that it turns each portion of the atmosphere it is mixed with into a perfect blackbody. A gas which is fairly transparent to the incoming shortwave stellar radiation but which interacts strongly with the outgoing (generally infrared) emitted radiation is called a greenhouse gas. Carbon dioxide, water vapor and methane are some examples of greenhouse gases, and the molecular properties that make a substance a good



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.

greenhouse gas will be discussed in Chapter 4. If one imagines slicing the atmosphere into a number layers so thin that they are essentially isothermal, then each layer with pressure greater than or equal to p_{rad} radiates like an ideal blackbody at its own temperature, but it is only the topmost of these layers that determines the radiation loss to space, since radiation from all the others is absorbed before it reaches the topmost layer. Since the topmost layer has temperature $T(p_{rad})$ and higher altitude layers are assumed transparent to infrared, the OLR is $\sigma T(p_{rad})^4$, which is less than σT_s^4 to the extent that $p_{rad} < p_s$. As shown in Figure 3.5, a greenhouse gas acts like an insulating blanket, reducing the rate of energy loss to space at any given surface temperature. The equilibrium surface temperature of a planet with a greenhouse gas in its atmosphere must be greater than that of a planet without a greenhouse gas, in order to radiate away energy at a sufficient rate to balance the absorbed solar radiation.

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



Figure 3.7: 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.

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

Figure 3.7 illustrates the reduction of infrared emission caused by the Earth's atmosphere. At every latitude, the observed OLR is much less than it would be if the planet radiated to space at its observed surface temperature. At the Equator the observed OLR is $238W/m^2$, 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.

3.3. RADIATION BALANCE OF PLANETS

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

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

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

	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	$47 \mathrm{K}$

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.

The concept of radiating level and radiating temperature also enables us to make sense of the way energy balance constrains the climates of gas giants like Jupiter and Saturn, which have no distinct surface. The essence of the calculation we have already done for rocky planets is to use the top of atmosphere energy budget to determine the parameters of the adiabat, and then extrapolate temperature to the surface along the adiabat. For a non-condensing adiabat, the atmospheric profile compatible with energy balance is $T(p) = T_{rad}(p/p_{rad})^{R/c_p}$. This remains the appropriate temperature profile for a (noncondensing) convecting outer layer of a gas giant, and the only difference with the previous case is that, for a gas giant, there is no surface to act as a natural lower boundary for the adiabatic region. At some depth, convection will give out and the adiabat must be matched to some other temperature model in order to determine the base of the convecting region, and to determine the temperature of deeper regions. There is no longer any distinct surface to be warmed by the greenhouse effect, but the greenhouse gas concentration of the atmosphere nonetheless affects T(p) through p_{rad} . For example, adding some additional 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 noncondensing adiabat is a one-parameter family of temperature profiles. The saturated adiabat for a mixture of condensing and noncondensing gases is also a one parameter family, defined by Eq. 2.30, and can therefore be treated similarly. If the appropriate adiabat for the planet had more than one free parameter, additional information beyond the energy budget would be needed to close the problem. On the other hand, a single component condensing atmosphere such as described by Eq. 2.24 yields a temperature profile with no free parameters that can be adjusted so as to satisfy the energy budget. The consequences of this quandary will be taken up as part of our discussion of the runaway greenhouse phenonenon, in Chapter 4.

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

3.3. RADIATION BALANCE OF PLANETS

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 ⁴. This is opposite from the imbalance that would be caused by an internal heating. It is a direct consequence of the rapid rise of CO_2 and other greenhouse gases, caused by the bustling activities of Earth's human inhabitants. The rapid greenhouse gas increase has cut down the OLR, but because of the time required to warm up the oceans and melt ice, the Earth's temperature has not yet risen enough to restore the energy balance.

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

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

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

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

 $^{^{4}}$ At the time of writing, top-of-atmosphere satellite measurements are not sufficiently accurate to permit direct observation of this imbalance

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

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

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 $T_s = T_{freeze}$

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.

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

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

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

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

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

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

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

This determines T_s as before, with the important difference that the Solar absorption on the left hand side is now a function of T_s instead of being a constant. Analogously to Fig. 3.5, the equilibrium surface temperature can be found by plotting the absorbed Solar radiation and the OLR vs. T_s on the same graph. This is done in Fig. 3.8, for four different choices of L_{\odot} . In this plot, we have taken $OLR = \sigma T^4$, which assumes no greenhouse effect ⁵. In contrast with the fixedalbedo case, the ice-albedo feedback allows the climate system to have multiple equilibria: there can be more than one climate compatible with a given Solar constant, and additional information is required to determine which state the planet actually settles into. The nature of the equilibria depends on L_{\odot} . When L_{\odot} is sufficiently small (as in the case $L_{\odot} = 1516W/m^2$ in Fig. 3.8) there is only one solution, which is a very cold globally ice-covered Snowball state, marked Sn_1 on the graph. Note that the Solar constant that produces a unique Snowball state exceeds the present Solar constant at Earth's orbit. Thus, were it not for the greenhouse effect, Earth would be in such a state, and would have been for its entire history. When L_odot is sufficiently large (as in the case $L_{\odot} = 2865 W/m^2$ 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_{\odot} , there are three solutions: a Snowball state (Sn_2) , a partially ice covered state with a relatively large ice sheet (e.g. A), and a warmer state (e.g. B) which may have a small ice sheet or be ice free, depending on the precise value of L_{\odot} . In the intermediate range of Solar constant, the warmest state is suggestive of the present or Pleistocene climate when there is a small ice-cap, and suggestive of Cretaceous-type hothouse climates when it is ice-free. In either case, the frigid Snowball state is available as an alternate possibility.

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

 $^{^{5}}$ 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.

a system changes discontinuously as some control parameter is continuously varied, is an example of a *bifurcation*.

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

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

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

which is positive by assumption. Thus,

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

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

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

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

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



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

Moreover, if the net flux G(T) is continuous 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 = 0should 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 < 0at 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, $dG/dT \ge 0$ at T_b . Thus, the state T_b is either stable or marginally stable, which proves our result. The proof goes through similarly if T_a is unstable. Note that we didn't actually need to make use of the condition that dG/dT be continuous everywhere: it's enough that it be continuous near the equilibria, so we can actually tolerate a few kinks in the curve.

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

3.4.1 Faint Young Sun, Snowball Earth and Hysteresis

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

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

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

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

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



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



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

3.4. ICE-ALBEDO FEEDBACK

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

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

The fact that the freeze-thaw cycle can exhibit hysteresis as atmospheric composition changes is at the heart of the Snowball Earth phenomenon. An initially warm state can fall into a globally glaciated Snowball if the atmospheric composition changes in such a way as to sufficiently weaken the greenhouse effect. Once the threshold is reached, the planet can fall into a Snowball relatively quickly – in a matter of a thousand years or less – since sea ice can form quickly. However, to deglaciate the Snowball, the greenhouse effect must be increased far beyond the threshold value at which the planet originally entered the Snowball state. Atmospheric composition must change drastically in order to achieve such a great increase, and this typically takes many millions of years. When deglaciation finally occurs, it leaves the atmosphere in a hyper-warm state, which only gradually returns to normal as the atmospheric composition evolves in such a way as to reduce the greenhouse effect. As discussed in Chapter 1, there are two periods in Earth's past when geological evidence suggests that one or more Snowball freeze-thaw cycles may have occurred. The first is in the Paleoproterozoic, around 2 billion years ago. At this time, $L_{\odot} \approx 1170 W/m^2$, 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, $L_{\odot} \approx 1290 W/m^2$ 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 9.

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

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

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

Another general lesson to be drawn from the preceding discussion is that the state with a stable, small icecap is very fragile, in the sense that the planetary conditions must be tuned rather precisely for the state to exist at all. For example, with the present Solar constant, the stable small icecap solution first appears when p_{rad} falls below 690mb. However, the icecap shrinks to zero as p_{rad} is reduced somewhat more, to 615mb. Hence, a moderate strengthening in the greenhouse effect would, according to the simple energy balance model, eliminate the polar ice entirely and throw the Earth into an ice-free Cretaceous hothouse state. The transition to an ice-free state of this sort is continuous 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 9. It is worth noting that small-icecap states like those of the past two million years appear to be relatively uncommon in the most recent half billion years of Earth's history, for which data is good enough to render a judgement about ice cover. The typical state appears to be more like the warm relatively ice-free states of the Cretaceous, and perhaps this reflects the fragility of the small-icecap state.

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

3.4.2 Climate sensitivity, radiative forcing and feedback

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

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

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

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

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

With this expression, Eq. 3.14 can be rewritten

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

where

$$\Phi = \frac{1}{4} L \frac{\frac{\partial \alpha}{\partial T}}{\frac{\partial OLR}{\partial T}}$$
(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 $\Phi \rightarrow -1$, 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 $L \approx 1330W/m^2$, $T \approx 277K$, $\Phi \rightarrow -1$ and the sensitivity becomes infinite. This divergence merely reflects the fact that the temperature curve is vertical at the bifurcation point. Near such points, the temperature change is no longer linear in radiative forcing. It can easily be shown that the temperature varies as the square root of radiative forcing near a bifurcation point, as suggested by the plot.

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

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

$$\Phi = \frac{\partial G}{\partial R} \frac{\partial R}{\partial T} \frac{\partial G}{\partial T}$$
(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

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

128

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

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

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

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

The Planck function is unambigously the natural choice of a weighting function for defining the mean emissivity \bar{e} for an object with temperature T. There is no such unique choice for defining the mean absorptivity over all frequencies and directions. The appropriate weighting function is determined by the frequency and directional spectrum of the incident radiation which requires a detailed knowledge of its source. If the incident radiation is a blackbody with temperature T_{source} then \bar{a} should be defined with a formula like Eq. 3.20, using $B(\nu, T_{source})$ as the weighting function. Note that the weighting function is defined by the temperature of the source rather than by the temperature of the the object doing the absorbing. As was the case for mean emissivity, the temperature dependence of the weighting function implies that \bar{a} will vary with T_{source} even if $a = a(\nu)$ and is not explicitly dependent on temperature.

Absorptivity and emissivity might appear to be independent characteristics of an object, but observations and theoretical arguments reveal an intimate relation between the two. This relation, expressed by *Kirchhoff's Law of Radiation* is a profound property of the interaction of radiation with matter that lies at the heart of all radiative transfer theory. Kirchoff's Law states that the emissivity of a substance at any given frequency equals the 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



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 $e = \bar{e}(T_1)$.

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 he presented a formal justification that eliminated the involvement of hypothetical ideal selective mirrors. The physical content of Hilbert's proof is that one doesn't need an ideal mirror, if one requires that a sufficient variety of materials with different absorbing and emitting properties will all come into an isothermal state at equilibrium. Hilbert's deriviation nonetheless relied on an assumption that radiation would come into equilibrium with matter at each individual wavelength considered separately. While Kirchoff did the trick with mirrors, Hilbert, in essence, did the trick with axioms instead, leaving the microscopic justification of Kirchoff's Law equally obscure. It is in fact quite difficult to provide a precise and concise statement of the circumstances in which a material will comply with Kirchoff's Law. Violations are quite commonplace in nature and in engineered materials, since it is quite possible for a material to store absorbed electromagnetic energy and emit it later, perhaps at a quite different frequency. A few examples that come to mind are phosphorescent ("glow in the dark") materials, fluorescence (e.g. paints that glow when exposed to ultraviolet, or "black" light), frequency doubling materials (used in making green laser pointers), and lasers themselves. In Nature, such phenomena involve insignificant amounts of energy, and are of no known importance in determining the energy balance of planets. We will content ourselves here with the statement that all known liquid and solid planetary materials, as well as the gases making up atmospheres, conform very well to Kirchoff's Law, except perhaps in the most tenuous outer reaches of atmospheres where the gas itself is not in thermodynamic equilibrium.



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.

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

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

Hence,

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

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

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

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

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



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

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

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

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

The stratosphere in the preceding calculation differs from the observed stratosphere of Earth in that it is isothermal rather than warming with altitude. The factor we have left out is that real stratospheres often contain constituents that absorb solar radiation. To rectify this shortcoming, let's consider the effect of solar absorption on the temperature of the skin layer. Let e_{ir} be the infrared emissivity, which is still assumed small, and a_{sw} be the shortwave (mostly visible) absorptivity, which will also be assumed small. Note that Kirchhoff's Law does not require $e_{ir} = a_{sw}$, as the emissivity and absorptivity are at different wavelengths. The solar absorption of incident radiation is $a_{sw}S$. We'll assume that the portion of the solar spectrum which is absorbed by the atmosphere is absorbed so strongly that it is completely absorbed before reaching the ground. This is in fact the typical situation for solar near-infrared and ultraviolet. In this case, one need not take into account absorption of the upwelling solar radiation reflected from the surface.

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

The energy balance for the skin layer now reads

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

Hence,

$$T = T_{skin} \left(1 + \frac{a_{sw}}{e_{ir}} \frac{S}{OLR}\right)^{\frac{1}{4}}$$
(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 26Kbelow the skin temperature. If anything, one might have expected the tropical temperatures to exceed the global mean skin temperatures, because the local tropospheric temperatures are warmer

	Skin temperature
Venus	213.
Earth	214.
Mars $(255K \text{ sfc})$	214.
Jupiter	106.
Titan	72.

Table 3.3: Computed skin temperatures of selected planets.

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

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

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

We have been using the term "stratosphere" rather loosely, without having attempted a precise definition. It is commonly said, drawing on experience with Earth's atmosphere, that a stratosphere is an atmosphereic layer within which temperature increases with height. This would be an overly restrictive and Earth-centric definition. The dynamically important thing about a stratosphere is that it is much more stably stratified than the troposphere, i.e. that its temperature goes down less steeply than the adiabat appropriate to the planet under consideration. The stable stratification of a layer indicates that convection and other dynamical stirring mechanisms are ineffective or absent in that layer, since otherwise the potential temperature would become well mixed and the temperature profile would become adiabatic. An isothermal layer is stably stratified, because its potential temperature increases with height; even a layer like that of Mars' upper atmosphere, whose temperature decreases gently with height, can be stably stratified. We have shown that an optically thin stratosphere is isothermal in the absence of solar absorption. Indeed, this is often taken as a back-of-the envelope model of stratospheres in general, in simple calculations. In the next chapter, we will determine the temperature profile of stratospheres that are not optically thin.

3.7. FOR FURTHER READING

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

We conclude this chapter with a few comparisons of observed tropopause heights with the predictions of the optically thin limit. We'll leave Venus out of this comparison, since its atmosphere is about as far from the optically thin limit as one could get. On Mars, using the dry adiabat for CO_2 and a 5mb surface pressure puts the trop pause 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 trop pause 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 7, but we can already estimate the effect of using the moist adiabat. Using the computation of the water-vapor/air moist adiabat described in Chapter 2, the tropopause rises to 157mb, based on a typical tropical surface temperature of 300K and the skin temperature estimated in Table 3.3. This is much closer to the observed tropopause (defined as the temperature minimum in the sounding), with the remaining mismatch being accounted for by the fact that the minimum temperature is appreciably colder than the skin temperature.

3.7 For Further Reading
Chapter 4

Radiative transfer in temperature-stratified atmospheres

4.1 Overview

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

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

In contrast, planets do not emit significant amounts of radiation in the shortwave spectrum, though shortwave scattering by molecules and clouds is invariably significant; absorption of shortwave radiation arises from quite different molecular processes than those involved in infrared absorption, and its wavelength-dependence has a correspondingly different character. Moreover, solar radiation generally reaches the planet in the form of a nearly parallel beam, whereas infrared from thermal emission by the planet and it's atmosphere is more nearly isotropic. The approximations pertinent to shortwave radiation will be taken up in Chapter 5, where we will also consider the effects of scattering on thermal infrared.

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

4.2 Basic Formulation of Plane Parallel Radiative Transfer

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

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

4.2.1 Optical thickness and the Schwarzschild equations

Although the radiation field varies in space only as a function of pressure, p, its intensity depends also on direction. Let $I(p, \hat{n}, \nu)$ be the flux density of electromagnetic radiation propagating in direction \hat{n} , measured at point p. This density is just like the Planck function $B(\nu, T)$, except that we allow it to depend on direction and position. The technical term for this flux density is *spectral irradiance*. Now we suppose that the radiation propagates through a thin layer of atmosphere of thickness δp as measured by pressure. The absorption of energy at frequency ν is proportional to the number of molecules of absorber encountered; assuming the mixing ratio of the absorber to be constant within the layer for small δp , the number of molecules encountered will be proportional to δp , in accord with the hydrostatic law. By Kirchoff's law, the absorptivity and emissivity of the layer are the same; we'll call the value $\delta \tau_{\nu}$, and keep in mind that in general it will be a function of ν . Let θ be the angle between the direction of propagation \hat{n} and the vertical, as shown in Figure 4.1. Now, let $\Delta \tau_{\nu}^*$ be the emissivity (and absorptivity) of the layer for radiation propagating in the direction $\theta = 0$. We may define the proportionality between emissivity and pressure through the relation $\delta \tau_{\nu}^* = -\kappa \Delta p/g$ where g is the acceleration of gravity and κ is an absorption coefficient. It has units of area per unit mass, and can be thought of as an absorption cross-section per unit mass – in essence, the area taken out of the incident beam by the absorbers contained in a unit mass of atmosphere. In general κ is a function of frequency, pressure, temperature and the mixing ratios of the various greenhouse gases in the atmosphere. Passing to the limit of small δp , we can define an *optical thickness* coordinate through the differential equation

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

Since pressure decreases with altitude, τ_{ν}^* increases with altitude. Radiation propagating at an angle θ relative to the vertical acts just like vertically propagating radiation, except that the thickness of each layer through which the beam propagates, and hence the number of absorbing molecules encountered, is increased by a factor of $1/\cos\theta$. Hence, the optical thickness for radiation propagating with angle θ is simply $\tau_{\nu} = \tau_{\nu}^*/\cos\theta$. The equations of radiative transfer can be simplified by using either τ_{ν} or τ_{ν}^* in place of pressure as the vertical coordinate.

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

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

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

Eq. 4.1 defines the optical thickness $\tau_{\nu}^{*}(p_{1}, p_{2})$ for the layer between pressures p_{1} and p_{2} . Unless κ is constant, it is not proportional to $|p_{1} - p_{2}|$, but it is a general consequence of the definition that $\tau_{\nu}^{*}(p_{1}, p_{2}) = \tau_{\nu}^{*}(p_{1}, p') + \tau_{\nu}^{*}(p', p_{2})$ if p' is between p_{1} and p_{2} . Consider an atmosphere with a single greenhouse gas having concentration q(p). Then, if κ_{G} is independent of p the optical thickness can be expressed as $\tau_{\nu}^{*}(p_{1}, p_{2}) = \kappa_{G}\ell$ where ℓ is the *path*, defined by

$$\ell(p_1, p_2) = -\int_{p_1}^{p_2} q(p) \frac{dp}{g}$$
(4.3)

The boundaries of the layer are generally chosen so as to make the path and optical thickness positive. The path is the mass of greenhouse gas in the layer, per square meter of the planet's surface, and in *mks* units has units of kg/m^2 . If the greenhouse gas is well mixed then $\ell = q \cdot (p_1 - p_2)/g$. Now, it often happens that κ_G increases linearly with pressure – a phenomenon known as *pressure broadening* (alternatively *collisional broadening* for reasons that will eventually become clear. If we write $\kappa_G(p) = \kappa_G(p_o) \cdot (p/p_o)$, then we can define an *equivalent path*

$$\ell_e = -\int_{p_1}^{p_2} q(p) \frac{p}{p_o} \frac{dp}{g}$$
(4.4)

such that $\tau_{\nu}^{*}(p_1, p_2) = \kappa_G(p_o)\ell_e$ much as before. The equivalent path still has units of mass per unit area, but because of the pressure weighting will differ from the actual path. For example, if the greenhouse gas is well-mixed then

$$\ell_e = \frac{1}{g} q \frac{1}{2p_o} (p_1^2 - p_2^2) = q \frac{p_1 - p_2}{g} \frac{p_1 + p_2}{2p_o}$$
(4.5)



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

The equivalent path is thus the actual path weighted by the ratio of the mean pressure to the reference pressure.

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

$$I(\tau_{\nu}^{*} + \delta \tau_{\nu}^{*}, \hat{n}, \nu) = (1 - \frac{\delta \tau_{\nu}^{*}}{\cos \theta})I(\tau_{\nu}^{*}, \hat{n}, \nu) + B(\nu, T(\tau_{\nu}^{*}))\frac{\delta \tau_{\nu}^{*}}{\cos \theta}$$
(4.6)

or, passing to the limit of small $\delta \tau_{\nu}^*$,

$$\frac{d}{d\tau_{\nu}^{*}}I(\tau_{\nu}^{*},\hat{n},\nu) = -\frac{1}{\cos\theta}[I(\tau_{\nu}^{*},\hat{n},\nu) - B(\nu,T(\tau_{\nu}^{*}))]$$
(4.7)

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

We can derive an equation for the net upward flux per unit frequency, I_+ , by multiplying Eq. 4.7 by $\cos \theta$ and integrating over all solid angles in the upward-pointing hemisphere. Integrating over the downward hemisphere yields the net downward flux I_- . However, because of the factor $1/\cos \theta$ on the right hand side of Eq. 4.7, the hemispherically averaged intensity appearing on the right hand side is not I_+ . Instead, it is $\int I(\tau^*, \hat{n}, \nu) d\Omega$, or equivalently $\int_0^{\pi/2} 2\pi I(\tau^*, \theta, \nu) \sin \theta d\theta$. One cannot proceed further without some assumption about the angular distribution. If we assume that the distribution remains approximately isotropic, by virtue of the isotropic source B, then

 $I(\tau^*, \theta, \nu)$ is independent of θ , and hence the problematic integral becomes $2\pi I \int_0^{\pi/2} \sin \theta d\theta$ which is equal to $2I_+$ under the assumption of isotropy. This result yields a closed equation for I_+ . It states that, if the radiation field remains approximately isotropic, the decay rate is the same as for unidirectional radiation propagating with an angle $\bar{\theta}$ such that $\cos \bar{\theta} = \frac{1}{2}$, i.e. $\bar{\theta} = 60^{\circ}$. From now on we will deal only with this approximate angle-averaged form of the equations, and use $\tau_{\nu} = \tau^*/\cos \bar{\theta}$ as our vertical coordinate. The choice $\cos \bar{\theta} = \frac{1}{2}$ is by no means a unique consequence of the assumption of isotropy. The fact is that an isotropic distribution is not an exact solution of Eq. 4.7 except in a few very special limits, so that the choice we make is between different errors of roughly the same magnitude. If we had calculated $\cos \bar{\theta}$ by multiplying Eq. 4.7 by $(\cos \theta)^2$ instead of $\cos \theta$ before averaging over angles, we would have concluded $\cos \bar{\theta} = \frac{2}{3}$ and this would be an equally valid choice within the limitations of the isotropic approximation. Sometimes, a judicious choice of $\cos \bar{\theta}$ is used to maximize the fit to an angle-resolved calculation in some regime of particular interest. For the most part we will simply use $\cos \bar{\theta} = \frac{1}{2}$ in our calculations unless there is a compelling reason to adopt a different value.

In terms of τ_{ν} , the equations for the upward and downward flux are

$$\frac{d}{d\tau_{\nu}}I_{+} = -I_{+} + \pi B(\nu, T(\tau_{\nu}))$$

$$\frac{d}{d\tau_{\nu}}I_{-} = I_{-} - \pi B(\nu, T(\tau_{\nu}))$$
(4.8)

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

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

$$I_{+}(\tau_{\nu},\nu) = I_{+}(0)e^{-\tau_{\nu}} + \int_{0}^{\tau_{\nu}} \pi B(\nu, T(\tau_{\nu}'))e^{-(\tau_{\nu} - \tau_{\nu}')}d\tau_{\nu}'$$

$$I_{-}(\tau_{\nu},\nu) = I_{-}(\tau_{\infty})e^{-(\tau_{\infty} - \tau_{\nu})} + \int_{\tau_{\nu}}^{\tau_{\infty}} \pi B(\nu, T(\tau_{\nu}'))e^{-(\tau_{\nu}' - \tau_{\nu})}d\tau_{\nu}'$$
(4.9)

where τ'_{ν} is a dummy variable and τ_{∞} is the optical thickness of the entire atmosphere, i.e. $\tau^*_{\nu}(p_s, 0)/\cos \bar{\theta}$. Note that τ_{∞} depends on ν in general, though we have suppressed the subscript for the sake of readability. The top of the atmosphere (p = 0) is at τ_{∞} . The physical content of these equations is simple: $I_+(\tau_{\nu}, \nu)$ consists of two parts. The first is the portion of the emission from the ground which is transmitted by the atmosphere (the first term in the expression for I_+). The second is the radiation emitted by the atmosphere itself, which appears as an exponentially-weighted average (the second term in the expression for I_+) of the emission from all layers below τ_{ν} , with more distant layers given progressively smaller weights. Similarly, $I_-(\tau_{\nu}, \nu)$ is an exponentially-weighted average of the emission from all layers above τ_{ν} , plus the transmission of incident downward flux. The atmospheric emission to space will be most sensitive to temperatures fairly opaque to the radiation emitted from the ground, whereas the transmitted ground emission will dominate when the atmosphere is fairly transparent. The downward radiation into the ground will be most sensitive to temperatures nearest the ground.

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

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

The weighting function appearing in the integrands in Eq. 4.9 is the *transmission function*. Written as a function of pressure, it is

$$\mathfrak{T}_{\nu}(p_1, p_2) = e^{-|\tau_{\nu}(p_1) - \tau_{\nu}(p_2)|} \tag{4.10}$$

 $\mathfrak{T}_{\nu}(p_1, p_2)$ is the proportion of incident energy flux at frequency ν which is transmitted through a layer of atmosphere extending from p_1 to p_2 ; whatever is not transmitted is absorbed in the layer. Note that $\mathfrak{T}_{\nu}(p, p')d\tau'_{\nu} = d\mathfrak{T}_{\nu}$ (with p held constant), if p < p', and $\mathfrak{T}_{\nu}(p, p')d\tau'_{\nu} = -d\mathfrak{T}_{\nu}$ if p > p'. Using this result Eq. 4.9 can be re-written

$$I_{+}(\tau_{\nu},\nu) = I_{+,s}(\nu)\mathfrak{T}_{\nu}(p,p_{s}) - \int_{p'=p}^{p_{s}} \pi B(\nu,T(p'))d\mathfrak{T}_{\nu}(p,p')$$

$$I_{-}(\tau_{\nu},\nu) = I_{-,\infty}(\nu)\mathfrak{T}_{\nu}(0,p) + \int_{p'=0}^{p} \pi B(\nu,T(p'))d\mathfrak{T}_{\nu}(p,p')$$
(4.11)

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

$$I_{+}(p,\nu) = \pi B(\nu,T(p)) + (I_{+,s}(\nu) - \pi B(\nu,T_{sa}))\mathfrak{T}_{\nu}(p,p_{s}) + \int_{p}^{p_{s}} \pi \mathfrak{T}_{\nu}(p,p') dB(\nu,T(p'))$$

$$I_{-}(p,\nu) = \pi B(\nu,T(p)) + (I_{-,\infty}(\nu) - \pi B(\nu,T_{\infty}))\mathfrak{T}_{\nu}(0,p) - \int_{0}^{p} \pi \mathfrak{T}_{\nu}(p,p') dB(\nu,T(p'))$$
(4.12)

Neither of these forms of the solution is particularly convenient for analytic work, but either one can be used to good advantage when carrying out approximate integrations via the trapezoidal rule (see Section 4.4.6). For analytical work, and some kinds of numerical integration, it helps to

rewrite the integrand using dB = (dB/dT)(dT/dp')dp'. The result is

$$I_{+}(p,\nu) = \pi B(\nu,T(p)) + (I_{+,s}(\nu) - \pi B(\nu,T_{sa}))\mathfrak{T}_{\nu}(p,p_{s}) + \int_{p}^{p_{s}} \pi \mathfrak{T}_{\nu}(p,p') \frac{dB}{dT}|_{T(p')} \frac{dT}{dp'} dp'$$

$$I_{-}(p,\nu) = \pi B(\nu,T(p)) + (I_{-,\infty}(\nu) - \pi B(\nu,T_{\infty}))\mathfrak{T}_{\nu}(0,p) - \int_{0}^{p} \pi \mathfrak{T}_{\nu}(p,p') \frac{dB}{dT}|_{T(p')} \frac{dT}{dp'} dp'$$
(4.13)

A considerable advantage of any of the forms in Eq. 4.11, 4.12 or 4.13 is that the integration variable p' is no longer dependent on frequency. This will prove particularly useful when we come to consider real gases, for which the optical thickness has an intricate dependence on frequency. The first two terms in the expression for the fluxes in either Eq. 4.12 or 4.13 give the exact result for an isothermal atmosphere; in each case, the first of the two terms represents the contribution of the local blackbody radiation, whereas the second accounts for the modifying effect of the boundaries. The boundary terms vanish at points far from the boundary, where \mathfrak{T} is small. Note that the boundary term for I_+ vanishes identically if the upward flux at the boundary has the form of blackbody radiation with temperature equal to the surface air temperature. For a planet with a solid or liquid surface, this would be the case if the ground temperature equals the surface air temperature and the ground has unit emissivity.

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

$$\mathfrak{H}_{\nu} = -\frac{d}{d\tau_{\nu}} (I_{+}(\tau_{\nu},\nu) - I_{-}(\tau_{\nu},\nu))$$
(4.14)

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

$$H_{\nu} = g \frac{d}{dp} (I_+ - I_-) = g \frac{d\tau_{\nu}}{dp} \frac{d}{d\tau_{\nu}} (I_+ - I_-) = \frac{\kappa}{\cos\bar{\theta}} \mathfrak{H}_{\nu}$$
(4.15)

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

4.2.2 Some special solutions of the Two-Stream equations

Beer's law

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

Infinite isothermal medium

Consider next an unbounded isothermal medium. In this case, it is readily verified that $I_{+} = I_{-} = \pi B(\nu, T)$ is an exact solution to 4.9. The right hand sides of the equations vanish, but the derivatives on the left hand sides vanish also, because T is independent of τ_{ν} . Hence, in an unbounded isothermal medium, the radiation field reduces to uniform blackbody radiation.

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

Exercise 4.2.1 Derive this result from Eq. 4.9; from Eq. 4.13.

Finite-thickness isothermal slab

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

$$I_{+}(\tau_{\nu},\nu) = [1 - \exp(-(\tau + \frac{\tau_{\infty}}{2}))]\pi B$$

$$I_{-}(\tau_{\nu},\nu) = [1 - \exp(+(\tau - \frac{\tau_{\infty}}{2}))]\pi B$$
(4.16)

Exercise 4.2.2 Derive this result from Eq. 4.13.

The radiation emitted out the top of the layer is $I_+(\tau_{\infty},\nu)$, or $(1 - \exp(-\tau_{\infty}))\pi B$, which reduces to the blackbody value πB when the layer is optically thick for the frequency in question, i.e. $\tau_{\infty}(\nu) \gg 1$. The same applies for the emission out of the bottom of the layer, *mutatis mutandum*. Note that in the optically thick limit, $I_+ = I_- = \pi B$ through most of the layer, and the inward-directed intensities only fall to zero in the two relatively thin skin layers near the top and bottom of the slab.

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

Substituting into Eq. 4.14, the heating rate for this solution is

$$\mathfrak{H}_{\nu} = -[\exp(-\tau) + \exp(\tau)] \exp(-\frac{\tau_{\infty}}{2})\pi B \tag{4.17}$$

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

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

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

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

Optically thick limit

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

The approximate form of the fluxes in the optically thick limit can be most easily derived from the integral expression in the form given in Eq. 4.9. Consider first the expression for I_+ . Away from the immediate vicinity of the bottom boundary, the boundary term proportional to $\mathfrak{T}_{\nu}(p, p_s)$ is exponentially small and can be dropped. To simplify the integral, we note that $\mathfrak{T}_{\nu}(p, p')$ is very small unless p' is close to p. Therefore, as long as the temperature gradient is a continuous function of p', it varies little over the range of p' for which the integrand contributes significantly to the integral. Hence dT/dp' can be replaced by its value at p, which can then be taken outside the integral. Likewise, dB/dT can be evaluated at T(p), so that this term can also be taken outside the integral. Finally, if one is not too close to the bottom boundary,

$$\int_{p}^{p_{s}} \mathfrak{T}_{\nu}(p',p) dp' \approx \int_{p}^{\infty} \mathfrak{T}_{\nu}(p',p) dp' = \frac{g\cos\bar{\theta}}{\kappa} \int_{\tau}^{\infty} \mathfrak{T}_{\nu}(\tau',\tau) d\tau' = \frac{g\cos\bar{\theta}}{\kappa}, \tag{4.18}$$

whence the upward flux in the optically thick limit becomes

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

$$\tag{4.19}$$

Near the bottom boundary, the neglected boundary term would have to be added to this expression. In addition, Eq. 4.18 would need to be corrected to allow for the fact that there is not room for $\int \mathfrak{T}$ to integrate out to its asymptotic value for an infinitely thick layer.

Using identical reasoning, the downward flux becomes

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

$$\tag{4.20}$$

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

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

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

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

$$I_{-,s}(\nu) - B(\nu, T_g) = I_{-,s}(\nu) - B(\nu, T_{sa}) = -\pi \frac{g \cos \bar{\theta}}{\kappa} \frac{dB}{dT} |_{T_{sa}} \frac{dT}{dp} |_{p_s}$$
(4.21)

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

In the optically thick limit, the net flux is

$$I_{+} - I_{-} = 2\pi \frac{g\cos\theta}{\kappa} \frac{dB}{dT} \frac{dT}{dp}$$
(4.22)

whence the radiative heating rate is

$$H_{\nu} = \frac{d}{dp} [D(\nu, p) \frac{dT}{dp}]$$
(4.23)

where

$$D(\nu, p) = 2\pi \frac{g^2 \cos \bar{\theta}}{\kappa} \frac{dB}{dT}|_{T(p)}$$
(4.24)

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

Consider an atmosphere which is transparent to solar radiation, and within which heat is redistributed only by infrared radiative transfer. Eq 4.15 then requires that the net upward flux $I_+ - I_-$ must be independent of altitude when integrated over all wavenumbers. This constant flux is nonzero, since the infrared flux through the system is set by the rate at which infrared escapes from the top of the atmosphere – namely, the *OLR*. Integrating Eq. 4.22 over the infrared yields an expression for dT/dp in terms of the *OLR* and the frequency-integrated diffusivity; because both *OLR* and diffusivity are positive, it follows that dT/dp > 0 for an optically thick atmosphere in pure infrared radiative equilibrium – that is, the temperature decreases with altitude. The more optically thick the atmosphere becomes, the smaller is *D*, and hence the stronger is the temperature variation in equilibrium. Pure radiative equilibrium will be discussed in detail in Sections 4.3.4 and 4.8, and the optically thick limit is explored in Problem ??.

Optically thin limit

The optically thin limit is defined by $\tau_{\infty} \ll 1$. Since $\tau_{\nu} \leq \tau_{\infty}$ and $\tau'_{\nu} \leq \tau$ in Eq. 4.9, all the exponentials in the expression for the fluxes are close to unity. Moreover, the integral is carried out over the small interval $[0, \tau_{\nu}]$, and hence is already of order τ_{∞} or less. It is thus a small correction to the first term, and we may set the exponentials in the integrand to unity and still have an expression that is accurate to order τ_{∞} . The boundary terms are not integrated, though, so we must retain the first two terms in the Taylor series expansion of the exponential to achieve the same accuracy. With these approximations, the fluxes become

$$I_{+}(\tau_{\nu}) = (1 - \tau_{\nu})I_{+,s} + \int_{0}^{\tau_{\nu}} \pi B(\nu, T(\tau_{\nu}'))d\tau_{\nu}'$$

$$I_{-}(\tau_{\nu}) = (1 - (\tau_{\infty} - \tau_{\nu}))I_{-,\infty} + \int_{\tau_{\nu}}^{\tau_{\infty}} \pi B(\nu, T(\tau_{\nu}'))d\tau_{\nu}'$$
(4.25)

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

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

$$B(\bar{T}_{\nu},\nu) = \frac{1}{\tau_{\infty}} \int_{0}^{\tau_{\infty}} B(\nu, T(\tau_{\nu}')) d\tau_{\nu}'$$
(4.26)

With this definition, the boundary fluxes are

$$I_{+,\infty} = (1 - \tau_{\infty})I_{+}(0) + \tau_{\infty}\pi B(T_{\nu},\nu)$$

$$I_{-,s} = (1 - \tau_{\infty})I_{-,\infty}) + \tau_{\infty}\pi B(\bar{T}_{\nu},\nu)$$
(4.27)

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

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

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

$$(4.28)$$

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

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

4.3 The Grey Gas Model

We will see in Section 4.4 that for most atmospheric gases κ , and hence the optical thickness, has an intricate dependence on wavenumber. This considerably complicates the solution of the radiative transfer equations, since the fluxes must be solved for individually on a very dense grid of wavenumbers, and then the results integrated to yield the net atmospheric heating, which is the quantity of primary interest. The development of shortcuts that can improve on a brute-force integration is an involved business, which in some respects is as much art as science, and leads to equations whose behavior can be difficult to fathom. The radiative transfer equations become much simpler if the optical thickness is independent of wavenumber. This is known as the grey gas approximation. For grey gases, the Schwartzschild equations can be integrated over wavenumber, yielding a single differential equation for the net upward and downward flux. More specifically, we shall assume only that the optical thickness is independent of wavenumber within the infrared spectrum, and that the temperature of the planet and its atmosphere is such that essentially all the emission of radiation lies in the infrared. Instead of integrating over all wavenumbers, we integrate only over the infrared range, thus obtaining a set of equations for the net infrared flux. Because

4.3. THE GREY GAS MODEL

of the assumption regarding the emission spectrum, the integrals of the Planck function $\pi B(\nu, T)$ over the infrared range can be well approximated by σT^4 .

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

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

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

$$\frac{d}{d\tau}I_{-} = I_{-} - \sigma T(\tau)^{4}$$
(4.29)

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

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

The OLR and surface back-radiation for an optically thin grey atmosphere are obtained by integrating Eq. 4.27 over all frequencies. The result is

$$I_{+,\infty} = (1 - \tau_{\infty})I_{+}(0) + \tau_{\infty}\sigma T^{4}$$

$$I_{-,s} = (1 - \tau_{\infty})I_{-,\infty} + \tau_{\infty}\sigma \bar{T}^{4}$$
(4.30)

where the mean atmospheric emission temperature is given by

$$\bar{T}^4 = \frac{1}{\tau_\infty} \int_0^{\tau_\infty} T^4 d\tau \tag{4.31}$$

The first term in the expression for $I_{+,\infty}$ is the proportion of upward radiation from the ground which escapes without absorption by the intervening atmosphere, while the second is the emission to space added by the atmosphere. In the expression for $I_{-,s}$ the first term is the proportion of incoming infrared flux which reaches the surface without absorption, while the second is the downward emission from the atmosphere. Note that for an optically thin atmosphere the atmospheric emission to space is identical to the atmospheric emission to the ground; in this regard the atmosphere radiates like an isothermal slab with temperature \overline{T} . According to Eq. 4.27, a nongrey atmosphere behaves similarly, if it is optically thin for all frequencies.

4.3.2 Radiative properties of an all-troposphere dry atmosphere

Let's consider an atmosphere for which the convection is so deep that it establishes a dry adiabat throughout the depth of the atmosphere. Thus, $T(p) = T_s(p/p_s)^{R/c_p}$ all the way to p = 0. We wish to compute the OLR for this atmosphere, which is done by substituting this T(p) into the grey-gas form of Eq. 4.9 and evaluating the integral for I_+ at $\tau = \tau_{\infty}$, i.e. the top of the atmosphere. Since the temperature is expressed as a function of pressure, it is necessary to substitute for pressure in terms of optical thickness in order to carry out the integral. We'll suppose that κ is a constant, so that $\tau_{\infty} - \tau = \kappa p/g = \tau_{\infty} p/p_s$. Using this to eliminate pressure from T(p), the integral for OLRbecomes

$$OLR = I_{+,s}e^{-\tau_{\infty}} + \int_{0}^{\tau_{\infty}} \sigma T_{s}^{4} (\frac{\tau_{\infty} - \tau'}{\tau_{\infty}})^{4R/c_{p}} e^{-(\tau_{\infty} - \tau')} d\tau'$$

$$= I_{+,s}e^{-\tau_{\infty}} + \int_{0}^{\tau_{\infty}} \sigma T_{s}^{4} (\frac{\tau_{1}}{\tau_{\infty}})^{4R/c_{p}} e^{-\tau_{1}} d\tau_{1}$$

$$= I_{+,s}e^{-\tau_{\infty}} + \sigma T_{s}^{4} \tau_{\infty}^{-4R/c_{p}} \int_{0}^{\tau_{\infty}} \tau_{1}^{4R/c_{p}} e^{-\tau_{1}} d\tau_{1}$$
(4.32)

The second line is derived by introducing a new dummy variable $\tau_1 = \tau_{\infty} - \tau'$. This is the the optical depth measured relative to the top of the atmosphere, and the re-expressed integral is computed by integrating from the top down, rather than from the ground up. The first term on the right hand side of Eq. 4.32 represents the proportion of the upward surface radiation which survives absorption by the atmosphere and reaches space. The second term is the net emission from the atmosphere. In the optically thin limit, the integral becomes small and the exponential in the first term approaches unity; thus, the *OLR* approaches the emission from the ground, $I_{+,s}$. As the atmosphere is made more optically thick, the boundary term becomes exponentially small, and the integral becomes more and more dominated by the emission from the upper reaches of the atmosphere. However, to obtain the optically thick limit, we cannot use the grey gas form of Eq 4.19, since dT/dp becomes infinite at p = 0 when the dry adiabat extends all the way to the top of the atmosphere.

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

$$OLR = \sigma T_s^4 \tau_\infty^{-4R/c_p} \Gamma(1 + \frac{4R}{c_p})$$
(4.33)

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

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

$$\sigma T_s^4 (\frac{p_{rad}}{p_s})^{4R/c_p} = OLR = \sigma T_s^4 \tau_{\infty}^{-4R/c_p} \Gamma(1 + \frac{4R}{c_p})$$
(4.34)

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

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

$$T_s = (S/\sigma)^{\frac{1}{4}} \Gamma (1 + \frac{4R}{c_p})^{-\frac{1}{4}} \cdot \tau_{\infty}^{R/c_p}$$
(4.35)

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

Exercise 4.3.1 This exercise illustrates the fact that if the Earth's atmosphere acted like a grey gas, then a doubling of CO_2 would make us toast! Using Eq. 4.35, find the τ_{∞} that yields a surface temperature of 285K for the Earth's absorbed solar radiation (about $270W/m^2$ allowing a crude correction for net cloud effects). Now suppose we double the greenhouse gas content of the atmosphere. If the Earth's greenhouse gases were grey gases, this would imply doubling the value of τ_{∞} from the value you just obtained. What would the resulting temperature be? Note that this rather alarming temperature doesn't even fully take into account the amplifying effect of water vapor feedback.

An examination of the radiative heating rate profile for the all-troposphere case provides much insight into the processes which determine where the troposphere leaves off and where a stratosphere will form. We'll assume that $I_{-,\infty} = 0$ and that the turbulent heat transfer at the ground is efficient enough that $T_{sa} = T_q$. Consider first the optically thin limit, for which the grey gas version of Eq. 4.28 is

$$H_{\nu} = \frac{\kappa}{\cos\bar{\theta}} \cdot [\sigma T_g^4 - 2\sigma T(p)^4]$$
(4.36)

assuming the stated boundary conditions. Since the radiative heating rate is nonzero, the temperature profile will not be in a steady state unless some other source of heating and cooling is provided to cancel the radiative heating. According to Eq. 4.36, the atmosphere is cooling at low altitudes, where $T > T_g/2^{\frac{1}{4}}$, i.e. where the local temperature is greater than the skin temperature. The cooling will make the atmosphere's potential temperature lower than the ground temperature, which allows the air in contact with the ground to be positively buoyant. The resulting convection brings heat to the radiatively cooled layer, allowing a steady state to be maintained if the convection is vigorous enough. However, in the upper atmosphere, where $T < T_g/2^{\frac{1}{4}}$ the atmosphere is being *heated* by upwelling infrared radiation, and there is no obvious way that convection could provide the cooling needed to make this region a steady state. Instead, the atmosphere in this region is expected to warm until a stratosphere in pure radiative equilibrium forms. Indeed, the tropopause as estimated by the boundary between the region of net heating and net cooling is located at the point where T(p) equals the skin temperature; this is precisely the same result as we obtained in the steady state model of the tropopause for an optically thin atmosphere, as discussed in Section 3.6.

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

$$I_{+} - I_{-} = 2 \frac{g \cos \bar{\theta}}{\kappa} (4\sigma T^{3}) \frac{dT}{dp} = 8\sigma T_{g}^{4} \frac{R}{c_{p}} \frac{g^{2} \cos \bar{\theta}}{\kappa p_{s}} (\frac{p}{p_{s}})^{4R/c_{p}-1}$$
(4.37)

in the all-troposphere grey-gas case. Recall that this expression breaks down in thin layers within roughly a unit optical depth of the bottom and top boundaries. The formula shows that whether the bulk of an optically thick atmosphere is heating or cooling depends on the lapse rate. The formula is valid even if κ depends on pressure and temperature. For constant κ , if $4R/c_p > 1$ the optically thick net flux decreases with height, and most of the atmosphere is *heated* by infrared radiation, and hence we expect a deep stratosphere and shallow troposphere. If $4R/c_p < 1$, corresponding to a weaker temperature lapse rate, most of the atmosphere instead experiences infrared radiative cooling, so we expect a deep troposphere. For constant κ , most gases would produce a deep stratosphere in the optically thick limit. This conclusion is greatly altered by pressure broadening. If $\kappa(p) = (p/p_s)\kappa(p_s)$, then the appearance of the pressure factor in the denominator alters the flux profile. Specifically, we now only require that $4R/c_p < 2$ in order for the flux to increase with height, yielding radiative cooling throughout the depth where the approximation is valid. Most atmospheric gases satisfy this criterion, and therefore most gases ought to yield a deep troposphere in the optically thick limit.

The above expectations regarding tropopause height will be confirmed in Chapter 7. Real gases are typically optically thick at some wavenumbers but optically thin at others, and we shall see in Section 4.9 how the competition plays out

Figure 4.2 shows numerically computed profiles of net infrared flux (I_+-I_-) without pressure broadening, for a range of optical thicknesses, with $R/c_p = \frac{2}{7}$. In this case, $4R/c_p > 1$, and we expect deep heating in the optically thick limit. For $\tau_{\infty} = 50$ the profile does follow the optically thick approximate form over most of the atmosphere, and exhibits a decrease in flux with height, implying deep heating. There is a thin layer of cooling near the ground, where the optically thick formula breaks down. When $\tau_{\infty} = 10$, the flux only conforms to the optically thick limit near the center of the atmosphere; there is a region of infrared cooling that extends from the ground nearly to 70% of the surface pressure. The case $\tau_{\infty} = 1$ looks quite like the optically



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

thin limit, with the lower half of the atmosphere cooling and the upper half heating. Numerical computations incorporating pressure broadening confirm the predictions of the optically thick formula. Specifically, the boundary between cooling and heating is unchanged for optically thin atmospheres, but rises to $p/p_s = .24$ for $\tau_{\infty} = 10$ and to $p/p_s = .11$ for $\tau_{\infty} = 50$. The profiles are not shown here, but are explored in Problem ??.

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

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

4.3.3 A first look at the runaway greenhouse

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

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

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

Substituting the one-component T(p) into the integral giving the solution to the Schwarzschild

equation, and substituting for pressure in terms of optical thickness, we find

$$OLR = I_{+}(0)e^{-\tau_{\infty}} + \int_{0}^{\tau_{\infty}} \sigma \frac{T_{o}^{4}}{(1 - \frac{RT_{o}}{L}\ln\frac{p}{p_{o}})^{4}} e^{-(\tau_{\infty} - \tau')} d\tau'$$

$$= I_{+}(0)e^{-\tau_{\infty}} + \sigma T_{o}^{4} \int_{0}^{\tau_{\infty}} \frac{1}{(1 - \frac{RT_{o}}{L}\ln\tau_{1})^{4}} e^{-\tau_{1}} d\tau_{1}$$
(4.38)

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

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

Figure 4.3 shows some results from a numerical evaluation of the integral in Eq. 4.38. For small surface temperatures, there is little atmosphere, and the OLR increases like σT_s^4 . As the surface temperature is made larger, the atmosphere becomes thicker and the OLR eventually asymptotes to a limiting value, as predicted. In accordance with the argument given above, the limiting OLR should be slightly less than the blackbody flux corresponding to the temperature T_o found one optical depth down from the top of the atmosphere. This temperature depends on g/κ , which is the pressure one optical depth down from the top. For $g/\kappa = 100Pa$, solving the simplified Clausius-Clapeyron relation for T at 100Pa yields $T_0 = 250.3K$, whence $\sigma T_o^4 = 222.6W/m^2$; for $g/\kappa = 1000Pa$, $T_o = 280.1K$ and $\sigma T_o^4 = 349.0W/m^2$. These values are consistent with the numerical results shown in the graph.

Note that for a given atmospheric composition (which determines κ) the Kombayashi-Ingersoll limit depends on the acceleration of gravity. A planet with stronger surface gravity will

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



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

have a higher Kombayashi-Ingersoll limit than one with weaker gravity. An explicit formula for the dependence on g/κ is obtained by substituting for T_o using the formula for the single-component saturated adiabat, Eq. 2.24. Thus, the limiting OLR can be written in the form

$$OLR_{\infty} = A(L/R)\sigma T_{o}^{4} = A(L/R)\frac{\sigma(L/R)^{4}}{\ln(\kappa p^{*}/g)^{4}}$$
(4.39)

where A(L/R) is the order unity constant discussed previously and $p^* = p_{ref} \exp(L/RT_{ref})$. In the formula for p^* , (p_{ref}, T_{ref}) is any point on the saturation vapor pressure curve, for example the triple point temperature and pressure. p^* is an enormous pressure $(2.3 \cdot 10^{11} Pa$ for water vapor), so Eq. 4.39 predicts that the Kombayashi-Ingersoll limit increases as surface gravity increases, since increasing g makes the logarithm in the denominator smaller. The apparent singularity when $\kappa p^*/g = 1$ is spurious, as the approximations we have made break down long before that value is reached.

We are now prepared to describe the runaway greenhouse phenomenon. Let S be the absorbed solar radiation per unit surface area of the planet, and let the limiting OLR computed above be OLR_{max} . If $S < OLR_{max}$ the planet will come to equilibrium in the usual way, warming up until it loses energy by infrared radiation at the same rate as it receives energy from its star. But what happens if $S > OLR_{max}$? In this case, as long as there is still an ocean or other condensed reservoir to feed mass into the atmosphere, the planet cannot get rid of all the solar energy it receives no matter how much it warms up; hence the planet continues to warm until the surface temperature becomes so large that the entire ocean has evaporated into the atmosphere. The temperature at this point depends on the mass and composition of the volatile reservoir. For example, the Earth's oceans contain enough mass to raise the surface pressure to about 100bars if dumped into the atmosphere in the form of water vapor. The ocean has been exhausted when



Figure 4.4: Schematic picture of the termination of a runaway greenhouse upon depletion of the volatile reservoir.

the saturation vapor pressure reaches this value. Using the simplified exponential form of the Clausius-Clapeyron relation to extrapolate the vapor pressure from the sea level boiling point (1 bar at 373.15K), we estimate that this vapor pressure is attained at a surface temperature of about 550K. This estimate is inaccurate, because the latent heat of vaporization varies appreciably over the range of temperatures involved. A more exact value based on measurements of properties of steam is 584K, but the grim implications for survival or emergence of life as we know it are largely the same.

At temperatures larger than that at which the ocean is depleted, the mass of the atmosphere becomes fixed and no longer increases with temperature. The greenhouse gas content of the atmosphere – which in the present case is the entire atmosphere – no longer increases with temperature. As a result, the OLR is once more free to increase as the surface becomes warmer, and the planet will warm up until it reaches an equilibrium at a temperature warmer than that at which the ocean is depleted. The additional warming required depends on the gap between the Kombayashi-Ingersoll limit and the absorbed solar radiation. This situation is depicted schematically in Figure 4.4. Once the ocean is gone, the lower atmosphere is unsaturated and air can be lifted some distance before condensation occurs. The resulting atmospheric profile is on the dry adiabat in the lower atmosphere, transitioning to the moist adiabat at the altitude where condensation starts. The situation is identical to that depicted for CO_2 in Fig. 2.6. Rain will still form in the condensing layer. Much of it will evaporate in the lower noncondensing layer; some of it may reach the ground, but the resulting puddles would tend to rapidly evaporate back into the highly undersaturated lower atmosphere. As surface temperature is made larger, the altitude where condensation sets in moves higher, until at very large temperatures the atmosphere behaves like a noncondensing dry system (albeit one where the entire atmosphere may consist of water vapor).

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

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

We will revisit the runaway greenhouse using more realistic radiation physics in Section 4.7. The effects of the stratosphere and of clouds will be brought into the picture in Chapter 7.

The runaway greenhouse phenomenon is usually thought of in conjunction with water vapor, but the concept applies equally well to any situation where there is a volatile reservoir of greenhouse gas, whether it be in solid or liquid form. For example, one could have a runawy greenhouse in association with the sublimation of a large CO_2 ice cap, or in association with the evaporation of a Methane or Ammonia ocean. In fact, the Kombayashi-Ingersoll limit determines whether a planet would develop a reservoir of condensed substance at its surface (a glacier or ocean), given sustained outgassing of that substance in the absence of any chemical sink. As the gas builds up in the atmosphere, the pressure increases and it would eventually tend to condense at the surface, preventig any further gaseous accumulation. However, the greenhouse effect of the gas warms the surface,which increases the saturation vapor pressure. The Kombayashi-Ingersoll limit tells us which effect wins out as surface pressure increases. Earth is below the threshold for water, so we have a water ocean. Venus is above the threshold for water and CO_2 , so both accumulate as gases in the atmosphere (apart from possible escape to space). When we revisit the problem with real-gas radiation, we will be able to say whether CO_2 would form a condensed reservoir on Earth or Mars, or CH_4 on Titan, given sustained outgassing in the absence of a chemical sink.

4.3.4 Pure radiative equilibrium for a grey gas atmosphere

For the temperature profiles discussed in Sections 4.3.2 and 4.3.3, the net infrared radiative heating computed from Eq. 4.14 is nonzero at virtually all altitudes; generally the imbalance acts to cool the lower atmosphere and warm the upper atmosphere. In using such solutions to compute OLR and back-radiation, we are presuming that convective heat fluxes will balance the cooling and keep the troposphere in a steady state. The upper atmosphere will continue to heat, and ultimately reach equilibrium creating a stratosphere, but in the all-troposphere idealization we presume that the stratosphere is optically thin enough that it doesn't much affect the OLR.

Now, we'll investigate solutions for which, in contrast, the net radiative heating vanishes individually at each altitude. Such solutions are in *pure radiative equilibrium*, as apposed to *radiative-convective equilibrium*. First we'll consider the case in which the only radiative heating

4.3. THE GREY GAS MODEL

is supplied by infrared; later we'll bring heating by atmospheric solar absorption into the picture. Pure radiative equilibrium is the opposite extreme from the all-troposphere idealization, and tells us much about the nature of the stratosphere, and the factors governing the tropopause height.

Assuming the atmosphere to be transparent to solar radiation, pure radiative equilibrium requires that the frequency-integrated longwave radiative heating \mathfrak{H} vanish for all τ . From the grey gas version of Eq. 4.14, we then conclude that $I_+ - I_-$ is independent of τ . Applying the upper boundary condition, we find that this constant is $I_+(\tau_{\infty})$, which is the *OLR*. Now, by taking the difference between the equations for I_+ and I_- we find

$$0 = \frac{d}{d\tau}(I_{+} - I_{-}) = -(I_{+} + I_{-}) + 2\sigma T^{4}$$
(4.40)

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

$$\frac{d}{d\tau}(I_+ + I_-) = -(I_+ - I_-) \tag{4.41}$$

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

$$2\sigma T^4 = (I_+ + I_-) = (1 + \tau_\infty - \tau)OLR \tag{4.42}$$

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

Eq. 4.42 also gives us the upward and downward fluxes, since we now know both $I_+ - I_$ and $I_+ + I_-$ at each τ . In particular, the downward flux into the ground is

$$I_{-}(0) = \frac{1}{2}((I_{+} + I_{-}) - (I_{+} - I_{-})) = \frac{1}{2}((1 + \tau_{\infty})OLR - OLR) = \frac{1}{2}\tau_{\infty}OLR$$
(4.43)

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

$$\sigma T_s^4 = (1 - \alpha)S + I_-(0) = (1 - \alpha)S \cdot (1 + \frac{1}{2}\tau_\infty)$$
(4.44)

where T_s is the surface temperature. This, together with the temperature profile determined by Eq. 4.42, determines what the thermal state of the system would be in the absence of heat transport mechanisms other than radiation. For an optically thin atmosphere, the surface temperature is only slightly greater than the no-atmosphere value. As τ_{∞} becomes large, the surface temperature increases without bound. Note that, while this formula yields a greenhouse warming of the surface, the relation between surface temperature and τ_{∞} is different from that given by the all-troposphere

radiative convective calculation in Eq. 4.35, because the pure radiative equilibrium temperature profile is different from the adiabat which would be established by convection.

Let's now compare the surface temperature with the temperature of the air in immediate contact with the surface. From Eq. 4.42 we find that the low level air temperature is determined by $\sigma T(0)^4 = (1 - \alpha)S \cdot (\frac{1}{2} + \frac{1}{2}\tau_{\infty})$ Taking the ratio,

$$\frac{T(0)}{T_s} = \left(\frac{\frac{1}{2} + \frac{1}{2}\tau_{\infty}}{1 + \frac{1}{2}\tau_{\infty}}\right)^{1/4} \tag{4.45}$$

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

In pure radiative equilibrium, the surface heating inevitably gives rise to convection. However, it is also possible that the temperature profile in the interior of the atmosphere may become unstable to convection, even without the benefit of a surface. This is a particularly important possibility to consider for gas giant planets, which have no distinct surface to absorb solar radiation and stimulate convection. To determine stability, we must compute the lapse rate dT/dp in radiative equilibrium, and see if it is steeper than that of the adiabat (moist or dry) appropriate to the atmosphere. Taking the derivative of 4.42 with respect to optical thickness, we find

$$8\sigma T^3 \frac{dT}{d\tau} = -OLR \tag{4.46}$$

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

$$\frac{d\ln T}{d\ln p} = -\frac{1}{4} \frac{1}{(1+\tau_{\infty}-\tau)} p \frac{d\tau}{dp}$$

$$\tag{4.47}$$

Stability is determined by comparing the slope of the adiabat to the radiative-equilibrium slope we have just computed. For the dry adiabat, the atmosphere is stable where

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

$$\tag{4.48}$$

The factor p appearing on the right hand side of this equation guarantees that the upper portion of the atmosphere will always be stable, unless $d\tau/dp$ blows up like 1/p or faster as $p \to 0$. Moreover, optically thin atmospheres are always stable throughout the depth of the atmosphere. This is so because the denominator is close to unity in the optically thin limit, while $-pd\tau/dp = \kappa p/(g \cos th e ta) < \tau_{\infty} \ll 1$. Since optically thin atmospheres are nearly isothermal in pure radiative equilibrium, it is hardly surprising that they are statically stable.

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

$$\frac{R}{c_p} \ge \frac{1}{4} \frac{\tau_\infty - \tau}{(1 + \tau_\infty - \tau)} \tag{4.49}$$

The right hand side has its maximum at the ground $\tau = 0$, and the maximum value is $\frac{1}{4}\tau_{\infty}/(1+\tau_{\infty})$. The more optically thick the atmosphere is, the more unstable it is near the ground. For large optical thickness, the stability criterion becomes the remarkably simple statement $4R/c_p \geq 1$. Dry Earth air, with $R/c_p = 2/7$, just misses being unstable by this criterion, and pure noncondensing water vapor is almost precisely on the boundary. Pure noncondensing CO_2 , NH_3 and CH_4 just barely satisfy the condition for instability near the ground when the atmosphere is optically thick.

Typical atmospheres, however, will be more unstable than the constant κ calculation suggests. As will be explained in Section 4.4 *,collisional broadening* typically causes the absorption coefficient to increase linearly with pressure, out to pressures of several bars. With collisional broadening, $\kappa(p) = \kappa(p_s) \cdot (p/p_s)$ for a combination of well-mixed greenhouse gases with pressure-independent concentrations. In this case $-pd\tau/dp = 2\tau_{\infty}$ at the surface, so that the dry adiabatic stability condition in the optically thick limit becomes $2R/c_p \geq 1$. The extra factor of 2 compared to the case without pressure broadening destabilizes all well-mixed atmospheres, provided they are sufficiently optically thick near the ground. The maximum value of the stability parameter occurs for mono-atomic gases like Helium, which has $2R/c_p = .8$ and easily meets the criterion for instability.

Other processes can destabilize the atmosphere as well. For example, on the moist adiabat the slope $-d \ln T/d \ln p$ is always less than the dry adiabatic slope, which deepens the layer within which the radiative-equilibrium atmosphere is unstable. In addition, a sharp decrease of optical thickness with height tends to destabilize the atmosphere, particularly if it occurs in a place where the atmosphere is optically thick. This happens whenever there is a layer where the concentration of absorbers decreases strongly with height. Because of Clausius-Clapeyron, this situation often happens in the lower portions of atmospheres in equilibrium with a reservoir of condensible greenhouse gas (water vapor on Earth or Methane on Titan, for example). In this case, the condensible substance has a destabilizing effect through its influence on the radiative equilibrium, as well as through its effect on the adiabat. Condensed water is a good infrared absorber, so radiative equilibrium drives cloud tops to be unstable. In contrast, the atmosphere is stabilized where the absorber concentration increases strongly with height; this situation is less typical, but can happen at the bottom of a water cloud.

4.3.5 Effect of atmospheric solar absorption on pure radiative equilibrium

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

In the Earth's stratosphere, solar absorption is largely due to the absorption of ultraviolet by ozone. On Earth as well as other planets having appreciable water in their atmospheres, absorption of solar near-infrared by water vapor and water clouds is important. CO_2 also has significant near-

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

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

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

$$-\frac{d}{d\tau}(I_{+} - I_{-}) + \frac{d}{d\tau}F_{\odot} = 0$$
(4.50)

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

$$(I_+ - I_-) - F_{\odot} = 0 \tag{4.51}$$

At the top of the atmosphere, this reduces to $OLR - (1 - \alpha)S = 0$, which is the requirement for top of atmosphere energy balance. Because the solar absorption does not change the infrared Schwartzschild equations, Eq. 4.41 is unchanged from the case of pure radiative equilibrium without solar absorption. Substituting Eq 4.51 and integrating, we obtain

$$I_{+} + I_{-} = \int_{\tau}^{\tau_{\infty}} F_{\odot}(\tau') d\tau' + (1 - \alpha)S$$
(4.52)

In writing this expression we have made use of the boundary condition $I_+ - I_- = OLR = (1 - \alpha)S$ at the top of the atmosphere. The heat balance equation 4.40 needs to be slightly modified, since the infrared cooling now balances the solar heating, instead of being set to zero. Thus,

$$\frac{d}{d\tau}F_{\odot} = \frac{d}{d\tau}(I_{+} - I_{-}) = -(I_{+} + I_{-}) + 2\sigma T^{4}$$
(4.53)

from which we infer

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

$$(4.54)$$

This gives the vertical profile of temperature in terms of the vertical profile of the solar flux; the previous case (without solar absorption) can be recovered by setting $F_{\odot} = const = (1 - \alpha)S$. At the top of the atmosphere, the integral in Eq. 4.53 vanishes, and the temperature becomes identical to the temperature of a skin layer heated by solar absorption, derived in Chapter 3 (Eq. 3.27).

4.3. THE GREY GAS MODEL

Taking the derivative with respect to τ yields

$$\frac{d}{d\tau}2\sigma T^4 = \frac{d^2}{d\tau^2}F_{\odot} - F_{\odot}$$
(4.55)

This equation provides a simple criterion determining when the solar absorption causes the temperature to increase with height. When there is no absorption, F_{\odot} is a constant and since it is positive the temperature decreases with height. The quantity $\sqrt{|F_{\odot}^{-1}d^2F_{\odot}/d\tau^2|}$ is local solar flux decay rate, expressed in units of infrared optical depth. Where the local solar decay rate is less than unity – meaning solar flux is attenuated at a lower rate than infrared – the radiative equilibrium temperature decreases with height. Where the local solar decay rate is greater than unity, the temperature increases with height. Note that it is the solar extinction rate *relative to the infrared extinction rate* that counts. One can make the temperature increase with height either by increasing the solar opacity or decreasing the infrared opacity (generally by decreasing the greenhouse gas concentration). The profile of solar absorption is also sensitive to the vertical distribution of solar absorbers. Where solar absorbers increases sharply with height, as is the case for ozone on Earth or organic haze on Titan, the stratospheric temperature increases with height.

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

$$2\frac{\sigma T^4}{(1-\alpha)S} = 1 + \tau_S + (\frac{1}{\tau_S} - \tau_S)e^{-(\tau_\infty - \tau)/\tau_S}$$
(4.56)

If $\tau_S > 1$ the temperature decreases with height, and if $\tau_S < 1$ the temperature increases with height. Defining the skin temperature as $T_{skin} \equiv (\frac{1}{2\sigma}(1-\alpha)S))^{1/4}$ the temperature at the top of the atmosphere is $(1+1/\tau_S)T_{skin}$, which reduces to the skin temperature when τ_S is large and becomes much greater than the skin temperature when τ_S is small. If the atmosphere is deep enough that essentially all solar radiation is absorbed before reaching the ground, then the exponential term vanishes in the deep atmosphere and the deep atmosphere becomes isothermal with temperature $(1 + \tau_S)T_{skin}$. Thus, when τ_S is small, all the solar radiation is absorbed within a thin layer near the top of the atmosphere. The temperature increases rapidly with height in this layer, but the bulk of the atmosphere below is approximately isothermal at the skin temperature. The strong solar absorption causes the deep atmosphere, and the ground (if there is one), to be colder than it would have been in the absence of an atmosphere. This *anti-greenhouse effect* arises because the deep atmosphere is heated only by downwelling infrared emitted by the solar-absorbing layer. This downward radiation equals the upward radiation loss to space, which must equal $(1-\alpha)S$ to satisfy the top of atmosphere balance. The deep atmosphere falls to the skin temperature because it is being illuminated by this flux from one side, but is radiating from both its top and bottom. This limit is relevant to the *nuclear winter* phenomenon, in which energetic explosions and fires loft a global or hemispheric solar-absorbing soot and dust cloud to high altitudes. The same situation would occur in the aftermath of a large bolide (asteroid or comet) impact. In either case, the atmosphere below the soot layer would become as frigid as the depth of winter, but moreover the relaxation to a uniform temperature state would shut off the convection which in large measure drives the hydrological cycle.

It can happen that the atmosphere is deep enough to absorb all solar radiation before it

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

The solution given in Eq. 4.56 shows that one can account for the temperature increase in the Earth's stratosphere if the upper atmosphere strongly absorbs solar radiation. As a model of the Earth's atmosphere, however, it has the shortcoming that if one makes the stratospheric absorption strong enough to yield a temperature increase with height, essentially all the solar beam is depleted in the stratosphere, leaving an isothermal lower atmosphere that won't convect and create a troposphere. What happens in Earth's real stratosphere is that ozone is a good absorber only in the ultraviolet part of the Solar spectrum. Once the ultraviolet is depleted, the remaining flux making its way into the lower atmosphere is only weakly absorbed by the atmosphere. In terms of Eq. 4.55, the solar decay rate $\sqrt{|F_{\odot}^{-1}d^2F_{\odot}/d\tau^2|}$ is large in the upper atmosphere, where there is still plenty of ultraviolet to absorb; in consequence, the temperature increases with height there. In the lower atmosphere the solar decay rate becomes small, so that the radiative equilibrium temperature decreases with height. There is also plenty of solar radiation left to heat the ground, destabilize the atmosphere, and create a troposphere. A variant of the solution Eq. 4.56 which captures this behavior is explored in Problem **??**.

The solution in Eq. 4.56 also explains why human-caused increases in CO_2 over the past century have led to tropospheric warming but stratospheric cooling, as illustrated in Figure 1.17. Increasing the greenhouse gas concentration is equivalent to increasing τ_{∞} . If one plots temperature as a function of pressure for a sequence of increasing τ_{∞} , the phenomenon is immediately apparent in cases where the upper level solar absorption is sufficiently strong. The behavior is explored in Problems ?? and ??. Without solar absorption, increasing τ_{∞} warms the atmosphere at every level, though the amount of warming decreases with height as the temperature asymptotes to the skin temperature. With solar absorption, however, the increased infrared cooling of the upper atmosphere offsets more and more of the warming due to solar absorption, leading to a cooling there. In the real atmosphere convection modifies the temperature profile in the lower atmosphere. Further, one must take into account real gas infrared and solar absorption in order to quantitatively account for the observed temperature trends. Nonetheless, the grey-gas pure radiative equilibrium calculation captures the essence of the mechanism.

The general lesson to take away from this discussion is that solar absorption near the top of the atmosphere stabilizes the atmosphere, reduces the greenhouse effect, and cools the lower portion of the atmosphere and also the ground. This is important in limiting the effectiveness of greenhouse gases like CH_4 and SO_2 , which significantly absorb solar radiation when their concentration becomes very high. It is also the way high altitude solar-absorbing haze clouds on Titan and perhaps Early Earth act to cool the troposphere. The soot and dust clouds lofted by an asteroid impact act similarly. In contrast, solar absorption concentrated near the ground has an effect which is not much different from simply reducing the albedo of the ground itself.

4.4 Real gas radiation: Basic principles

4.4.1 Overview: OLR through thick and thin

It would be exceedingly bad news for planetary habitability if real greenhouse gases were grey gases (see Exercise 4.3.1). Greenhouse gas concentrations would have to be tuned exceedingly accurately to maintain a planet in a habitable temperature range, and there would be little margin for error. Thus, it is of central importance that, for real gases, OLR varies much more gradually with greenhouse gas concentration than it would for an idealized grey gas². This is another area in which the quantum nature of the Universe directly intervenes in macroscopic phenomena governing planetary climate.

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

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

Since the main point is to understand how the wavenumber dependence of absorptivity affects the sensitivity of OLR to greenhouse gas concentration, we'll begin with a discussion of the spectrum of outgoing longwave radiation in an idealized case. Let's consider a planet whose surface radiates like an ideal blackbody in the infrared, having an atmosphere whose air temperature at the surface T_{sa} is equal to the ground temperature T_g . The temperature T(p) is monotonically decreasing with height in the troposphere, and is patched continuously to an isothermal stratosphere having temperature T_{strat} . The atmosphere consists mostly of infrared-transparent N_2 and O_2 with a surface pressure of $10^5 Pa$, like Earth. Unlike Earth, the only greenhouse gas is a mythical substance (call it Oobleck), which is a bit like CO_2 , but much simpler to think about. It has the same molecular weight as CO_2 , but it's absorption coefficient $\kappa_{Oob}(\nu)$ has an absorption band centered on wavenumber $\nu_o = 600 cm^{-1}$. Within $100 cm^{-1}$ of ν_o , κ_{Oob} has the constant value κ_o . Outside of this limited range of wavenumbers, Oobleck is transparent to infrared, i.e. $\kappa_{Oob} = 0$. To make life even simpler for the atmospheric physicists of this planet, κ_{Oob} is independent of both temperature and pressure. Like real CO_2 , the specific concentration of Oobleck (q_{Oob}) is constant

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



Figure 4.5: The OLR spectrum for a hypothetical gas which has a piecewise constant absorption coefficient. The dashed-dotted lower curve is the blackbody spectrum corresponding to the stratospheric temperature T_{strat} while the dashed upper curve is the blackbody spectrum corresponding to the surface temperature T_s . The calculation was carried out for $T_s = 280K$ and $T_{strat} = 200K$, and with a greenhouse gas concentration sufficient to make the optical thickness ≈ 10 in the central absorption band. Left panel: The gas has an absorption coefficient of $1m^2/kg$ within a single absorption band extending from 500 to 700 cm⁻¹. Right Panel: The gas has additional weak absorption bands from 300 to 500 cm⁻¹ and 700 to 900 cm⁻¹, within which the absorption coefficient is $.125m^2/kg$.

throughout the depth of the atmosphere.

What does the spectrum of OLR look like for this planet? The answer is shown in the left panel of Figure 4.5. In this figure, we have assumed that the Oobleck molecule has an absorptivity of $1m^2/kg$. Then, with a molar concentration of 300ppmv (like CO_2 in the 1960's), the specific concentration is $4.6 \cdot 10^{-4}$ and the optical thickness $\kappa_o q_{Oob} p_s/g \cos \bar{\theta}$ is 9.4 within the absorption band. Since the atmosphere is optically thick in this wavenumber region, infrared radiation in this part of the spectrum exits the atmosphere with the temperature of the stratosphere. This is exactly what we see in the graph. Outside the absorption band, the atmosphere is transparent, and hence infrared leaves the top of the atmosphere at the much higher temperature of the ground. The overall appearance of the OLR spectrum is that the greenhouse gas has "dug a ditch" in the spectrum of OLR, or perhaps "taken a bite" out of it. The ditch in the spectrum reduces the total OLR of the planet, but not so much so as if the absorption were strong throughout the spectrum, as would be the case for a grey gas. This is the typical way that real greenhouse gases work: they make the atmosphere optically thick in a limited part of the spectrum, while leaving it fairly transparent elsewhere. The strength of the greenhouse effect is not so much a matter of how deep the ditch, but how wide.

Oobleck is a very contrived substance, but the above exercise gives us a fair idea of what to look for when interpreting real observations of the spectrum of OLR. Figure 4.6, giving the OLRspectrum of Mars observed at two times of day by the TES instrument on Global Surveyor, is a case in point. Mars has an essentially pure CO_2 atmosphere complicated only be optically thin ice clouds and dust clouds (which can be very thin between major dust storms). The planet thus provides perhaps the purest illustration of the CO_2 radiative effect available in the Solar system. In Figure 4.6 a CO_2 "ditch" centered on about $650cm^{-1}$ is evident both in the afternoon and



Figure 4.6: Some representative OLR spectra for Mars, observed by the Thermal Emission Spectrometer on Mars Global Surveyor at various times of day.

sunset spectra. At the trough of this ditch, the radiation exits the atmosphere with a radiating temperature of about 170K both in the afternoon and sunset cases. This temperature is similar to the coldest temperature encountered in the upper atmosphere of Mars in the Summer (see Fig 2.2), and is compatible with the strong decrease of temperature with height seen in the soundings. Away from CO_2 ditch, the atmosphere appears transparent, and the emission resembles the blackbody emission from a land surface having temperature 265K in the afternoon case and 212K in the sunset case. These numbers are compatible with the observed range of ground temperature on Mars, cross-checked by near-surface data from landers.

In a situation like that shown in the left panel of Figure 4.5, the OLR is as low as it is going to get, provided the stratospheric temperature is held fixed. Increasing the greenhouse gas concentration q_G cannot lower the OLR further since, in the spectral region where the gas is radiatively active the atmosphere is already radiating at the coldest available temperature ³. Suppose, however, that instead of the gas being transparent outside the central absorption band, there is a set of weaker absorption bands waiting in the wings on either side of the primary band – a gas we may call "Two-Band Oobleck." In this case, illustrated in the right panel of Figure 4.5, the effect of the weaker bands on OLR is not yet saturated, and increases in q_G will cause the OLR to go down until these bands, too, are saturated. But what if there are yet-weaker absorption bands waiting a bit farther out? Then further increases of q_G will yield additional decreases in the OLR. One can imagine making the process continuous by making the width of the bands smaller, and the jump in absorption coefficients between adjacent bands smaller. Real greenhouse gases act very much like this, as they almost invariably have absorption whose overall strength decays strongly with distance in wavenumber from a central peak. The rate at which the absorption decays with

³This example is somewhat contrived, since increasing the concentration of a greenhouse gas generally cools the stratosphere. However, it serves to illustrate the way additional weak absorption bands influence the OLR. The additional OLR reduction from cooling of the stratosphere as q_G increases is a secondary effect. Since the temperature there is already so low, it wouldn't throw off the result very much to simply replace the OLR at the depths of the ditch to zero.

distance determines the rate at which the OLR decreases as the greenhouse gas concentration is made larger.

From Eq. 4.11, 4.12 or 4.13, if we know the transmission function, we can carry out the integral needed to obtain the radiative fluxes. As we shall see shortly, in most cases the dependence of κ on wavenumber is so intricate that solving the problem by doing a brute-force integral over wavenumber is prohibitive if one aims to do the calculation enough times to gain some insight from modelling a climate (even in a single dimension). In any event, doing the calculation with enough spectral resolution to directly resolve all the wiggles in $\kappa(\nu)$ provides much more information about spectral variability than is needed in most cases. What we really want is to understand something about the properties of the transmission function averaged over a finite sized spectral region of width Δ , centered on a given frequency ν . Specifically, let's choose Δ to be small enough that the Planck function B and its derivative dB/dT are both approximately constant over the spectral interval of width Δ . In that case, when the solution for the flux given in Eq. 4.12 or its alternate forms is averaged over Δ , B can be treated as nearly independent of ν and taken outside the average. In consequence, the resulting band-averaged equations have precisely the same form as the original ones, save that the fluxes are replaced by average fluxes like

$$\bar{I}_{+}(\nu,p) = \frac{1}{\Delta} \int_{\nu-\Delta/2}^{\nu+\Delta/2} I_{+}(\nu',p) d\nu'$$
(4.57)

and the transmission function is replaced by

$$\bar{\mathfrak{T}}_{\nu}(p,p') = \frac{1}{\Delta} \int_{\nu-\Delta/2}^{\nu+\Delta/2} \mathfrak{T}_{\nu'}(p,p') d\nu'$$
(4.58)

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

The transmission function satisfies the *multiplicative property*, that

$$\mathfrak{T}_{\nu}(p_1, p_2) = \mathfrak{T}_{\nu}(p_1, p')\mathfrak{T}_{\nu}(p', p_2) \tag{4.59}$$

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

4.4.2 The absorption spectrum of real gases

We will now take a close look at the absorption properties of CO_2 , in order to introduce some general ideas about the nature of the absorption of infrared radiation by molecules in a gas. Continuing to use CO_2 as an example, these ideas will be developed in Sections 4.4.3,4.4.4 and 4.4.6 into a computationally efficient means of calculating infrared fluxes in a real-gas atmosphere. A survey of the spectral characteristics of selected other greenhouse gases will be given in Sections 4.4.7 and 4.4.8.

Figure 4.7 shows the absorption coefficient of CO_2 as a function of wavenumber, for pure CO_2 gas at a pressure of 1bar and a temperature of 296K. In some spectral regions, e.g. 1700-1800 cm^{-1} , CO_2 at this temperature and pressure is essentially transparent. This is a window region through which infrared can easily escape to space if no other greenhouse gas intervenes. For a 285K blackbody, $60W/m^2$ can be lost through this window. There are two major bands in which absorption occurs. For Earthlike temperatures, the lower wavenumber band, from about 450 to $1100 \ cm^{-1}$ is by far the most important. At 285K the blackbody emission in this band is $218W/m^2$ out of a total of $374W/m^2$, so the absorption in this band is well tuned to intercept terrestrial infrared and to thus reduce OLR. The blackbody emission in the higher wavenumber band, from 1800 to $2500 \ cm^{-1}$, is only $6W/m^2$. This band has a minor effect on OLR for Earth, but it can become important for much hotter planets like Venus, and even for Earth is important for the absorption of solar near-infrared. Within either band, the absorption coefficient varies by more than eight orders of magnitude.

The absorption does not vary randomly. It is arrange around six peaks (three in each major band), with the overall envelope of the absorption declining approximately exponentially with distance from the peak. However, there is a great deal of fine-scale variation within the overall envelope. Zooming in on a typical region in the inset to Figure 4.7 we see that the absorption can vary by an order of magnitude over a wavenumber range of only a few tenths of a cm^{-1} . Most significantly, the absorption peaks sharply at a discrete set of frequencies.

Why does the absorption peak at preferred frequencies? In essence, molecules are like little radio receivers, tuned to listen to light only at certain specific frequencies. Since energy is conserved, the absorption or emission of a photon must be accompanied by a change in the internal energy state of the molecule. It is a consequence of quantum mechanics that the internal energy of a molecule can only take on values drawn from a finite set of possible energy states, the distribution of which is determined by the structure of the molecule. If there are N state, there are N(N-1)/2 possible transitions, and each one leads to a possible absorption/emission line as illustrated in Figure 4.8. Transitions between different energy states of a molecule's electron configuration do not significantly contribute to infrared absorption in most planetary situations. The energy states involved in infrared absorption and emission are connected with displacement of the nuclei in the molecule, and take the form of vibrations or rotations. Every molecule has an equilibrium configuration, in which each nucleus is placed so that the sum of the electromagnetic forces from the other nuclei and from the electron cloud sum up to zero. A displacement of the nuclear positions will result in a restoring force that brings the system back toward equilibrium, leading to vibrations. The nucleii can be thought of as being connected with quantum-mechanical springs (one between each pair of nucleii) of different spring constants, and the vibrations can be thought of as arising from a set of coupled quantum-mechanical oscillators. Rigid molecules, held together by rigid rods rather than springs, would have rotational states but not vibrational states. The fact that molecules are not rigid causes the rotational states to couple to the vibrational states, through the coriolis and centrifugal forces.

Noble gases (*He*, Ar, etc.) are monatomic, have only electron transitions, and are not active in the infrared. A diatomic molecule (Fig. 4.9) has a set of energy levels associated with the oscillation caused by pulling the nuclei apart and allowing them to spring back; it also has a set of energy levels associated with rotation about either of the axes perpendicular to the line joining the nuclei. Centrifugal force couples the stretching to the rotation. Triatomic molecules (Fig. 4.10) have an even richer set of vibrations and rotations, especially if their equilibrium state is bent rather than linear (Fig. 4.11). Polyatomic molecules like CH_4 , NH_3 , SF_6 and the chlorofluorcarbons



Figure 4.7: The absorption coefficient vs. wavenumber for pure CO_2 at a temperature of 293K and pressure of $10^5 Pa$. This graph is not the result of a measurement by a single instrument, but is synthesized from absorption data from a large number of laboratory measurements of spectral features, supplemented by theoretical calculations. The inset shows the detailed wavenumber dependence in a selected spectral region.



Figure 4.8: Schematic of emission of a photon by transition from a higher energy state to a lower energy state



Figure 4.9: Vibration and rotation modes of a diatomic molecule made of a pair of identical atmos, with associated charge distributions

(e.g. CFC-12, which is CCl_2F_2) have yet more complex modes of vibration and rotation. As the set of energy states becomes richer and more complex, the set of differences between states fills in more and more of the spectrum, making the molecule a better infrared absorber.

For a molecule to be a good infrared absorber and emitter, it is not enough that it have transitions whose energy corresponds to the infrared spectrum. In order for a photon to be absorbed or emitted, the associated molecular motions must also couple strongly to the electromagnetic field. Although the quantum nature of radiation is crucial for many purposes, when it comes to the interaction of infrared or longer wavelength radiation with molecules, one can productively think of the interaction in semiclassical terms. The reason is that the wavelength of infrared is on the order of 10 *micrometers*, which is two to three orders of magnitude larger than the size of the molecules we will be considering. Thus, one can think of the infrared light as providing a large scale fluctuating electric and magnetic field which alters the environment in which the molecule finds itself, and exerts a force on the constituent parts of the molecule. This force displaces the nuclei and electron cloud, and excites vibration or rotation. Conversely, a vibrating or rotating molecule creates a moving charge distribution, which classically radiates an electromagnetic wave.
174CHAPTER 4. RADIATIVE TRANSFER IN TEMPERATURE-STRATIFIED ATMOSPHERES



Figure 4.10: Vibration and rotation modes of a linear symmetric triatomic molecule (like CO_2), with associated charge distributions



Figure 4.11: Some polar triatomic molecules. Two different modes of rotation are indicated for the H_2O molecule. There is a third mode of rotation about an axis perpendicular to the page.

While one must fully take into account quantum effects in describing molecular motion, one need not for our purposes confront the much harder problem of quantizing the electromagnetic field as well (the problem of "quantum field theory"). The only way in which we make use of the quantum nature of the electromagnetic field is in converting the energy difference $E_u - E_\ell$ into a frequency of light, via $\Delta E = h\nu$.

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

Many common atmospheric molecules have no dipole moment in their unperturbed equilibrium state. Such nonpolar molecules can nonetheless couple strongly to the electromagnetic field. They do so because vibration and rotation can lead to a dipole moment through distortion of the equilibrium positions of the electron cloud and the nucleii. As illustrated in Figure 4.9, diatomic molecules made of two identical atoms, do not acquire a dipole moment under the action of either rotation or stretching. Symmetric diatomic molecules, such as N_2 , O_2 and H_2 in fact have plenty of rotational and vibrational transitions that are in the infrared range. Because the associated molecular distortions have no dipole moment, however, these gases are essentially transparent to infrared unless they are strongly perturbed by frequent collisions. This is why the most common gases in Earth's atmosphere – N_2 and O_2 – do not contribute to Earth's greenhouse effect. However, it is important to recognize that situations in which diatomic molecules become good greenhouse gases are in fact quite common in planetary atmospheres. When there are frequent collisions, such as happen in the high density atmospheres of Titan and on all the giant planets, diatomic molecules acquire enough of a dipole moment during the time collisions are taking place that the electromagnetic field can indeed interact with their transitions quite strongly. This makes N_2 and H_2 the most important greenhouse gases on Titan, and H_2 a very important greenhouse gas on all the gas giant planets. In terms of volume of atmosphere affected, Hydrogen is by far the most important greenhouse gas in the entire Solar System. Collision-induced absorption of this type forms a *continuum* in which the absorption is a very smooth function of wavenumber, without any significant line structure. Polyatomic molecules can also have significant continua, existing alongside the line spectra. Continuum absorption will be discussed in Section 4.4.8.

 CO_2 is a linear molecule with the two oxygens symmetrically disposed about the central carbon, as illustrated in Figure 4.10. A uniform stretch of such a molecule does not create a dipole moment, but a vibrational mode which displaces the central atom from one side to the other does. In addition, bending modes of CO_2 have a fluctuating dipole moment, which can in turn be further influenced by rotation. Both these modes are illustrated schematically in Figure 4.10. Modes of this sort make CO_2 a very good greenhouse gas - the more so because the typical energies of the transitions involved happen to correspond to frequencies near the peak of the Planck function for Earthlike temperatures.

Some molecules – called *polar* have a dipole moment even in their undisturbed state. Most common diatomic gases made of two different elements – notably HF and HCl – are polar, and their vibrational and rotational modes cause fluctuations in the dipole which make them quite

good infrared absorbers. They are not commonly thought of as greenhouse gases, because they are highly chemically reactive and do not appear in radiatively significant quantities in any known planetary atmosphere. However, one must keep an open mind about such things. Most triatomic atmospheric gases $(H_2O, SO_2, O_3, NO_2 \text{ and } H_2S)$, among others) are polar. CO_2 , a symmetric linear molecule with the carbon at the center, is a notable exception. Ammonia (NH_3) is also polar, having its three hydrogens sticking out on one side like legs of a tripod attached to the nitrogen atom at the other side. Polar molecules couple strongly to the electromagnetic field, and their asymmetry also gives them a rich set of coupled rotation and vibration modes with many opportunities for transitions corresponding to the infrared spectrum. The spectrum is enriched because rotation about the axis with the largest moment of inertia (shown as the vertical axis for the water molecule in Figure 4.11) causes the wing molecules to fling outwards, changing the bond angle and the dipole moment. The molecule can also rotate about an axis perpendicular to the plane of the Figure, leading to distinct set of energy levels. Further, energy can be stored in rotations about the axis with minimum moment of inertia (shown as horizonal in the Figure). For a linear molecule like CO_2 , rotation about the corresponding axis has essentially no energy.

Let's return now to the matter of how the OLR varies as a function of the greenhouse gas content of an atmosphere. Essentially, we revisit the discussion of Figure 4.5, but this time in the context of the actual absorption spectrum of CO_2 instead of the hypothetical gas discussed earlier. We use the same temperature profile as in Figure 4.5, but the OLR is computed using the fully resolved absorption coefficient as a function of wavenumber, reproduced in the lower panel of the Figure 4.12. The horizontal lines in this panel indicate the spectral regions that are optically thick for CO_2 paths of $\frac{1}{10}$, 1, and 1000 kg/m^2 . The upper panel shows the corresponding OLR for the same three paths. The OLR was computed at the full spectral resolution of the lower panel, but was smoothed over bands of width $10cm^{-1}$ to make the pattern easier to see. The smoothing is done in such a way that the integral of the smoothed OLR curve over wavenumber yields the same net value as the integral over the original unsmoothed curve. This calculation of OLR is still not completely realistic, since, to keep things simple, it is carried out as if the absorption coefficient were uniform throughout the depth of the atmosphere. In reality, κ_{CO2} varies with both temperature and pressure, though we'll see eventually that to a good approximation this variation can be handled through the introduction of an *equivalent path*, which is generally somewhat more than half the actual full-atmosphere path, if the reference pressure at which absorption coefficients are stated is taken to be the surface pressure.

Figure 4.12 explains why the OLR reduction is approximately logarithmic in greenhouse gas concentration for CO_2 and similar greenhouse gases. The key thing to note is that the absorption coefficient in the principal band centered on $675 cm^{-1}$ decays exponentially with distance from the center. Hence, as the CO_2 path is increased by a factor of 10, from $\frac{1}{10}$ to $1 kg/m^2$, the width of the ditch within which the radiating temperature is reduced to cold stratospheric values increases only like the logarithm of the ratio of paths. This is true for paths as small as $.01kg/m^2$ and as large as $100kq/m^2$. However, when the path gets as large as $1000kq/m^2$, the weak absorption bands on the shoulder, near 950 and 1050 cm^{-1} start to become important, and enhance the optical thickness beyond what one would expect on the basis of the central absorption peak. 1000 kg/m^2 corresponds to a partial pressure of CO_2 of about 100mb for Earth's gravity, or equivalently a molar mixing ratio of about 10 % for Earth's current surface pressure. This is far in excess of any CO_2 concentration on Earth likely to have been attained in the past 300 million years, but is well within the range of what has been contemplated at the end of a Neoproterozoic Snowball episode. or earlier during the Faint Young Sun period. Many greenhouse gases also have a central absorption peak with exponential skirts, and these will also exhibit a nearly logarithmic dependence of OLRon the concentration of the corresponding greenhouse gas.



Figure 4.12: Lower panel: The absorption coefficient for CO_2 at 1 bar and 300K, in the wavenumber range of interest for Earthlike and Marslike planets. The horizontal lines show the wavenumber range within which the optical thickness exceeds unity for CO_2 paths of $\frac{1}{10}$, 1 and 1000 kg/m^2 . Upper panel: The corresponding OLR for the three path values, computed for the same temperature profiles as in Figure 4.5. The OLR has been averaged over bands of width 10 cm^{-1} .

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

4.4.3 I walk the line

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

$$\kappa_G(\nu, p, T) = \frac{S}{\gamma} f(\frac{\nu - \nu_c}{\gamma})$$
(4.60)

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

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

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

The type of line broadening of primary interest in planetary climate problems is *collisional* broadening, alternatively called *pressure broadening*. Collisional broadening arises because the kinetic energy of a molecule is not quantized, and therefore if a molecule has experienced a collision sufficiently recently, energy can be borrowed from the kinetic energy in order to make up the difference between the photon's energy and the energy needed to jump one full quantum level. The theory of this process is exceedingly complex, and in many regards incomplete. There is a simple semi-classical theory that predicts that collision-broadened lines should have the Lorentz line shape $f(x) = 1./(\pi \cdot (1+x^2))$, and this shape seems to be supported by observations, at least within a hundred widths or so of the line center. For the Lorentz shape, absorption decays rather slowly with distance from the center; 10 half-widths γ from the center, the Lorentz absorption has decayed to only $\frac{1}{101}$ of its peak value, whereas the Gaussian doppler-broadened line has decayed to less that 10^{-43} of its peak. There are both theoretical and observational reasons to believe that the very far tails of collision broadened lines die off faster than predicted by the Lorentz shape. A full discussion of this somewhat unsettled topic is beyond the level of sophistication which we aspire to here, but the shape of far-tails has some important consequences for the continuum absorption, which will be taken up briefly in Section 4.4.8.

In the simplest theories leading to the Lorentz line shape, the width of a collision-broadened line is proportional to the mean collision frequency, i.e. the reciprocal of the time between collisions. The Lorentz shape is valid in the limit of infinitesimal duration of collisions; it is the finite time colliding molecules spend in proximity to each other that leads to deviations from the Lorentz shape in the far tails, but there is at present no general theory for the far-tail shape. For many common planetary gases the line width is on the order of a tenth of a cm^{-1} when the pressure is 1 bar and the temperature is around 300K. For fixed temperature, the collision frequency is directly proportional to pressure, and laboratory experiment shows that the implied proportionality of line width to pressure is essentially exact. Holding pressure fixed, the density goes down in inverse proportion to temperature while the mean molecular velocity goes up like the square root of temperature. This should lead to a collision frequency and line width that scales like $1/\sqrt{T}$. Various effects connected with the way the collision energy affects the partial excitation of the molecule lead to the measured temperature exponent differing somewhat from its ideal value of $\frac{1}{2}$. Putting both effects together, if the width is known at a standard state (p_o, T_o) , then it can be extrapolated to other states using

$$\gamma(p,T) = \gamma(p_o, T_o) \frac{p}{p_o} (\frac{T_o}{T})^n \tag{4.61}$$

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

Another complication is that the collision-broadened line width depends on the molecules doing the colliding. Broadening by collision between molecules of like type is called *self-broadening*, while that due to dissimilar molecules is called *foreign broadening*. The simple Lorentz theory would suggest a proportionality to collision rate, which is a simple function of the ratio of molecular weights. Not only does the actual ratio of self to foreign broadened width deviate from what would be expected by this ratio, but the ratio actually varies considerably from one absorption line to another. For CO_2 , for example, some self-broadened lines have essentially the same width as for the air-broadened case, whereas others can have widths nearly half again as large. For water vapor, the disparity is even more marked. Evidently, some kinds of collisions are better at partially exciting energy levels than others. There is no good theory at present that enables one to anticipate such effects. Standard spectral databases tabulate the self-broadened and air-broadened widths at standard temperature and pressure, but if one were interested in broadening of water vapor by collisions with CO_2 (important for Early Mars) or broadening of NH_3 by collisions with H_2 (on Jupiter or Saturn), one would have to either find specialized laboratory experiments or extrapolate based on molecular weights and hope for the best.

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

$$S(T) = S(T_o)(\frac{T}{T_o})^n \exp(-\frac{h\nu_\ell}{k}(\frac{1}{T} - \frac{1}{T_o}))$$
(4.62)

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

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

$$\bar{\mathfrak{T}}(p_1, p_2) = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} [\exp(-\frac{1}{g\pi} \int_{p_1}^{p_2} \frac{S(T)\gamma q}{\nu'^2 + \gamma^2} dp)] d\nu'$$
(4.63)

where $\nu' = \nu - \nu_c$. The argument of the exponential is just the optical thickness of the layer between p_1 and p_2 , and to keep the notation simple we will assume the integral to be taken in the sense that makes it positive. The double integral and the nonlinearity of the exponential make this a hard beast to work with, but there are two limits in which the result becomes simple. When the layer of atmosphere between p_1 and p_2 is optically thin even at the center of the line, where absorption is strongest, the line is said to be in the *weak line regime*. All lines are in this regime in the limit $p_2 \rightarrow p_1$, though if the line is very narrow or the intensity is very large, the atmospheric layer might have to be made exceedingly small before the weak line limit is approached. For weak lines the exponential can be approximated as $\exp(-\delta\tau) \approx 1. - \delta\tau$, whence

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

$$= 1 - \frac{1}{\Delta} \frac{1}{g\cos\bar{\theta}} \int_{p_1}^{p_2} S(T) q dp$$

$$= 1 - \frac{1}{\Delta} S(T_o) \ell_w$$
(4.64)

where T_o is a constant reference temperature and the *weighted path* for strong lines is

$$\ell_w(p_1, p_2) \equiv \frac{1}{g \cos \bar{\theta}} \int_{p_1}^{p_2} \frac{S(T(p))}{S(T_o)} q(p) dp$$
(4.65)

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

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

$$\bar{\mathfrak{T}}(p_1, p_2) \approx \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left[\exp\left(-\frac{1}{\nu'^2} \frac{1}{g\pi \cos\bar{\theta}} \int_{p_1}^{p_2} S(T)\gamma q dp \right) \right] d\nu' \\
= \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \exp\left(-\frac{X}{\nu'^2}\right) d\nu' \\
= \frac{1}{2\zeta_m} \int_{-\zeta_m}^{\zeta_m} \exp\left(-\frac{1}{\zeta^2}\right) d\zeta$$
(4.66)

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

$$\ell_s \equiv \frac{1}{g\cos\bar{\theta}} \int_{p_1}^{p_2} \frac{S(T(p))}{S(T_o)} (\frac{T_o}{T})^n \frac{p}{p_o} q(p) dp \tag{4.67}$$

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

$$\int_{0}^{\zeta_m} \exp(-\frac{1}{\zeta^2}) d\zeta \approx \zeta_m - \sqrt{\pi}$$
(4.68)

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

$$\bar{\mathfrak{T}}(p_1, p_2) \approx 1 - \frac{1}{\Delta} 2\sqrt{S(T_o)\gamma(p_o)\ell_s}$$
(4.69)

For strong lines the equivalent width is $W \equiv 2\sqrt{S(T_o)\gamma(p_o)\ell_s}$. In this case, the width of the chunk taken out of the spectrum increases like the square root of the path because the absorption coefficient decreases like $1/\nu'^2$ with distance from the line center, implying that the width of the spectral region within which the atmosphere is optically thick scales like the square root of the path. Unlike weak lines, strong lines really do take almost all of the energy out of a limited segment of the spectrum. The multiplicative property for transmission is equivalent to an additive property for equivalent width. The nonlinearity of the square root linking path to equivalent width in the strong line case thus means that the band-averaged transmission has lost the multiplicative property. As in our earlier general discussion of this property, the loss stems from the progressive depletion of energy in parts of the spectrum near the line center.

The pressure-weighting of the strong-line path reflects the fact that, away from the line centers, the atmosphere becomes more optically thick as pressure is increased and the absorption is spread over a greater distance around each line. Note that if we choose as the reference pressure p_o any pressure that remains between p_1 and p_2 , then $\ell_s \to \ell_w$ as $p_1 \to p_2$. In this case, one can use the strong line path ℓ_s regardless of the pressure range, since the strong line path reduces to the correct weak line path for thin layers where weak line approximation becomes valid. A common choice for the reference pressure is the average $(p_1 + p_2)/2$ but one could just as well choose one of the endpoints of the interval instead. In the case of a well-mixed greenhouse gas (constant q_G) for a nearly isothermal layer, the equivalent path becomes $1/\cos \bar{\theta}$ times $\frac{1}{2g}q_G(p_2^2 - p_1^2)/p_o$, which reduces to the correct transmission by using the conventional mass path with absorption coefficients computed for the average pressure of the layer. This is known as the *Curtiss-Godson* approximation.

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

The additivity of strong-line equivalent widths breaks down at large paths. Since each W_i increases like the square root of the path, eventually the sum exceeds Δ , leading to the absurdity of a negative transmission. What is going wrong is that, as the equivalent widths become large, the absorption regions associated with each line start to overlap. One is trying to take away the same chunk of the spectrum more than once. This doesn't work for spectra any more than it works for ten hungry people trying to eat an eighth of a pizza each. One approach which has met with considerable success is to assume that the lines are randomly placed, so that the transmission functions due to each line are uncorrelated. This is *Goody's Random Overlap Approximation*. For

uncorrelated transmission functions, the band-averaged transmissions can be multiplied, yielding

$$\bar{\mathfrak{T}} \approx (1 - \frac{W_1}{\Delta})(1 - \frac{W_2}{\Delta})...(1 - \frac{W_N}{\Delta}) = \exp\sum \ln(1 - \frac{W_i}{\Delta}) \approx \exp(-\frac{1}{\Delta}\sum W_i)$$
(4.70)

The last step follows from the assumption that each individual equivalent width is small compared to Δ . Note that when the sum of the equivalent widths is small compared to Δ , this expression reduces to the previous expression given for individual or non-overlapping lines.

The line parameters laid out above – position, width, strength and temperature scaling – lie at the heart of most real-gas radiative transfer calculations. There being thousands of spectral lines for dozens of substances of interest in planetary climate, teasing out the data one needs from the original literature is a daunting task. Fortunately, a small but dedicated group of spectroscopists ⁴ have taken on the task of validating, cross-checking and assembling the available line data into a convenient database known as HITRAN. It is suitable for most planetary calculations, though it must sometimes be supplemented with information on absorption that is not associated with spectral lines (the *continuum* absorption), and with additional data on weak lines which are important in the extremely hot, dense atmosphere of Venus. Instructions for obtaining the HITRAN database, along with sources for additional spectral data of use on Venus, Titan and Jupiter, are given in the references section at the end of this chapter, and the software supplement to this book provides a simple set of routines for reading and performing calculations with the HITRAN database.

4.4.4 Behavior of the band-averaged transmission function

Although the absorption spectrum has very complex behavior, the band-averaged transmission function averages out most of the complexity. The definition of the transmission guarantees that it decays monotonically as $|p_1 - p_2|$ increases and the path increases, but in addition the decay is invariably found to be smooth, proceeding without erratic jumps, kinks or other complex behavior. This smoothness is what makes computationally economical radiative transfer solutions possible, and the various schemes for carrying out the calculation of fluxes amount to different ways of exploiting the smoothness of the band-averaged transmission function.

By way of example, the band-averaged transmission function for CO_2 is shown for three different bands in Fig. 4.13. The calculation of $\bar{\mathfrak{T}}_{\nu}(p_1, p_2)$ was carried out using a straightforward – and very time consuming – integration of the transmission over frequency; at each frequency in the integrand, one must do an integral of $\kappa_{CO2}(\nu, p)$ over pressure, and each of those κ must be evaluated as a sum over the contributions of up to several hundred lines. Temperature was held constant at 296K and a constant mass-specific concentration of .0005 (330ppmv) of CO_2 mixed with air was assumed. The pressure p_1 was held fixed at 100mb, while p_2 was varied from 100mb to 1000mb. This plot thus gives an indication of the upward flux transmitted from each layer of the atmosphere, as seen looking down from the Earth's tropical tropopause. The results are plotted as a function of the pressure-weighted strong-line path, which for constant q and T is $q \cdot (p_2^2 - p_1^2)/(2gp_o \cos theta)$, where the reference pressure p_o is taken to be 10^5Pa . Plotting the results this way makes it easier to compare them with theoretical expectations, and also makes it easier to generalize the results to transmission between different pairs of pressure levels, which will

⁴May they live forever!

have different amounts of pressure broadening. The rationale for using the strong-line path is that the lines are narrow enough that almost all parts of the spectrum are far from the line centers in comparison to the width, and in such cases the collision-broadened absorption coefficient increases linearly with pressure almost everywhere. This behavior is incorrect near the line centers, but the error in the transmission introduced by this shortcoming is minimal, since the absorption is so strong there the contribution to the transmission is essentially zero anyway. This reasoning – based directly on what we have learned from the strong-line limit – is at the basis of most representations of pressure-broadening effects in radiative calculations. Here, we only are using it as a graphical device, since the transmission itself is computed without approximation. Note that the strong line path becomes proportional to the (weak-line) mass path $q \cdot (p_2 - p_1)/(g \cos \bar{\theta})$ when $p_2 \rightarrow p_1$, with proportionality constant p_1/p_o . In the present calculation, when p_2 is at its limit of 1000mb, the path is about $5kg/m^2$, which is about half the unweighted mass path over the layer. This reflects the fact that the lower pressure over most of the layer weakens the absorption relative to the reference value at $p = p_o$.

Apart from noticing that the transmission function is indeed smooth, we immediately remark that the transmission first declines sharply, as portions of the spectrum with the highest absorption coefficient are absorbed. At larger paths, the spectrum becomes progressively more depleted in easily-absorbed wavenumbers, and the decay becomes slower. For the two strongly absorbing bands in the left panel, the transmission curve becomes nearly vertical at small paths, as suggested by the square-root behavior of the strong line limit. There is guaranteed to be a weak-line region at sufficiently small paths, where the slope becomes finite, but in these bands the region is so tiny it is invisible. In fact, the strong line transmission function in Eq. 4.69 fits the calculated transmission in the 575-600 cm^{-1} band almost exactly throughout the range of paths displayed, when used with the random-overlap modification in Eq. 4.70. For the more strongly absorbing 600-625 cm^{-1} band the fit is very good out to paths of 1.5 kg/m^2 , but thereafter the actual transmission decays considerably more rapidly than the strong-line form. This mismatch occurs because the derivation of the strong-line transmission function assumes that the absorption coefficients within the band approach zero arbitrarily closely: as more and more radiation is absorbed, there is always some region where the absorption coefficient is arbitrarily close to zero, which leads to ever-slower decay. In reality, overlap between the skirts of the lines leads to finite-depth valleys between the peaks (see the inset of Fig. 4.7), and the absorption is bounded below by a finite positive value. The decay of the transmission at large paths is determined by the local minima in the valleys, and will tend toward exponential decay, rather than the slower decay predicted by the strong line approximation.

For the weakly absorbing band shown in the right panel of Fig. 4.13, a hint of weakline behavior can be seen at small values of the path, with the result that the behavior diverges noticeably from the best strong-line fit. The representation of the transmission can be improved by adopting a two-parameter fit tailored to give the right answer in both the weak and strong limits. The *Malkmus model* is a handy and widely-used example of this approach. It is defined by

$$\sum W_i = 2\frac{R^2}{S} \frac{p_1}{p_o} \left(\sqrt{1 + \frac{S^2}{R^2} (\frac{p_o}{p_1})^2 \ell_s - 1} \right)$$
(4.71)

where R and S are the parameters of the fit ⁵. The parameters can be identified with characteristics of the absorption spectrum in the band by looking at the weak line (small ℓ_s) and strong line (large ℓ_s) limits. For small ℓ_s , the sum of the equivalent widths is $S \cdot (p_o/p_1)\ell_s = S\ell_w$, so by comparing

⁵The factor p_1/p_o deals with the difference between the strong line and weak line paths, and is necessary so that the limits work out properly for small and large path. There is some flexibility in defining this factor. It is common to use $\frac{1}{2}(p_1 + p_2)/p_o$ to make things look more symmetric in p_1andp_2 . This slightly changes the way the function interpolates between the weak and strong limits, without changing the endpoint behavior



Figure 4.13: The band averaged transmission as a function of path, for the three different bands, as indicated. In each case, the transmission is computed between a fixed pressure $p_1 = 100mb$ and a higher pressure p_2 ranging from 100mb to 1000mb. Calculations were carried out assuming the temperature to be constant at 296K, with a constant CO_2 specific concentration of q = .0005, and assuming a mean propagation angle $\cos theta = \frac{1}{2}$. Results are plotted as a function of the pressure-weighted path for strong lines, $q \cdot (p_2^2 - p_1^2)/(2gp_o \cos theta)$, where $p_o = 1000mb$. In the left panel, the best fit to the strong-line transmission function is shown as a dashed curve; the fit is essentially exact for the $575 - 600cm^{-1}$ band, so the fitted curve isn't visible. For the weaker absorption band in the right panel, fits are shown both for the strong line and the Malkmus transmission function, but the Malkmus fit is essentially exact and can't be distinguished.

with Eq. 4.64 we identify S as the sum of the line intensities. For large ℓ , the sum is $2\sqrt{R^2\ell}$, whence on comparison with Eq. 4.69 we identify R^2 as the sum of $\gamma_i(p_o)S_i$ for all the lines in the band. The parameters R and S can thus be determined directly from the database of line intensities and widths, though in some circumstances it can be advantageous to do a direct fit to the results of a line-by-line calculation like that in 4.13 instead. One uses the Malkmus equivalent-width formula with the random-overlap transformation given in Eq. 4.70, so as to retain validity at large paths. With the Malkmus model, the transmission function in the weakly absorbing 550-575 cm^{-1} band can be fit almost exactly. Since the Malkmus model reduces to the strong line form at large paths, it fits the transmission functions in the left panel of Fig. 4.13 at least as well as the strong line curve did. However, it does nothing to improve the fit of the strongly absorbing case at large paths, since that mismatch arises from a failure of the strong-line assumption itself.

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

Empirical fits to the transmission function are a time-honored and effective means of dealing with infrared radiative transfer. This approach has a number of limitations, however. We have already seen some inadequacies in the Malkmus model when the path gets large; patching up these problems leads to fits with more parameters, and finding fits that are well-tailored to the characteristics of some new greenhouse gas one wants to investigate can be quite involved. It also complicates the implementation of the algorithm to have to use different classes of fits for different gases, and maybe even according to the band being considered. A more systematic and



Figure 4.14: Evaluation of the area under a curve by Lebesgue integration

general approach is called for. The one we shall pursue now, known as *exponential sums*, has the additional advantage that it can be easily generalized to allow for the effects of scattering, which is not possible with band-averaged fits like the Malkmus model. As a gentle introduction to the subject, let's consider the behavior of the integral

$$\bar{\mathfrak{T}}(\ell) = \frac{1}{\Delta} \int_{\nu_o - \Delta/2}^{\nu_o + \Delta/2} e^{-\kappa_G(\nu)\ell} d\nu$$
(4.72)

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

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

$$\int_{0}^{1} f(x)dx \approx \sum_{1}^{N} f(\frac{x_{j} + x_{j+1}}{2})(x_{j+1} - x_{j})$$
(4.73)

The problem with this approach is that a great many rectangles are needed to represent the complex area under the curve f(x). Instead, we may define the function H(a), which is the sum of the lengths of the intervals for which $f(x) \leq a$, as illustrated in Fig. 4.14. Now, the integral can be approximated instead by the sum

$$\int_{0}^{1} f(x)dx = \int_{f=f_{min}}^{f_{max}} fdH(f) \approx \sum_{1}^{M} \frac{f_j + f_{j+1}}{2} (H(f_{j+1}) - H(f_j))$$
(4.74)

where we have divided the range of the function f (i.e. $[f_1, f_2]$) into M partitions. This representation can be very advantageous if H(f) is a much more smoothly varying function than f(x). To mathematicians, this form of the approximation of an integral by a sum is the first step in the magnificent apparatus of *Lebesgue integration*, leading onwards to what is known as measure theory, which forms the basis of rigorous real analysis.

The idea is to apply the Lebesgue integration technique to the transmission function defined in Eq 4.72, with the absorption coefficient κ_G playing the role of f and the frequency ν playing



Figure 4.15: Cumulative probability function of the natural log of the absorption coefficient for CO_2 . Results are given for the 600-625 cm^{-1} and 575-600 cm^{-1} bands, and were computed at a pressure of 100mb and a temperature of 296K.

the role of x. Thus, if H(a) is the sum of the lengths of the frequency intervals in the band for which $\kappa_G \leq a$, then H(a) = 0 when a is less than the minimum of κ_G and H(a) approaches the bandwidth Δ when a approaches the maximum of κ_G . The transmission function can then be written approximately as

$$\bar{\mathfrak{T}}(\ell) = \int_{\kappa_{min}}^{\kappa_{max}} e^{-\kappa_G \ell} dH(\kappa_G) \approx \sum_{1}^{M} e^{-(\kappa_{j+1} + \kappa_j)\ell/2} (H(\kappa_{j+1}) - H(\kappa_j))$$
(4.75)

This is the exponential sum formula. It can be regarded as an M term fit to the transmission function, much as the Malkmus model is a two-parameter fit. The Lebesgue integration technique amounts to a simple reshuffling of the terms in the integrand: we collect together all wavenumbers with approximately the same κ , compute the transmission for that value, and then weight the result according to "how many" such wavenumbers there are.

Because the absorption coefficient varies over such an enormous range, it is more convenient to work with $H(\ln \kappa)$ rather than $H(\kappa)$. A typical result for CO_2 is shown in Fig. 4.15, computed for two bands at a pressure of 100mb. The function is quite smooth, and can be reasonably well characterized by ten points or less. In contrast, given that the typical line width at 100mb is only $.01cm^{-1}$, evaluation of the transmission integral in the Riemann form, Eq. 4.73, would require at least 25000 points in a band of width $25cm^{-1}$. Thus, the exponential sum approach is vastly more economical of computer time than a direct line-by-line integration would be.

The decay of the transmission with path length described by Eq. 4.75 is exactly analogous to the decay in time of the concentration of a mixture of radioactive substances with different half-lives. The short-lived things go first, leading to rapid initial decay of concentration; as time goes on, the mixture is increasingly dominated by the long-lived substances, and the decay rate is correspondingly slower. The way the transmission function converges as additional terms are included in the exponential sum formula is illustrated in Figure 4.16. Specifically, we divide the range of absorption coefficients into 20 bins equally spaced in log κ_G , and then truncate H so as to keep only the N largest absorption coefficients, with N ranging from 1 (retaining only the strongest absorption) to 20 (retaining all absorption coefficients including the weakest). When



Figure 4.16: Convergence of exponential sum representation of the band averaged transmission function as the number of terms is increased. The calculation is for CO_2 in a wavenumber band near the peak absorption.

only the strongest absorptions are included, the steep decay of transmission for small paths is correctly represented, but the transmission function decays too strongly at large paths. As more of the weaker absorption terms are included, the weaker decay of the transmission is well represented out to larger and larger paths. The ability to represent the decay rate of transmission over a very large range of paths is one of the two advantages of exponential sums over the Malkmus approach, the other advantage being the ability to incorporate scattering effects. An analytical example exploring related features of the exponential integral in Eq. 4.75 is explore in Problem ??.

If it weren't for the dependence of absorption coefficient on pressure and temperature, the exponential sum representation would be exact in the limit of sufficiently many terms. The computational economy of exponential sums comes at a cost, however, which is scrambling the information about which absorption coefficient corresponds to which frequency. This is not a problem if one is dealing with a layer with essentially uniform pressure and temperature, but it becomes a cause for concern in the typical atmospheric case where one is computing the transmission over a layer spanning considerable variations in temperature and pressure. The problem is that changing pressure or temperature changes the shape of the distribution $H(\kappa)$, and there is no rigorously correct way to deal with this within the exponential sum framework. In the discussion of line shapes, for example, we learned that increasing p reduces the peak absorption, but increases absorption between peaks. In terms of the probability distribution of κ , this means that the largest and smallest values of κ become less prevalent at the same time that the intermediate values become more prevalent. At very large values of pressure where the lines become extremely broad, κ becomes a smooth function of frequency within each band and the probability distribution becomes concentrated on a single mean value of κ . The effect of temperature on the shape of $H(\kappa)$ can be even more complex, since the temperature-dependence coefficients of line strength can differ greatly even for neighboring lines.

All these problems notwithstanding, experience has shown that one can obtain a reasonably accurate approximation to the band-averaged transmission function by assuming that all the absorption coefficients within a band have separable scaling of the form

$$\kappa_G(p,T) = \kappa_G(p_o,T_o)F(p/p_o,T/T_o) \tag{4.76}$$

Given scaling of this sort, one can compute the transmission for a path through an inhomogeneous atmosphere using Eq 4.75, by defining a suitable equivalent path. For example, if the specific

concentration of the greenhouse gas is q_G , the absorption scaling is linear in pressure, and the temperature dependence scales according to a function $S(T/T_o)$, then the equivalent path for the layer between p_1 and p_2 is

$$\ell = \frac{1}{g} \int_{p_1}^{p_2} q_G(p) \frac{p}{p_o} S(\frac{T(p)}{T_o}) dp$$
(4.77)

The temperature scaling function S would be computed separately for each band. The use of linear scaling in pressure is justified by the fact that most of the spectrum is far from line centers, so absorption scales like the strong-line approximation as long as the pressure is not so large that line widths become comparable to the width of the band under consideration. The use of a single temperature-scaling function is harder to justify theoretically, but seems to be supported by numerical experiment.

When q_G is not small, there is one additional complication to take into account when defining the equivalent path. Namely, the absorption coefficient for self-broadened collisions is generally different from that for foreign-broadened collisions, so one must scale the absorption coefficient according to the proportion of self vs. foreign collisions. Like the temperature scaling factor, the ratio of self to foreign broadening can vary considerably even amongst nearby lines. Typically, though, one simply chooses a representative ratio of self to foreign broadening which is assumed constant within each band over which the distribution H is computed. Let's call this ratio a_{self} . The molar concentration of the greenhouse gas is $(\bar{M}/M_G)q_G$, where \bar{M} is the mean molecular weight of the mixture; hence the proportion of collisions which are self-collisions is $(\bar{M}/M_G)q_G$ while the proportion of foreign collisions is $1 - (\bar{M}/M_G)q_G$. Then, if H is computed using the foreign-broadened absorption at the standard pressure, the appropriate equivalent path to use in computing the transmission is

$$\ell = \frac{1}{g} \int_{p_1}^{p_2} (1 + \frac{\bar{M}}{M_G} q_G(p) (a_{self} - 1)) q_G(p) \frac{p}{p_o} S(\frac{T(p)}{T_o}) dp$$
(4.78)

For a pure one-component atmosphere, $(\bar{M}/M_G)q_G = 1$ and one simply uses the self-broadened absorption coefficients in preparing the distribution $H(\kappa)$, rather than going through the intermediary of defining a_{self} for each band.

Because the scaling of absorption coefficient with pressure and temperature is only approximate, it is important to compute H for a reference pressure and temperature that is characteristic of the general range of interest for the atmosphere under consideration, so as to minimize the amount of scaling needed. Typically, one might use a half or a tenth of the surface pressure as the reference pressure, and a mid-tropospheric temperature as the reference temperature; if one were primarily interested in stratospheric phenomena, or if one were computing OLR on a planet like Venus where most of the OLR comes from only the uppermost part of the atmosphere, pressures and temperatures characteristic of a higher part of the atmosphere would be more appropriate.

4.4.5 Dealing with multiple greenhouse gases

We now know how to efficiently compute the band-averaged transmission function for a single greenhouse gas acting alone. It is commonly the case, however, that two or more greenhouse gases are simultaneously present – CO_2 , CH_4 and water vapor in Earth's case, for example. How do we compute the averaged transmission function is this situation? The issues are closely related to thosed discussed in Section 4.4.1 in connection with the loss of the multiplicative property in band-averaged transmission functions. Similar reasoning shows that the average transmission function for two greenhouse gases acting together generally differs from the product of the averaged

4.4. REAL GAS RADIATION: BASIC PRINCIPLES

transmission functions of each of the individual gases taken alone. Fortunately, however, the special circumstances under which the multiplicative property holds for multiple gases are expected to be fairly common.

By way of illustration, let's consider an idealized greenhouse gas A with transmission function $\mathfrak{T}_A(\nu)$ in a wavenumber band of width Δ , with the property that $\mathfrak{T}_A = a$ on a set of wavenumber subintervals with length adding up to $r \cdot \Delta$, but with $\mathfrak{T}_A = 1$ elsewhere; naturally, we require a < 1 and $r \leq 1$. Consider a second greenhouse gas B with $\mathfrak{T}_B = b$ on a set of wavenumber subintervals with length adding up to $s \cdot \Delta$, and with $\mathfrak{T}_B = 1$ elsewhere. The band-averaged transmission for gas A in isolation is $r \cdot a + (1 - r)$, and for B in isolation is $s \cdot b + (1 - s)$. What is the transmission when both gases act in combination?

The answer depends on how the regions of absorption of gas A – spectral regions where $\mathfrak{T}_A < 1$ – line up with those of gas B. We can distinguish three limiting cases. First, the regions of absorption of the two gases may be perfectly correlated, in which case r = s, $\mathfrak{T}_A(\nu) = a$ precisely where $\mathfrak{T}_B(\nu) = b$, and

$$\bar{\mathfrak{T}}(\nu) = \frac{1}{\Delta} \int_{\nu-\Delta/2}^{\nu+\Delta/2} \mathfrak{T}_A(\nu') \mathfrak{T}_B(\nu') d\nu'$$

$$= rab + (1-r)$$

$$= \bar{\mathfrak{T}}_A \bar{\mathfrak{T}}_B - r(1-r)(a+b-ab-1)$$
(4.79)

Further, a + b - ab = a + (1 - a)b, and so this expression has a value lying between a and b, both of which are less then unity. Hence (a + b - ab - 1) < 0 and $r(1 - r)(a + b - ab - 1) \leq 0$, given that $0 \leq r \leq 1$. We conclude that the mean transmission function for the two gases acting in concert is always greater than or equal to the product of the individual transmission functions, the equality applying only when r = 0 or r = 1, i.e. when there is no absorption or completely uniform absorption. For example, if we take $r = \frac{1}{2}$ and a = b = frac110, the mean transmission function for the two gases acting together is $\frac{101}{200}$, or just a bit over a half, whereas the individual transmission functions are $\frac{11}{20}$ each, multiplying out to $\frac{121}{400}$ or just over a quarter. When the absorption regions of the two gases coincide, the gases acting together transmit considerably more radiation than one would infer by allowing each gas to act independently in sequence. This happens because one of the gases uses up some of the frequencies that the other gas would like to absorb.

Exercise 4.4.1 Derive the final equality in Eq. 4.79. Sketch graphs of transmission functions vs. frequency for the two gases for *two different* cases illustrating perfectly correlated absorption regions. Evaluate the mismatch between $\bar{\mathfrak{T}}$ and $\bar{\mathfrak{T}}_A \bar{\mathfrak{T}}_B$ for $a = b = r = \frac{1}{2}$. Allowing a, b and r to vary over all possible values, what is the greatest possible mismatch?

At the other extreme, the absorption regions of the two gases may be completely *disjoint*, so that $\mathfrak{T}_A = 1$ wherever $\mathfrak{T}_B = b < 1$ and $\mathfrak{T}_B = 1$ wherever $\mathfrak{T}_A = a < 1$. For this to be possible, we require $r + s \leq 1$. In the disjoint case,

$$\bar{\mathfrak{T}}(\nu) = \frac{1}{\Delta} \int_{\nu-\Delta/2}^{\nu+\Delta/2} \mathfrak{T}_A(\nu') \mathfrak{T}_B(\nu') d\nu'$$

$$= ra + sb + (1 - (r+s))$$

$$= \bar{\mathfrak{T}}_A \bar{\mathfrak{T}}_B - rs(1-a)(1-b)$$
(4.80)

In the disjoint case, then, the transmission for the two gases acting together is always *less than* the compounded transmission of the two gases acting independently.

Exercise 4.4.2 Derive the final equality in Eq. 4.80. Sketch some transmission functions illustrating the disjoint case. Put in a few numerical values for a, b, r and s to show the size of the mismatch between $\bar{\mathfrak{T}}$ and $\bar{\mathfrak{T}}_A \bar{\mathfrak{T}}_B$. What is the greatest possible mismatch in the disjoint case?

As the final limiting case, suppose that the absorption of the two gases is *uncorrelated*, so that at any given frequency the probability that $\mathfrak{T}_A = a$ is r regardless of the value of \mathfrak{T}_B there, and the probability that $\mathfrak{T}_B = b$ is s regardless of the value of \mathfrak{T}_A there. This situation is also known as the *random overlap* case. In this case

$$\bar{\mathfrak{T}}(\nu) = \frac{1}{\Delta} \int_{\nu-\Delta/2}^{\nu+\Delta/2} \mathfrak{T}_A(\nu') \mathfrak{T}_B(\nu') d\nu'$$

= $r(1-s)a + s(1-r)b + rsab + (1-r)(1-s)$
= $\bar{\mathfrak{T}}_A \bar{\mathfrak{T}}_B$ (4.81)

The reasoning behind the second line is that r(1-s) is the probability that only the first gas is absorbing, s(1-r) is the probability that only the first gas is absorbing, rs is the probability that both gases are absorbing, and (1-r)(1-s) is the probability that neither gas is absorbing. Multiplying out the terms in the product of \mathfrak{T}_A and \mathfrak{T}_B we find that in the random overlap case, the mean transmission of the two gases acting together is precisely the same as the compounded transmission of the two gases acting independently.

The properties illustrated by the three cases just discussed can be generalized to an arbitrary set of transmission functions. Let \mathfrak{T}_A and \mathfrak{T}_B be any two transmission functions, and define the fluctuation

$$\mathfrak{T}_A' = \mathfrak{T}_A - \mathfrak{T}_A \tag{4.82}$$

and similarly for \mathfrak{T}_B . Then,

$$\overline{\mathfrak{T}_A \mathfrak{T}_B} = \overline{\mathfrak{T}}_A \overline{\mathfrak{T}}_B + \overline{\mathfrak{T}'_A \mathfrak{T}'_B} \tag{4.83}$$

From this we conclude that the transmission of the two gases acting in concert is greater than the product of the individual transmissions if the two transmissions are positively correlated, less than the product of the individual transmissions of the two transmissions are negatively correlated, and equal to the product if the two transmissions are uncorrelated.

In fact, by exercising just a little more mathematical sophistication, it is possible to go further and put an upper bound on the amount by which the mean transmission function for joint action by the two gases deviates from the product of the individual transmission functions. The key is to use a handy and powerful relation known as the *Schwarz Inequality*, which states that for any two functions f(x) and g(x), $(\overline{fg})^2 \leq \overline{f}^2 \overline{g}^2$, where an overbar indicates an average over x. The equality applies only when f(x) is proportional to g(x). Applying the Schwarz Inequality to Eq. 4.83, we find that the deviation satisfies the inequality

$$\left|\overline{\mathfrak{T}'_{A}\mathfrak{T}'_{B}}\right| \leq \sqrt{(\overline{\mathfrak{T}'_{A}})^{2}}\sqrt{(\overline{\mathfrak{T}'_{B}})^{2}} \tag{4.84}$$

In other words, in the worst case the deviation can become as large as the product of the standard deviations of the two individual transmission functions. Since $0 \leq \mathfrak{T}'_{A,B} \leq 1$, the maximum standard deviation is $\frac{1}{2}$, occuring when each transmission function is zero for half of the frequencies in the band and unity for the other half; the error in random overlap in this case is $+\frac{1}{4}$ if the transmissions are perfectly correlated and $-\frac{1}{4}$ if the transmissions are perfectly anticorrelated. These errors should be compared to the random-overlap value for the limiting case, which is $\frac{1}{4}$. The effects of non-random overlap are potentially severe.

4.4. REAL GAS RADIATION: BASIC PRINCIPLES

Fortunately, the situation is rarely as bad as the worst-case suggests. The positions of spectral lines are a sensitive function of molecular structure, so it is a reasonable guess that the absorption spectra of dissimilar molecules should be fairly uncorrelated. Thus, in most circumstances one can get a reasonable approximation to the joint transmission function by computing the individual transmission functions for each gas and taking the product of the individual transmission functions. It is fairly easy, if computationally expensive, to test the accuracy of the random-overlap assumption in a given band by computing the correlations of the full frequency-dependent transmission functions in the band. However, finding a general characterization of the correction to random-overlap, and the way the correction depends on the concentrations of the individual gases, is an intricate art which we will not pursue here.

4.4.6 A homebrew radiation model

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

- A means of computing the band-averaged transmission over a specified wavenumber range
- The band-averaged integral (Eq. 4.11,4.12, or 4.13) giving the band-averaged solution to the Schwartzschild equation in terms of the preceding transmission functions

and the recipe is:

- Divide the spectrum into bands of a suitable width
- Prepare in advance: Malkmus coefficients or exponential sum coefficients $H(\ln \kappa)$ for each band, for each greenhouse gas present in significant quantities in the atmosphere
- Program up a function to compute the band averaged transmission in each band, using the coefficients prepared in the previous step.
- If there are multiple greenhouse gases, do the preceding for each individual greenhouse gas and combine the resulting transmission functions, allowing suitably for the nature of the overlap between absorption bands of the competing gases (for advanced chefs only!)
- Use the resulting transmission in a numerical approximation to the integral in Eq. 4.11,4.12 or 4.13 in each band to get the band-averaged fluxes.
- Sum up the fluxes in each band to get the total flux
- Serve up the fluxes to the rest of the climate model and enjoy

In the typical climate simulation application, one is given a list of values of temperature and greenhouse gas concentrations tabulated on a finite grid of pressure levels p_j for j = 0, ...N, and one must compute the fluxes based on this information. Either of Eq. 4.11 or 4.12 provides a suitable basis for numerical evaluation when one is working from atmospheric data tabulated on a grid. In writing down the approximate expressions for the flux, we will adopt the convention that j = 0 at the top of the atmosphere and that j = N represents the ground. We shall use the superscript (k) to refer to quantities averaged or integrated over the band k, centered on frequency $\nu^{(k)}$ and having width $\Delta^{(k)}$. Let's define the gridded quantities:

$$B_{j} \equiv B(\nu^{(k)}, \frac{1}{2}(T_{j} + T_{j+1}))\Delta^{(k)}$$

$$\bar{\mathfrak{T}}_{ij}^{(k)} \equiv \bar{\mathfrak{T}}^{(k)}(p_{i}, p_{j})$$

$$e_{ij} \equiv \bar{\mathfrak{T}}_{ij}^{(k)} - \bar{\mathfrak{T}}_{i(j+1)}^{(k)}$$

(4.85)

The trapezoidal-rule approximation to the the expression for upward flux in band (k), based on Eq. 4.11, is then simply

$$I_{+}^{(k)}(p_i) = I_{+,s}^{(k)} \bar{\mathfrak{T}}_{iN}^{(k)} + \sum_{j=i}^{N} B_j e_{ij}$$
(4.86)

The expression for the downward flux follows a similar form. B_j is the blackbody emission from layer j, and the flux at a given level is a weighted sum of the emissions from each layer below (for upward flux) or above (for downward flux) layer i. The weighting coefficient e_{ij} characterizes the joint effects of the emissivity at layer j and the absorptivity by all layers between i and j.

Exercise 4.4.3 Write down the analogous trapezoidal-rule approximation to I_{-}

Exercise 4.4.4 Write down analogous trapezoidal-rule approximations to I_+ and I_- based on the form of the solution given in Eq. 4.12. What would be the advantages of using this form of solution?

To implement Eq. 4.86 and its variants as a computer algorithm, one generally writes a function which computes the transmission between levels p_i and p_j . The rest of the algorithm is independent of the form this function takes, and so one can easily switch from one representation to another (e.g. Malkmus to exponential sums) by simply switching functions. One can equally easily use different representations for different bands. The transmission function requires as arguments the transmission parameters for the band under consideration (e.g. R and S parameters for Malkmus, or the H distribution for exponential sums), as well as enough information to compute the equivalent path. For a well-mixed greenhouse gas, if we are ignoring temperature scaling effects the equivalent path is simply $q_G \frac{1}{2} |(p_1^2 - p_2^2)| / (p_o g \cos \bar{\theta})$, and one can simply make the concentration q_G and the pressures p_1 and p_2 arguments of the transmission function. For an inhomogeneous path, arising when q_G varies with height or one needs to take into account temperature scaling which also varies in height, the path is determined by an integral. In this case, it is inefficient to recompute the path from scratch each time. Since the equivalent path can be computed incrementally using $\ell(p_1, p_2 + \delta p) = \ell(p_1, p_2) + \ell(p_2, p_2 + \delta p)$, it is better to use the equivalent path as an argument to the transmission function, and compute the path from layer i to each layer j iteratively in the same loop in Eq. 4.86 where the weighted emission is computed.

In the preceding algorithm, we have used exponential sums to represent the transmission function appearing in the integral form of the solution to the Schwartzschild Equations. However, because the equations are linear in the fluxes, and because the exponential sum method is a weighted sum of calculations for a number of different absorption coefficients, exactly the same results can be obtained by organizing the calculation in a quite different way. Namely, instead of working from the integral form of the solution, we can work directly with a set of independent Schwartschild equations (Eq. 4.8) – one for each κ going into the exponential sum for a given band; as usual, the band would be chosen narrow enough that $B(\nu, T)$ could be assumed independent of frequency within the band, so we wouldn't need to know anything about which set of wavenumbers each κ corresponded to. With a 10-term exponential sum, for example, we would solve the Schwartschild equation for each of the 10 values of κ , then form a weighted sum of the 10 resulting fluxes. This alternate formulation is not available with band-averaged transmission function models such as the Malkmus model. The weighted differential-equation approach offers a number of advantages over the transmission-function approach. A single solution gives the fluxes at all levels, making optimal use of calculations for preceding levels. This is useful when computing heating rate profiles. Moreover, it is easy to use a high-order integrator to obtain high accuracy with fewer levels. These two advantages make the method computationally attractive, but there is a further advantage that is even more compelling for our purposes: it is straightforward to extend this method to incorporate scattering, whereas it is essentially impossible to do so with band-averaged transmission approaches. We will carry out this program in Chapter 5. One could well ask why the weighted differential-equation approach hasn't completely taken over the business of radiation modelling. There is some evidence that, when scattering is unimportant, pressure and temperature scaling can be done more accurately in band-averaged transmission models, but in large measure the transmission models are a holdover from an earlier day when many radiation calculations were done on paper, and when slow computers required highly tuned special approximations in order to speed up the calculation. It does seem that the exponential sum (and its close cousin the *correlated-k*) approach are gradually taking over. We have nonetheless chosen to introduce the transmission function approach first, because it corresponds better to what is going on in most existing radiation models, and because the form of the solution gives considerable direct insight into the factors governing the fluxes at a given level.

4.4.7 Spectroscopic properties of selected greenhouse gases

Now we will provide a survey of the infrared absorption properties of a few common greenhouse gases, with particular emphasis on the "big 3" that determine much of climate evolution of Earth, Mars and Venus from the distant past through the distant future: CO_2 , H_2O and CH_4 . This section will deal only with that part of absorption that can be identified as being caused by nearby spectral lines associated with energy transitions of the molecule in question; the *continuum absorption* which is not so directly associated with spectral lines will be discussed separately. Although CH_4 is a greenhouse gas on Titan, most of its contribution there comes from continuum absorption rather than its line spectrum.

Although we focus on some of the more conventional examples, the reader is encouraged to keep an open mind with regard to what might be a greenhouse gas. At present, NO_2 and SO_2 do not exist in sufficient quantities on any known planet to be important as a greenhouse gas, but with different atmospheric chemistries occuring in the past or on as-yet undiscovered planets, the situation could well be different. For that matter, things like SiO_2 that we consider rocks on Earth could be gases and clouds on "roasters" – extrasolar gas giants in near orbits – and there one ought to give some thoughts to their effect on thermal infrared.

When interpreting the absorption spectra to be presented below, it is useful to keep the Planck function in mind. Absorption is not very important where there is little flux to absorb, so the relevant part of the absorption curve varies with the temperature of the planet under consideration. For Titan at 100K, $\frac{3}{4}$ of the emission is at wavenumbers below $350 \ cm^{-1}$. For Earth at 280K, $\frac{3}{4}$ of the emission occurs below $1000 \ cm^{-1}$. Mars is slightly colder, and the threshold wavenumber is therefore a bit less. For Venus at 737K, $\frac{3}{4}$ of the emission is below $2550 \ cm^{-1}$, but moreover the flux beyond this wavenumber amounts to over $4000 \ W/m^2$. This near-infrared flux is vastly in excess of the mere $170W/m^2$ of absorbed solar radiation which maintains the Venusian climate. For Venus one needs to consider the absorption out to higher wavenumbers than for Earth or Mars,



Figure 4.17: The minimum, 25^{th} percentile, median, 75^{th} percentile, and maximum absorption coefficients in bands of width 50 cm^{-1} , computed for CO_2 in air from spectral line data in the HITRAN database. The left panel shows results at an air pressure of 100mb, whereas the right panel gives results at 1bar. Both are calculated at a temperature of 260K.

and given the large fluxes involved and the huge mass of CO_2 present, exquisite attention to detail and to the effect of weak lines that can normally be ignored is required. For the aforementioned roaster planets, the "thermal" emission extends practically out to the visible range.

An ideal greenhouse gas absorbs well in the thermal infrared part of the spectrum but is transparent to incoming solar radiation. The following survey concentrates on thermal infrared – the part of the spectrum that is involved in OLR– but it should be remembered that some of the gases discussed, such as Methane, absorb quite strongly in the solar near-infrared spectrum. This can compromise their effectiveness as greenhouse gases when they reach concentrations high enough that the solar absorption becomes significant. Even the weaker solar absorption from CO_2 and H_2O can affect the thermal structure of atmospheres in significant ways. The implications of real-gas solar absorption will be taken up in more detail in Chapters 5 and 7.

Carbon Dioxide

We have already introduced a fair amount of information about the absorption properties of CO_2 but here we will present the data in a more systematic and general way, which will make it easier to compare CO_2 with other greenhouse gases. In a line graph of a wildly varying function like the absorption, such as shown in Figure 4.7, only only sees the maximum and minimum defining the envelope of the absorption. There is no useful information about the relative probabilities of the values in between. To get around these problems, we divide up the spectrum into slices of a fixed width (50 cm^{-1} in the results presented throughout this section), and then compute the minimum, maximum, median and the 25^{th} and 50^{th} percentiles of the log of the absorption in each band. By plotting these statistics vs. wavenumber, it is possible to present a fairly complete picture of the probability distribution of the absorption. To make the absorption data easier to interpret, we exponentiate the median , quartiles, min and max, and plot the resulting values on a logarithmic vertical axis. Plots of this sort for CO_2 at a temperature of 260K are shown in Figure 4.17. The left panel is computed at a pressure of 100mb while the right panel is computed at 1000mb. Both results are for air-broadened lines.

4.4. REAL GAS RADIATION: BASIC PRINCIPLES

For Earthlike and Marslike temperatures, only the lower frequency absorption region, from about $400 - 1100cm^{-1}$ affects the OLR, since there is little emission in the range of the higher frequency band. For Earth, the higher frequency band can contribute to Solar absorption, since the Solar radiation reaching the Earth in that part of the spectrum amounts to $7.2W/m^2$ (or about $3W/m^2$ at the orbit of Mars). However, most of the absorption of Solar near-infrared, whether by water vapor or CO_2 , occurs at higher frequencies. Venus, however, has significant surface emission in the range of the higher frequency group, which therefore has a significant effect on that planet's OLR.

The absorption spectrum depends on temperature and pressure, and on whether the lines are broadened by collision with air or by collisions with molecules of the same type (e.g. CO_2 with CO_2 in the present case); rather than present page upon page of graphs, the strategy is to show the results for a single standard temperature and pressure, and make use of appropriate temperature and pressure weighted equivalent paths to apply the standard absorption to a wider variety of atmospheric circumstances. We will adopt T = 260K, p = 100mb and air-broadening as our standard conditions when discussing other greenhouse gases in this section. The comparison of the two panels of Figure 4.17 illustrates the nature of the pressure scaling. Increasing the pressure by a factor of 10 shifts the minimum, median and quartiles upward by a similar factor. The main exception to the scaling is the maximum absorption in each band, which is reduced and becomes more tightly clustered to the median. This is the expected behavior, since increased pressure reduces the absorption near the centers of absorption lines. Since this part of the spectrum is rare, and absorbs essentially everything hitting it anyway, the errors in pressure scaling near the line centers are of little consequence. A bit of numerical experimentation indicates that the linear pressure scaling works quite well for pressures below 1 or 2 bars. At higher pressures (certainly by 10 bar) the lines overlap to such an extent that the whole probability distribution collapses on the median, in which case the absorption can be described as a smooth function of the wavenumber, and loses its statistical character. The other gases to be discussed have similar behavior under pressure, and so we will not repeat the description separately for each gas.

To understand a planet's radiation balance, the main thing we need to understand is where in the spectrum the atmosphere is optically thick, and where it is optically thin. Pressure exerts by far the strongest modifying influence on the standard-state absorption coefficients given in Figure 4.17. Hence, we will first discuss the effect of CO_2 on the optical thickness of several planetary atmospheres in terms of the 260K air-broadened coefficients using pressure-adjusted paths. Later we will make a few remarks about how self-broadening and temperature affect the results.

The strong line pressure-weighted equivalent path for a well mixed greenhouse gas with specific concentration q is $\ell = \frac{1}{2}q(p_1^2 - p_2^2)/(p_o g \cos \bar{\theta})$. If we take the reference pressure $p_o = 100mb$, the equivalent paths can be multipled by the absorption coefficients in the left panel of Figure 4.17 to obtain optical thickness. Some typical values of the equivalent path are given in Table 4.1. Although Mars at present has much more CO_2 per square meter in its atmosphere than modern Earth, the equivalent paths are quite similar in the two cases because the total pressure on Mars is so much lower. For Earth and Mars temperatures only the lower frequency absorption group is of interest, and within this group it is only the dominant spike in absorption near $675cm^{-1}$ that contributes significantly to the absorption. Both atmospheres have optical thicknesses exceeding unity within the wavenumber band from roughly $610 - 750cm^{-1}$ and are optically thin outside this band. The weak absorption shoulder occurring to the high-wavenumber side of $870cm^{-1}$ has little effect for modern Earth or Mars. For any given wavenumber range, the 10% of the atmosphere nearest the ground (i.e. the lowest 100mb on Earth) is one fifth as optically thick as the atmosphere as a whole, meaning that a substantial part of the back-radiation of infrared to the surface attibutable to CO_2 comes from the lowest part of the atmosphere. In contrast, because

196CHAPTER 4. RADIATIVE TRANSFER IN TEMPERATURE-STRATIFIED ATMOSPHERES

	Mod. Earth	Early Earth	Mod. Mars	Early Mars	Venus
Whole atm	47.	15000.	30.	10^{6}	10^{9}
Bottom 10%	9.	2900.	5.	$2 \cdot 10^5$	$2 \cdot 10^8$
Top 10%	.5	160.	.3	10^{4}	10^{7}
Top 1%	.005	1.5	.003	100.	10^{5}

Table 4.1: Pressure-weighted equivalent paths for various planetary situations, in units of kg/m^2 . The weighted path is based on a 100mb reference pressure, so these paths are intended to be used with the absorption coefficients in the left-hand panel of Figure 4.17. A mean slant path $\cos \bar{\theta} = \frac{1}{2}$ is assumed. The Modern Earth case is based on 300ppmv CO_2 in a 1bar atmosphere, while the Early Earth case assumes 10% (molar) CO_2 in a 1bar atmosphere. Modern Mars is based on a 10mb pure CO_2 atmosphere, while Early Mars assumes a 2bar pure CO_2 atmosphere. The Venus case consists of a 90bar pure CO_2 atmosphere.

of low pressure aloft, the uppermost 10% of the atmosphere, which may loosely be thought of as the stratosphere, is comparatively optically thin. If CO_2 were the only factor in play, the greenhouse effect of the present Earth and Mars atmospheres would be quite similar. The two planets are rendered qualitatively different primarily by the greater role of water vapor in the warmer atmosphere of modern Earth, and by differences in vertical temperature structure between the two atmospheres (arising largely from differences in solar absorption).

On Early Earth – either during the high CO_2 phase after a long-lived Snowball, or during the Faint Young Sun period – the CO_2 may have been 10% or more of the atmosphere by molar concentration. In this case, the equivalent paths are vastly greater than at present. The Earth atmosphere in this case is optically thick from about $520 - 830cm^{-1}$, and the higher wavenumber shoulder just barely begins to be significant. It becomes rapidly more so as the CO_2 molar concentration is increased beyond 10%. This is good to keep in mind when doing radiation calculations at very high CO_2 , since many approximate radiation codes designed for the modern Earth only incorporate the effect of the principal absorption feature centered on $675cm^{-1}$. Even with such high CO_2 concentrations, the atmosphere is optically thick in only a limited portion of the spectrum, so that the net greenhouse effect will be modest unless other greenhouse gases come into play. This remark is particularly germane to Snowball Earth, where the cold temperatures allow little water vapor in the atmosphere. Without help from water vapor, CO_2 has only limited power to warm up a Snowball Earth to the point of deglaciation.

On a hypothetical Early Mars with a 2bar pure CO_2 atmosphere, the equivalent path is orders of magnitude greater than the Early Earth case. This renders the atmosphere nearly opaque within the wavenumber range $500 - 1100cm^{-1}$. However a great deal of OLR can still escape through the CO_2 window regions, so continuum absorption, water vapor and other greenhouse gases will play a key role in deciding whether the gaseous greenhouse effect can explain a warm, wet Early Mars climate.

Since the equivalent path assumes absorption increases linearly with pressure, the equivalent paths given for the bottom 10% and for the entirety of the Venus atmosphere yield overestimates of the true optical depths, given that the increase of absorption with pressure weakens substantially above 10bars. Even assuming that the equivalent paths are overestimated by a factor of 10, the implied optical depths for the lowest 9bars of the Venus atmosphere, and for the entire Venus atmosphere, remain so huge that these layers can be considered essentially completely opaque to infrared outside the CO_2 window regions. In fact even the top 1bar (about 1%) of the Venus atmosphere has an optical thickness greater than unity throughought the 500 – $1100cm^{-1}$ and $1800 - 2500cm^{-1}$ spectral regions. Within these regions, essentially all the OLR comes from the

relatively cold top 1bar of the atmosphere.

The window region, where CO_2 has no absorption lines, presents a challenge to the explanation of the surface temperature of Venus, particularly since the peak of the Planck function for the observed Venusian surface temperature lies in this region. In fact, if the $1200 - 1700 cm^{-1}$ band were really completely transparent to infrared, then emission through this region alone would reduce the Venus surface temperature to a mere 355K (assuming a globally averaged absorbed solar radiation of $163W/m^2$). Some energy also leaks out through the low frequency window region, which would reduce the temperature even further. We must plug the hole in the spectrum if we are to explain the high surface temperature of Venus. The high altitude sulfuric acid clouds of Venus play some role, by reflecting infrared back to the surface; there may also be some influence of the less common isotopes of CO_2 , since asymmetric versions of the molecule (e.g. made with one ${}^{16}O$ and one ^{18}O can have lines in the window region. The principal factor at play is the *continuum* absorption to be discussed in Section 4.4.8. Figure 4.17 was computed taking into account only the relatively nearby contributions of each absorption line (within roughly 1000 line widths), whereas collisions can allow some absorption to occur much farther from the line centers. This far-tail absorption spills over into the window region, particularly at high pressure. It cannot be reliably described in terms of the Lorentz line shape, and therefore requires separate consideration.

The difficulties posed by Venus do not end with the window region; For planets having Earthlike temperatures, the thermal radiation beyond $2300 cm^{-1}$ is insignificant, so we needn't be too concerned about the absorption properties there. However, for a planet with a surface temperature of 700K, the ground emission on the shortwave side of $2300 cm^{-1}$ is $3853 W/m^2$, compared to under $200W/m^2$ of absorbed solar radiation for Venus. In these circumstances it simply won't do to assume the atmosphere transparent at high wavenumbers The HITRAN spectral database is a monumental accomplishment, but it is still rather Earth-centric, and lacks the weak lines needed to deal with high temperature atmospheres; there may also be continuum absorption in the shortwave region, especially at high pressures. In order to deal with the high wavenumber part of the Venus problem, one needs to employ specialized high-temperature CO_2 databases, which are less verified and more in a state of flux than HITRAN. Among specialized databases in use for Venus, the HITEMP database (described in the supplementary reading at the end of this Chapter) is particularly convenient, because it at least uses the same data format as HITRAN. It will be left to the reader to explore the use of this database. Suffice it to say that there appears to be sufficient absorption in the high wavenumber region to raise the radiating level to altitudes where the temperature is low enough that one may not need to consider shortwave emission in computing OLR.

The preceding discussion was based on air-broadened absorption at 260K, whereas selfbroadened data would be more appropriate to the pure- CO_2 atmospheres and in all cases one must think about whether the increase of line strength with temperature substantially alters the picture presented. A re-calculation of Figure 4.17 for self-broadened pure CO_2 indicates that the self-broadened absorption is generally about 30% stronger than the air-broadened case, though there are a few bands in which the enhancement is as little as 13%. This is quantitatively significant, but the enhancement factor is too small to alter the general picture presented above. The temperature dependence can have a more consequential effect. In the strong line approximation, valid away from line centers, the temperature affects the absorption only in the form of the product of line strength and line width, $S(T)\gamma(p,T)$, which yields a temperature dependence of the form $\kappa_{CO2} \sim$ $\exp(-T^*/T)$ for some coefficient T^* . The coefficient differs from line to line, but we can still attempt to fit this form to the computed temperature dependence of the median absorption within each $50cm^{-1}$ band. The result is shown in Figure 4.18. This still only gives an incomplete picture of the effect of temperature on absorption, since the other quartiles may have different scaling



Figure 4.18: Left panel: Temperature dependence coefficient for air-broadened CO_2 at 100mb. The coefficient is computed for the median absorption in bands of width $50cm^{-1}$, as shown in Figure 4.17. The median absorption within each band increases with temperature in proportion to $\exp(-T^*/T)$. Right panel: The ratio of self-broadened to air-broadened median absorption coefficients.

coefficients. Within an exponential-sum framework, however, one can do little else that pick a base temperature most appropriate to the planet under consideration, compute the probability distribution for that case, and then assume that each absorption coefficient in a band scales with the same function of temperature. This procedure in any event gives an estimate of the magnitude of the temperature effect. From the Figure, it is seen that the temperature dependence varies greatly with wavenumber. It is low near the principal absorption spike at $675 cm^{-1}$, with T^* as low as 1000K. This value increases the absorption by a factor of 1.7 going from 260K to 300K, and decreases absorption by a factor of 2 going from 260K to 220K. Where $T^* = 3000K$, the corresponding ratios are 4.7 and .14. Therefore CO_2 is significantly more optically thick than our previous estimates in the warmer-lower reaches of the present Earth atmosphere, and significantly less optically thick in the cold trop pause regions. For the 737K surface temperature of Venus, the absorption is enhanced by a factor of 12 when $T^* = 1000K$ but by a substantial factor of 1750 when $T^* = 3000 K$. This just makes an already enormously optically thick part of the atmosphere even thicker. Outside the window regions, most of the atmosphere of Venus is in the optically thick limit where very slow radiative diffusion transfers heat; infrared radiative cooling in the deeper parts of the Venus atmosphere is determined almost exclusively by what is going on in the window regions.

Water Vapor

Figure 4.19 shows the standard-state absorption spectrum for water vapor. Unlike CO_2 , the H_2O molecule, which has more complex geometry, has lines throughout the spectrum, so there is no completely transparent window region. Water vapor nonetheless has two window regions where the absorption is very weak; it will turn out that continuum absorption from far tails excluded from the computation in the figure substantially increases the absorption in these window regions (Section 4.4.8). The peak absorption coefficient for water vapor has a similar magnitude to that for CO_2 , but water vapor absorbs well over a far broader portion of the spectrum than CO_2 . In particular, the H_2O absorption has a peak within both the $1000cm^{-1}$ and longwave window regions of CO_2 . This critically affects the greenhouse warming on Earth and on Early Mars, but it plays little role on present-day Venus, which has little water vapor in its atmosphere. It should not be concluded that water vapor overwhelms the greenhouse effect of CO_2 , however. It would be more precise to say that the water vapor greenhouse effect complements that of CO_2 . CO_2 absorbs



Figure 4.19: As in Figure 4.17, but for H_2O . The continuum regions marked on the figure are regions where the measured absorption substantially exceeds that computed from the spectral lines.

strongly near the peak of the Planck function for Earthlike temperatures, but the water vapor absorption is nearly two orders of magnitude weaker there. Further, for Earthlike planets, water vapor condenses and therefore disappears in colder regions of the planet; it is only the long-lived CO_2 greenhouse effect that can persist in cold parts of the atmosphere.

On a planet without a substantial condensed water reservoir, water vapor could be a wellmixed noncondensing greenhouse gas much like CO_2 on modern or early Earth. In most known cases of interest, though, atmospheric water vapor is in equilibrium with a reservoir –an ocean or glacier – which fills the atmosphere to the point that the atmospheric water vapor content is limited by the saturation vapor pressure. The prime cases of interest are water vapor feedback on Earth and on Early Mars, and the runaway greenhouse on Early Venus. The runaway greenhouse is also relevant to the ultimate far-future fate of the Earth, and the evolution of hypothetical water-rich extrasolar planets. The status of water vapor as a greenhouse gas whose concentration is limited, via condensation, by temperature does not derive from any special properties of water. One tends to focus on a role of this sort for water simply because planets that are "habitable" to the only example of life with which we are presently familiar seem to require a planet with liquid water and an operating temperature range in which the saturation vapor pressure is high enough for water vapor to be present in sufficient quantities to be active as a greenhouse gas. On present and past Mars, as well on a hypothetical early Snowball Earth, CO_2 can be limited by condensation, and on Titan today CH_4 condenses, while NH_3 and other gases have condensation layers on the gas giant planets.

First, let's think about the effect of water on OLR, supposing that the atmosphere is saturated at each altitude. The water vapor greenhouse effect is determined by a competition between two factors. Water vapor causes the greatest optical thickness near the ground, where both pressure-broadening and saturation vapor pressure are highest. However, a strong greenhouse effect requires optical thickness at higher altitudes, where the temperature is substantially colder than the surface, in order to reduce the radiating temperature of the planet. For this reason, water vapor in the mid to upper troposphere is more important than water vapor near the ground. Consider a typical Earth tropical case, on the moist adiabat with 300K temperature near the ground. If the low level air is saturated, then the equivalent path of the lowest 100mb of the atmosphere is about $400kg/m^2$. This is sufficient to make the lower atmosphere optically thick except within the window regions, so that the atmosphere would radiate to the ground at the near-surface air temperature except within the windows. Most of the infrared cooling of the ground occurs in the windows, but we'll see eventually that at temperatures of 300K and above, the continuum substantially closes off the window region as well. This near-surface opacity doesn't much affect the OLR, however. Near 400mb, the temperature is about 260K, and in saturation the equivalent path between 400mb and 500mb is only $25kg/m^2$. This relatively small amount of water vapor is still sufficient to make the layer optically thick between $1350cm^{-1}$ and $1900cm^{-1}$, as well as on the low frequency side of $450 cm^{-1}$. This substantially reduces the OLR. At the tropopause level, in contrast, the temperature is about 200K, the pressure is 100mb, and the equivalent path from 100mb to 200mb is only $.01kq/m^2$. At such low concentrations, water vapor is optically thin practically throughout the spectrum. At lower surface temperatures, the dominant greenhouse effect of water vapor comes from correspondingly lower altitudes. With a 273K surface temperature, the 400mb temperature on the dry air adiabat is only 210K, and the water vapor opacity is inconsequential there. One has to go down to about 600 - 700mb, where the path is about $2kq/m^2$, to get a significant water vapor greenhouse effect. On a low CO_2 Snowball Earth soon after freeze-up, where the tropical surface temperature is under 250K, the water vapor greenhouse effect is essentially negligible. The water vapor greenhouse effect only starts to play a role when the planet has warmed to the point that the tropical temperatures approach the melting point.

As was the case for CO_2 , the absorption coefficient for water vapor decays roughly exponentially with distance in wavenumber space from each peak of absorption. This has similar consequences for OLR as were discussed previously in connection with CO_2 . Because of the exponential envelope of absorption, doubling or halving the water vapor content of a layer of the atmosphere has approximately the same effect on the optical thickness of that layer regardless of whether the base amount being doubled or halved is very large (say 200 kg/m^2) or very small (say 2 kg/m^2). There may not be much water to work with in the Earth's mid troposphere, but nonetheless halving or doubling the amount would have a significant effect on OLR. This remark is particularly significant because there are dynamical effects which in fact keep the Earth's midtroposphere substantially undersaturated (see Chapter 7). Although there are regions with relative humidity as low as 10%, it would still significantly increase the OLR if the relative humidity were reduced still more to 5%, and conversely it would significantly decrease the OLR if the relative humidity were increased to 20%. Because of the spectral position of the absorption peaks relative to the shape of the Planck curve, the effect of water vapor concentration on OLR is not as precisely logarithmic as is the case for CO_2 . Nonetheless, it is fair to say that the change in water vapor content relative to the amount initially present gives a more true idea of the radiative impact of the change than would the change in the absolute number of kilograms of water present in a layer.

Figure 4.20 shows the temperature dependence and the ratio of self to air-broadened absorption for water vapor. The general range of temperature sensitivity is much the same as it was for CO_2 . However, whereas self-broadened absorption for CO_2 is only a few tens of percents stronger than air-broadened absorption, the self-broadened H_2O absorption is fully five to seven times stronger than the air-broadened case. This is extremely important to the runaway greenhouse, which involves portions of the atmosphere which consist largely of water vapor. Moreover, when a species has a molar concentration in air of, say, 10% or less one wouldn't ordinarily have to worry much about self-broadening, since collisions with air are so much more common than self-collisions. However, because of the great amplification of self-broadened absorption for water vapor, the self-broadening in fact starts to become dominant even at molar concentrations of around 10%. At the Earth's surface, this concentration is achieved at a temperature of 320K. With a dry air partial pressure of 100mb, this concentration would be achieved at temperatures near 280K; this situation is relevant to a hot planet on the verge of a runaway greenhouse. The strong enhancement of self-broadened absorption relative to air-broadened absorption is far in excess of



Figure 4.20: Left panel: Temperature dependence coefficient for air-broadened H_2O at 100mb, as defined in Fig. 4.18. Right panel: Ratio of self-broadened to air-broadened median absorption for H_2O at 100mb and 260K. As usual, medians are computed in bands of width $50cm^{-1}$.

what would be anticipated from simple effects associated with the different molecular weights of the colliders. The sensitivity to the nature of the colliding molecule raises interesting questions about the effect of collisions with other molecules. Would CO_2 broadened coefficients be more like the air-broadened or self-broadened case? The answer to this question has some impact on the climates of Early Mars and Early Earth, which are often assumed to have had substantial amounts of CO_2 in their atmospheres. Unfortunately, laboratory measurements bearing on the subject are hard to come by.

Next we turn our attention to aspects of the absorption which govern the runaway greenhouse effect for a planet with a water-saturated atmosphere. Specifically, we revisit the question of the Kombayashi-Ingersoll limit, which is the limiting OLR such a planet can have in the limit of large surface temperature. As discussed for the grey gas case in Section 4.3.3, the limiting OLRis approximately determined by the temperature at the pressure level where the optical thickness between that level and the top of the atmosphere becomes unity. For a grey gas this characteristic pressure was independent of wavenumber, whereas for a real gas it is quite strongly wavenumber dependent. To keep things simple, we'll consider a saturated pure water vapor atmosphere. In this case, if the temperature at a given altitude is T, the corresponding pressure is $p_{sat}(T)$, determined by Clausius-Clapeyron. The equivalent path from this altitude to zero pressure is $a_{self} \frac{1}{2} p_{sat}(T)^2 / (p_o g \cos \bar{\theta})$, where p_o is the standard reference pressure (100mb for use with Figure (4.19) and a_{self} is the ratio of self-broadened to air-broadened absorption (about 6). By using this path together with the absorption coefficients in Figure 4.19, we can estimate the maximum effective radiating temperature as a function of wavenumber. Based on the median absorption in each band, the radiating temperature varies from about 245K at $100cm^{-1}$ to 278K at $500cm^{-1}$ to 350K at the valley of the window region near $1000cm^{-1}$. The high values of radiating temperature in the window region lead to large estimates of the Kombayashi-Ingersoll limit. As a crude estimate, if we assume that the planet radiates at 350K in the window regions between 544 and 1314 cm^{-1} , and on the high wavenumber side of $1950cm^{-1}$, then the OLR would be about $520W/m^2$ even if the planet didn't radiate at all in the rest of the spectrum. The Kombayashi-Ingersoll limit is strongly affected by the absorption in the window regions, and we will see in Section 4.4.8 that the absorption here is dominated by a continuum which is not captured by the nearby line contribution. The estimate of the Kombayashi-Ingersoll limit for H_2O (and CO_2) will be completed in Section 4.7, after we have discussed the continuum.

202CHAPTER 4. RADIATIVE TRANSFER IN TEMPERATURE-STRATIFIED ATMOSPHERES

Methane

Figure 4.21 shows the standard-state absorption spectrum for methane. From the standpoint of OLR on Modern and Early Earth, and perhaps on Early Mars, the most important absorption feature is that near 1300 cm^{-1} , which occurs in a part of the spectrum where water vapor and CO_2 absorption are weak, but where the Planck function still has significant amplitude at Earthlike temperatures. In contrast with water vapor, the very long wave absorption (below 1000 cm^{-1}) is so weak that it does not significantly affect OLR in any atmosphere likely to have existed on an Earthlike or Marslike planet. Titan has extremely large amounts of methane in its atmosphere, which could in principle make the very longwave group important. Even there, however, the weakening of the absorption due to the very cold temperatures makes this absorption group fairly insignificant. For atmospheres containing appreciable amounts of oxygen, methane oxidizes rather quickly to CO_2 , so it is hard to build up very large concentrations. The Earth's pre-industrial climate had about 1ppmv of methane in it, and the intensive agriculture of the past century may eventually come close to doubling this concentration. The associated equivalent paths are quite small – on the order of $.06kg/m^2$. For paths this small, Earth's atmosphere has an optical thickness of only .14 based on the median absorption coefficient occurring the $1300 cm^{-1}$ peak. For methane paths typical of oxygenated atmospheres, one gets significant absorption only from the upper quartile of absorption coefficients, and a short distance from the dominant peak one only gets significant absorption very near the line centers. In this case, the ditch in OLR dug by methane is a very narrow feature centered on $1300 cm^{-1}$. In an anoxic atmosphere, as Earth's is likely to have been earlier than about 2.7 billion years ago, the rate of methane destruction is much lower, and it is believed that production of methane by methanogenic (methane-producing) bacteria could have driven methane concentrations to quite large values. With 100ppmv of methane in an Earthlike atmosphere, the equivalent path is about $.6kg/m^2$, and based on the median absorption coefficient the atmosphere becomes optically thick from about $1200 - 1400 cm^{-1}$. If the methane concentration builds up to 1% of the atmosphere, then the equivalent path is nearly $600 kg/m^2$, and the atmosphere becomes optically thick from $1150 - 1750 cm^{-1}$; for such high concentrations, the shoulder to the right of the absorption peak starts to become important. There are also some speculations that abiotic processes could have led to high Methane concentrations on Early Mars, or on the prebiotic Earth.

Beyond what is shown in Figure 4.21, Methane has strong absorption bands that extend well into the Solar near-IR. These are not terribly important at concentrations up to a few hundred ppmv in a 1bar atmosphere, but when Methane makes up a percent or so of the atmosphere, it can absorb most of the incident solar energy between 2500 and 9000 cm^{-1} . At higher concentrations, significant absorption can extend even into the visible range.

It is often said that, molecule for molecule, CH_4 is a better greenhouse gas than CO_2 . However, this is more a reflection of the relative abundances of CH_4 and CO_2 in the present Earth atmosphere than it is a statement about any intrinsic property of the gases; in fact, the absorption coefficients for the two gases are quite similar in magnitude, and CH_4 absorbs in a part of the spectrum that is less well placed to intercept outgoing terrestrial radiation than is the case for CO_2 . The high effectiveness of CH_4 relative to CO_2 in the present atmosphere of Earth stems from the fact that currently there is rather a lot of CO_2 in the air (380ppmv and rising) but rather little CH_4 (1.7ppmv and also rising). In a situation like this, one has already depleted infrared of those frequencies that are most strongly absorbed by CO_2 , so when adding CO_2 one is adding "new absorption" in spectral regions where the absorption is relatively weak. Hence, it takes a large amount of the gas to have much radiative effect. In contrast, when starting with a small amount of CH_4 , when one adds more, one adds "new absorption" where the absorption



Figure 4.21: As in Figure 4.17, but for CH_4 .



Figure 4.22: Left panel: Temperature dependence coefficient for air-broadened CH_4 at 100mb, as defined in Fig. 4.18. Right panel: Ratio of self-broadened to air-broadened median absorption for CH_4 at 100mb and 260K. As usual, medians are computed in bands of width $50cm^{-1}$.

coefficient is quite strong, since the strongly absorbing part of the spectrum is not yet depleted. This behavior depends crucially on the lack of significant overlap between the Methane and CO_2 absorption regions.

The temperature scaling coefficient for CH_4 is shown in Figure 4.22. It is in the same general range as for the cases discussed previously. The ratio of self to air broadening for CH_4 is similar to that for CO_2 , throughout the spectral range of most importance for OLR. In the solar near-IR the self-broadened absorption can be nearly twice the air-broadened absorption. For Methane, the self-broadening is mostly of academic interest, since the methane concentration is too low for self-collisions to be significant in atmospheres encountered or envisioned so far. Titan and similar cryogenic atmospheres are potentially an exception to this remark, but there the absorption is dominated by a continuum that is not clearly related to the absorption lines under consideration here.

Because Titan has a surface temperature on the order of 100K, the peak of the Planck function occurs at about the third of the wavenumber where the peak is for Earthlike temperatures. In consequence, there is little thermal emission in the vicinity of the dominant $1300cm^{-1}$ absorption group. It is only the longwave absorption group that is potentially of interest. The lower atmosphere of Titan contains up to $20\% CH_4$, which with Titan's low surface gravity yields an actual (*not* equivalent) mass path of nearly $15000kg/m^2$. However, due to the strong reduction of line strength with temperature, the median absorption is well under $10^{-6}m^2/kg$ in the longwave group, even when broadened by a pressure of 1.5bar. The radiative effect of Methane on Titan arises mainly from a collisional continuum of the sort described in Section 4.4.8.

4.4.8 Collisional continuum absorption

Diatomic molecules and general considerations

A Nitrogen molecule N_2 in isolation does not interact to any significant extent with infrared light; one might think that collisions do not change this picture, as N_2 has no lines to be broadened by collisions. Nonetheless, during the time a collision is taking place the pair of colliding molecules momentarily behaves somewhat like a more complex four-atom molecule, which has transitions that can indeed absorb and emit infrared radiation. This leads to collision-induced absorption, whose associated absorption coefficient is generally a smooth function of wavenumber. Because of the lack of line structure, such absorption is referred to as a *continuum*. There are many possible processes through which collisions can induce absorption. The collision can impart a temporary dipole moment to a rotation or vibration that ordinarily had none, allowing it to absorb or emit a photon. The collision can break a symmetry, allowing transitions that are otherwise "forbidden" by symmetry principles. Colliding molecules can form *dimers*, which are short-lived complexes which nevertheless persist long enough to have radiatively active transitions not present in the colliding molecules. Most of the "non-absorbing" diatomic molecules, including N_2 and H_2 , exhibit significant collision-induced continuum absorption at sufficiently high densities. These are mostly associated with the induced-dipole mechanism, and therefore can to some extent be anticipated on the basis of the underlying transitions of the diatomic molecule.

Collision-induced absorption can be thought of as a trinary chemical reaction involving the two colliding molecules and a photon. The rate of "reaction" (i.e. absorption) is proportional to the product of the concentrations of the two colliding species with the photon concentration, the latter being proportional to the radiation flux. The absorption coefficient is the rate constant for the reaction. Unlike the case of collision-broadened line absorption, in collision-induced absorption there is no physical distinction between the "absorbing" molecule and the "perturbing" molecule. Both are equal partners in the process allowing absorption or emission of a photon. For this reason, it is most natural to describe collision-induced absorption in terms of a binary absorption coefficient, which expresses the proportionality between the product of the concentrations of the two colliding species and the rate of absorption of radiation. Nonetheless, in order to facilitate comparison with the previously defined line absorption coefficients, and in order to make it easier to incorporate collisional continuum absorption in radiation calculations which also take into account line absorption, it is convenient to characterize the collision-induced absorption by massspecific absorption coefficients in which one of the colliding molecules is arbitrarily designated the "absorber," whose absorption is enhanced in proportion to the partial pressure of the "collider". For example, for an N_2 - H_2 collision in a box of gas with uniform temperature T and uniform N_2 partial pressure p, the optical thickness can be expressed as

$$\tau = \frac{p_{N2}}{p_o} \kappa_{H2}(\nu, p_o, T) \ell_{H2}$$
(4.87)

where ℓ_{H2} is the mass path of Hydrogen in the box, in kg/m^2 , and p_o is a standard pressure. The coefficient κ_{H2} has dimension m^2/kg and can be used in precisely the same way as the absorption



Figure 4.23: The $N_2 - N_2$ and $N_2 - H_2$ collision-induced continuum absorption coefficients as a function of temperature (indicated on curves). The coefficients are given at an N_2 partial pressure of 100mb.

coefficients we defined earlier for use with the line spectrum. To relate this to the binary absorption coefficient commonly defined in the spectroscopy literature, let z be the length of the path in meters, so that $\ell_{H2} = z \cdot p_{H2}/(R_{H2}T)$. Then

$$\tau = \frac{\kappa_{\nu, p_o, H2}}{p_o R_{H2} T} p_{N2} p_{H2} z \equiv \kappa_{H2-N2}(\nu, T) p_{N2} p_{H2} z \tag{4.88}$$

which defines the binary coefficient in the case where the collider amounts are specified as partial pressures. It is more common to specify the collider amounts in terms of densities or molar densities, but the alternate forms can be readily derived from the preceding by use of the ideal gas law.

Continuum absorption may be difficult to understand from *a priori* physical principles, and difficult to measure accurately in the laboratory, but by definition the absorption coefficient for the continuum is a smoothly varying function of wavenumber. Therefore, it is relatively easy to incorporate into radiative transfer models. One only needs to determine the absorption coefficients and their pressure and temperature scaling in a set of relatively broad bands, and multiply the transmission computed from the line absorption (if any) by the corresponding exponential decay factor.

The continuum arising from diatomic molecule collisions becomes particularly important for dense, cold, massive atmospheres, of which Titan's is probably the best studied example. Figure 4.23 shows the $N_2 - N_2$ and $N_2 - H_2$ collision induced absorption coefficients in the temperature range prevailing in Titan's atmosphere. These coefficients are based on laboratory measurements made at somewhat higher temperatures, extrapolated to colder values using a theoretical model with a few empirical coefficients fit to the data. (See the paper by Courtin *et al.* listed in the Further Readings section of this chapter). The equivalent path for Titan based on N_2 partial pressure is about $10^6 kg/m^2$, which yields a peak optical thickness of 40 for temperatures near those prevailing at Titan's surface. The H_2 content of Titan's atmosphere is less well constrained, but plausible estimates suggest that this gas, too, can contribute significantly to the infrared opacity of Titan's atmosphere. Note that the absorption decreases sharply with increasing temperature; this is partly due to the decrease in density with temperature, but is also affected by the shorter duration of high-velocity collisions, which apparently are less effective at inducing a dipole moment. The $N_2 - N_2$ continuum is unimportant for Earthlike collisions because of the higher temperatures on Earth, and because Earth's atmosphere is much less massive than Titan's, per unit surface area.



Figure 4.24: The $N_2 - CH_4$ and $CH_4 - CH_4$ collision-induced continuum absorption coefficients at 100K, assuming a collider partial pressure of 100mb. For $N_2 - N_2$ the "collider" is N_2 , while for $CH_4 - CH_4$ the "collider" is CH_4 .

Methane Continuum on Titan

Methane also significantly affects the infrared opacity of Titan's atmosphere, though its effects on OLR are rather less than N_2 or H_2 since methane is concentrated in the warmer, lower layers of the atmosphere. Essentially all of the infrared absorption due to methane on Titan comes from a collision-induced continuum. The N_2 -induced and self-induced absorption coefficients at 100K are shown in Figure 4.24. Like the other continuum coefficients, these too become weaker with increasing temperature. Assuming Methane to be in saturation with temperature given by the Methane-Nitrogen moist adiabat in Titan conditions, the pressure-weighted path for self-induced absorption is over 170000kg/m². The self-absorption yields a peak optical thickness of about 5, while the foreign absorption gives a peak optical thickness of about 20, dropping to 10 in the higher frequency shoulder near 200cm⁻¹

Carbon Dioxide Continuum

Given the importance of the CO_2 window regions to the high- CO_2 climates of Venus and Early Mars, it is rather surprising that the CO_2 continuum has been so little studied. The coefficients in use in most models stem from limited experiments and there is little agreement on the theoretical basis for this continuum or its temperature scaling. At the time of writing, it appears that the subject has not been re-examined in laboratory experiments since the late 1970's. The discussion below is based on absorption coefficients reported in the literature cited in the Further Readings section of this chapter.

The measured CO_2 continuum absorption, rescaled to 100mb is shown in Figure 4.25. The values shown are for collisions of CO_2 in air; the self-induced continuum absorption is generally assumed to be 1.3 times that of the foreign-induced continuum. Referring to the equivalent paths in Table 4.1, we see that the continuum absorption is large enough to make the top one bar of the atmosphere of Venus optically thick throughout most of the window region. The continuum absorption is strong enough to be important in the thick atmosphere of Early Mars, but only marginally so for the more moderate CO_2 levels present on Early Earth.

When incorporating the continuum in radiation models, it may be more convenient to work



Figure 4.25: The air-induced CO_2 continuum absorption (solid lines) compared with the bandwiseminimum absorption computed from the line spectrum (dashed lines).

from a curve-fit rather than tabulated data. The absorption coefficient for the CO_2 continuum can be fit with the function

 $\begin{aligned} \kappa_{CO2}(\nu, 300K, 100mb) &= \exp(-8.853 + 0.028534\nu - 0.00043194\nu^2 + 1.4349 \cdot 10^{-6}\nu^3 - 1.5539 \cdot 10^{-9}\nu^4) \\ \text{(4.89)} \end{aligned}$ from 25 to 450 cm⁻¹ and by

 $\kappa_{CO2}(\nu, 300K, 100mb) = \exp(-537.09 + 1.0886\nu - 0.0007566\nu^2 + 1.8863 \cdot 10^{-7}\nu^3 - 8.2635 \cdot 10^{-12}\nu^4)$ (4.90)

from 1150 to 1800 cm^{-1} where ν is measured in cm^{-1} . The continuum absorption coefficients weaken with increasing temperature, according to the empirical power law $(300./T)^n$, with n = 1.7 for wavenumbers greater than $190cm^{-1}$, increasing to 1.9 at 130 cm^{-1} , 2.2 at 70 cm^{-1} and 3.4 at 20 cm^{-1} .

Water vapor continuum

Since water vapor has absorption lines throughout the spectrum, it is hard to unambiguously define the continuum. Laboratory measurements clearly show, however, that in the window regions indicated in Figure 4.19 the net absorption is far in excess of what can be accounted for by the contributions of nearby lines. The prevailing view currently is that this excess absorption is due to the very far tails of the stronger absorption bands flanking the window regions, rather than dimers, forbidden transitions, or collision-induced dipole moments. The theoretical and observational basis for this viewpoint is exceedingly weak, however. In the following we will confine ourselves to empirical descriptions of the laboratory measurements, without reference to underlying mechanisms. Comparisons with direct measurements of transmission in the Earth's atmosphere have confirmed that the laboratory measurements provide an adequate basis for modelling water vapor absorption in the window regions. The laboratory measurements show that the air-broadened or N_2 broadened water vapor continuum is very weak, so that the window region absorption is by far dominated by self-collisions of water vapor. The following discussion will therefore be limited to self-induced absorption; the characterization of foreign-induced absorption by CO_2 appears to be an open question at present, though it is potentially of importance to water- CO_2 atmospheres



Figure 4.26: The water vapor self-induced continuum near $1000 \ cm^{-1}$, compared with the median absorption coefficient computed from the Lorentz line contribution within 1000 line widths of the line centers in the HITRAN database. The continuum curve given is based on laboratory observations in saturation at 296K, scaled up to what they would be at a water vapor partial pressure of 100mb. See citations in the Further Readings section of this chapter for data sources.

such as might have occurred on Early Mars.

There is some ambiguity in the spectroscopic literature as to how to define the water vapor continuum, given that in analyzing measurements one must be careful not to be thrown off by the strong absorption near the centers of individual lines in the window regions. Most useful definitions of the continuum amount to reading the absorption at the minima "between the lines." The results of such a measurement of the self-induced continuum are shown in Figure 4.26. The measurements were made for water vapor in saturation at 296K, with water vapor partial pressure of about 28mb, but have been scaled to a standard water vapor partial pressure of 100mb for the sake of discussion. We'll focus on the lower frequency of the two window regions, since that is by far the most important for planetary climate calculations. Similar data exists for the higher frequency window. From the figure, we see that the measured continuum absorption is several orders of magnitude stronger than the typical line contribution. To get an idea of the significance of the water vapor continuum, let's consider a layer of air of depth z, with uniform temperature T, within which the water vapor is at the saturation vapor pressure corresponding to T. Since the water vapor continuum is dominated by self-collisions, it matters little what the background air pressure is in this layer. The equivalent path for this layer is $(p_{sat}(T)/p_o)(p_{sat}(T)/(R_wT))z$; the first factor gives the degree of pressure-induced enhancement of absorption relative to the standard. while the second factor is the density of water vapor in the layer. Note that the equivalent path is *quadratic* in the water vapor partial pressure. For this reason, the optical thickness in the continuum region grows very rapidly with temperature. At 300K, then, with a layer depth of 1kmthe equivalent path is $9.3kg/m^2$. Since the minimum absorption coefficient in the window regions about $.1m^2/kg$, this path gives the layer an optical thickness of unity or more in the window region. Since the absorption is even stronger outside the window region, at the lowest layer of the Earth's atmosphere acts practically like an ideal blackbody at tropical temperatures. At 310Kthe equivalent path increases to $27kg/m^2$, so the window region closes off even more. This has profound consequences for the runaway greenhouse. In fact, there would be essentially no prospect for a runaway greenhouse even in Venusian conditions were it not for the water vapor continuum.

Over the range of wavenumbers shown in the figure, the water vapor continuum absorption

can be fit by the polynomial

$$\kappa_{H2O}(\nu, 296K, 100mb) = \exp(12.167 - 0.050898\nu + 8.3207 \cdot 10^{-5}\nu^2 - 7.0748 \cdot 10 - 8\nu^3 + 2.3261 \cdot 10^{-11}\nu^4)$$

$$(4.91)$$

Like the other continua, the water vapor continuum absorption becomes weaker as temperature increases. Data on the temperature dependence is sparse, but suggests a temperature dependence of the form $(296/T)^{4.25}$. For temperatures much colder than 300K, the saturation vapor pressure is so low that the details of the temperature dependence are unimportant. As temperature increases beyond 300K, the exponential growth of saturation vapor pressure is far more important to the optical thickness than the rather mild decline of the continuum absorption coefficient with temperature.

Water vapor has another continuum region at shorter wavelengths, in the vicinity of $2500cm^{-1}$. This is not important for Earthlike temperatures, but it is a very significant factor for the hotter temperatures encountered in runaway greenhouse calculations. At temperatures much in excess of 320K, there is enough emission in this region that it accounts for a significant part of the infrared cooling if the continuum is not included. The $2500cm^{-1}$ continuum, covering $2100cm^{-1} \le \nu \le 3000cm^{-1}$ can be represented by the polynomial fit

 $\kappa_{H2O}(\nu, 296K, 100mb) = \exp(-6.0055 - 0.0021363x + 6.4723 \cdot 10^{-7}x^2 - 1.493 \cdot 10 - 8x^3 + 2.5621 \cdot 10^{-11}x^4 + 7.328 \cdot 10^{-14}x^5)$ (4.92)

where $x = \nu - 2500 cm^{-1}$. The scaling in pressure and temperature can be taken to be the same as for the longer wavelength continuum, though the experimental support for the temperature dependence is somewhat weak.

The importance of the water vapor continuum to climate when temperatures approach or exceed 300K is demonstrated in Figure 4.27. Here we present calculations of the spectrum of OLR and of surface back-radiation using the exponential sum radiation code described previously, but modified to take into account the vertical variation of water vapor concentration and the continuum. With the continuum included, the low layer atmosphere radiates to the surface practically like a blackbody; in fact, if one increases the surface air temperature slightly, to 310K, the back radiation becomes indistinguishable from the blackbody spectrum corresponding to the surface air temperature. In contrast, without the continuum, there is essentially no back-radiation in the window region, allowing the surface to cool strongly through the window. Likewise, without the continuum, the atmosphere can radiate to space very strongly through the window, whereas the cooling to space is very much reduced if the continuum absorption is included. These calculations were carried out for an Earthlike water-air atmosphere on a planet with $g = 9.8m/s^2$. On a planet with weaker surface gravity, the water vapor continuum would become important at lower temperatures, because a given partial pressure corresponds to a greater mass of water. Conversely, on a planet with stronger surface gravity, the water vapor window closes at higher temperatures.

4.4.9 Condensed substances: Clouds

Clouds are made of particles of a condensed substance, which may be in a liquid or solid (e.g. ice) phase. The molecules of a condensed substance are in close proximity to one another, and at typical atmospheric temperatures the collisions are so frequent that no line structure survives in the spectrum. In consequence, the absorption coefficient for a condensed substance is generally a very smoothly varying function of wavenumber. Absorption by condensed substances behaves rather like the gaseous continuum absorption we discussed in the preceding section.

Water clouds are of particular interest, since they are by far the dominant type of cloud on


Figure 4.27: The spectra of surface back radiation and OLR computed using an exponential sum radiation code including the effects of the longwave water vapor continuum (left panel) and excluding the continuum (right panel). The calculation was done with the temperature profile on the water-air moist adiabat corresponding to 300K surface temperature, assuming the water vapor partial pressure to be saturated at all levels. This calculation does not take into account the temperature variation of absorption coefficients or the enhancement of self-induced absorption in the line contribution.

Earth. They would also occur on any world habitable for life as we know it, since such a world would have a repository of liquid water somewhere, and condensation of water vapor somewhere in the atmosphere would then be practically inevitable. Water clouds would also form in the course of a runaway greenhouse on a world with a water ocean, such as the primordial Venus. The absorption coefficient for liquid water is shown over the infrared range in Figure 4.28. For comparison, we show the median absorption coefficient for water vapor at 100mb pressure and 260K temperature. Keep in mind that the absorption for liquid water is a true continuum, so that, unlike the median absorption curve shown for the vapor phase, the curve for liquid water displays the full wavenumber variability of the absorption.

We see that a kilogram of water in the liquid phase is a far better absorber than the same kilogram in the form of vapor. Near the peak absorption wavenumbers of water vapor, the difference can be as little as a factor of ten, but in the window regions liquid water has an absorption coefficient many thousands of times that of water vapor. The absorption coefficient for liquid water varies little enough that over the infrared range it can be quite well approximated as a grey gas (or more properly, a grey liquid). In fact, with a typical absorption coefficient of $100m^2/kg$, it takes a layer of liquid only 10^{-5} meters to have unit optical thickness and to begin to behave like a grey body. This is the depth of penetration of atmospheric infrared back-radiation into the surface of a lake or ocean, and it is the depth whose temperature directly determines the infrared radiation from the surface of a lake or ocean. Water ice is somewhat more transparent to infrared than liquid water, but the general features of the behavior still apply.

Because of the nonlinearity of the exponential function which determines the amount of absorption suffered by infrared as it passes through a number of particles, it matters greatly how the mass of water is distributed amongst particles of various sizes. For example, at $400 \ cm^{-1}$ liquid water has an absorption coefficient of 171 m^2/kg ; since water has a density of $1000 kg/m^3$ that means that a layer of liquid water of depth 5.8 *micrometers* has optical thickness of unity, and attenuates incident infrared by a factor of 1/e. That means, roughly speaking, that a spherical droplet of radius r will remove essentially all the infrared hitting it $-\pi r^2$ times the incident flux –



Figure 4.28: The absorption coefficient for liquid water. The median absorption coefficient for water vapor at (100mb, 260K) is reproduced from Fig 4.19 for the purposes of comparison.

as long as r is 5 micrometers or more. A mere 10 grams of water is sufficient to make $1.9 \cdot 10^{10}$ particles of radius 5micrometers, which would have a total cross section area of $1.5 m^2$. Distributed randomly within a column of air having base $1 m^2$ these particles would be sufficient to remove essentially all of the flux of infrared entering the base of the column. In other words, a mass path of liquid water as little as 10 grams per square meter is sufficient to make an optically thick cloud, provided the water takes the form of sufficiently small droplets. However, if we take the same mass of water and gather it up into a single drop of radius 1.3 cm – it would only intercept an insignificant .0005 m^2 of the incident light.

This estimate of the absorption by cloud droplets is not quite correct, because it fails to take into account the extent to which electromagnetic radiation pentrates the droplet as opposed to being diffracted around it. The calculation will be done more precisely in Chapter 5, but the simple estimate gives the right answer to within a factor of 2 or better.

The net result is that, for the typical droplet size found in Earth's water clouds, a cloud layer containing anything more than about 10 grams per square meter of condensed water acts essentially like a blackbody in the infrared. Water is not at all typical in this regard. Other condensed cloud-forming substances, including liquid methane and CO_2 ice, are far more transparent in the infrared, and have a qualitatively different effect on planetary energy balance. The effect of such clouds will be taken up in Chapter 5.

4.5 Real gas *OLR* for all-troposphere atmospheres

Calculation of OLR is one of the most fundamental steps in determining a planet's climate. Now that we are equipped with an ability to compute the OLR for real gases, we can revisit some of our old favorite problems – Snowball Earth, the Faint Young Sun, Early Mars, and so forth – but this time relate the results to the actual atmospheric composition. In this section we present results for the all-troposphere model introduced in Section 4.3.2, occasionally limiting the upper air temperature drop by patching the adiabat to an isothermal stratosphere.

The homebrew exponential sum radiation model described in the preceding sections has the advantage of simplicity, generality and understandability. We will use it wherever it is sufficiently accurate to capture the main phenomena under discussion. However, professionally written terrestrial radiation codes are the product of a great deal of attention to detail, particularly with regard to temperature scaling and the simultaneous effects of multiple greenhouse gases. They can give highly accurate results provided one does not stray too far from the Earthlike conditions for which they have been optimized. In the following, and at various places in future chapters, we will have recourse to one of these standard radiation models, produced by the National Center for Atmospheric Research as part of the Community Climate Model effort. We'll refer to this model as the ccm radiation model. Although it uses a good many special tricks to achieve accuracy at high speed, and a lot of detailed bookkeeping to deal with the properties of a half dozen different greenhouse gases of interest on Earth, what is going on inside this rather massive piece of code is not fundamentally different from the Malkmus type band models and the exponential sum model described previously.

A detailed discussion of surface back-radiation for real gases will be deferred to Chapter 6. Some aspects of the infrared cooling profile for real gas atmospheres will be touched on in Section 4.8. The effect of clouds on OLR and on shortwave albedo will be discussed in Chapter 5.

4.5.1 CO_2 and dry air

First we'll compute the OLR for a mixture of CO_2 in dry air, with the temperature on the dry air adiabat $T(p) = T_g \cdot (p/p_s)^{2/7}$ and ground temperature equal to surface air temperature. This is a real-gas version of the calculation leading to Eq. 4.33; it amounts to a canonical OLR computation which serves as a simple basis for intercomparison of different radiation models. Performed for other greenhouse gases, it also can provide a basis for comparing the radiative effects of the gases. The results presented here are carried out with Earth gravity and 1 bar of air partial pressure, but can easily be scaled to other conditions. The CO_2 path is inversely proportional to gravity, so 100ppmvof CO_2 on Earth is equivalent to 1000ppmv on a planet with ten times the Earth's surface gravity or 10ppmv on a planet with a tenth of the Earth's surface gravity. For fixed CO_2 concentration, surface pressure has a quadratic effect on the path, since the mass of CO_2 in the atmosphere (given fixed concentration) increases in proportion to pressure, but one gets an additional pressure factor in the equivalent path from pressure broadening. Thus, 100ppmv of CO_2 in 1bar of air is equivalent to 1ppmv of CO_2 in 10bars of air.

In Figure 4.29 we show how OLR varies with CO2 concentration for a fixed surface temperature of 273K. This curve gives the amount of absorbed solar radiation needed to maintain the surface temperature at freezing. The CO_2 amount is expressed as molar concentration in *ppmv*, but for the range of concentrations considered, the difference between concentration and mixing ratio is not very significant. Results are shown for both the ccm model and the simplest form of the homebrew exponential sums radiation code. The homebrew calculation employed 10-term sums with coefficients computed at 260K. The path was pressure-weighted to reflect collisional broadening, but temperature weighting was neglected. Over the range of CO_2 covered, only the principal absorption region centered on $650cm^{-1}$ needs to be taken into account. The homebrew calculation deviates by up to $10 W/m^2$ from the more comprehensive ccm calculation, but this is quite good agreement in view of the fact that the homebrew code takes up barely a page and gener-



Figure 4.29: The OLR vs CO2 concentration (measured in ppmv) for CO_2 in a dry air atmosphere with temperature profile given by the dry adiabat. The surface air temperature and the ground temperature are both 273K, and the acceleration of gravity is $9.8m/s^2$. Results are shown for a simple exponential sum radiation code without temperature weighting, and for the comprehensive ccm radiation code.

alizes easily to any greenhouse gas, in contrast to the rather Earth-specific ccm code, which involves several thousand lines of rather unpretty FORTRAN. Most of the mismatch arises from the neglect of temperature scaling in the homebrew code. The homebrew code slightly overestimates OLRfor low CO_2 where the radiating level is low in the atmosphere where warmer temperatures ought to increase the infrared opacity. It underestimates OLR because the higher, colder atmosphere is assumed more optically thick than it should be. If one complicates the homebrew code very slightly to incorporate a band-independent temperature weighting of the form $\exp -(T^*/T - T^*/T_o)$ in the path computation, then one can reduce the mismatch to under $2W/m^2$ with $T^* = 900K$.

As anticipated from the shape of the CO_2 absorption spectrum, the OLR goes down approximately in proportion to the logarithm of the CO_2 concentration. Between 10ppmv and 1000ppmveach doubling of CO_2 reduces OLR by about $4W/m^2$ based on the ccm model. At large CO_2 concentrations, the logarithmic slope becomes somewhat greater, as the weaker absorption bands begin to come into play. At 10000ppmv, a doubling reduces OLR by $6.8W/m^2$. CO_2 becomes a somewhat more effective greenhouse gas at high concentrations, but never approaches the potency of a grey gas, for which each doubling would more than halve the OLR once the optically thick limit is reached. Were it not for the relatively gentle dependence of OLR on CO_2 caused by the highly frequency-selective nature of real-gas absorption, modest fluctuations in the atmosphere's CO_2 content would lead to wild swings of temperature and almost certainly render the planet uninhabitable.

Calculations show that for fixed CO_2 concentration the OLR increases very nearly like the fourth power of temperature, just as in the grey gas result in Eq. 4.33. In effect, the radiating pressure remains nearly fixed as temperature varies. This makes it easy to do planetary temperature calculations. For example, let's compute what temperature the Earth would have if the CO_2 concentration were at the pre-industrial value of 280ppmv, but there were no other greenhouse gases in the atmosphere. The OLR for a surface temperature of 273K is $267W/m^2$. Balancing against an absorbed solar radiation of $240W/m^2$, the temperature is determined by $267 \cdot (T/273)^4 = 240$, yielding T = 263K. Without the additional greenhouse effect of water vapor, Earth would be a

very chilly place. According to Figure 4.29, for a dry Earth CO_2 would have to be increased all the way to 24000ppmv just to bring the temperature up to the freezing point.

Exercise 4.5.1 What temperature would Venus have if it had a 1bar air atmosphere mixed with 280ppmv of CO_2 ? Assume that the planetary albedo is 30%, like that of Earth. How does this temperature compare with the temperature Venus would have without any greenhouse gases in its atmosphere?

One can't get the Earth's temperature right without water vapor, but one can still make a decent estimate of the amount of CO_2 increase needed to offset the reduction of Solar forcing in the Faint Young Sun era, or due to a global Snowball Earth glaciation. A 25% reduction of Solar absorption at the Earth's orbit during the Faint Young Sun amounts to $60 W/m^2$, assuming an albedo of .3. This is equivalent to the radiative forcing caused by increasing CO_2 from 100 ppmv to $10^5 ppmv$ (about 100mb CO₂ partial pressure, or 10% of the atmosphere). From this we include that it's likely that it would take somewhat over 100mb of CO_2 to keep Earth unfrozen when the Sun was dim. Next let's take a look at what it takes to deglaciate a Snowball Earth. Let's suppose that for a Neoproterozoic solar constant, the tropics freeze over when the CO_2 is reduced to 100ppmv (this is not far off estimates based on comprehensive climate models). Icing over the Earth increases the albedo to about .6, leading to a reduction of almost $100W/m^2$ in absorbed solar radiation. Hence, restoring the Tropics to the melting point would require well in excess of 100mb of CO_2 in the atmosphere; calculations with the homebrew model at higher CO_2 concentrations than shown in Figure 4.29 indicate that fully a bar of CO_2 mixed with a bar of air would be needed. Deglaciation might not require restoring the full $100W/m^2$ of lost Solar forcing, but it is still clear that a great deal of CO_2 is needed to deglaciate a Snowball.

These estimates are crude, but they do get across one central idea: that because of the logarithmic dependence of OLR on greenhouse gas concentration, it takes a huge increase in the mass of CO_2 to make up for rather moderate changes in albedo or solar output. Aside from neglect of overlapping water vapor absorption, these estimates somewhat overstate the effect of CO_2 on OLR because they employ the dry adiabat rather than the less steep moist adiabat. We'll revisit the estimates shortly, after we bring water vapor into the picture.

4.5.2 Pure CO₂ atmospheres: Present and Early Mars, and Venus

Figure 4.30 shows the OLR as a function of surface pressure for a pure CO_2 atmosphere subject to Martian gravity. The results span the range of surface pressure from those similar to the thin atmosphere of present Mars up to the thick atmospheres commonly hypothesized for Early Mars ⁶. The calculations were carried out for a fixed surface temperature of 270K, since we are primarily interested in the question of how much CO_2 there would have to be in order to warm Early Mars up to near the freezing point and permit the widespread liquid water at the surface that is seemingly demanded by the surface geology of the ancient Martian terrain. Results are shown for two different variants of the all-troposphere model. In the first, the atmosphere is on the dry CO_2 ideal gas adiabat throughout its depth. This profile is inconsistent at high surface pressure, however, since it becomes supersaturated aloft. For this reason, we also include results in which the temperature profile is on the one-component condensing CO_2 adiabat, which is on the dry adiabat where unsaturated but pinned to the Clausius-Clapeyron result when it becomes saturated (as in Fig. 2.6). With condensation, the surface pressure cannot be increased beyond

⁶There is no strong reason to exclude the possibility of a substantial amount of N_2 in the Early Martian atmosphere. Addition of N_2 to the atmosphere would increase surface pressure and enhance CO_2 absorption.

35.4bar at a surface temperature of 270K, since the surface becomes saturated at that point and no further CO_2 can be added to the atmosphere without causing condensation. This does not pose a very significant constraint on the climate of Early Mars, however; a more important limitation is the amount of mass that could plausibly be lost from the primordial Martian atmosphere in the past four billion years. The effect of condensation aloft on OLR, in essence, increases the amount of CO_2 needed to warm Early Mars to the point where it is unclear that so much atmosphere could be lost.

At surface pressures comparable to that of Present Mars, the CO_2 greenhouse effect reduces the OLR by $35W/m^2$, and it would take $267W/m^2$ of absorbed solar radiation to maintain a surface temperature of 270K. The required solar heating is well below the $440W/m^2$ solar forcing at the subsolar point. Assuming the OLR to scale with the fourth power of temperature for fixed surface pressure, this would support a temperature of 301K at the subsolar point, in contrast with a temperature of 292K in the absence of the atmospheric greenhouse effect. The atmosphere exerts only a modest warming effect on the surface temperature of Present Mars. To be sure, most of the planet is much colder; the absorbed solar flux averaged over the surface of the planet is only $110W/m^2$, which supports a temperature of 216K with the greenhouse effect and 210K without. Recall, however, that the planetary mean budget is not very meaningful on a planet like Present Mars with no ocean and little atmosphere to average out the diurnal variations. A thin layer of the rocky surface will be quite warm within a circle centered on the subsolar point, but the nightside surface falls to temperatures well below 216K.

At higher CO_2 , the OLR decreases, approximately logarithmically in surface pressure for pressures above $10^4 Pa$. At pressures above 1bar, however, condensation becomes important, and the consequent increase in temperature aloft limits the decline of OLR. This limitation is quite important for the climate of Early Mars. Taking into account the relatively high albedo caused by molecular reflection from a thick CO_2 atmosphere, the absorbed solar radiation for Early Mars at a time when the solar flux is 30% reduced compared to today is about $70W/m^2$. This would be sufficient to sustain a 270K surface temperature with a surface pressure of 3.6bar if it weren't for the effects of condensation. When condensation is taken into account, however, fully 10 bars of CO_2 are necessary to bring the surface temperature up to 270K⁷. In fact, given the increase in albedo associated with scattering of solar radiation in a 10bar atmosphere, even 10bars is likely to prove insufficient. The effect of atmospheric scattering on albedo will be quantified in Chapter 5. In that chapter we will also discuss the potential for the scattering greenhouse effect from CO_2 ice clouds to warm Early Mars. As time progresses toward the present, and the Sun gets brighter, it becomes progressively easier to warm Mars to the point where liquid water can persist at the surface. The climate history of Mars is a race between the brightening Sun and the loss of atmosphere, which seems to have been lost by the latter.

The continuum absorption in the CO_2 window region is extremely important to these results. Without continuum absorption, the OLR for a 2bar atmosphere would be over $50W/m^2$ higher. The temperature scaling of the continuum affects the results by $10W/m^2$ or more. Whether or not one can account for prevalent liquid water on Early Mars by a gaseous CO_2 greenhouse effect hinges on a matter of $10W/m^2$ of flux or so, and therefore the importance of the continuum is disconcerting. To settle this question, one must get the CO_2 continuum right, and that is far from clear at this point, in view of the rather sparse experimental and theoretical results on the subject.

Now what about Venus? Are we finally equipped to say that we can account for the high

⁷The implications of CO_2 condensation for the gaseous greenhouse effect on Early Mars were first discussed in: Kasting JF 1991, *Icarus* **94**. The reader is also referred there for a more comprehensive treatment of the radiative transfer problem, including the effects of water vapor and a stratosphere. The conclusions are broadly similar to those based on our homebrew radiation model



Figure 4.30: The OLR vs surface pressure for a pure CO2 atmosphere. The results were done with a 10-term exponential sum code based on a 100mb reference pressure, and including the temperature scaling of both the continuum and line absorption. Calculations were done for a ground temperature of 270K, for Martian gravity. The dashed line shows the OLR for the case in which the temperature profile is on the dry ideal gas CO_2 adiabat, while the solid line incorporates the effect of condensation on the temperature profile.

surface temperature of Venus in terms of the CO_2 greenhouse effect? Unfortunately, not quite. The problem is that the high albedo of Venus means that the climate is maintained by a relative trickle of absorbed solar radiation, while the high surface temperature means that the infrared emission at wavenumbers higher than $2300 cm^{-1}$ would exceed $3800 W/m^2$ if the atmosphere were transparent in that spectral region. Unlike the Earthlike case, the atmospheric opacity there matters very much; however the HITRAN database does not include the weak absorption lines that determine the atmospheric opacity at high wavenumbers. Extensions to the database suitable for use in Venusian conditions are described in the Further Readings at the end of this chapter. We will not pursue detailed calculations with the extended database, since one must, after all, stop somewhere. Instead, we will make use of a highly simplified treatment of the short wave thermal emission which at least tells us how close we are to being able to explain the temperature of Venus. Specifically, we use exponential sums based on HITRAN for the spectral region with lower wavenumbers than $2300 cm^{-1}$, but represent the emission from higher wavenumbers by assuming that there is a radiating pressure p_{rad} such that the atmosphere radiates to space like a blackbody with temperature $T(p_{rad})$ throughout the high wavenumber spectral region. This is essentially the same approach as we took in formulating the simplest model of the greenhouse effect in Chapter 3, except that this time we apply the radiating-level concept only to the high wavenumber part of the emission. It is equivalent to stating that the absorptivity in the high wavenumber region is sufficiently large to make the layer of the atmosphere between p_{rad} and the ground optically thick throughout the high wavenumber region. Because of the shape of the Planck function, as p_{rad} is made smaller and $T(p_{rad})$ is made colder, the peak emission shifts to lower wavenumbers and in consequence the shortwave emission is sharply curtailed.

With these approximations the OLR can be written as $OLR_{<}(T_g) + OLR_{>}(T_g, p_{rad})$ where $OLR_{<}$ is computed for the low wavenumber spectral region alone using exponential sums and

$$OLR_{>} = \int_{\nu_{1}}^{\infty} \pi B(\nu, T(p_{rad}, T_g)) d\nu$$

$$(4.93)$$

where ν_1 is the frequency cutoff for the high wavenumber region, B is the Planck function and $T(p_{rad}, T_g)$ is computed using the dry CO_2 adiabat. The OLR thus computed must be balanced against the absorbed solar radiation to determine the surface temperature. At present, the solar radiation absorbed by Venus amounts to $163W/m^2$. Assuming a 93bar surface pressure, this is in balance with a surface temperature of 652K if there is no emission at all from the high wavenumber region, i.e. if $p_{rad} = 0$. This is a limiting case giving the maximum temperature that can be obtained with CO_2 alone regardless of how optically thick the high wavenumber region may be. The fact that this is still somewhat short of the observed 720K surface temperature of Venus means that a modest additional source of atmospheric opacity other than CO_2 is still needed to close the remaining gap in explaining the surface temperature. If p_{rad} is increased to 10 bars, the equilibrium temperature only drops to 633K, so it is only necessary for CO_2 to be essentially opaque in the high wavenumber region at pressures of 10 bars or greater. On the other hand, if CO_2 were really transparent at high wavenumbers, i.e. $p_{rad} = 93bar$, then the surface temperature would drop all the way to 461K, which is well below the observed value. Detailed radiation modeling of the high wavenumber region is consistent with a value $p_{rad} \approx 10bar$, so it appears that the CO_2 greenhouse effect alone gets us almost all the way to explaining the high surface temperature. The remaining opacity needed to bring the surface temperature up to 720Kis provided by Venus' high sulfuric acid clouds, the trace of water vapor in the atmosphere, and sulfur dioxide (in order of importance). The sulfur dioxide clouds are not good infrared absorbers, and exert their greenhouse effect through infrared scattering, as discussed in Section 5.

4.5.3 Water vapor feedback

For a planet like the Earth which has a substantial reservoir of condensed water at the surface (be it ocean or glacier), if left undisturbed for a sufficiently long time water vapor would enter the atmosphere until the atmosphere reached a state where the water vapor pressure was equal to the saturation vapor pressure at all points. In a case like this, if anything happens to increase the temperature of the atmosphere, then the water vapor content will eventually increase; since water vapor is a greenhouse gas, the additional water vapor will lead to an additional greenhouse gas, warming the planet further beyond the initial warming. Amplification of this sort is known as *water vapor feedback*, and works to amplify cooling influences as well. In this section we'll examine some quantitative models of real-gas water vapor feedback.

It turns out that atmospheric motions have a drying effect which keeps the atmosphere from reaching saturation. This will be discussed at greater length in Chapter 7, but suffice it to say for the moment that comprehensive simulations of the Earth's atmosphere suggest that the situation can be reasonably well represented by keeping the relative humidity fixed at some subsaturated value as the climate warms or cools. That is the approach we shall adopt here, and it still yields an atmosphere whose water vapor content increases roughly exponentially with temperature.

In these all-troposphere models, we shall also assume that the temperature profile is given by the moist adiabat. Observations of the Earth's tropics show this to be a good description of the tropospheric temperature profile even where the atmosphere is unsaturated and not undergoing convection. (See Problem ??). Evidently, the regions that are undergoing active moist convection control the lapse rate throughout the tropical troposphere; there are also theoretical reasons for believing this to be the case, but they require fluid dynamical arguments that are beyond the scope of the present volume. The situation in the midlatudes is rather less clear, but we will use the moist adiabat there as well, because it is hard to come up with something better in a model without any atmospheric circulation in it. Results are presented for Earth surface pressure and gravity, though we will make a few remarks on how the results scale to planets with greater or

lesser gravity.

Since most of the big questions of Earth climate and climate of habitable Earthlike planets involve water vapor feedback in conjunction with one or more other greenhouse gases, it is in this section that we will for the first time get fairly realistic answers regarding the Faint Young Sun problem, and so forth.

The case of a saturated pure water vapor atmosphere will be treated in Section 4.7 as part of our treatment of the runaway greenhouse for real gas atmospheres. Here will will begin our discussion with water vapor mixed with Earth air, with the temperature profile on the moist adiabat for an air-water mixture. The OLR curve for this case is shown in Figure 4.31. Since there is no other greenhouse gas, the OLR for the dry case (zero relative humidity) is just σT_q^4 . For ground temperatures below 240K there is so little water vapor in the air that the exact amount of water vapor has little effect on the OLR. At larger temperatures, the curves for different relative humidity begin to diverge. Even for relative humidity as low as 10%, the water vapor greenhouse effect is sufficient to nearly cancel the upward curvature of the dry case; at 320K the OLR is reduced by over $130W/m^2$ compared to the dry case. At larger relative humidites, the curvature reverses, and the OLR as a function of temperature shows signs of flattening at high temperature, in a fashion reminiscent of that we saw in our discussion of the Kombayashi-Ingersoll limit for grey gases. This is our first acquaintance with the essential implication of water vapor feedback: the increase of water vapor with temperature reduces the slope of the OLR vs. temperature curve, making the climate more sensitive to radiative forcing of all sort – whether it be changes in the Solar constant, changes in surface albedo, or changes in the concentration of CO_2 . To take but one example, increasing the absorbed solar radiation from $346W/m^2$ to $366W/m^2$ increases the ground temperature from 280K to 283K in the dry case. When the relative humidity is 50%, it takes only $290W/m^2$ of absorbed solar energy to maintain the same 280K ground temperature, but now increasing the solar absorption by $20W/m^2$ increases the surface temperature to 288K. Water vapor feedback has approximately doubled the climate sensitivity, which is a typical result for Earthlike conditions. The increase in climate sensitivity due to water vapor feedback plays a part in virtually any climate change phenomenon that can be contemplated on a planet with a liquid water ocean.

The influence of water vapor is strongest at tropical temperatures, but is still significant even at temperatures near freezing. It only becomes negligible at temperatures comparable to the polar winter.

One sometimes hears it remarked cavalierly that water vapor is the "most important" greenhouse gas in the Earth's atmosphere. The misleading nature of such statements can be inferred directly from Figure 4.31. Let's suppose the Earth's climate to be in equilibrium with $256W/m^2$ of absorbed solar radiation, averaged over the Earth's surface. This corresponds to an albedo of 25%, which we take to be somewhat smaller than the actual observed albedo to account crudely for the fact that part of the cloud albedo effect is canceled by cloud greenhouse effects which we do not take into account in the Figure. If water vapor were the only greenhouse gas in the Earth's atmosphere, the temperature would be a chilly 268K, and that's even before taking ice albedo feedback into account, which would most likely cause the Earth to fall into a frigid Snowball state. We saw earlier that the Earth would also be uninhabitably cold if CO_2 were the only greenhouse gas in the atmosphere. With regard to Earth's habitability, it takes two to tango. In order to maintain a habitable temperature on Earth without the benefit of CO_2 , the Sun would have to be 13% brighter. It will take well over a billion more years before the Sun will become this bright.

Now let's add some CO_2 to the atmosphere. Figure 4.32 shows how the OLR curve changes if we add in 300ppmv of CO_2 – slightly more than the Earth's pre-industrial value. The general



Figure 4.31: OLR vs. surface temperature for water vapor in air, with relative humidity held fixed. The surface air pressure is 1bar, and Earth gravity is assumed. The temperature profile is the water/air moist adiabat. Calculations were carried out with the ccm radiation model.

pattern is similar to the water-only case, but shifted downward by an amount that varies with temperature. At 50% relative humidity, the addition of CO_2 reduces the OLR at 280K by a further $36W/m^2$ below what it was with water vapor alone. Because of the additional greenhouse effect of CO_2 , the same $256W/m^2$ of absorbed solar radiation we considered previously can now support a temperature of 281K when the relative humidity is 50%. Note that without the action of CO_2 , the atmosphere would be too cold to have much water vapor in it, so one would lose much of the greenhouse effect of water vapor as well. It appears that the actual surface temperature of Earth can be satisfactorily accounted for on the basis of the CO_2 greenhouse effect supported by water vapor feedback.

Exercise 4.5.2 For the four moisture conditions in Figure 4.32, determine how much absorbed solar radiation would be needed to support a surface temperature of 280K. For each case, use the graph to estimate how much the surface temperature would increase if the absorbed solar radiation were increased were increased by $20W/m^2$ over the original value. How does the amplication due to water vapor feedback compare with the results obtained without CO_2 in the atmosphere?

Next, let's take a look at how the OLR curve varies as a function of CO_2 , with relative humidity held fixed at 50%. The results are shown in Fig. 4.33. The addition of CO_2 to the atmosphere lowers all the curves, and the more CO_2 you add, the lower the curves go. CO_2 is planetary insulation: adding CO_2 to a planet reduces the rate at which it loses energy, for any given surface temperature, just as adding fiberglass insulation to a house reduces the rate at which the house loses energy for a fixed interior temperature (thus reducing the fuel that must be burned in order to maintain the desired temperature). Another thing we note is that when the CO_2 concentration becomes very large, the curve loses its negative curvature and becomes concave upward, like σT^4 . This happens because the CO_2 greenhouse effect starts to dominate the water vapor greenhouse effect, so that the flattening of the OLR curve due to the increase of water vapor with temperature becomes less effective over the temperature range shown. Even for very high CO_2 , water vapor eventually would assert its dominance as temperatures are raised in excess of 320K, causing the curve once more to flatten as the Kombayashi-Ingersoll limit is approached.



Figure 4.32: As in Figure 4.31, but with 300ppmv of CO_2 included. Note that the "Dry" case excludes only the radiative effects of water vapor; the moist adiabat is still employed for the temperature profile.



Figure 4.33: OLR vs surface temperature for various CO_2 concentrations, at a fixed relative humidity of 50%. Other conditions are the same as for Figure 4.31.

As an example of the use of the information in Fig. 4.33, let's start with a planet with 100ppmv of CO_2 in its atmosphere, together with a sufficient supply of water to keep the atmosphere 50% saturated in the course of any climate change. From the graph, we see that absorbed solar radiation of 257 W/m^2 would be sufficient to maintain a mean surface temperature of 280K. From the graph we can also see that if the absorbed solar radiation is held fixed, increasing CO_2 tenfold to 1000ppmv would increase the temperature to 285K once equilibrium is re-established, and increasing it another tenfold to 10000ppmv (about 1% of the atmosphere) would increase the temperature to 293K. A further increase to 100000ppmv (10% of the atmosphere) increases the temperature to 309K. These represent substantial climate changes, but not nearly so extreme as they would have been if CO_2 increase is needed to maintain the same 280K surface temperature with a dimmer sun. Reading vertically from the intersection with the line $T_g = 280K$, we find that this temperature can be maintained with an absorbed solar radiation of $195W/m^2$ if the CO_2 concentration is 100000ppmv. Thus, an increase in CO_2 by a factor of 1000 can make up for a Sun which is 25% dimmer than the base case.

Finally, we'll take a look at how the OLR varies with CO_2 for a fixed surface temperature (Fig. 4.34). This figure is a more Earthlike version of the results in Fig. 4.29, in that the effects of water vapor on radiation and the adiabat have been taken into account. As in the dry case, there is broad range of CO_2 concentrations – about 5ppmv to 5000ppmv – within which the OLR decreases very nearly like the logarithm of CO_2 concentration. The slope depends only weakly on the relative humidity, especially if one leaves out the completely dry case. This suggests that the effects of water vapor and CO_2 on the OLR are approximately additive in this range. We note further that the logarithmic slope of OLR vs CO_2 becomes steeper at very high CO_2 , since one begins to engage more of the outlying absorption features of the CO_2 spectrum; again, CO_2 becomes an increasingly effective greenhouse gas at high concentrations. Conversely, at very low concentrations the logarithmic slope is reduced, as CO_2 absorption comes to be dominated by a relatively few dominant, narrow absorption features.

For any given CO_2 value, increasing the moisture content reduces the OLR, as would be expected from the fact that water vapor is a greenhouse gas. A little bit of water goes a long way. With 100ppmv of CO_2 in the atmosphere, going from the dry case to 10% relative humidity reduces the OLR by $36 W/m^2$. To achieve the same reduction through an increase of CO_2 , one would have to increase the CO_2 concentration all the way from 100ppmv to 10000ppmv. Clearly, water vapor is a very important player in the radiation budget, though we have already seen that because of the thermodynamic control of water vapor in Earthlike conditions, CO_2 nonetheless remains important. As the atmosphere is made moister, the further effects of water vapor are less dramatic. Increasing the relative humidity from 10% to 50% only brings the OLR down by $17W/m^2$, and going all the way to a saturated atmosphere only brings OLR down by a further 12 W/m^2 . Further, at very high CO_2 concentrations the OLR becomes somewhat more insensitive to humidity, as CO_2 begins to dominate the greenhouse effect.

The information presented graphically in Figures 4.32, 4.33 and 4.34 amount to a miniature climate model, allowing many interesting questions about climate to be addressed quantitatively without the need to perform detailed radiative and thermodynamic calculations; it's the kind of climate model that could be printed on a wallet-sized card and carried around everywhere. Calculations using this information can be simplified by presenting the data as polynomial fits, which eliminates the tedium and inaccuracy of measuring quantities off graphs. Then, what we have is a miniature climate model that can be programmed into a pocket calculator. Polynomial fits allowing the OLR to be calculated as a function of temperature, CO_2 and relative humidity are tabulated in Tables 4.2 and 4.3. Values of OLR for parameters intermediate between the tabulated



Figure 4.34: OLR vs CO_2 for a fixed surface temperature $T_g = 280K$, for various values of the relative humidity. Other conditions are the same as for Figure 4.31.

CO_2	a_o	a_1	a_2	a_3
10ppmv	258.56	2.5876	-0.0059165	-0.00013402
100ppmv	246.13	2.5056	-0.0034095	-0.00010672
1000ppmv	232.51	2.3815	-0.0015855	-8.3397e-05
10000ppmv	215.74	2.1915	0.00056634	-5.0508e-05
100000ppmv	189.06	1.8554	0.0044094	1.0735e-05

Table 4.2: Coefficients for polynomial fit $OLR = a_o + a_1x + a_2x^2 + a_3x^3$, where $x = T_g - 275$. Calculation carried out with rh = .5.

one can easily be obtained by interpolation.

Using these polynomial fits, we can put numbers to some of our old favorite climate questions by solving $OLR(T, CO_2) = S$ for T, given various assumptions about CO_2 and the absorbed solar radiation S. To wit –

• Global Warming: If we assume an albedo of 22.5%, the absorbed solar radiation is 265 W/m^2 . For 50% relative humidity, and a pre-industrial CO_2 concentration of 280ppmv, the corresponding equilibrium temperature is 285K, which is close to the pre-industrial global mean temperature. The albedo we needed to assume to get this base case is somewhat smaller than the Earth's observed albedo, because a portion of the cloud albedo is offset by the cloud greenhouse effect. Now, if we double the CO_2 to 560ppmv, the new temperature is

rh	a_o	a_1	a_2	a_3
Dry	313.8	-6.275	-0.36107	-0.019467
.1	277.28	-5.9416	-0.35596	-0.020237
.5	259.52	-5.5332	-0.33915	-0.018932
Saturated	249.1	-5.183	-0.32187	-0.017367

Table 4.3: Coefficients for polynomial fit $OLR = a_o + a_1x + a_2x^2 + a_3x^3$, where $x = \ln(CO_2/100)$, with CO_2 in *ppmv*. Calculation carried out with $T_g = 280K$ for the indicated moisture conditions.

287K – a two degree warming. This is essentially the same answer as obtained by Manabe and Wetherald in their pioneering 1967 calculation, and was obtained by essentially the same kind of calculation we have employed. If we double CO_2 once more, to 1120ppmv, then the temperature rises to 289K, a further two degrees of warming. The fact that each *doubling* of CO_2 gives a fixed additional increment of warming reflects the logarithmic dependence of OLR on CO_2 ; until one gets to extremely high concentrations, each doubling reduces the OLR by approximately $4W/m^2$.

- Pleistocene Glacial-Interglacial Cycles: In the depths of an ice age, the CO_2 drops to 180ppmv. Using the same base case as in the previous example, the temperature drops to 284K, about a one degree cooling relative to the base case. This by no means accounts for the full amount of ice age cooling, but it is significant enough to imply that CO_2 is a major player. In the Southern Hemisphere midlatitudes, away from the direct influence of the growth of major Northern Hemisphere ice sheets, the CO_2 induced cooling is a half to a third of the total, indicating that either the ice sheet influence is propagated into the Southern Hemisphere or ocean, or that the cooling we have calculated has been further enhanced by feedbacks due to clouds or sea ice.
- The PETM warming: How much CO_2 would you have to dump into the ocean-atmosphere system to account for the PETM warming discussed in Section 1.9.1? One answers this by first deciding how much the atmospheric CO_2 concentration needs to increase, and then making use of information about the partitioning of carbon between the atmosphere and ocean. The PETM warming has been conservatively estimated at 4C, and has nearly the same magnitude in the tropics as in the Arctic. The PETM event starts from an already warm hothouse climate, so for the sake of argument let's assume that the CO_2 starts at four times the pre-industrial value, yielding a starting temperature of 289K. We need to increase the CO_2 from 280ppmv to 1090ppmv (just under two doublings) to achieve a warming to 293K. This amounts to an addition of 1744 qigatonnes of C to the atmosphere as CO_2 , which is comfortably within the limits imposed by the ${}^{13}C$ data, but not all carbon added to the atmosphere stays in the atmosphere. Over the course of a thousand years, approximately 80% of atmospheric carbon will be absorbed into the ocean, and over longer periods the ocean may be able to take up even more. Thus, to sustain the warming more than a millennium, one would need to add at least 5 times the nominal value, or 8720 gigatonnes to the oceanatmosphere system. It would be hard to reconcile such a large addition of carbon with the isotopic record. This is the essential puzzle of the PETM, which seems to call for some kind of strong destabilizing feedback in the climate system.
- Deglaciation of Snowball Earth: If we increase the Earth's albedo to 60% (in accord with the reflectivity of ice) and reduce the solar constant by 6% (in accord with the Neoproterozoic value) the absorbed solar radiation is only 128 W/m^2 . With CO_2 of 280 ppmv, the equilibrium global mean temperature is a chilly 228K, more or less independent of what we assume about relative humidity. To determine the deglaciation threshold, we'll assume generously that the Equator is 20K warmer than the global mean, so that we need to warm the global mean to 254K to melt the tropics. Assuming 50% relative humidity, increasing CO_2 all the way to 200,000 ppmv (about 20% of the atmosphere) still only brings the global mean up to 243K, which is not enough to deglaciate. From this we conclude that without help from some other feedback in the system, CO_2 would have to be increased to values in excess of 20% of the atmosphere to deglaciate. In fact, detailed climate model calculations indicate that it is even harder to deglaciate a snowball than this calculation suggests
- Temperature of the post-Snowball hothouse Let's assume that somehow or other the Snowball

does deglaciate when CO_2 builds up to 20%. After deglaciation, the albedo will revert to 22.5%, and the absorbed solar radiation to 249 W/m^2 . When the planet re-establishes equilibrium, the temperature will have risen to 311K. This is hot, and the tropics will be hotter than the global mean. However, the planet does not enter a runaway greenhouse and these temperature are well within the survival range of heat-tolerant organisms, especially since the polar regions would probably be no warmer than today's tropics.

- Faint Young Sun: Let's consider a time when the Sun was 25% fainter than today, reducing the absorbed solar radiation by 66 W/m^2 . How much would CO_2 have to increase relative to pre-industrial values in order to keep the global mean temperature at 280K and prevent a freeze-out? We'll address this question by using the fit in Table 4.3. For 280 ppmv of CO_2 the OLR is 253 W/m^2 assuming 50% relative humidity. We need to bring this down by 66 W/m^2 to make up for the faint Sun. Using the fit, this can be done by increasing CO_2 to 240,000 ppmv (24% of the atmosphere) if we keep the relative humidity at 50%. If we on the other hand assume that for some reason the atmosphere becomes saturated with water vapor, then it is only necessary to increase the CO_2 to 15% of the atmosphere.
- The Earth in one billion years: According to Eq. 1.1, the solar constant will have increased to 1497 W/m^2 , increasing the absorbed solar radiation per unit surface area to 290 W/m^2 . If CO_2 is held fixed at the pre-industrial value, the Earth will warm to a global mean temperature of 296 K if relative humidity is held fixed at 50%. In order for silicate weathering to restore a temperature of 287K, the weathering would have to bring the CO_2 all the way down to 10ppmv, at which point photosynthesis as we know it would probably become impossible.
- Temperature of Gliese 581c and 581d: The planets Gliese 581c and 581d are in close orbits around a dim M-dwarf star. The redder spectrum of an M-dwarf would have some effect on the planetary energy balance, through changes in the proportion of solar energy absorbed directly in the atmosphere. Neglecting this effect, though we can estimate the temperatures of these planets assuming them to have an Earthlike atmosphere consisting of water vapor, CO_2 and N_2/O_2 . Gliese 581c is in an orbit where it would absorb about 583 W/m^2 , assuming the typical albedo of a rocky planet with an ocean. Gliese 581d would absorb only 50 W/m^2 . Even if CO_2 were 20% of the atmosphere, the OLR would be $82 W/m^2$, so Gliese 581d is likely to be an icy Snowball. On the other hand even with only 1ppmv of CO_2 in the atmosphere the OLR at 330K would be 351 W/m^2 , far below the absorbed solar radiation for Gliese 581c. Thus, if Gliese 581c has an ocean, it is very likely to be in a runaway state – something we'll confirm when we re-examine the runaway greenhouse for real-gas atmospheres. There's an additional wrinkle to the Gliese system, though, in that these planets are more massive than Earth and have higher surface gravity. The higher gravity somewhat reduces the water vapor greenhouse effect, since for a given vapor pressure the corresponding amount of mass in the atmosphere is lower, according to the hydrostatic relation. This turns out to cool Gliese 581c somewhat, but still not enough to save it from a runaway.

Exercise 4.5.3 About how much carbon would need to be added to the atmosphere to achieve a 4K PETM warming if the initial CO_2 at the beginning of the event were only twice the pre-industrial value? If the initial CO_2 were eight times the pre-industrial value? (Note that in the first case we are implicitly assuming some unknown process keeps the late Paleocene warm even with relatively little help from extra CO_2)

The above calculations include the effects of water vapor feedback, and also include the effects of changes in albedo due to ice cover, where explicitly mentioned. However, they do not

incorporate any feedbacks due to changing cloud conditions. Cloud changes could either amplify or damp the climate change predicted on the basis of clear-sky physics, according to whether changes in the cloud greenhouse effect or the cloud albedo effect win out. Unfortunately, there is no simple thermodynamic prescription that does for cloud feedbacks what the assumption of fixed relative humidity does for water vapor feedbacks. We will learn more about the factors governing cloud radiative forcing in Chapter 5, but idealized conceptual models for prediction of cloud feedbacks remain elusive.

The problem of whether elevated CO_2 can account for hothouse climates such as the Cretaceous and Eocene is considerably more challenging than the other problems we have discussed above, since it requires one to answer a regional climate question: Under what conditions can we suppress the formation of polar ice? We already saw in Chapter 3 that a rather small change in radiation balance can make the difference between a planet with a small polar ice sheet and a planet which is globally ice-free. Suffice it to say at this point that an increase of CO_2 to 16 times pre-industrial values – the upper limit of what is plausibly consistent with proxy data – would yield a global mean warming of 10K. Would this be enough to suppress formation of sea ice in the Arctic, and keep the mean Arctic temperatures around 10C? Would the associated tropical temperatures be too hot to be compatible with available proxy data? We'll have to return to these questions in Chapter 8, where we discuss the regional and seasonal variations of climate.

4.5.4 Greenhouse effect of CO_2 vs CH_4

4.6 section:OLR.CO2vsCH4

There is considerable interest in the idea that on the Early Earth methane may have taken over much of the role of CO_2 in offsetting the Faint Young Sun. In part this interest is due to rather sketchy geochemical evidence that at some times in the Archaean CO_2 concentrations may not have been high enough to do the trick, but regardless of whether the evidence actually *demands* a relatively low- CO_2 atmosphere, possibilities abound that in an anoxic atmosphere methane could build up to high concentrations. Even on an abiotic planet, there are possibilities for direct volcanic outgassing of methane, at a rate dependent on the state of oxygen in the planet's interior. Once biology comes on the scene, methanogens can convert volcanic CO_2 and H_2 to CH_4 , or can make CH_4 by decomposing organic matter produced by anoxygenic photosynthesis.

Methane cannot build up to very high concentrations in a well-oxygenated atmosphere, but the relatively small amounts of methane in the atmosphere today (about 1.7 ppmv) nevertheless contribute significantly to global warming. There is, however, a widespread misconception that methane is in some sense an intrinsically better greenhouse gas than CO_2 . A few simple calculations will serve to clarify the true state of affairs.

In order to compare the relative effects of CH_4 and CO_2 on a planet's radiation budget, we calculate the OLR for each case in what we have been calling the canonical atmosphere – a mixture of each gas into a dry atmosphere consisting of 1 bar of Earth air with temperature profile on the dry air adiabat, carried out with Earth's surface gravity. Results for a fixed surface temperature of 280K, computed using the homebrew radiation model employing a constant temperature scaling coefficient $T^* = 900K$, are shown in Fig. 4.35. This graph in essence gives the amount of greenhouse gas needed to sustain a surface temperature of 280K, given any specified amount of solar absorption. For example, with an absorbed solar radiation of $300W/m^2$ the surface temperature can be sustained with either 464ppmv of CO_2 or 35,600 ppmv of CH_4 (3.56% of the atmosphere



Figure 4.35: **CAPTION

by mole fraction). These results somewhat overestimate the effect of each gas as compared to an actual moist atmosphere, since a moist atmosphere would be on the less steep moist adiabat and in a moist atmosphere the water vapor absorption would compete to some extent with the CO_2 and CH_4 absorption. Still, as an estimate of the relative effect of the two gases, the story is pretty clear. Methane is, intrinsically speaking, a considerably worse greenhouse gas than CO_2 . The OLR curve for for methane is everywhere well above the curve for CO_2 , so that it takes more methane than CO_2 to achieve a given reduction of OLR.

The common statement that methane is, molecule for molecule, a better greenhouse gas than CO_2 is true only for situations like the present where methane is present in far lower concentrations than CO_2 . In this situation, the greater power of a molecule of CH_4 to reduce the OLR results simply from the fact that the greenhouse effect of both CH_4 and CO_2 are approximately logarithmic in concentration. Reading from Fig. 4.35, we see that for methane concentrations of around 1ppmv, each doubling of methane reduces OLR by about $2W/m^2$. On the other hand, for CO_2 concentrations near 300 ppmv, each doubling of CO_2 reduces the OLR by about 6 W/m^2 . Hence, to achieve the same OLR reduction as a doubling of CO_2 one needs three doublings of methane, but since methane starts from a concentration of only 1ppmv, this only takes the concentration to 8ppmv, and requires only $\frac{7}{300}$ as many molecules to bring about as was needed to achieve the same reduction using a doubling of CO_2 . Equivalently, we can say that adding 1ppmv of methane yields as much reduction of OLR as adding 75ppmv of CO_2 . The logarithmic slopes in this example are exaggerated compared to the appropriate values for Earth's actual atmosphere, because of the use of the dry adiabat and because of inaccuracies in the simple temperature scaling used in the homebrew radiation code. Using the ccm radiation code on the moist adiabat, with water vapor at 50% relative humidity, we find instead that each doubling of methane near 1ppmv reduces OLRby 0.77 W/m^2 , while each doubling of CO_2 near 300ppmv reduces OLR by 4.3 W/m^2 ; in this case adding 1ppmv of methane reduces the OLR by as much as adding 38ppmv of CO_2 . Nonetheless, the principle remains the same: If methane were the most abundant long-lived greenhouse gas in our atmosphere, and CO_2 were present only in very small concentrations, we would say instead that CO_2 is, molecule for molecule, the better greenhouse gas.

Kirschvink and others have proposed that the Makganyene Snowball came about through a methane catastrophe, in which oxygenation converts methane to CO_2 and reduces the greenhouse effect sufficiently to precipitate a snowball. A methane crash, due to a reduction in methanogenic



Figure 4.36: Total OLR reduction for the canonical atmosphere with a mixture of CH_4 and CO_2 . Each curve gives the OLR reduction relative to a transparent atmosphere for a fixed sum of CH_4 and CO_2 molar concentrations, indicated on the curve in units of *ppmv*. The results are plotted as a function of the CH_4 molar concentration to the total molar concentration for the two gases. The ratio is equal to the ratio of atmospheric carbon in the form of CH_4 to total atmospheric carbon. Larger values of the OLR reduction correspond to a stronger greenhouse effect.

activity of some other mechanism, has sometimes been proposed as a trigger for the Neoproterozoic Snowballs as well. It is by no means easy to make these scenarios play out as they are supposed to, since methane contributes much less to the greenhouse effect than CO_2 when the two gases have similar abundances in the atmosphere. Conversion of methane to CO_2 will only reduce the greenhouse effect if methane is initially present in sufficiently small concentrations – but if there is too little methane present, the contribution of methane to the total greenhouse effect is too small to make much difference. Let's use the data in Fig. 4.35 to illustrate how the conversion of methane to CO_2 would affect climate in a few illustrative cases. The detailed numbers would change with a more accurate radiation model, or if the effect of water vapor were brought in, but the basic conclusions would remain much the same.

For example, suppose we started out with an atmosphere that contained 36,650 ppmv of methane, which would be sufficient to maintain a 280K surface temperature given 300 W/m^2 of absorbed solar radiation. If this were all converted to CO_2 by oxidation, then, according to Fig. 4.35 the OLR would plunge to 255 W/m^2 . In order to re-establish radiation balance, the planet would have to warm up to a temperature well in excess of the the initial 280K. Far from causing a Snowball, in this case the oxidation of methane would cause a hot pulse, followed by gradual recovery to the original temperature as the CO_2 is drawn down by silicate weathering.

Next let's consider a more general situation, and identify the conditions necessary for a conversion of CH_4 to CO_2 to substantially reduce the greenhouse effect. Because the absorption features of CH_4 and CO_2 do not overlap significantly for the range of concentration under consideration, the combined effects of the two gases can be obtained by summing the OLR reduction ΔOLR for each of the gases taken in isolation. We wish to ask the question: if we have a given number of carbon atoms to use in supplying the atmosphere with greenhouse gases, how does the net greenhouse effect depend on the way we divvy up those atoms between CH_4 and CO_2 ? Since each molecule has a single carbon, this question can be addresed by varying the molar concentration of CH_4 while keeping the sum of the molar concentrations of CH_4 and CO_2 fixed. Results of

a calculation of this type are shown in Fig. 4.36. For any fixed total atmospheric carbon content, the OLR reduction has a broad maximum when plotted as a function of the methane ratio, and varies little except near the extremes of an all-methane or all- CO_2 atmosphere. The only case in which one can get a substantial reduction in greenhouse effect by oxidizing methane into CO_2 is when the initial CO_2 concentration is very high, the initial CH_4 fraction is between about 10% and 90% of the total, and the CH_4 is almost entirely converted to CO_2 . For example, with a total carbon concentration of 10000ppmv, reducing the CH_4 concentration from 1000ppmv to 1ppmvreduces the greenhouse effect from $104W/m^2$ to 80 W/m^2 . Because the curve is so flat, starting from an atmosphere which is 80% methane works almost as well: in that case the greenhouse effect is reduced from 101 W/m^2 to 80 W/m^2 . If we have a total of 100,000 ppmv of carbon in the atmosphere, then the maximum greenhouse effect occurs for an atmosphere which is about 25% methane, and has a value of 151 W/m^2 . Reducing the methane to 1 ppmv brings down the greenhouse effect 36 W/m^2 , to 115 W/m^2 . In the Paleoproterozoic or Archaean, when the net greenhoue effect needed to be high to offset the Faint Young Sun, it is possible that a methane crash could have reduced the greenhouse effect enough to initiate a snowball, but it is essential that in a methane crash, the methane concentration be brought almost all the way down to zero; a reduction of methane from 50% of the atmosphere to 10% of the atmosphere would not do much to the greenhouse effect. By the time of the Neoproterozoic, when the solar luminosity is higher and less total greenhouse effect is needed to maintain open water conditions, it is far less likely that a methane catastrophe could have initiated glaciation. Some further remarks on atmospheric transitions that could initiate a Snowball will be given in Chapter 9.

4.7 Another look at the runaway greenhouse

We are now equipped to revisit the runaway greenhouse phenomenon, this time using the absorption spectrum of actual gases in place of the idealized grey gas employed in Section 4.3.3. The setup of the problem is essentially the same as in the grey gas case. We consider a condensible greenhouse gas, optionally mixed with a background gas which is transparent to infrared and noncondensing. A surface temperature T_g is specified, and the corresponding moist adiabat is computed. The temperature and the greenhouse gas concentration profiles provide the information necessary to compute the OLR, in the present instance using the homebrew exponential sums radiation model in place of the greygas OLR integral. As before, the OLR is plotted as a function of T_g for the saturated atmosphere, and the Kombayashi-Ingersoll limit is given by the asymptotic value of OLRat large surface temperature.

We'll begin with water vapor. Figure 4.37 shows the results for a pure water vapor atmosphere, computed for various values of the surfaced gravity. The overall behavior is very similar to the grey gas result shown in Fig. 4.3: the OLR attain a limiting value as temperature is increased, and the limit – defining the absorbed solar radiation above which the planet goes into a runaway state – becomes higher as the surface gravity is increased, and for precisely the same reasons as invoked in the grey gas case. The result, however, is now much easier to apply to actual planets, since with the real gas calculations we have the real numbers in hand for water vapor, and not for some mythical gas characterized by a single absorption coefficient. Several specific applications will be given shortly.

As in the grey-gas case the limiting OLR increases as surface gravity is increased.

Now let's generalize the calculation, and introduce a noncondensing background gas which is transparent in the infrared; we use N_2 in this example, though the results are practically the same if we use any other diatomic molecule. The background gas affects the Kombayashi-Ingersol



Figure 4.37: OLR vs surface temperature for a saturated pure water vapor atmosphere. The numbers on the curve indicate the planet's surface gravity. The calculation was done with the homebrew exponential sums radiation code, incorporating both the 1000 cm^{-1} and 2200 cm^{-1} continua, but neglecting temperature scaling of absorption outside the continua. 20 terms were used in the exponential sums, and wavenumbers out to 5000 cm^{-1} were included; the atmosphere was considered transparent to higher wavenumbers

limit in two ways: First, the pressure broadening increases absorption, which should lower the limit. Secondly, the background gas shifts the lapse rate toward the dry adiabat, which is much steeper than the single-component saturated adiabat. The increase in lapse rate in principle could enhance the greenhouse effect, but given the condensible nature of water vapor, it actually reduces the greenhouse effect, because the low temperatures aloft sharply reduce the amount of water vapor there. If the background gas were itself a greenhouse gas, this effect might play out rather differently. At sufficiently high temperatures, water vapor will dominate the background gas and so the limiting OLR at high temperature will approach the pure water vapor limit shown in Figure 4.37. However, for intermediate temperatures, the background gas can modify the shape of the OLR curve.

Results for various amounts of N_2 are shown in Fig. 4.38.

- Runaway greenhouse on Earth: With present absorbed solar radiation (adjusted for net cloud effects) of 265 W/m^2 , the Earth at present is comfortably below the Kombayashi- Ingersoll limit for a planet of Earth's gravity. According to Eq. 1.1, as the solar luminosity continues to increase, the Earth will pass the 291 W/m^2 threshold where a runaway becomes possible in about 700 million years. In 1.7 billion years, it will pass the $310W/m^2$ threshold where a runaway becomes inevitable for an atmosphere with 1 bar of N_2 and no greenhouse gases other than water vapor.
- Venus: **
- Gliese 581c: **
- Evaporation of icy moons in Earthlike orbit: **
- Lifetime of a post-impact steam atmosphere: [**Do Earth case here. Mars as an exercise]



Figure 4.38: As for Fig. 4.37 but for a mixture of water vapor in N_2 on the saturated moist adiabat. Calculations were carried out with a surface gravity of $20m/s^2$, for the indicated values of N_2 partial pressure at the ground.

• Freeze-out time of a magma ocean:**

4.8 Pure radiative equilibrium for real gas atmospheres

Pure radiative equilibrium amounts to an all-stratosphere model of an atmosphere, and is a counterpoint to the all-troposphere models we have been discussing. Real atmospheres sit between the two extremes, sometimes quite near one of the idealizations. In this section we will focus on pure infrared radiative equilibrium. The effects of solar absorption in real gases will be taken up in Chapter 5.

From simple analytic solutions, we know essentially all there is to know about pure radiative equilibrium for grey gases. It is important to understand these things because the structure of atmospheres results from an interplay of convection and pure radiative equilibrium. A thorough understanding of pure radiative equilibrium provides the necessary underpinning for determining where the Stratosphere starts, and its thermal structure. We will now examine how the key elements of the behavior of pure radiative equilibrium differ for real gases. The specific issues to be addressed are:

- For grey gases in radiative equilbrium, the minimum temperature is the skin temperature based on OLR, and is found in the optically thinnest part of the atmosphere in the absence of atmospheric solar absorption. For real gases, can the radiative equilibrium temperature be much lower than the skin temperature?
- For a grey gas atmosphere with a given vertical distribution of absorbers, the radiative equilibrium temperature profile is uniquely determined once the OLR is specified. Specifically, one can determine the temperature profile of the radiative-equilibrium stratosphere without needing to know anything about the troposapheric temperature structure. To what extent is this also true for real gases?



Figure 4.39: OLR vs surface temperature for a saturated pure CO_2 atmosphere. Calculations were performed with the values of surface gravity indicated on each curve. The scale at the top gives the surface pressure corresponding to the temperature on the lower scale.

- For a grey gas atmosphere with a given vertical distribution of absorbers, the normalized temperature profile $T(p)/T_g$ is independent of the ground temperature. This means that the radiative equilibrium temperature profile has the same shape, regardless of the magnitude of the solar radiation with which the atmosphere is in balance. How much does this result change for real gases?
- For a grey gas, the temperature jump at the ground is greatest when the atmosphere is optically thin, and vanishes as the atmosphere becomes optically thick. For real gases, the atmosphere is optically thick in some parts of the spectrum, and optically thin in others. What determines the temperature jump in these circumstances?
- Grey gases are most unstable to convection where they are optically thick. In the optically thick limit, the slope $d \ln T/d \ln p$ equals $\frac{1}{4}$ without pressure broadening, and $\frac{1}{2}$ with pressure broadening. Hence, a dry adiabat can go unstable near the ground only if $R/c_p < \frac{1}{4}$ without pressure broadening, or $R/c_p < \frac{1}{2}$ with pressure broadening. How do these thresholds differ for a real gas?

All of the issues except for the behavior of the static stability near the ground are well illustrated by the *semi-grey* model, which we referred to more poetically as "one-band Oobleck" in Section 4.4.1. In this model, we assume κ to be constant within a band of width Δ centered on frequency ν_o , and zero elsewhere. To keep the algebra simple we will make the additional assumption that Δ is small enough that B can be considered essentially frequency-independent within the band; this assumption can be easily dispensed with at the cost of a slightly more involved calculation. What makes the semigrey model tractable is that the infrared heating is due solely to the flux within the absorbing band, so that one can still deal with a single optical thickness without any need to sum or integrate over frequency to get the net heating.

First, let's consider the semigrey radiative equilibrium in the optically thin limit. From the results in Section 4.2.2, the infrared radiative cooling at any level is simply $2\pi B(\nu_o, T)\Delta$, whereas the heating by absorption of upwelling infrared from the ground is $\pi B(\nu_o, T_q)\Delta$. Balancing the

two, gives the equation determining the atmospheric temperature:

$$B(\nu_o, T) = \frac{1}{2}B(\nu_o, T_g)$$
(4.94)

Since the equation for T is independent of level, we conclude that in the optically thin case, the atmosphere is isothermal in infrared radiative equilibrium, just as it is for the optically thin grey case. The resulting atmospheric temperature is always lower than the ground temperature, and may be called the *semigrey skin temperature*. For a grey gas, $T_{skin}/T_g = 1/2^{1/4} \approx .84$, but for the semigrey case, the ratio depends on the frequency of the absorbing band. From the form of the Planck function, it can easily be shown that T_{skin}/T_g depends on frequency only through the ratio $h\nu_o/kT$. A simple analytic calculation shows that $T_{skin}/T_g \rightarrow \frac{1}{2}$ for small values of $h\nu_o/kT$ and $T_{skin}/T_g \rightarrow 1$ when $h\nu_o/kT$ becomes large; a simple numerical solution shows that the ratio increases monotonically between the two limiting cases (see Problem ??). The grey-gas skin temperature ratio sits in the middle of this range, and indeed for frequencies near the peak of the Planck function, the semigrey ratio does not differ greatly from the grey value. For example, at $650cm^{-1}$ and a ground temperature of 280K, the semigrey skin temperature is 233K, whereas the grey skin temperature is 235K

When deriving the radiative equilibrium for grey gases, we specified the OLR and used that as a boundary condition for determining the thermal structure; the ground temperature is then computed from the resulting lower boundary fluxes. For real gases, it proves more convenient to specify the ground temperature T_g , and find the resulting temperature structure and OLR. We adopt this approach not only in our analytic solution of the semigrey case, but also in our numerical solutions for actual gases. When fixing T_g and increasing optical thickness, the OLR goes down, and so the amount of absorbed solar radiation needed to maintain the stated T_g decreases.

The derivation of the full infrared radiative equilibrium solution for the semigrey case is identical to that we used in the grey case, with the following substitutions: (1) The optical thickness τ is based on the value of κ in the absorbing band alone, (2) σT^4 is replaced by $\pi B(\nu_o, T)$ (3) The fluxes I_+ and I_- represent the flux integrated over the band Δ alone, and (4) The *OLR* appearing in the top boundary condition is no longer the net *OLR* emitted by the planet, but only the portion of *OLR* (call it OLR_{Δ}) emitted in the absorbing band. The assumption of infrared equilibrium still implies that the flux $I_+ - I_-$ is constant and equal to OLR_{Δ} , but since this flux is only the portion of *OLR* in the band, it is no longer determined a priori by planetary energy balance. Instead, it must be determined by making use of both boundary conditions, that $I_- = 0$ at the top of the atmosphere, and $I_+ = \pi B(\nu_o, T_g)$ at the bottom of the atmosphere (assuming the ground to have unit emissivity). The result of applying both boundary conditions is

$$I_{+} - I_{-} = \frac{2\pi}{2 + \tau_{\infty}} B(\nu_{o}, T_{g})$$
(4.95)

This is equal to the ground emission in the optically thin case, and approaches zero as the atmosphere is made more optically thick. Note, however, that it is only the OLR in the band that approaches zero; the net emission approaches a finite, and possibly quite large, lower bound, because the emission from the rest of the spectrum where the atmosphere is transparent is simply the ground emission. By substituting the expression for the net upward flux into the expression for $I_+ + I_-$, we find the following expression determining $T(\tau)$

$$B(\nu_{o}, T) = \frac{1 + \tau_{\infty} - \tau}{2 + \tau_{\infty}} B(\nu_{o}, T_{g})$$
(4.96)

where τ is the optical thickness within the band where κ is nonzero. At the top of the atmosphere, this reduces to $B(\nu_T) = B(\nu_o, T_g)/(2 + \tau_\infty)$ Note that this is always less than the semi-grey skin

temperature, and becomes progressively colder in comparison the the skin temperature as τ_{∞} is made large. For the semigrey case, then, the stratospheric temperature differs from the grey gas case in two important ways. First, as the atmosphere is made optically thick in the absorbing band, the stratospheric temperature approaches zero even though the net OLR remains finite (owing to the emission through the transparent part of the spectrum). Thus, the stratosperic temperature can be much colder than the grey-gas skin temperature, resolving the quandary raised in Chapter 3. Moreover, as the optical thickness of the atmosphere is increased, the stratosphere actually becomes colder than even the semigrey skin temperature; this contrasts with the grey case, where the temperature of the uppermost part of the atmosphere always approaches the skin temperature, regardless of how optically thick the rest of the atmosphere is. This important difference arises because, in the semigrey case, the optical thickness of the lower part of the atmosphere causes the upwelling spectrum illuminating the stratosphere to be depleted in those frequencies where the stratosphere absorbs best. Nonetheless, the stratosphere continues to *emit* effectively at those frequencies, leading to very cold temperature. This depletion also has the important consequence that the stratospheric temperature is no longer independent of the existence of a troposphere, or of the tropopause height and thermal structure of the troposphere. This has the potential to affect the calculation of the tropopause height when we put radiative equilibrium together with convection.

The temperature jump at the ground is determined by $B(\nu_o, T_{sa}) = ((1+\tau_\infty)/(2+\tau_\infty))B(\nu_o, T_g)$, which has the same form as the corresponding expression for the grey gas case, save for the appearance of the Planck function in place of σT^4 . As in the grey gas case, the jump is a maximum for optically thin atmospheres, and vanishes in the optically thick limit. The main difference here is that the jump can be made to vanish by making the atmosphere optically thick in just a limited part of the spectrum, even though the atmosphere is optically thin (in this case absolutely transparent) elsewhere. This feature will reappear in our discussion of radiative equilibrium for an atmosphere in which infrared absorption is provided by CO_2 . As the atmosphere becomes optically thick in the absorption band, the unstable temperature jump at the ground diminishes, but what happens to the interior static stability of the air near the ground? To determing this, we take the derivative of Eq. 4.96 and multiply by p_s/T_{sa} to obtain the logarithmic radiative equilibrium lapse rate at the ground:

$$\frac{d\ln T}{d\ln p}\Big|_{p_s} = \frac{B(\nu_o, T_g)}{T_{sa}\frac{dB}{dT}(\nu_o, T_{sa})}\frac{\kappa(p_s)p_s/g}{2+\tau_\infty}$$
(4.97)

In the optically thick limit, the second factor is unity without pressure broadening, or $\frac{1}{2}$ with pressure broadening. When the atmosphere is optically thick, the first factor can be evaluated with $T_{sa} = T_g$. With a little algebra, the first term can be re-written as $(\exp(u) - 1)/(u \exp(u))$, where $u = h\nu_o/(kT_g)$; this term has a maximum value of unity at u = 0 (high frequencies or low temperatures) and decays to zero like 1/u at large u (low frequencies or high temperatures). For u = 1 which puts the absorption band near the maximum of the Planck function, the value is about .63. For the semi-grey model then, we conclude that the degree of instability of radiative equilibrium in the optically thick limit is bounded; for example, with u = 1 and incorporating pressure broadening, the radiative equilibrium near the ground is statically unstable when $R/c_p >$.315 (vs. a threshold value of .5 for a grey gas). Hence, the semigrey case has a somewhat enhanced instability at the ground as compared to the grey case, but the difference is not great. We'll see shortly that this is one regard in which the qualitative behavior of a real gas differs significantly from the semigrey case.

Finally, let's look at how the radiative equilibrium profile for the semigrey atmosphere changes if we change T_g and leave everything else fixed. For a grey gas, the function T/T_g is invariant because both the surface emission and the interior atmospheric emission increase by a



Figure 4.40: Left panel: Infrared radiative equilibrium temperature profiles for one and two-band Oobleck, and for a CO_2 -air mixture at 300ppmv and 3000ppmv, subject to a fixed 280K ground temperature. For the Oobleck cases, the absorption coefficients were chosen such that the optical depth of the atmosphere as a whole is 10 in the strong absorption band $(650 - 700cm^{-1})$ and 1 in the weak bands on the flanks $(600 - 650cm^{-1} \text{ and } 700 - 750cm^{-1})$. Right panel: The corresponding logarathmic slope, $d \ln T/d \ln p$

factor b^4 if we replace T_g by $b \cdot T_g$. For the semigrey case, the emission is given by $B(\nu_o, T)$, which is no longer a simple power of T. This means that the ratio T/T_g is no longer independent of T_g . The nature of the dependence is left for the reader to explore in Problem ??.

Working our way up the ladder to reality, we'll now present some numerical solutions for 2-band Oobleck and for mixtures of CO_2 in air. The latter are computed using our homebrew exponential sum radiation code, incorporating the effect of pressure broadening but without taking temperature scaling into account. This is sufficient to show how the extreme range of absorption coefficients in a typical real gas affect the radiative equilibrium. The equilibria were found by a simple time stepping method with fixed T_q . For any given initial temperature profile T(p) one can calculate the infrared fluxes, and hence, by differencing in the vertical, the infrared heating rates. These are used to update T(p), and the whole process is repeated until equilibrium is achieved, where the infrared heating is zero and the temperature no longer changes. Since we are only interested in the equilibrium and not the time course of the approach to equilibrium, we can afford to be somewhat sloppy in our time-stepping method, so long as it is stable enough to yield an equilibrium at the end. Figure reffig:RealGasRadEq shows the resulting equilibrium profiles for two-band Oobleck with the secondary-band absorption coefficient one tenth the value of the primary, and for a CO_2 -air mixture in an Earthlike atmosphere, at 300ppmv and 3000ppmv. The left panel shows the temperature profile, while the right panel gives the logarithmic slope of temperature, which determines the static stability of the atmosphere. For comparison, the analytically derived results for one-band pressure-broadened Oobleck (the semigrey model) are also shown. These calculations were carried out with a ground temperature $T_g = 280K$. Some key characteristics of the results are summarized in Table 4.4.

Comparing the curves for 1-band and 2-band Oobleck shows that adding in the weakly absorbing bands reduces the vertical temperature gradient except in a thin layer near the ground. In this sense, most of the atmosphere acts as if it were made more optically thin, despite the fact that we have actually made the atmosphere more opaque to infrared by adding new absorption in additional bands without taking away absorption in the original band. Indeed, the OLR goes down from 333 W/m^2 in the 1-band case to 309 W/m^2 in the 2-band case, despite the warmer temperatures aloft in the latter. The key to this behavior was pointed out back at the end of Section 4.2.2: in an atmosphere which is optically thin in some parts of the spectrum but optically thick in others, the heating rate (which in turn determines the radiative equilibrium) is dominated by the parts of the spectrum where the optical thickness is nearest unity. This is the reason the weaker absorption bands control the behavior of the temperature in the interior of the atmosphere. The associated reduction in temperature gradient warms the atmosphere aloft, at pressures below 700mb. In essence, the weak bands allow the upper reaches of the atmosphere to capture more of the infrared upwelling from below; spreading the absorption over a somewhat broader range of the spectrum in essence makes the problem a bit more like a grey gas, and makes the upper air temperature somewhat closer to the grey gas skin temperature.

While the addition of the weak bands reduces the temperature gradient aloft and hence stabilizes the atmosphere against convection, this comes at the expense of increasing the gradient near the ground, and destabilizing the layer there. This stabilization/destabilization pattern shows up clearly in the plot of the logarithmic temperature derivative shown in the right panel of Figure 4.40. Based on $R/c_p = 2/7$, the one-band Oobleck profile is unstable for pressures higher than 450mb, whereas the two-band case is only unstable for pressures higher than 760mb, though within the unstable layer the 2-band case is considerably more unstable than the one-band case. In the twoband case, more of the net vertical temperature contrast of the atmosphere is concentrated in a thin layer near the ground. Overall, the atmosphere acts as if it were optically thick near the ground, but relatively optically thin aloft. The behavior near the ground results from the extremely strong absorption near the center of the CO_2 absorption band. This spectral region captures the upwelling radiation from the ground, which is not yet depleted in the strongly absorbing wavenumbers. If there were much temperature discontinuity near the ground, the absorption would lead to strong radiative heating of the low level air; hence the only way to be in radiative equilibrium is for the air temperature to approach the ground temperature. More mathematically speaking, the phenomenon arises because the low level heating is controlled largely by the boundary terms in the flux integral in Eq 4.9, which in turn is dominated by the strongly absorbing spectral regions.

The real-gas CO_2 results have many features in common with 2-band Oobleck, notably the weak temperature gradient in the interior of the atmosphere and the enhancement of temperature gradient near the ground. The strong destabilization near the ground is even more pronounced for CO_2 , because it has a far stronger peak absorption than the Oobleck case we considered, and because the absorption varies over a greater range of values. The associated optical thickness near the ground also keeps the unstable temperature jump between the ground and the overlying air small. An optically thin grey-gas would have a large unstable temperature jump at the ground. The strong absorption bands in a real gas smooth out this discontinuity and move it into a finite width layer in the interior of the atmosphere. From the standpoint of the convection produced, there is little physical difference between the two cases.

The main differences between the CO_2 cases and the Oobleck cases shows up in the upper atmosphere, where the CO_2 cases show steep declines of temperature with height – though not so steep as to destabilize the upper atmosphere. As at the bottom boundary, the culprit is the spectral region where absorption coefficients peak. These regions lead to very strong emission in the upper layers of the atmosphere, which are poorly compensated by absorption since the upwelling infrared

236CHAPTER 4. RADIATIVE TRANSFER IN TEMPERATURE-STRATIFIED ATMOSPHE

	$OLR, W/m^2$	T_{skin}	T(0)	$T(p/p_s = .5)$	$T_g - T(p_s)$
1-band Oobleck	333	232K	$161 \mathrm{K}$	206K	$6.5 \mathrm{K}$
2-band Oobleck	309	228K	182K	215K	$12.75 \mathrm{K}$
$CO_2/\text{air } 300ppmv$	295	226K	126K	214K	$2.8 \mathrm{K}$
CO_2 /air 3000ppmv	275	222K	115K	214K	$1.75 \mathrm{K}$
Mars, $CO_2, p_s = 7mb$	303	227K	128K	215K	2.9 K
Mars, $CO_2, p_s = 7mb, T_g = 250K$	192	203K	126K	196K	$4.4 \mathrm{K}$
Mars, $CO_2, p_s = 2bar$	86	166K	102K	208K	1.3K
$Venus, CO_2, p_s = 90bar, T_a = 700K$	55	148K	77K	500K	.1K

Table 4.4: Summary properties of infrared radiative equilibrium solutions. Calculations were done with $T_g = 280K$ unless otherwise noted.

reaching the upper atmosphere has been depleted in most wavenumbers that absorb at all well. The strong emission causes the temperature near the top to fall far below the skin temperature. Based on the net OLR, the grey gas skin temperature for the two CO_2 cases is somewhat above 220K, whereas the actual temperature at the 1mb level is 126K in the 300ppmv case and 115K in the more optically thick 3000ppmv case. Finally, we note that increasing CO_2 by a factor of 10 has relatively little effect on the radiative equilibrium temperature profile, despite the fact that the increase lowers the OLR by nearly $20W/m^2$. The changes are principally seen near the ground, where the increased optical thickness in the wings has reduced the surface temperature jump. While the two-band Oobleck model reproduces the near-ground destablization present in the real CO_2 calculation, it is unable to simultaneously represent the temperature jump at the ground. Lacking the extremely strong peak absorption of CO_2 , the addition of the weak absorption wings in 2-band Oobleck makes the surface budget act like an optically thinner atmosphere, increasing the surface jump.

To illustrate how the radiative equilibrium solution scales with T_g , we show the profiles of T/T_g for surface temperatures ranging from 240K to 320K. Only the case of 300ppmv CO_2 is shown, though the other atmospheres treated in Figure 4.40 yield similar results. For a grey gas, all the curves for a given atmospheric composition would collapse onto a single universal profile; for the reasons discussed in the semigrey case, this is no longer true for real gases. However, while the temperature aloft does not scale precisely with the ground temperature, the deviations are modest enough that one can still get useful intuition about the behavior of the system by assuming that radiative equilibrium temperature scales with the ground temperature. For CO_2 , the actual temperature aloft is always somewhat colder than that which one would estimate by proportionately scaling the temperature upward from a colder to warmer ground temperature. Recall that these calculations are done without temperature scaling of the absorption coefficient, so the effect shown is purely due to the shape of the Planck function.

The general features encountered in the terrestrial calculations discussed above carry over to the pure CO_2 atmospheres characteristic of Present and (possibly) Early Mars. Martian radiative equilibrium solutions with a 7mb thin atmosphere or 2bar thick atmosphere are shown in Figure 4.42. The most striking feature of these solutions is that increasing the mass of the atmosphere by a factor of nearly 300 causes very little increase in the vertical temperature contrast. This stands in sharp contrast to the grey gas case, for which the enormous increase in optical depth going from the 7mb case to the 2bar case would cause the temperature in the latter to drop to nearly zero within a short distance of the ground. As before, the reason for the relative insensitivity of the temperature profile is that the radiative heating is determined largely by the part of the spectrum where the optical depth is of order unity. For the present Mars case, this occurs in the near wings



Figure 4.41: Variation of the shape of the temperature profile as a function of ground temperature. Results are shown for 300ppmv of CO_2 in air.



Figure 4.42: Infrared radiative equilibrium for pure CO_2 atmospheres with Martian gravity. Results are shown for an Early Mars case with a 2 *bar* surface pressure and 280*K* ground temperature, and for Present Mars cases with 700*mb* surface pressure and 250*K* and 280*K* ground temperatures. To make it possible to compare the cases, the temperatures have been plotted as a function of p/p_s . The 2*bar* case includes the CO_2 continuum absorption.

of the principle absorption peak, whereas for Early Mars it occurs within the continuum window region. The shift allows both cases to act roughly like a case with order unity optical depth, apart from the thin radiative boundary layers near the ground and the top. The temperature profiles are similar despite the fact that it would require $303W/m^2$ of absorbed Solar radiation to maintain a surface temperature of 280K in the thin present Martian atmosphere, but only $86W/m^2$ in a 2bar atmosphere. For the Present Mars case, we have also included a calculation with with a realistically cold daytime surface temperature, the equilibrium temperature aloft is too cold in comparison with observations. This suggests once more an important role for solar absorption in determining the temperature structure of the present Martian atmosphere.

It is only when we go to massive atmospheres like that of Venus that the atmosphere becomes optically thick throughout the spectrum, allowing the vertical temperature contrast to increase dramatically. A simplified calculation without temperature scaling, and ignoring emission beyond $2300 cm^{-1}$, shows that with a 700K ground temperature, the radiative equilibrium temperature drops to 500K at the midpoint of the atmosphere and all the way to 80K at the 100mb level. This vields a very strong greenhouse effect: it takes only a trickle of $55W/m^2$ of absorbed solar radiation to maintain the torrid ground temperature. Note, however, that it is important that a fair amount of this trickle actually be absorbed at the ground, and not in the upper reaches of the atmosphere: otherwise the deep atmosphere becomes isothermal, and can in fact become as cold as the skin temperature in extreme cases, as discussed in Section 4.3.5. So far as the maintainence of its thermal structure is concerned, the troposphere of Venus is more like the Antarctic glacier than it is like the Earth's troposphere. The trickle of heat escaping the Earth's interior beneath the glacier - a mere $30 \ mW/m^2$ - is sufficient to raise the basal ice temperature to the melting point and create subglacial Lake Vostok precisely because the diffusivity of heat in ice is so small. So it is, too, with the atmosphere of Venus; the extremely optically thick tropophere renders the radiative diffusivity of heat very small, and allows the tiny trickle of solar radiation reaching the surface to accumulate in the lower atmosphere and raise the temperature to extreme values. Unlike the glacier, however, when the lower atmosphere becomes hot enough, it can start to convect. Convection supplants the radiative heat flux, but also establishes the adiabat, allowing the surface to be much hotter than the radiating level.

Different noncondensible greenhouse gases differ somewhat in details of their radiativeequilibrium profiles, but the general picture does not differ greatly from what we have learned by looking at CO_2 . When the greenhouse gas can condense near the ground, however, the situation becomes quite different. The case of water vapor in air provides a prime example. If the ground temperature is high enough that the amount of water vapor present in saturation makes the lower atmosphere optically thick, then the temperature will decline rapidly with height above the ground, because that is what optically thick atmospheres do in radiative equilibrium. Water vapor exhibits this effect particularly strongly, since it easily makes the atmosphere optically thick everywhere outside the window regions of the spectrum, and even the windows close off above 300K. As the temperature decreases, however, the water vapor content decreases in accordance with the limits imposed by Clausius-Clapeyron. Within a small distance above the ground, the air is so cold that there is little water vapor left, and the atmosphere further aloft becomes optically thin. As a result, most of the variation in optical thickness of the atmosphere is concentrated into a thin, radiative boundary layer near the ground, and the optical thickness (and hence the temperature) varies greatly within this layer. Because of the strong temperature gradient and high optical thickness of the boundary layer, a strong greenhouse effect is generated entirely within the boundary layer, leading to low OLR. If one imposes equilibrium with an Earthlike absorbed solar radiation, the ground temperature must increase to temperature well in excess of 320K to achieve balance, and the steep increase of saturation vapor pressure with temperature further exacerbates the high temperature gradient in the radiative boundary layer. It is a bit as if the



Figure 4.43: Basic features of real gas pure infrared radiative equilibrium.

entire optically thick atmosphere of Venus were squeezed into a boundary layer having a depth of a kilometer or less. Adding a noncondensing gas like CO_2 to the mix alters the temperature profile above the boundary layer, but does not eliminate the basic pathology of the situation. The equilibrium profile in this case is of little physical consequence because the slightest convection or other turbulent mixing would mix away the thin radiative boundary layer, warming and moistening a much deeper layer of the atmosphere The radiative equilibrium solutions we have been studying for the noncondensing case are worthy of protracted consideration because they provide some useful insight as to the stratospheric temperature for more realistic atmospheres in which there is some low level convection. The same cannot be said for pure radiative equilibrium in the condensing water-vapor/air system, which is an exercise in pathology having little or no bearing on the operation of real atmospheres.

Though real gas radiative equilibrium is not amenable to the kind of complete solution we enjoyed for the grey gas case, its behavior can be reasonably well captured by a few generalities, summarized in Figure 4.43. For real gases as for grey gases heated by infrared emission from the lower boundary, the temperature decreases with distance from the boundary. This can be viewed as a kind of thermal diffusion in the sense that heat transfer is down the temperature gradient, though the process is only described by a true local diffusion in the optically thick limit. Real gases behave as if they are optically thick near the ground, exhibiting strong convectively unstable temperature gradients there and little temperature jump between the ground and overlying air. The temperature gradient weakens in the interior, but there is generally a region of strong, though stable, temperature decline near the top of the atmosphere. The upper atmosphere is considerably colder than the grey gas skin temperature, since (by Kirchoff's law) the atmosphere radiates efficiently in the strongly absorbing parts of the spectrum, but the radiation illuminating the upper atmosphere is depleted in this portion of the spectrum. In contrast to the grey gas case, the contrast in temperature across the depth of the atmosphere is relatively insensitive to the amount of greenhouse gas in the atmosphere. As long as there are some spectral regions where

240CHAPTER 4. RADIATIVE TRANSFER IN TEMPERATURE-STRATIFIED ATMOSPHERES

the atmosphere is optically thick, and some where the atmosphere is optically thin, the radiative cooling tends to be dominated by the intermediate spectral regions. As a result, temperature tends to drop by a factor of two to three between the ground and the upper atmosphere, with only slight increases even when the greenhouse gas content is increased by many orders of magnitude. This behavior persists until there is so much greenhouse gas present that the atmosphere becomes very optically thick throughout the thermal infrared spectrum, as is the case on Venus.

The upshot of all this is that atmospheres whose temperature is maintained by absorption of upwelling infrared from a blackbody surface will never exhibit pure radiative equilibrium. There will always be a layer near the surface which is unstable to convection. If the atmosphere is optically thin, the instability is generated by a temperature jump at the surface. If the atmosphere is optically thick and subject to pressure broadening, the instability is generated by strong temperature gradients in the interior of the atmosphere near the surface. This remark even applies qualitatively to gas giants which have no surface, as the deep atmosphere is dense enough that it can begin to act like a blackbody even though there is no distinct surface. An atmosphere can be stabilized throughout its depth, however, if it is subject to atmospheric solar heating which increases with altitude in a suitable fashion.

4.9 Tropopause height for real gas atmospheres

The radiative equilibrium solutions discussed in the preceding section are all unstable near the ground. As convection sets in, it will mix away the unstable layer and replace it by an adiabat; the well-mixed region is the troposphere. The change in lower level temperature profile, however, will alter the upward radiation which heats the stratosphere, and therefore cause temperature changes even above the layers reached directly by convection. When all this sorts itself out, how deep is the troposphere? This is the problem of tropopause height, which we have already touched on briefly for grey gases, and which we will consider in depth and generality in Chapter 7. Here we will offer a taste of a few of the most important aspects of real gas behavior, to tide the reader over until the more complete treatment to be introduced later. In this section we will assume, as in the all-troposphere model, that turbulent fluxes couple the ground so tightly to the overlying air that there is no discontinuity at the ground. This assumption will be relaxed in Chapters 6 and 7.

For a grey gas, the problem of finding the tropopause height is relatively simple. Since the radiative equilibrium profile depends only on OLR – and that only via a simple formula – one starts with the radiative equilibrium profile for the desired OLR, picks a guess for the tropopause pressure, and then replaces the temperature between there and the ground with the adiabat for the gas under consideration. One then computes the actual OLR for the resulting profile, and generally will find that it is generally somewhat different from the OLR assumed in computing the radiative equilibrium. To make the solution consistent, one then adjusts the tropopause height until the computed OLR including the troposphere is the same as the target OLR within some desired accuracy. This is a simple problem in root-finding for a function of a single variable (the tropopause pressure), and can be solved by any number of means, Newton's Method and bisection being among the most commonly employed.

For a real gas, the radiative equilibrium in the upper atmosphere depends on the spectrum of the infrared upwelling from below, so we no longer have the luxury of assuming that the stratospheric temperature profile remains fixed as we vary the estimate of the tropopause height. Instead, one must simultaneously solve for both the tropopause height and the corresponding equilibrium profile aloft. This is most easily done by a modification of the time-stepping method we employed to compute the pure radiative equilibrium solutions. As in that case, it is somewhat awkward to



Figure 4.44: Radiative-convective equilibrium for an Earthlike dry atmosphere with $300ppmv CO_2$. The left panel shows the temperature profile in comparison with pure radiative equilibrium and the dry adiabat. The right panel shows the radiative heating rates for the radiative-convective solution and for the all-troposphere model. In all cases the ground temperature is 280K.

pick an OLR and find the corresponding ground temperature T_g . Instead, we fix T_g , and compute the corresponding OLR. This can be done for a range of ground temperatures, whereafter the ground temperature in equilibrium with any specified solar absorption can be determined. We are back in the familiar business of computing the $OLR(T_g)$ curve, much as we did for the alltroposphere model, but this time taking into account the effect of a self-consistent stratosphere on the OLR.

The general problem of representing convection in climate models is a very challenging one, about which entire volumes have been written. For the problem at hand, there are a number of simplifying assumptions which allow us to avoid some of the more subtle aspects of the subject. First, we will be content to assume that convection instantaneously resets the profile to an adiabatic profile. Next, given where instability occurs in the pure infrared radiative equilibrium profiles, it is safe to assume that convection occurs in a single layer extending from the ground to the tropopause height, without any possibility of multiple interleaved internal convecting and radiative equilibrium layers. Further, we will only seek an equilibrium solution, without attempting to accurately represent the approach to equilibrium. Finally, we carry out the calculation by holding T_q fixed and allowing the rest of the atmosphere to relax to the corresponding equilibrium. Under these circumstances, the elimination of unstable layers by convective mixing can be carried out through the following simple modification to the pure radiative equilibrium time-stepping algorithm: One calculates the adiabat $T_{ad(p)}$ corresponding to the ground temperature T_q and surface pressure T_s . Then at each timestep, wherever $T(p) < T_{ad}(p)$, the temperature is instantaneously reset to T_{ad} . The rationale for doing this is that convection is a much faster process than radiative relaxation, and that wherever the temperature is below the adiabatic temperature, air parcels starting at the ground have enough buoyancy to reach that level, mixing air all along the way.

Results for an Earthlike air/CO_2 atmosphere are shown in Figure 4.44. The convection

reaches to 380mb, where the temperature is about 200K; above that, the atmosphere is in radiative equilibrium, which defines the stratosphere. Note that the temperature continues to decline even above the maximum height reached by convection, because the infrared radiative equilibrium profile also has decreasing temperature, so long as part of the spectrum is optically thick. This is the case also for Earth's real stratosphere, even though ozone heating eventually causes the upper stratospheric temperature to turn around and begin to increase. Thus, one shouldn't take temperature decline as a signature of convection, and in a case where atmospheric heating causes upper stratospheric temperature to increase the temperature minimum will generally be above the top of the convective layer.

The OLR for all three calculations is similar: $296.4 W/m^2$ for the radiative-convective model vs. $295 W/m^2$ for pure radiative equilibrium and $296.6 W/m^2$ for the all-troposphere model. Thus, if the atmosphere is maintained by $295 W/m^2$ of absorbed solar radiation, neither the formation of a troposphere by convection nor the formation of a stratosphere by upper level radiative heating has much effect on the surface temperature, though the effects on the atmospheric profile are considerable.

The heating rate profile shown in the right panel of Figure 4.44 sheds some light on the basic mechanism maintaining convection in the troposphere. The entire troposphere is subject to radiative cooling. This feature is guaranteed by the construction of the solution, since any positive heating would warm the atmosphere, causing it to exceed the adiabatic temperature and shutting off convection. Suppose convection has just occurred and reset the temperature to the adiabat. Then, in the next small interval of time, radiative cooling will cause the temperature to fall below the adiabat, triggering convection once more, which adds back the heat lost by radiative cooling and restores the adiabat. The heat is supplied by parcels of air that pick up heat from the ground, become buoyant, and carry the heat upward to the level where it is needed. The thermal balance in the troposphere is between radiative cooling and convective heating. The addition of atmospheric solar absorption to the troposphere would have no effect on the tropospheric temperature or the tropopause height, so long as it doesn't turn the net radiative cooling at any level into a net radiative heating. Short of that happening, the sole effect of tropospheric solar absorption is to reduce the convective heating, and hence the frequency or vigor of convection. However, since infrared cooling is weakest just below the tropopause, solar absorption near the tropopause level can easily move the tropopause downward.

Using Figure 4.44 we can compare two simple estimates of the tropopause height with the actual value. Looking at where the adiabat intersects the pure radiative equilibrium, we find an estimate of 205mb. This is somewhat too high in altitude, since the formation of the troposphere changes the upward flux and warms the stratosphere. The other way to estimate tropopause height is to look at the heating profile computed for the dry adiabat, shown in the right panel of the Figure. Identifying the region of radiative heating with the stratosphere yields an estimate of 325mb, which is closer to the true value, but still too high in altitude. In the light of the discussion surrounding Fig. fig:AllTropNetIRFluxGrey, we can say that the real gas atmosphere behaves rather like a pressure-broadened grey-gas atmosphere with optical depth somewhat greater than unity.

The tropopause height and temperature in this simple radiative-convective calculation are quite similar to the Earth's actual midlatitude values, despite the neglect of stratospheric ozone heating and tropospheric water vapor. Why is the tropical tropopause higher (near 100mb)? The main factor governing the tropopause height is the lapse rate. If the lapse rate is weaker, then one has to go to higher altitudes in order to intersect the radiative equilibrium profile. In the warm tropics, the moist adiabat has significantly weaker gradient than the dry adiabat. In fact, a calculation with the dry CO_2 /air atmosphere, but employing the moist adiabat in the temperature profile, yields a tropopause height of **mb. This suggests that the effects of moisture on lapse rate are more important than the radiative effects of tropospheric moisture in elevating the tropical tropopause. This calculation is carried out in Problem ??

4.10 The lesson learned

This has been a rather arduous chapter (certainly for the author, and no doubt for the reader as well, but hopefully to a lesser extent). The basic lesson, however, can be summed up in a few pithy remarks. The greenhouse effect relies on infrared optical thickness of the atmosphere and temperature decline with height. Real greenhouse gases do not make the atmosphere optically thick uniformly throughout the infrared spectrum. Rather, the optical thickness is concentrated in preferred gas-dependent spectral bands, and the main way the greenhouse effect gets stronger as the gas concentration increases is through the spread of the optically thick regions to every greater portions of the spectrum. There are two basic ways to get the temperature decline which is necessary to translate optical thickness into OLR reduction: radiative equilibrium in an atmosphere that is optically thick through at least part of the spectrum, or convection in an atmosphere where the radiative equilibrium is statically unstable either at the surface or internally. The tropopause height determines the blend of the two mechanisms in force. Both mechanisms can vield a surface temperature very much in excess of the no-atmosphere blackbody temperature. but a radiatively-dominated atmosphere is a very different place from a convectively-dominated atmosphere, since the latter has vigorous vertical mixing that can give rise to a stew of small scale turbulent phenomena. A mostly radiative-equilibrium atmosphere is a more quiescent place, in which mixing is dominated by the more ponderous large scale fluid motions. A central remaining question is how the tropopause height behaves as an atmosphere is made more optically thick. When do we approach an all-troposphere atmosphere, and when do we approach an all-stratosphere atmosphere? This question will be taken up in detail in Chapter 7.

4.11 For Further Reading

The HITRAN spectroscopic database is described in

 Rothman LS, et al. 2005: The HITRAN 2004 molecular spectroscopic database, J. Quant. Spectroscopy and Radiative Transfer, 96, pp 139-204.

In a book that one hopes will stick around for a while, there is always some risk in referring to specific means of obtaining digital data. The HITRAN database is so valuable, however, that it is sure to be available in some form more or less indefinitely. At the time of writing, the HITRAN data can be obtained over the Internet at the URL http://cfa-www.harvard.edu/hitran. The 1970's era *Goody and Yung* book on atmospheric radiation refers to obtaining the data on "AFGL tapes," an no doubt earlier books made reference to things like "punch cards" or "paper tapes." No doubt, our reference to the "Internet" will seem similarly quaint within a few years.

The HITRAN database does not include the very weak CO_2 absorption lines that become important for extremely massive atmospheres such as that of Venus, and moreover, the temperature dependence data of the lines that are included becomes somewhat inaccurate at Venusian temperatures. There are two databases that extend the CO_2 absorption database to cover the Venusian regime, both of which use the same data format as HITRAN. The first is the HITEMP database. At

244CHAPTER 4. RADIATIVE TRANSFER IN TEMPERATURE-STRATIFIED ATMOSPHERES

the time of writing, there is neither a convenient published document describing the database nor a generally accessible download site, but an updated version of the HITEMP database and expanded documentation are expected to be made available through the HITRAN site in the near future. In the meantime, information about the existing database can be found in

• Rothman LS, et al. 1995: HITRAN, HAWKS, and HITEMP High-Temperature Molecular Database, Proc.Soc.Photo-Optical Instrumentation Engineers 2471 105-111.

and the original version of the database can be downloaded by contacting the managers of the HITRAN site. A similar high-temperature, high-pressure database is described in

• Tashkun SA, et al. 2003: CDSD-1000, the high-temperature carbon dioxide spectroscopic databank, J. Quantitative Spectroscopy and Radiative Transfer, 82, pp 165-196.

It is available online via ftp.iao.ru/pub/CDSD-1000.

Information on the CO_2 collision-induced continuum is very sparse. The modeling of the CO_2 continuum used throughout this book (and incorporated in the software supplement) is based on a polynomial fit to absorption coefficients described in

• Kasting JF, Pollack JB and Crisp D 1984: Effects of high C02 levels on the surface temperature and atmospheric oxidation state of the early Earth, J. Atmos. Chem, 1, pp 403-428.

References to the laboratory measurements upon which the parameterization is based amount to one published paper, one NASA technical report, and one unpublished personal communication; these may be found in the above referenced article. Some theoretical developments, which have been incorporated in a few of the more recent representations of the far-infrared continuum, are described in

• Gruszka M and Borysow A 1997: Far Infrared Collision-Induced Absorption of CO2 for the Atmosphere of Venus at Temperatures from 200K to 800K, *Icarus*, **129**, pp 172-177.

but there seem to have been no new laboratory measurements since those discussed in the former paper.

The collision induced continua of H_2 , CH_4 and N_2 relevant to the atmosphere of Titan are given in

• Courtin R, 1988: Pressure-Induced Absorption Coefficients for Radiative Transfer Calculations in Titan's Atmosphere, *Icarus*, **75**, pp 245-254.

The water vapor continuum is described in the following two papers:

- Clough SA, Kneizys FX and Davies RW ,1989: Line shape and the water vapor continuum, *Atmospheric Research*, **23**, pp 229-241.
- Grant WB,1990: Water vapor absorption coefficients in the 8-12 μ m spectral region: A critical review, Applied Optics, **29**, pp 451-462.

The first of these is considered the standard reference at time of writing, but one must take care in reading it, as there are a certain number of typographical errors and mislabeled figures.

4.11. FOR FURTHER READING

The full-featured ccm radiation code is described in complete and somewhat intimidating detail as part of the general description of the NCAR Community Atmospheric Model (CAM) in NCAR Technical Note TN-464+STR, available at the time of writing at http://www.ccsm.ucar.edu/models/atm-cam An accessible overview of an earlier version of the radiation model can be found in

• Kiehl J and Briegleb B 1992: Comparison of the Observed and Calculated Clear Sky Greenhouse-Effect - Implications for Climate Studies, J. Geophys. Res, 97 (D9), 10037-10049.

A version of this radiation code with a simple Python user interface is distributed as part of the software supplement to this book. The Python interface makes it easy to use the code to compute OLR and heating rates within a Python script, and eliminates the need for any familiarity with the FORTRAN language in which the underlying computation is written.
$246 CHAPTER \ 4. \ RADIATIVE \ TRANSFER \ IN \ TEMPERATURE-STRATIFIED \ ATMOSPHERES$

Chapter 5

Scattering

5.1 Overview

In the atmospheres considered so far, the blackbody source term adds new radiation to the atmosphere travelling in all directions, but once present in the atmosphere radiation travels in a fixed direction; it can be absorbed as it travels, but it does not scatter into other directions. In this class of problems, the two-stream approximation consists entirely in doing the calculation for a single equivalent propagation angle $\bar{\theta}$, but does not change the essential structure of the full problem. If one wanted more information about the angular distribution, one would simply do the same problem over several times, with different θ , and average the results to get the net upward and downward fluxes; each calculation is independent, and has the form of a simple first order ordinary differential equation for the flux. With scattering the situation is very different, as the scattering couples the flux at one angle with the fluxes at all other angles. The full problem now takes the form of a computationally demanding integro-differential equation, with the derivative of the flux at a given angle expressed as a weighted integral over the fluxes at all other angles.

Light with wavelengths in the near-infrared or shorter is significantly scattered from molecules, though molecules are too small to appreciably scatter thermal infrared or longer wavelengths. Many atmospheres (Earth's included) contain very fine aerosol particles with diameters on the order of a micrometer or less; they are typically made of mineral dust, or of condensed substances such as sulfuric acid or other sulfur compounds. They are very powerful scatterers of solar radiation, and therefore can significantly affect a planets's albedo even when the total mass of aerosols is quite small. Cloud particles made of various condensed substances have typical diameters of 10-100 micrometers. Because water clouds like Earth's absorb so strongly in the infrared, cloud scattering is often thought of primarily in terms of the solar spectrum. However, taking a broader view, cloud substances commonly found on other planets have a very important thermal infrared scattering effect. Water clouds are the exception, rather than the rule, but because of their importance on Earth, thermal infrared scattering by clouds is a far less developed subject than is shortwave scattering.

Clouds, in their many and varied manifestations, pose one of the greatest challenges to the understanding of Earth and planetary climate. On Earth, water clouds reflect a great deal of sunlight but also have a considerable greenhouse effect. The net cloud effect is a fairly small residual of two large and uncertain terms, and the way the two effects play out against each other plays a central role in climate change problems extending from the Early Earth to Cretaceous Warmth, to ice ages, to global warming, and the distant-future fate of our climate. The high albedo of Venus is caused largely by clouds made of sulfur dioxide and sulfuric acid droplets, but the very same clouds are effective infrared scatterers and help to increase the planet's greenhouse effect. On an Early Mars with a 2 bar CO_2 atmosphere, formation of clouds of CO_2 ice would play an important role in the planet's climate, both in the infrared nd solar spectrum. Titan's present-day methane clouds affect the satellite's radiation budget both through infrared and visible scattering; Neptune is cold enough that methane ice clouds can form in parts of its atmosphere. The swirling psychedelic colors of Jupiter and Saturn arise from clouds of a dozen or more different types, which no doubt also affect the radiation budget. It will turn out that the radiative effects of clouds are highly sensitive to the size of the particles of which they are composed. This leads to the disconcerting conclusion that the climate of an object as large as an entire planet can be strongly affected by poorly-understood processes happening on the scale of a few micrometers.

Scattering calculations play a critical role not only in determining planetary radiation balance, but also in interpreting a wide range of observations of the Earth, Solar System planets, and extrasolar planets. There is no doubt that if justice were to be served (and the reader had unlimited time) scattering should be a treatment at least as in-depth as that which we have accorded to purely absorbing/emitting atmospheres. However, in order to maintain progress towards our primary goal of understanding the essentials of planetary climate, the treatment given in this chapter will be highly abbreviated, and focus on the minimal understanding of the subject needed to estimate planetary albedo, shortwave atmospheric heating, and the basic effect of clouds on outgoing infrared radiation and on solar absorption. In particular, we will leap directly into the two-stream approximation, without much discussion of the properties of the scattering equations in their full generality. Were it not for the position of clouds at the forefront of much research on planetary climate, we would be content to leave the discussion of scattering to a few brief remarks concerning planetary albedo.

5.2 Basic concepts

The atmosphere can be considered to be a mix of particles, some of which absorb, some of which scatter, and some of which do both. The particles could be molecules, or they could be macroscopic particles of a condensed substance, as in the case of cloud droplets or dust particles. One builds up the absorbing and scattering properties of the atmosphere as a whole from the absorbing and scattering properties of the individual particles. In keeping with the usage in the preceding chapters, we will ultimately characterize the effect of atmospheric composition on radiation in terms of scattering properties per unit mass of atmosphere, just as we did for the absorption coefficient.

Consider a parallel, monochromatic (single-frequency) beam of light with flux F in W/m^2 travelling in some specific direction. When the beam encounters a particle of finite extent, a certain amount of the flux will be absorbed, and a certain amount will be scattered into other angles. The rate at which energy is taken out of the beam by absorption and scattering can be characterized in terms of coefficients with dimensions of area, which are known as *cross-sections*. The rate of energy absorption is $F\chi_{abs}$, where χ_{abs} is the *absorption cross-section*, and the rate of energy scattering into other directions is $F\chi_{sca}$, where χ_{sca} is the *scattering cross-section*¹. The scattering and absorption cross-sections can be quite different from the actual cross-section area of the object. The cross-sections can be thought of as the cross-section areas of hypothetical equivalent objects

¹The more usual notation for the cross-section is σ , but in our subject matter that symbol has been reserved for the Stefan-Boltzman constant. One can think of χ as standing for $\chi \rho \omega \sigma \sigma$ -section.



Figure 5.1: Definition of propagation angles (left) and scattering angle (right).

which absorb or scatter all light hitting the object while leaving the rest of the beam to pass by undisturbed. The ratio of scattering cross-section to the actual cross-section area of the scatterer is called the scattering efficiency, Q_{sca} . For a spherical particle of radius r, $Q_{sca} = \chi_{sca}/(\pi r^2)$. The absorption efficiency is defined similarly. For spherical particles, the cross sections are independent of the angle at which the radiation is directed at the particle. For non-spherical particles the crosssections for an individual particle depend on angle, but the typical physical situation involves scattering off of an ensemble of particles presented with random orientations. In this case, we can average over all orientations and represent the mean scattering or absorption in terms of the cross-section for an equivalent sphere. This approach can break down if particles are not randomly oriented, as can be the case for plate-like ice crystals that become oriented through frictional drag forces as they fall. The single-scattering albedo for a particle is the ratio of flux of the incident beam lost via scattering to net flux lost. Using the notation ω_{po} for the single-scattering albedo of an individual particle, we have $\omega_{po} = \chi_{sca}/(\chi_{sca} + \chi_{abs})$. Later we will introduce the single-scattering albedo for the medium as a whole. The cross sections for particles or molecules can be measured in the laboratory, and often can be computed from basic physical principles.

Since the radiation fields we will deal with are generally distributed over a range of frequencies and direction, instead of being monochromatic and unidirectional, we will write our equations in terms of the spectral irradiance I introduced in Chapters 3 and 4. Recall that if the spectral irradiance is $I((\theta, \phi), \nu)$ at a given point, then $Id\Omega d\nu$ is the flux of radiation in frequency band $d\nu$ with directions of travel within a solid angle $d\Omega$ about the direction (θ, ϕ) , which passes through a plane perpendicular to the direction of travel. To apply the results of the preceding paragraph to smoothly distributed radiation, one needs only to substitute $Id\Omega d\nu$ for the incident flux F.

As in previous chapters, we'll make the plane-parallel assumption, and assume that I depends on position only through pressure. Suppose that in the vicinity of some pressure level p there are N scatterers of type i per unit mass of atmosphere, and that each scatterer has mass m_i . Suppose that the light impinging on the layer is traveling with angle θ to the vertical. Then, taking a layer of thickness dp which is small enough that multiple scattering can be neglected, the rate of energy lost by the incident beam due to absorption and due to scattering into different angles is

$$-\frac{dp}{g\cos\theta}N\cdot(\chi_{abs,i}+\chi_{sca,i})Id\Omega d\nu = -\frac{dp}{g\cos\theta}q_i\cdot(\frac{1}{m_i}\chi_{abs,i}+\frac{1}{m_i}\chi_{sca,i})Id\Omega d\nu$$
(5.1)

where q_i is the mass concentration of the particles in question. From this we can define the absorption coefficient of the substance $\kappa_i \equiv \chi_{abs,i}/m_i$ which has units of m^2/kg . This absorption coefficient is the same quantity we defined in Chapter 4 in connection with gaseous absorption. The additional term in the above equation characterizes the energy lost from the incident beam due to scattering. We won't introduce separate notation for this term since scattering is most commonly characterized in terms of the cross section itself.

If there is only one optically active substance i in the atmosphere, we define the optical depth in the vertical direction by the equation

$$\frac{d\tau^*}{dp} = -\frac{1}{g}(\kappa_i + \frac{1}{m_i}\chi_{sca,i})q_i \tag{5.2}$$

Because the absorbing and scattering properties typically depend on wavenumber, the optical depth is generally a function of wavenumber, though we will only append a wavenumber subscript to τ^* when we wish to call attention specifically to the wavenumber dependence. If there are many types of scatterers and absorbers – which could include particles of a single substance but with different sizes – then we define the optical depth by summing over all species. Thus

$$\frac{d\tau^*}{dp} = -\frac{1}{g} (\kappa + \sum_i \frac{1}{m_i} \chi_{sca,i} q_i)$$
(5.3)

where the net absorption coefficient is

$$\kappa \equiv \sum_{i} \kappa_{i} q_{i} \tag{5.4}$$

We then define the single-scattering albedo for the medium as a whole as

$$\omega_o \equiv \frac{\sum q_i \chi_{sca,i}/m_i}{\kappa + \sum q_i \chi_{sca,i}/m_i} \tag{5.5}$$

The pair (κ, ω_o) constitutes the basic description of the absorption and scattering properties of the medium. Both are typically functions of wavelength and altitude, and may also directly be functions of pressure and temperature. If the medium consists of only a single type of particle, and the gas in which the particles are suspended neither absorbs nor scatters, then $\omega_o = \omega_{po}$. In general, though, the single-scattering albedo of the medium depends on the mix of absorbers and scatterers. For example, an atmosphere may consist of a mix of cloud particles which are perfect scatterers ($\omega_{po} = 1$) with a strong greenhouse gas which is an absorber. In this case, ω_o will go down as the greenhouse gas concentration increases, even if the cloud particle concentration is kept fixed.

Using the definition of optical depth, Eq 5.1 for the rate of energy loss from the beam can be rewritten as simply $dI = -Id\tau^*/\cos\theta$. Since the vertical component of flux is $I\cos\theta$, this expression can be recast as an expression for the rate of loss of vertical flux, namely

$$dI\cos\theta = -Id\tau^* \tag{5.6}$$

The proportion of this lost to scattering is ω_o while the proportion lost to absorption is $1 - \omega_o$. The fate of the energy lost to absorption is different from the fate of that lost due to scattering. The former disappears into the pool of atmospheric heat, whereas energy lost to scattering from one beam reappears as flux in a range of other directions, so we need to keep track of the two loss mechanisms separately. The beam loss in a given direction is offset by two source terms: one due to thermal emission, and one due to scattering from other directions. The thermal emission term is

5.2. BASIC CONCEPTS

proportional to the Planck function, and can be treated in a fashion similar to that used in deriving the Schwartzschild equations. We'll leave the thermal emission out for now, and concentrate on scattering; the thermal emission term will be put back in in Section 5.5.

To understand better where the scattered flux goes, consider the energy budget for a box of thickness $d\tau^*$ in the vertical, shown from the side in Figure 5.2. Since the radiation field is independent of the horizontal dimensions, the flux entering the box from the side is the same as the flux leaving it from the side, and does not affect the budget. If the base of the box has area A, an amount $A \cdot I(\tau^*) \cos \theta$ enters the box from the bottom and a somewhat lesser amount $A \cdot I(\theta, \phi, \tau^* + d\tau^*) \cos \theta$ leaves the box from the top. Taking the difference gives the loss of energy from the beam per unit time, due to scattering and absorption. Using Eq. 5.6 this can be written as simply $A \cdot I(\theta, \phi, \tau^*) d\tau^*$; it doesn't matter whether I is evaluated at $d\tau^*$ or $\tau^* + d\tau^*$ in this expression, since $d\tau^*$ is presumed small. The energy per unit time scattered and restributed into all other directions is then $A \cdot \omega_0 I(\theta, \phi, \tau^*) d\tau^*$. Now, to write an equation for how the vertical component of flux changes between τ^* and $\tau^* + d\tau^*$, we need to find how much flux is added to the direction (θ, ϕ) by scattering from all other directions of radiation impinging on the layer. We can do this by considering the incident radiation one direction at a time, and summing up. Consider a beam of light traveling with direction (θ', ϕ') , having radiance $I(\theta', \phi', \tau^*)$. The scattering contributed to direction (θ, ϕ) comes from the scatterers in the shaded parallelogram shown in FIgure 5.2, which is greater than the amount of scatterer in a rectangular box by a factor of $1/\cos\theta$. Further, only a proportion of the radiation scattered from the contents of the parallelogram goes into the direction (θ, ϕ) . We will write this proportion as $P/4\pi$, where P depends on both the incident and scattered directions. Thus, the radianced contributed to direction (θ, ϕ) by scattering is $A \cdot (P/4\pi)\omega_0 I(\theta',\phi',\tau^*)d\tau^*/\cos\theta$, and the vertical component of this is obtained by multiplying by $\cos \theta$, yielding $A \cdot (P/4\pi) \omega_0 I(\theta', \phi', \tau^*) d\tau^*$. This is the vertical flux contributed by scattering, and is added in to the flux leaving the top of the box. The scattering acts as a source of radiation in direction (θ, ϕ) , which is added to the right hand side of Eq. 5.6. Dividing out the area of the base of the box, the flux balance for the box becomes

$$dI(\theta,\phi)\cos\theta = -I(\theta,\phi)d\tau^* + \frac{\omega_o}{4\pi}P(\theta,\phi,\theta',\phi')I(\theta',\phi')d\tau^*$$
(5.7)

if one considers only the flux contributed by scattering of a single direction (θ', ϕ') . To complete the equation, one must integrate over all incident angles (θ', ϕ') . To determine the radiation field in its full generality, it is necessary to satisfy the flux balance for each direction of propagation simultaneously. Before proceeding toward that goal, we'll check Eq. 5.7 to verify that the scattered energy is conserved. Applying the control volume sketch to the *incident* beam direction, we infer that the incident beam traveling in direction (θ', ϕ') deposits energy in the control volume at a rate $I(\theta', \phi')d\tau^*$ (per unit area). A proportion $\omega_o P/4\pi$ of this should show up as an increase in the energy in the box propagating in direction (θ, ϕ) , and that is precisely the source term appearing in Eq 5.7 The books are indeed balanced.

It is worth thinking quite hard about Figure 5.2, because the cosine terms that appear in such computations – and which are the source of most of the difficulties in writing two-stream approximations – can be quite confusing. The cosine weights play two quite different roles. In one guise, they express the number of scatterers or absorbers encountered along a slanted path, but in another guise they represent the projection of the flux on the vertical direction. Most confusion can be resolved by thinking hard about the energy budget of the control volume.

The quantity P introduced in Figure 5.2 is called the *phase function*, and describes how the scattered radiation is distributed over directions. For spherically symmetric scatterers, the phase function depends only on the angle Θ between the incident beam and a scattered beam (as depicted in Fig. 5.1). The phase function is usually expressed as a function of $\cos \Theta$. If \hat{n}_I is the



Figure 5.2: **CAPTION

unit vector in the direction of propagation of the incident beam, and \hat{n}_{sca} is the unit vector in the direction of propagation of some scattered radiation, then

$$\cos\Theta = \hat{n}_I \cdot \hat{n}_{sca} = \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\phi - \phi')$$
(5.8)

where θ and ϕ are the direction angles of the incident beam and θ' and ϕ' are the angles of the scattered beam under consideration. The phase function for the medium as a whole can be determined from the phase functions of the individual particles doing the scattering – remember that from ω_o and $d\tau^*$ we already know the amount of energy scattered out of a beam, so the phase function only needs to tell us how that energy is distributed amongst directions. The phase function for an individual particle is defined in such a way that the scattered flux within an element of solid angle $d\Omega'$ near direction (θ', ϕ') is $\chi_{sca}I(\theta, \phi)P(\cos\Theta(\theta, \phi, \theta', \phi'))d\Omega'/4\pi$. P is normalized such that $\int Pd\Omega' = 4\pi$, so that integrating the scattered flux over all solid angles yields $\chi_{sca}I$. Note further that

$$\int P(\cos\Theta)d\Omega = 2\pi \int_{-1}^{1} P(\cos\Theta)d\cos\Theta = \int P(\cos\Theta)d\Omega' = 4\pi$$
(5.9)

where solid angle integrals without limits specified explicitly denote integration over the entire sphere. The final equality is a matter of definition and the other two equalities follow because one is free to rotate the coordinate system so as to define the angles with respect to any chosen axis, if one is integrating over the entire sphere. Isotropic scattering, in which the scattered radiation is distributed uniformly over all angles, is defined by P = 1.

If the scatterers in the atmosphere are all identical particles, then the phase function for the medium is the same as the phase function for an individual particle. If the phase functions differ from one particle to another, then the phase function for the medium is simply the average of the individual particle phase functions, weighted compatibly with Eq 5.3. The averaging is particularly important when the particles are non-spherical. Though the phase function for any individual particle is not a function of $\cos \Theta$ alone, the particles are generally oriented in random directions, and the average phase function for an ensemble of randomly oriented particles acts like the phase function for an equivalent sphere.

If one divides Eq 5.7 by $d\tau^*$ and integrates over all incident directions (θ', ϕ') the equation for the vertical component of the flux due to radiation traveling in direction (θ, ϕ) is found to be

$$\frac{d}{d\tau^*}I(\cos\theta,\phi)\cos\theta = -I(\cos\theta,\phi) + \frac{\omega_o}{4\pi}\int P(\cos\Theta)I(\cos\theta',\phi')d\Omega'$$
(5.10)

where $\cos \Theta$ is given in terms of $(\theta, \theta', \phi - \phi')$ by Eq. 5.8. Thermal emission would add an additional source term $B(\nu, T(\tau^*))$ to the right hand side, but we shall leave that out for now. This is the full equation whose solutions give the radiation field. The integral couples together all directions of propagation; if one approximated the integral by a sum over 100 angles, for example, the equation would be the equivalent of solving a system of 100 coupled ordinary differential equations. While, with modern computers, this is not so overwhelming a task as it once might have seemed, it is still intractible in typical climate calculations, where one is doing the calculation for each of a large array of wavenumbers, at each time step of a radiative-convective model, and perhaps for each latitude and longitude grid point in a general circulation model as well. Moreover, it is always helpful to have a simplified form in hand if one's goal is understanding and not merely computing a number. Hence, our emphasis will be on reduction of the equation to an approximate set of equations for two streams of radiation, which may be thought of as the upward and downward streams. In this section we'll derive some exact constraints, which will be used to obtain two-stream closures of the problem in Section 5.5. We first need to define the upward and downward fluxes, which are

$$I_{+} \equiv \int_{\Omega^{+}} I(\cos\theta, \phi) \cos\theta d\Omega = \int_{\cos\theta=0}^{1} \int_{0}^{2\pi} I(\cos\theta, \phi) \cos\theta d\phi d\cos\theta$$
$$I_{-} \equiv -\int_{\Omega^{-}} I(\cos\theta, \phi) \cos\theta d\Omega = -\int_{\cos\theta=-1}^{0} \int_{0}^{2\pi} I(\cos\theta, \phi) \cos\theta d\phi d\cos\theta$$
(5.11)

The fluxes are defined in such a way that both are positive numbers. Given that $d\Omega$ can be written as $d\cos\theta \cdot d\phi$ it is convenient to write all the fluxes as a function of $\cos\theta$, as we have done here. Henceforth we shall use Ω^+ and Ω^- as shorthand for integral over the upward or downward hemisphere, respectively. With these definitions, the net vertical flux (positive upward) is $I_+ - I_- = \int I \cos\theta d\Omega$, the integral being taken over the full sphere.

Solar radiation enters the top of the atmosphere in the form of a nearly parallel beam of radiation, characterized by an essentially unique angle of propagation. It is gradually converted by scattering into radiation that is continuously distributed over angles. Because the incoming solar radiation has an angular distribution concentrated on a single direction of propagation, it is useful to divide the radiation up into a *direct beam* component propagating exactly in this direction, and a *diffuse* component, travelling over all angles. You can see the Sun as a sharply defined disk in clear sky, which shows that the direct-beam solar radiation isn't completely converted into diffuse radiation by scattering, except perhaps in heavily cloudy conditions. To define the direct beam flux, let L_{\odot} be the solar constant and ζ be the angle between the vertical and the line pointing toward the Sun; ζ is called the *zenith angle*. By convention, the zenith angle is defined as the angle of the vector pointing *toward* the Sun, rather than the direct of the rays coming *from* the Sun. Thus, if θ_{dir} is the angle of the direct-beam radiation in our usual angle coordinate system, the zenith angle is $\zeta = \pi - \theta_{dir}$. The azimuth angle of the direct beam radiation ϕ_{dir} is defined in the usual coordinate system.

Now, since the direct beam flux is concentrated in a single direction, there is essentially zero probability of any scattered flux contributing back into the exact direct beam direction. That would be like the exactly hitting an infinitesimal dot on a dartboard. Therefore, flux is scattered out of the direct beam but is never added into it, and the direct beam decays exponentially. Making use of the slant path, the direct beam flux is then $L_{\odot} \exp(-(\tau_{\infty}^* - \tau^*)/\cos \zeta)$. We rewrite the flux as the sum of a diffuse component and the direct beam:

$$I(\cos\theta,\phi) = I_{diff}(\cos\theta,\phi) + L_{\odot}\exp(-(\tau_{\infty}^* - \tau^*)/\cos\zeta)\delta(\theta - (\pi - \zeta))\delta(\phi - \phi_{dir})$$
(5.12)

where I_{diff} is the diffuse flux and δ is the Dirac delta-function. From now on, for economy of notation we'll drop the "diff" subscript on the diffuse radiation and simply write I for the diffuse component. In typical situations, the top-of-atmosphere boundary condition states that the radiance of all downward-directed angles of the diffuse component must vanish.

Substituting into Eq 5.10, the equation for the diffuse flux becomes

$$\frac{d}{d\tau^*} I(\cos\theta, \phi) \cos\theta = -I(\cos\theta, \phi) + \frac{\omega_o}{4\pi} \int P(\cos\Theta) I(\cos\theta', \phi') d\Omega' + L_{\odot} \frac{\omega_o}{4\pi} P(\cos\Theta(-\cos\zeta, \cos\theta, \phi - \phi_{dir}) \exp(-(\tau_{\infty}^* - \tau^*)/\cos\zeta)$$
(5.13)

The scattering from the direct beam acts as a source term for the diffuse radiation. Integrating over all angles yields the following exact expression for the net vertical diffuse flux

$$\frac{d}{d\tau^*}(I_+ - I_-) = -(1 - \omega_o) \int I(\cos\theta', \phi') d\Omega' + \omega_o L_\odot \exp\left(-(\tau^*_\infty - \tau^*)/\cos\zeta\right)$$
(5.14)

since $\int P(\cos \Theta) d\Omega = 4\pi$. In this expression, I_+ and I_- now represent just the diffuse part of the flux. Conservative scattering – that is, scattering without absorption – is defined by $\omega_o = 1$. For conservative scattering the first term on the right hand side of Eq. 5.14 vanishes. Integrating the direct beam term with respect to τ^* just multiplies it by $\cos \zeta$, whence we are left with the result that $I_+ - I_- - L_{\odot} \cos \zeta \exp(-(\tau_{\infty}^* - \tau^*)/\cos \zeta)$ is a constant. Thus, for conservative scattering the sum of the direct beam vertical flux – which is negative because it is downward – with the diffuse flux is independent of height. As the direct beam is depleted, the flux lost goes completely into the diffuse component. This is as it should be, because, in conservative scattering, the flux lost has no place else to go.

Eq. 5.14 provides the first of the two constraints needed to derive the two-stream approximations. The second constraint is provided by multiplying Eq 5.13 by a function $H(\cos \theta)$ which is antisymmetric between the upward and downward hemispheres, and then performing the angle integral. The rationale for multiplying by an antisymmetric function is that we already know something about $I_+ - I_-$ from the first constraint, and weighting by an antisymmetric functions gives us some information about $I_+ + I_-$. Multiplying by H and carrying out the angle integral, we get

$$\frac{d}{d\tau^*} \int I(\cos\theta, \phi) H(\cos\theta) \cos\theta d\Omega = -\int IH(\cos\theta) d\Omega + \omega_o \int G(\cos\theta') I(\cos\theta', \phi') d\Omega' + \omega_o L_{\odot}G(-\cos\zeta) \exp\left(-(\tau_{\infty}^* - \tau^*)/\cos\zeta\right)$$
(5.15)

where

$$G(\cos\theta') = \frac{1}{4\pi} \int H(\cos\theta) P(\cos\Theta(\cos\theta,\cos\theta',\cos\phi)) d\Omega$$
(5.16)

We were free to replace $\cos(\phi - \phi')$ in this expression by $\cos \phi$, since the integral is taken over all angles ϕ and so a constant shift of azimuth angle does not change the value of the integral. Sinced *H* is assumed antisymmetric, the function $G(\cos \theta')$ characterizes the up-down asymmetry of scattering of a beam coming in with angle θ' . The symmetry properties of $\cos \Theta$ imply that $G(-\cos \theta') = -G(\cos \theta')$.

Exercise 5.2.1 Derive the claimed antisymmetry property of G.

If the phase function satisfies $P(\cos \Theta) = P(-\cos \Theta)$ the scattering is said to be symmetric. For symmetric scatterers, there is no difference between scattering in the forward and backward directions. From Eq. 5.8 it follows that $\cos \Theta(\cos \theta, \cos \theta', \phi) = -\cos \Theta(-\cos \theta, \cos \theta', \phi + \pi)$. The antisymmetry of $H(\cos \theta)$ then implies that G vanishes if P is symmetric, since the contribution to the integral from $(\cos \theta, \phi)$ cancels the contribution from $(-\cos \theta, phi + \pi)$. For symmetric scattering, Eq. 5.15 takes on a particularly simple form, since both terms proportional to ω_o vanish. The physical content of this result is that symmetric scattering does not directly affect the asymmetric component of the diffuse radiation, since equal amounts are scattered into the upward and downward directions.

When the scattering isn't symmetric, the terms involving G do not vanish, and we need a way to characterize the asymmetry of the phase function. The most common measure of asymmetry is the cosine-weighted average of the phase function

$$\tilde{g} \equiv \frac{1}{2} \int_{\cos\Theta = -1}^{1} P(\cos\Theta) \cos\Theta d\cos\Theta$$
(5.17)

which goes simply by the name of the asymmetry factor. The asymmetry factor vanishes for symmetric scattering. All radiation is backscattered in the limit $\tilde{g} = -1$, as if the scattering

particles were little mirrors. When $\tilde{g} = 1$ there is no back-scatter at all, and all rays continue in the forward direction, though their direction of travel is altered by the particles, much as if they were little lenses.

The asymmetry factor \tilde{g} characterized forward-backward scattering asymmetry relative to the direction of travel of the incident beam, but some tedious manipulations with Eq. 5.8 allow one to show that the same factor characterizes cosine-weighted asymmetry in the upward-downward direction, regardless of the direction of the incident beam. Specifically, if the incident beam has direction (ϕ', θ'), then

$$\frac{1}{4\pi} \int P(\cos\Theta(\cos\theta,\cos\theta',\cos\phi))\cos\theta d\Omega = \tilde{g}\cos\theta'$$
(5.18)

where $d\Omega = d\phi \cdot d\cos\theta$ as usual. This leads to a particularly tidy result if we choose $H(\cos\theta) = \cos\theta$ in Eq. 5.15, since then $G(\cos\theta') = \tilde{g}\cos\theta'$ and the antisymmetric projection of the scattering equation becomes

$$\frac{d}{d\tau^*} \int I(\cos\theta, \phi) \cos^2\theta d\Omega = -(1 - \omega_o \tilde{g})(I_+ - I_-) + \omega_o L_{\odot} \tilde{g} \cos\zeta \exp\left(-(\tau_{\infty}^* - \tau^*)/\cos\zeta\right) \quad (5.19)$$

The integral appearing on the left hand side is sensitive only to the symmetric component of the radiance field. In order to obtain a two-stream closure, it is necessary to express the integral in terms of $I_+ + I_-$, which requires making an assumption about the angular distribution of the radiation. The same assumption applied to the right hand side of Eq. 5.14 allows one to estimate $\int I d\Omega$ in terms of $I_+ + I_-$. The different forms of two-stream approximations we shall encounter correspond to different assumptions about the angular distribution of radiance.

For other forms of H the asymmetry function $G(\cos \theta')$ has more complicated behavior that is not so simple to characterize. The other form of H we shall have occasion to deal with is

$$H(\cos\theta) = \begin{cases} 1 & \text{for } \cos\theta \ge 0, \\ -1 & \text{for } \cos\theta < 0, \end{cases}$$
(5.20)

which is used to derive the hemispherically-isotropic form of the two-stream equations. This choice is convenient because the left hand side of Eq. 5.15 reduces to the derivative of $I_+ + I_-$, but it is inconvenient because G no longer has a simple cosinusoidal dependence on the incident angle. One could simply compute G from the phase function for the medium and use this to form the weights in the scattering equation, but given the inaccuracies we already accept in reducing the problem to two streams, it is hardly worth the effort. Instead, we will *approximate* G as having a cosinusoidal dependance as it does in the previous case. This approximation is exact if the phase function has the form $P = 1 - \frac{1}{3}b + a\cos\Theta + b\cos^2\Theta$, and one can add a third and fourth order term without very seriously compromising the representation. By carrying out the integral defining the asymmetry factor, we find that $\tilde{g} = \frac{1}{3}a$. Then, evaluating G for the assumed form of phase function we find $G = \frac{1}{2}a\cos\theta' = \frac{3}{2}\tilde{g}\cos\theta'$. With this result the antisymmetric scattering equation projection becomes

$$\frac{d}{d\tau^*}(I_+ + I_-) = -(1 - \omega_o \frac{3}{2}\tilde{g})(I_+ - I_-) + \omega_o L_{\odot} \frac{3}{2}\tilde{g}\cos\zeta\exp\left(-(\tau_{\infty}^* - \tau^*)/\cos\zeta\right)$$
(5.21)

The right hand side becomes precisely the same as Eq. 5.19 if we redefine the asymmetry factor to be $\frac{3}{2}\tilde{g}$. Eq. 5.21 is already written in terms of the upward and downward stream, and needs no further approximation in order to be used to derive a two-stream approximation. To complete the derivation of the hemispherically isotropic two-stream equation, one need only write the integral $\int Id\Omega$ appearing in Eq. 5.14 in terms of $I_+ + I_-$ using the assumed angular distribution of radiance. If I is assumed hemispherically isotropic in forward and backward directions separately, then this integral is in fact $2(I_+ + I_-)$, which completes the closure of the problem.

There is one last basic quantity we need to define, namely the *index of refraction*, which characterizes the effect of a medium on the propagation of electromagnetic radiation. It will turn out that the index of refraction amounts to an alternate way of representing the information already present in the scattering and absorption cross-sections. For a broad class of materials – including all that are of significance in planetary climate – the propagation of electromagnetic radiation in the material is described by equations that are identical to Maxwell's electromagnetic equations, save for a change in the constant that determines the speed of propagation (the "speed of light"). In particular, the equations remain linear, so that the superposition of any two solutions to the wave equations is also a solution, allowing complicated solutions to be built up from solutions of more elementary form. The reduction in speed of light in a medium comes about because the electric field of an imposed wave induces a dipole moment in the molecules making up the material, which in turn gives rise to an electric field which modifies that of the imposed wave. The equations remain linear because the induced dipole moment for non-exotic materials is simply proportional to the imposed electric field. When the medium is nonabsorbing, the ratio of the speed in a vacuum to the speed in the medium is a real number known as the *index of refraction*.

The physical import of the index of refraction is that, at a discontinuity in the index such as occurs at the surface of a cloud particle suspended in an atmosphere, the jump in the propagation speed leads to partial reflection of light hitting the interface, and deflection (refraction) of the transmitted light relative to the original direction of travel. The larger the jump in the index of refraction, the larger is the reflection and refraction. To a considerable extent, the refraction of light upon hitting an interface can be understood in terms of a particle viewpoint. If one represents a parallel beam of light as a set of parallel streams of particles all moving at speed c_1 in the outer medium, then if the streams hit an interface with a medium where the speed it $c_2 < c_1$, then the streams that hit first will be slowed down first, meaning that the wave front will tilt and the direction of propagation of the beam will be deflected toward the normal, as shown in Figure 5.3. The classic analogy is with a column of soldiers marching in line, who encounter the edge of a muddy field which slows the rate of march. If Θ_1 is the angle of incidence relative to the normal to the interface, and Θ_2 is the angle of the refracted beam on the other side of the interface, then the deflection due to the change in speed is described by Snell's Law, which states $c_2 \sin \Theta_1 = c_1 \sin \Theta_2$, or equivalently $\sin \Theta_2 = (n_1/n_2) \sin \Theta_1$. Now, if a beam is traveling within the medium at angle Θ_2 and exits into a medium with lower index of refraction (e.g. glass to air), then the angle of the exiting beam is given by $\sin \Theta_1 = (n_2/n_1) \sin \Theta_2$; hence, the beam is deflected away from the normal, as indicated in the sketch. At such an interface, if $(n_2/n_1)\sin\Theta_2 > 1$ then there is no transmitted beam and the ray is refracted so much that it is totally reflected back into the medium - a phenomenon known as *total internal reflection*. In reality, there is always some partial reflection at an interface. Partial reflection, as well as many other phenomena we shall encounter, depends on the wave nature of light as described by Maxwell's equations, and cannot be captured by the "corpuscular" viewpoint. This was as much of a groundbreaking conceptual challenge for early optical theorists as blackbody radiation was for investigators presiding over the dawn of quantum theory.

The concept of index of refraction can be extended to absorbing media. Suppose that a plane wave propagating through the medium has spatial dependence $exp(2\pi i kx)$, where x is the distance measured in the direction of propagation. Then, the expression for the speed of the wave



Figure 5.3: Refraction of a beam of light at an interface between a medium with index of refraction n_1 and a medium with index of refraction n_2 . In the sketch, $n_2 > n_1$ so the speed of light is slower in the medium than in the surroundings, as is the case for glass or water in air.

in terms of its frequency and wavenumber becomes

$$\frac{\nu}{k} = \frac{1}{n}c\tag{5.22}$$

where c is the speed of light in a vacuum. Thus $k = (\nu/c)n$. Note that ν/c is the vacuum wavenumber we have been using all along to characterize radiation. For real n, k is the wavenumber in the medium, which is larger than the vacuum wavenumber by a factor of n. If we allow n to be complex, its imaginary part characterizes the absorption properties of the medium. To see this, write

$$k_R + ik_I = \frac{\nu}{c} n_R + i\frac{\nu}{c} n_I \tag{5.23}$$

Since the wave has spatial dependence $\exp(2\pi i kx) = \exp(w\pi i k_r x) \exp(-2\pi k_i x)$, the coefficient $2\pi k_i = 2\pi n_I(\nu/c)$ gives the attenuation of the light by absorption, per unit distance travelled. Note that because of the factor ν/c , the quantity $2\pi n_I$ gives the attenuation of the beam after it has traveled by a distance equal to one wavelength of the light. Hence $n_I = 1$ corresponds to an extremely strong absorption. Visible light traveling through such a medium, for example, would be almost completely absorbed by the time it had traveled one *micrometer*.

The absorption coefficient k_I is proportional to the absorption cross section per unit mass we introduced in Chapter 4, and which reappeared above in the context of absorption by particles. If the density of the medium is ρ , the corresponding mass absorption coefficient is $\kappa = 2\pi k_I / \rho = n \cdot (\nu/c) / \rho$. In *mks* units, this quantity has units of m^2/kg , and is thus an absorption cross section per unit mass.

The real index of refraction for some common cloud-forming substances is given in Table 5.1. The index of refraction for these and similar substances lies approximately in the range 1.25 to 1.5, and is only weakly dependent on wavenumber; data also shows the index of refraction to depend only weakly on temperature. The weak dependence of index of refraction on wavelength does give rise to a number of readily observable phenomena, such as separation of colors by a prism or the droplets that give rise to rainbows, but such phenomena, beautiful as they are, are of little importance to planetary energy balance. The one exception to the typically gradual variation

	Thermal-IR	Near-IR	Solar	UV-B
Liquid water	1.40	1.31	1.33	1.43
Water ice	1.53	1.29	1.31	1.39
CO_2 ice	1.45	1.40	1.41	1.54
Liquid CH_4	1.28	1.27	1.27	1.49
$H_2SO_4 \ 38\%$	1.56	1.36	1.38	1.53
$H_2SO_4 81\%$	1.41	1.51	1.44	1.58

Table 5.1: Real part of the index of refraction for selected condensed substances. Thermal-IR data is at 600 cm^{-1} , Near-IR is at 6000 cm^{-1} , Solar at 17000 cm^{-1} (.59 *micrometers*), and UV-B at 50000 cm^{-1} (.2 *micrometers*. Liquid water data was taken at a temperature of 293K, water ice at 273K, CO_2 ice at 100K, liquid methane at 112K and H_2SO_4 at approximately 270K. The percentage concentrations given for the latter are in weight percent.

of the real index of refraction occurs near spectrally localized absorption features; the real index also has strong variations in the vicinity of such points. In considering the scattering of light by particles suspenden in an atmospheric gas, the index of refraction of the gas can generally be set to unity without much loss of accuracy. A vacuum has n = 1, and gases at most densities we'll consider are not much different. Specifically, for a gas n - 1 is proportional to the density. At 293K and 1bar, Earth air has an index of refraction of 1.0003 in the visible spectrum. CO_2 in the same conditions has an index of 1.0004, and even at the 90 bar surface pressure of Venus has an index of only 1.016. The resultant refraction by the atmospheric gas can be useful in determining the properties of an atmosphere through observations of refraction from the visible through radio spectrum, but it has little effect on scattering by cloud particles.

Insofar as the real index of refraction goes, it would appear that it matters very little what substance a cloud is made of. The minor differences seen in Table 5.1 are far less important than the effects of cloud particle size and the mass of condensed substance in a cloud. The absorption properties, on the other hand, vary substantially from one substance to another, and these can have profound consequences for the effect of clouds on the planetary energy budget. The behavior of the imaginary index for liquid water, water ice, and CO_2 ice is shown in Figure 5.4. Water and waterice clouds are nearly transparent throughout most of the solar spectrum; for these substances, n_I is less than 10^{-6} for wavenumbers between $10000cm^{-1}$ and $48000cm^{-1}$ (wavelengths between 1 micrometer and .2 micrometer, though the absorption increases sharply as one moves into the far ultraviolet. In the thermal infrared spectrum, however, water and water-ice are very good absorbers, having n_I in excess of .1 between wavenumbers of 50 and 1000 cm^{-1} . Such a large value of n_I implies that most thermal infrared flux would be absorbed when passing through a cloud particle having a diameter of 10 micrometers. For this reason, infrared scattering by water and water-ice clouds can be safely neglected, such clouds being treated as pure absorbers and emitters of infrared. This is not the case for clouds made of CO_2 ice (important on Early Mars and perhaps Snowball Earth) or liquid CH_4 (important on Titan). CO_2 ice clouds are still quite transparent in the solar spectrum, apart from strong absorption in the far ultraviolet. In constrast to water clouds, however, they are largely transparent to thermal infrared. For CO_2 ice clouds, n_I is under 10^{-4} between 1000 and 2000 cm⁻¹, and even between 500 and 1000 cm⁻¹ n_I is generally below .01 except for a strong, narrow absorption feature near 600 cm^{-1} . Likewise, liquid methane has n_I well under .001 between 10 and 1200 cm^{-1} . In both cases the infrared scattering effect of clouds can have an important effect on the OLR, leading to a novel form of greenhouse effect. Concentrated sulfuric acid, which makes up aerosols on Earth and the clouds of Venus, is quite transparent for wavenumbers larger than 4000 cm^{-1} but the imaginary index of refraction increases greatly at smaller wavenumbers, and in the thermal infrared sulfuric acid absorbs nearly as well as water.



Figure 5.4: The imaginary index of refraction for liquid water, water ice, and CO_2 ice.

Nonetheless, the scattering by sulfuric acid clouds has a significant effect on the OLR of Venus, because Venus is so hot that it has considerable thermal emission at wavenumbers greater than $4000cm^{-1}$.

As a general rule of thumb, typical cloud-forming condensates tend to be very transparent in the visible and near-ultraviolet and quite transparent in the near-infrared, but vary considerably in their absorption properties in the thermal infrared. Most substances – whether gaseous or condensed – are very good absorbers in the very shortwave part of the ultraviolet spectrum, with wavelengths below .1 *micrometer*. For this reason, this part of the UV spectrum is often referred to as "vacuum UV," because it is essentially only present in the hard vacuum of outer space.

5.3 Scattering by molecules: Rayleigh scattering

Rayleigh scattering theory is a classical (i.e. non-quantum) electromagnetic scattering theory which began life as a theory for scattering of an electromagnetic plane wave from a small sphere with real index of refraction n. "Small" in this context means small compared to the wavelength of the light being scattered. The scattering calculation is quite simple in the Rayleigh limit because the incident electric field is nearly constant over the particle, which makes it simple to compute the induced electromagnetic field within the particle. In essence, the electric field of the incident wave causes charges within the particle to migrate so that positive charge accumulates on one side and negative charge on the other, leading to a dipole moment which oscillates with the same frequency as that of the incident wave. The index of refraction is in fact a measure of the polarizability of the medium – the proportionality between the strength of the electric field and the strength of the dipole moment induced. The scattered wave in the Rayleigh limit is then simply the electromagnetic radiation emitted by an oscillating dipole, which is one of the more elementary calculations that can be done in electromagnetic theory.

5.3. SCATTERING BY MOLECULES: RAYLEIGH SCATTERING

Perhaps surprisingly, the Rayleigh theory works quite well as a description for scattering of light from molecules, even though molecules are not dielectric spheres. It's true that the typical size of a molecule (e.g. 0003 microns for N_2) is much smaller than the wavelength of visible or even ultraviolet light, but one might have thought that the quantum response of the molecule might substantially affect the scattering. Certainly, Rayleigh theory does not provide a suitable basis for computing molecular absorption of radiation, which, as we have seen in Chapter 4, is inextricably linked to the quantum nature of the molecule. We will not go further into the reasons that a classical theory works so much better for molecular scattering than for molecular absorption, but it is indeed a convenient turn of events. In practice, it works fine to use spectroscopically measured absorption coefficients to compute gaseous absorption, together with Rayleigh scattering to compute gaseous scattering.

For a spherically symmetric scatterer, Rayleigh theory yields the following formula for the scattering cross-section:

$$\chi_{sca} = \frac{8\pi}{3} (\frac{2\pi}{\lambda})^4 \alpha_p^2 \tag{5.24}$$

where λ is the wavelength of the incident light in vacuum, and α_p is the polarizability constant of the scatterer, which expresses the proportionality between the electric field and the induced dipole moment. In practice, the polarizability constant is inferred from measurements of the scattering cross section itself. It is only a weak function of wavelength. The very strong dependence of Rayleigh scattering cross section on wavelength is notable; short waves (high wavenumbers) scatter much more strongly than long waves (low wavenumbers). The explanation of the blue skies of Earth is perhaps the most famous application of Rayleigh scattering: blue through violet light has shorter wavelength than the rest of the visible spectrum, and therefore dominates the diffuse radiation caused by scattering of the solar beam from air molecules. For scatterers that are not spherically symmetric – and this includes all the polyatomic molecules like N_2 , H_2 and CO_2 present in most of the atmospheres we have been considering – the dipole moment is not in the same direction as the imposed electric field, and this effect slightly alters the expression for the scattering cross section. Molecules in a gas are randomly oriented, and it can be shown that, averaged over all orientations, the modified cross section consists of the symmetric cross section in Eq 5.24 multiplied by $3(2+\delta)/(6-7\delta)$, where δ is the *depolarization factor*, which is a property of the molecule. The depolarization factor is zero for a spherically symmetric scatterer. For our purposes, the effect of the depolarization factor is not very consequential. It has a value of .054 for O_2 , of .0305 for N_2 and .0805 for CO_2 . These lead to only a minor increase in the scattering cross section.

Using Maxwell's equations, it can also be inferred that the index of refraction is related to the polarizability of the molecules making up the medium via the relation

$$n = 1 + 2\pi N \alpha_p \tag{5.25}$$

where N is the number of molecules per unit volume. This is a very useful relation, as it allows one to determine Rayleigh scattering cross sections through simple measurements of the refractive index, which can be carried out by straightforward measurement of the angle of deflection of light as it moves from a transparent solid container (e.g. glass) into the gas.

Table 5.2 gives the measured Rayleigh scattering cross section relative to H_2 for a number of common atmospheric gases, as well as the corresponding cross-section per unit mass. The absolute value of the cross section for H_2 is given for a number of wavelengths in the caption, allowing the actual cross sections for the other molecules to be readily computed; the values given for H_2 in the caption deviate somewhat from the $1/\lambda^4$ wavelength scaling because of the slight dependence of index of refraction on wavelength but generally speaking it is adequate to extrapolate to other wavelengths using the fourth-power law. He stands out as an exceptionally weak scatterer. Most

	H_2	He	air	N_2	O_2	CO_2	H_2O	NH_3	CH_4
χ_{sca}	1.	.0641	4.4459	4.6035	3.8634	10.5611	3.3690	7.3427	10.1509
χ_{sca}/m	1.	0.0321	0.3066	0.3288	0.2415	0.4800	0.3743	0.8638	1.2689
$ au_{ray}$	0.40653	0.01305	0.12464	0.13367	0.09818	0.19513	0.15216	0.35116	0.51585

Table 5.2: Rayleigh scattering cross sections, and cross sections per unit mass, relative to H_2 . These results are based on observations of the index of refraction, and do not take into account variations in the polarization factors. The scattering cross sections and cross sections per unit mass for H_2 are $1.4 \cdot 10^{-38}m^2$ ($4.215 \cdot 10^{-12}m^2/kg$) at a wavelength of 10 micrometers, $8.270 \cdot 10^{-33}m^2$ ($2.490 \cdot 10^{-6}m^2/kg$) at a wavelength of 1 micrometer, and $3.704 \cdot 10^{-28}m^2$ ($1.11 m^2/kg$) at a wavelength of .1 micrometers. τ_{ray} is the optical depth due to Rayleigh scattering at $\frac{1}{2}$ micrometer for a 1 bar atmosphere of the indicated gas under Earth gravity.

of the rest of the molecules have scattering cross-sections per unit mass which are moderately smaller than H_2 , with the exception of CH_4 , which is moderately larger.

To get an idea of how important the scattering is in various contexts, we can use the cross sections per unit mass to determine the optical depth of the entire column of an atmosphere. When the optical depth is small, the atmosphere scatters hardly at all, but when optical depth becomes large a significant amount of radiation will be scattered; in the case of the incident Solar radiation, this means a lot of the incident beam will be reflected back to space. The last line of Table 5.2 gives the optical depth for an atmosphere consisting of 1 bar of the given gas under Earth gravity. One can scale this up to other planets by multiplying by the appropriate surface pressure, and dividing by the planets' gravity relative to Earth gravity. The optical depth values are given at a wavelength of $\frac{1}{2}$ micrometer, in the center of the visible spectrum, which is also near the peak of the Solar spectrum. For Earth's present atmosphere, the optical depth is small, but not insignificant; Rayleigh scattering affects about 12% of the incident beam. For an Early Mars having 2 bar of CO_2 in its atmosphere, the Rayleigh scattering is quite strong, owing to the somewhat elevated scattering cross section of CO_2 relative to air, to the low gravity and to the extra surface pressure. The Rayleigh optical depth for Early Mars would be 1.03 in the visible. The associated reflection of solar radiation is a significant impediment to warming Early Mars with a gaseous CO_2 greenhouse effect. If one took away the reflective clouds of Venus, the CO_2 Rayleigh scattering would still make Venus quite reflective, since the optical depth of a 90 bar CO_2 atmosphere on Venus is nearly 20. A 1.5 bar N_2 atmosphere on Titan would have an optical depth of 1.45 in the absence of clouds. The top 10 bars of Jupiter's mostly H_2 atmosphere would have an optical depth of 1.6, and likewise scatter significantly.

The optical depths for other wavelengths can be obtained by scaling these results according to $1/\lambda^4$. Thus, at thermal infrared wavelengths which are 5 times or more greater than the one we have been considering, the optical depth is at least 625 times smaller; Rayleigh scattering is insignificant at these wavelengths, which is why it is safe to neglect gaseous scattering when doing computations of *OLR*. On the other hand, the Rayleigh scattering optical depths are at least 16 times greater for ultraviolet. We will learn how to turn these optical depth values into planetary albedos in Section 5.6.

Rayleigh scattering is not isotropic, but the phase function is symmetric between the forward and backward direction. Within the two-stream approximation, then, we do not really need any information beyond the scattering cross section. It is worth having a look at the phase function anyway, if only to get an idea of how close to isotropic it is. Given the depolarization factor δ , the Rayleigh phase function is

$$P(\cos\Theta) = \frac{3}{2(2+\delta)}(1+\delta+(1-\delta)\cos^2\Theta)$$
(5.26)

From this we see that Rayleigh scattering is mildly anisotropic, with stronger scattering in the forward and backward direction than in the direction perpendicular to the incident beam. For $\delta = 0$ the scattering is twice as strong in the forward and backward directions ($\Theta = 0, \pi$) as it is in the side lobes ($\Theta = \pi/2$). Increasing δ reduces the anisotropy. In fact, laboratory measurements of the intensity of side vs. forward scattering provide a convenient way to estimate the depolarization factor.

5.4 Scattering by particles

Rayleigh theory tells us everything we need to know about scattering from the gas making up an atmosphere, but to deal with cloud and aerosol particles, we need to know about scattering from objects that are not small compared to a wavelength, and indeed could be considerably larger than a wavelength, as is the case for visible light scattering from water or ice clouds on Earth. The answer is provided by Mie theory ², which is a general solution for scattering of an electromagnetic wave from a spherical particle having uniform complex index of refraction. Mie theory reduces to Rayleigh theory in the limit of small non-absorbing particles. Ice crystals and dust particles are not spherical, and while numerical solutions are available for complex particles, for our purposes it will prove adequate to treat such cases in terms of equivalent spheres.

The Mie solution is a solution to Maxwell's electromagnetic equations which is asymptotic to a plane wave at large distances from the particle, and satisfies appropriate continuity conditions on the electromagnetic field at the particle boundary, where the index of refraction is discontinuous. Since Maxwell's equations are linear, the solution can be built up from more elementary solutions to the equations, and this is how Mie theory proceeds. It furnishes the solution in terms of an infinite sum over spherical Bessel functions, and is a real tour-de-force of early 20^{th} century applied mathematics. The formula is very convenient for evaluation of scattering parameters on a computer, but it is too complicated to yield any insight as to the nature of the solution. For that reason, we do not bother to reproduce the formula here; it is derived and discussed in the references section for this chapter, and a routine for evaluating the Mie solution is provided as part of the software supplement. Here we will only present some key results needed to provide a basis for cloud and aerosol scattering in the solar spectrum and infrared.

Let n_o , assumed real, be the index of refraction of the medium in which the particle is suspended. Define the Mie parameter $\bar{r} = n_o r/\lambda$, where r is the particle radius and λ is the wavelength of the incident light in vacuum, measured in the same units as r. The relative index of refraction is $n_{rel} = n/n_o$, where n is the index of refraction of the substance making up the particle. The main things we wish to compute from the Mie solution are the phase function, the scattering efficiency, the absorption efficiency, and the asymmetry factor. These are all non-dimensional quantities, and depend only on \bar{r} and n_{rel} .

Let's first take a look at some scattering properties in the conservative case, $n_I = 0$. In this case, $Q_{abs} = 0$. Fig. 5.5 shows the scattering efficiency and asymmetry factor as a function of r/λ for several different values of the real index of refraction of the scatterer. Since $1/\lambda$ is the

 $^{^{2}}$ The theory is named for Gustav Mie (1869-1957), who published the solution in 1908, while he was a professor at Greifswald University in Germany.



Figure 5.5: Scattering efficiency (left panel) and asymmetry factor (right panel) for Mie scattering from a non-absorbing sphere. The index of refraction of the medium is unity. r is the radius of the particle, and λ is the wavelength of the incident light measured in the same units as r.

wavenumber, the graph can be thought of as displaying the scattering properties for increasing particle size with fixed wavenumber, or for increasing wavenumber with fixed particle size. Note that in the limit $n_R \rightarrow 1$ there should be no scattering at all, since in that case the particle is not optically distinct from the surrounding medium.

For any given n_R , the scattering efficiency becomes small when r/λ is sufficiently small; this is the Rayleigh limit. The scattering efficiency reaches its first peak at an order unity value of r/λ , and the position of the first peak gets closer to zero as $n_R - 1$ is made larger and the particle is made more refractive. In this sense, for a given size and wavelength, particles made of more refractive substances like CO_2 ice or concentrated H_2SO_4 act like smaller particles than particles made of less refractive substances like water or liquid CH_4 . The first peak represents the optimal conditions for scattering. At the first peak, the scattering cross section can be 4 times or more the actual cross-section area of the particle.

As r/λ is increased past the first peak, the scattering efficiency oscillates between values somewhat below 2 to somewhat above 2 through a number of oscillations of decreasing amplitude, asymptoting at a value of 2 when the particle is large compared to the wavelength. The limit of large r/λ is the *geometric optics* limit, familiar from schoolbook depictions of how lenses work. In the geometrics optic limit, a beam of light is represented as a bundle of independent parallel rays, each of which travels in a straight line unless deflected from its course by an encounter with the interface between the particle and the medium – once upon entering the particle, and once upon leaving it. It is surprising that, in this limit, the scattering efficiency should asymptote to 2, since one would be quite reasonable in thinking that rays that do not encounter the object would be unaffected, implying $Q_{sca} = 1$. What is missing from the geometric optics picture is diffraction. Light is indeed a wave, and this has consequences that cannot be captured by the ray-tracing on which geometric optics is based. The light encountering the sphere is a plane electromagnetic wave, and the scattering takes a circular chunk out of it: a wave with a "hole" in it is simply not a solution to Maxwell's equations, and as one proceeds past the obstacle the hole fills in with parts of the beam that never directly encountered the obstacle. It is the diversion of this part of the incident beam that accounts for the "extra" scattering cross section. The nature of diffraction is far more easily understood through examination of some elementary solutions to Maxwell's equations than

5.4. SCATTERING BY PARTICLES

it is through this rather cryptic explanation, and the reader in pursuit of deeper understanding is encouraged to study the treatment in the textbooks listed in the references to this chapter.

The main thing to take away from the preceding discussion is that refractive particles made of common cloud-forming substances become very good scatterers when their radius is comparable to or exceeds the wavelength of the light being scattered. The scattering cross section is two or more times the cross section area, and for particles whose radius is more than a few times the wavelength, the scattering efficiency is close to 2, more or less independent of what the particle is made of, and more or less independent of wavelength. This limit applies to visible light scattering off of typical cloud droplets, which have radii of 5 to 10 *micrometers*. For infrared light scattering off of cloud particles (e.g. methane clouds on Titan or dry-ice clouds on Early Mars), or for visible light scattering from micrometer-sized aerosol particles, the wavelength is comparable to the particle size, and one needs to take both wavelength and index of refraction into account in order to see how much the scattering efficiency is enhanced over the geometric-optic limit.

Turning to the right-hand panel of Figure 5.5, we see that the scattering becomes symmetric for particles small compared to a wavelength, but becomes extremely forward-peaked for particles of radius comparable to or larger than a wavelength. For $n_R = 1.25$, somewhat smaller than the value for water, the asymmetry factor is on the order of .85 for large particles. The reason for the strong forward bias is that the large particles are rather like spherical lenses, and can bend light somewhat from the oncoming path, but cannot easily reflect it into the backward direction. This feature of scattering is very important in the treatment of cloud effects on the radiation budget. The forward bias in scattering reduces the effectiveness of clouds as scatterers, and reduces their albedos well below what one would have for layers of symmetric scatterers having the same optical thickness. Without the forward scattering bias, clouds would be much more reflective of solar radiation, and planets would be much colder.

A better appreciation of the strongly forward-peaked nature of Mie scattering from large particles can be obtained by examining the phase functions shown in Figure 5.6. Almost nothing is back-scattered, and there is a sharp spike near $\Theta = 0$ which becomes sharper as the particle is made larger. For $r/\lambda = 4$, 40% of the scattered flux is in the forward peak with $\Theta < .1radians$. When the particle is comparable in size to a wavelength, the scattering is still forward-peaked, but much less so, with appreciable amounts of flux being scattered a half radian or more from the original direction. Multiple scattering in this case allows a considerable amount of light to be backscattered relative to its original direction, though it can take several bounces before the light turns the corner. This effect allows carbon dioxide ice clouds composed of particles of size comparable to an infrared wavelength to be quite good reflectors of infrared light trying to escape to space. Visible light scattering from such clouds is much more forward-peaked, and correspondingly less efficient.

Now let's compute the optical thickness of some typical cloud and aerosol layers. Suppose that the layer is made up of nonabsorbing particles with a scattering efficiency Q_{sca} . Recall that the scattering efficiency is close to 2 for particles large compared to the wavelength, and can be as large as 4 for particles with diameter comparable to a wavelength, but falls rapidly to zero as the particle size is made smaller. For light in the solar spectrum, particles with diameters of a half *micrometer* or more are very efficient scatterers, with particles at the small end of this range being the most efficient. If the density of the substance making up the particle is ρ and the particle radius is r, then $\chi_{sca}/m = 2\pi r^2 Q_{sca}/(\frac{4}{3}\pi r^3 \rho) = \frac{3}{2}Q_{sca}/(\rho r)$. This is the factor by which one multiplies the mass path of scatterer to get the optical depth. The formula implies that, for a given mass of scatterers, small particles lead to much more scattering than large particles. 1 kg of 1 micrometer cloud droplets. The difference in index of refraction between sulfuric



Figure 5.6: Scattering efficiency (left panel) and asymmetry factor (right panel) for Mie scattering from a non-absorbing sphere. The index of refraction of the medium is unity. r is the radius of the particle, and λ is the wavelength of the incident light measured in the same units as r.

acid aerosols and water droplets is of far less consequence than the difference in particle size. To proceed, assume $Q_{sca} = 2$ and $\rho = 1000 kg/m^3$. Then, for 1*micrometer* particles, $chi_{sca}/m = 3000$, so it takes a mere $\frac{1}{3000}$ of a kilogram – *a third of a gram* – of aerosol particles added to a column of atmosphere with a one square meter base to bring the optical depth up to unity. This is the reason that tiny amounts of aerosol forming compounds can have significant effects on planetary climate. The small particle size also tends to make the albedo effect of aerosols dominate their greenhouse effect, despite the fairly strong absorption coefficient of sulfuric acid in the thermal infrared. For cloud droplets with a radius of 10 *micrometers*, it would take about 3 grams of water to achieve the same optical thickness as for the smaller aerosol particles, but this is still a tiny fraction of the water content of the atmosphere. A 1 km column of air in saturation at 280K at Earth surface pressure contains 7.8 kg of water vapor per square meter, for example. A cloud need not weigh much in order to have a profound effect on the albedo of a planet!

Now we'll turn our attention to absorbing particles. Figure 5.7 shows the scattering and absorption efficiencies for particles with $n_R = 1$ having various nonzero values if n_I . For $n_I = .1$ and $n_I = .01$ the absorption efficiency increases monotonically with particle size, and approaches unity from below (for the latter of these cases, the absorption efficiency is .89 at $r/\lambda = 100$ and .997 at $r/\lambda = 500$, outside the range plotted in the graph). This is typical behavior for particles with n_I appreciably less than unity, for which the light takes several wavelengths to decay upon encountering the particle. Larger particles absorb more simply because the light travels a longer distance within the particle, and has more opportunity to decay. For similar reasons, when n_I is made smaller, the particle must be made larger in order for there to be appreciable decay. In any event, when the particle is made large enough, it essentially absorbs a portion of the incident beam with area equal to the cross section area of the particle. Perhaps more surprisingly, though, the scattering efficiency also rises with particle size – absorbing particle don't just absorb; they also deflect light from the incident direction. This is due to the same diffraction phenomenon we encountered previously. Taking a disk out of the incident beam inevitably causes the remaining part of the beam to be deflected from its original path. For large particles the scattering efficiency and absorption efficiency sum to 2, with half the intercepted beam being absorbed and the other half scattered.



Figure 5.7: Scattering efficiency (left panel) and absorption efficiency (right panel) for Mie scattering from a partially absorbing sphere. The real part of the index of refraction is held fixed at unity while the imaginary part is varied as indicated for each curve in the figure. Other parameters are defined as in Fig. 5.5.

When the particles are very strongly absorbing the behavior is somewhat different, as typified by the curve for $n_I = 1$ in Fig. 5.7. In this case the absorption efficiency actually overshoots unity for particles somewhat smaller than a wavelength. The particle is able to sweep up and absorb radiation from an area larger than its cross-section, owing to the distortion of the electromagnetic field caused by the particle itself. On the other hand, as the particle is made larger, the absorption efficiency goes down and asymptotes to a value somewhat less than unity, because the incident wave is not able to penetrate deeply into the particle, and instead skirts along near its surface. In compensation, the scattering efficiency for large particles becomes greater than unity.

The above results were for particles with $n_R = 1$, in which case there is neither scattering nor absorption when n_I approaches zero. The behavior for $n_R > 1$ is explored in Problem ??. When $n_R > 1$ then the scattering cross section behavior resembles the conservative case when n_I is small, but even a small n_I damps out the ripples in $q_{sca}(r)$ for sufficiently large particles. The asymptotic value of q_{abs} is still close to unity for sufficiently large particles throughout the range of n_I , but the overshoot properties differ somewhat from the case $n_I = 1$.

As a rule of thumb, then, we can say that when particles become larger than the characteristic decay length $\lambda/2\pi n_I$, they absorb essentially everything within a disk of area equal to the cross section area of the particle. The absorption efficiency is somewhat reduced for particles with $n_I > .1$, but even for $n_I = 1$, which is the largest value likely to be encountered, the reduction in efficiency is rather modest. Liquid water has $n_I > .1$ throughout the infrared, so any liquid water cloud droplet larger than about 10 micrometers in radius will absorb nearly all the infrared it encounters. In fact, for liquid water $n_I \approx 1$ throughout most of the infrared, so that even quite small particles are efficient absorbers. Water ice can have n_I as low as .05 in some parts of the infrared spectrum, so the particles need to be twice as large to be equally good absorbers in that part of the spectrum, but ice clouds observed on Earth do tend to have larger particle sizes than water clouds.

5.5 The two-stream equations with scattering

The two stream approximations to the full scattering equation are derived from Eq. 5.14 and Eq. 5.15 by constraining the angular distribution of the radiation in such a way as to allow all integrals appearing in these equations to be written in terms of either $I_+ + I_-$ or $I_+ - I_-$. In the resulting equations, flux in the upward stream is absorbed, or scattered into the downward stream, at a rate proportional to the upward stream intensity, and similarly for the downward stream. The two-stream approximations are an instance of what physicists euphemistically like to call "uncontrolled approximations," in that they are not actually exact in any useful limit but are nonetheless physically justifiable and perform reasonably well in comparison to more precise calculations. The two-stream approximiations have inevitable inaccuracies because it is not, in fact, possible to precisely determine the scattering or the absorption from knowledge of the upward and downward fluxes alone. The two-stream approximations can be thought of as the first term in a sequence of N-stream approximations which become exact as N gets large. Fortunately, N = 2 proves sufficiently accurate for most climate problems.

The general form of a two-stream approximation for diffuse radiation is

$$\frac{d}{d\tau^*}I_+ = -\gamma_1 I_+ + \gamma_2 I_- + \gamma_B \pi B(\nu, T(\tau^*_{\nu})) + \gamma_+ L_{\odot} \exp(-(\tau^*_{\infty} - \tau^*)/\cos\zeta)$$

$$\frac{d}{d\tau^*}I_- = \gamma_1 I_- - \gamma_2 I_+ - \gamma_B \pi B(\nu, T(\tau^*_{\nu})) - \gamma_- L_{\odot} \exp(-(\tau^*_{\infty} - \tau^*)/\cos\zeta)$$
(5.27)

where τ^* is the optical depth in the vertical direction (increasing upward) including scattering loss. The coefficients γ_j depend on frequency, on the properties of the scatterers, and on the particular assumption about angular distribution of radiation that was made in order to derive an approximate two-stream form from the full angular-resolved equation. We recover the hemispherically isotropic Schwartzschild equations used in previous chapters by taking $\gamma_2 = \gamma_+ = \gamma_- = 0$ and $\gamma_1 = \gamma_B = 2$. The terms proportional to γ_B represent the source due to thermal emission of radiation, while the terms proportional to γ_+ and γ_- represent the source of diffuse radiation caused by scattering of the direct beam. The direct beam is assumed to have flux L_{\odot} (generally the solar constant) in the direction of travel, and to travel at an angle ζ relative to the vertical. There is no upward direct beam term because it is assumed that all direct beam flux scattered from the ground scatters into diffuse radiation.

The symmetry between the coefficients multiplying I_+ and I_- is dictated by the requirement that the equations be invariant in form when one exchanges the upward and downward directions. γ_1 gives the rate at which flux is lost from the upward or downward radiation, while γ_2 gives the rate of conversion between upward and downward radiation by scattering. We can derive additional constraints on the γ_i . Subtracting the two equations gives us the equation for net vertical flux

$$\frac{d}{d\tau^*}(I_+ - I_-) = -(\gamma_1 - \gamma_2)(I_+ + I_-) + 2\gamma_B\pi B + +(\gamma_+ + \gamma_-)L_\odot \exp(-(\tau_\infty^* - \tau^*)/\cos\zeta) \quad (5.28)$$

First, we demand that in the absence of a direct-beam source, the fluxes reduce to black body radiation in the limit of an infinite isothermal medium. Since B is constant and $L_{\odot} = 0$ in this case, we may assume the derivative on the left hand side to vanish. Since $I_{+} = I_{-} = \pi B$ for blackbody radiation, we find $\gamma_{B} = \gamma_{1} - \gamma_{2}$. Comparing with Eq. 5.14 we also find that $\gamma_{+} + \gamma_{-} = \omega_{o}$. To further exploit Eq. 5.14 we must approximate $\int Id\Omega$ as being proportional to $I_{+} + I_{-}$. The constant of proportionality, which we shall call $2\gamma'$, depends on the angular distribution of radiation assumed. With this approximation it follows that $\gamma_{1} - \gamma_{2} = 2\gamma'(1 - \omega_{o})$. **Exercise 5.5.1** As a check on the above reasoning, show that in the conservative scattering limit $\omega_o = 1$ the sum of the diffuse vertical flux with the direct beam vertical flux is constant.

Next, we sum the equations for I_+ and I_- to obtain

$$\frac{d}{d\tau^*}(I_+ + I_-) = -(\gamma_1 + \gamma_2)(I_+ - I_-) + (\gamma_+ - \gamma_-)L_\odot \exp(-(\tau^*_\infty - \tau^*)/\cos\zeta)$$
(5.29)

This can be compared to the symmetric flux projection given by Eq. 5.15. Making an assumption about the angular distribution allows one to approximate $\int IH(\cos\theta)\cos\theta d\Omega$ as proportional to I_++I_- , and $\int IHd\Omega$ and $\int IHd\Omega$ each as being proportional to I_+-I_- . In consequence, $\gamma_1+\gamma_2 = 2\gamma \cdot (1-\hat{g}\omega_o)$, where γ is related to the proportionality coefficient and \hat{g} is a coefficient characterizing the asymmetry of the scattering. If $H = \cos\theta$, then \hat{g} is in fact the asymmetry factor \tilde{g} defined by Eq. 5.17, but other forms of H yield somewhat different asymmetry factors, though these tend to be reasonably close to \tilde{g} . For example, with the form of H given by Eq. 5.20, we showed that $\hat{g} = \frac{3}{2}\tilde{g}$ for phase functions that are truncated to their first three Fourier components. Finally, under circumstances when we can write $G = \hat{g}\cos\theta$, it follows that $\gamma_+ - \gamma_- = -2\gamma\omega_o\hat{g}$. This relation holds exactly when $H = \cos\theta$, and imposing it for other forms of H introduces errors that are no worse than other errors that are inevitable in reducing the full scattering equation down to two streams.

The general form of the set of two-stream coefficients satisfying all the above constraints is then

$$\gamma_{1} = \gamma \cdot (1 - \hat{g}\omega_{o}) + \gamma' \cdot (1 - \omega_{o})$$

$$\gamma_{2} = \gamma \cdot (1 - \hat{g}\omega_{o}) - \gamma' \cdot (1 - \omega_{o})$$

$$\gamma_{B} = 2\gamma' \cdot (1 - \omega_{o})$$

$$\gamma_{+} = \frac{1}{2}\omega_{o} - \gamma\omega_{o}\hat{g}\cos\zeta$$

$$\gamma_{-} = \frac{1}{2}\omega_{o} + \gamma\omega_{o}\hat{g}\cos\zeta$$
(5.30)

The coefficients γ and γ' are purely numerical factors that depend on the assumption about the angular distribution of radiation which is used to close the two-stream problem. All vertical dependence then comes in through ω_{o} , and possibly through \hat{g} if the asymmetry properties of scattering particles vary with height. There are three common closures in use. The first is the hemispherically isotropic closure, which we used earlier in deriving the two-stream equations without scattering. In this closure, it is assumed that the flux is isotropic (i.e. I is constant) in each of the upward and downward hemispheres, but with a different value in each hemisphere. The hemi-isotropic closure is derived by using the weighting function H defined by Eq. 5.20 and making use of Eq. 5.21. Given the isotropy of the blackbody source term, it is generally believed that the hemi-isotropic approximation is most appropriate for thermal infrared problems, with or without scattering. Another widely used closure is the *Eddington approximation*. The Eddington closure is obtained by taking $H = \cos\theta$ and making use of Eq. 5.19. To complete the closure, $\int I \cos^2\theta d\Omega$ is written in terms of $I_+ + I_-$ by assuming that the flux is truncated to the first two Fourier components, so $I = a + b \cos \theta$. This is probably the most widely used closure for dealing with solar radiation. It is generally believed that this closure is a good choice for dealing with both Rayleigh scattering and the highly forward-peaked scattering due to cloud particles, though the mathematical justification for this belief is not very firm. The quadrature approximation is similar, except that $\int I \cos^2 \theta d\Omega$ is evaluated using a technique known as *Gaussian quadrature*, which yields a different proportionality constant from the Eddington closure. The defining coefficients for the three closures are given in Table 5.3.

	γ	γ'
Hemi-isotropic	1	1
Quadrature	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{3}$
Eddington	$\frac{1}{4}$	1

Table 5.3: Coefficients for various two-stream approximations

When $\omega_o = 0$ there is no scattering and so the upward and downward streams should become uncoupled. From Eq. 5.30 we see that this decoupling happens only if $\gamma = \gamma'$, a requirement that is satisfied for the hemi-isotropic and quadrature approximations but not for the Eddington approximation. It follows that the Eddington approximation can incur serious errors when the scattering is weak, though it can nonetheless outperform the other approximations when scattering is comparable to or dominant over absorption.

5.6 Some basic solutions

When the scattering and absorption properties of the atmosphere are independent of τ^* , the two stream equations have simple exponential solutions. We'll begin with an elementary solution for conservative scattering, which provide quite useful estimates of the effect of clear-sky and cloudy atmospheres on the solar-spectrum albedo of planets. We shall specify an incoming direct-beam flux of solar radiation, and we seek the outgoing reflected flux at the same wavenumber. For conservative scattering, $\omega_o = 1$. In that case $\gamma_1 = \gamma_2$. Then, Eq 5.28 becomes

$$\frac{d}{d\tau^*}(I_+ - I_-) = L_\odot \exp(-(\tau_\infty^* - \tau^*)/\cos\zeta)$$
(5.31)

 \mathbf{SO}

$$I_{+} - I_{-} - L_{\odot} \cos \zeta \exp(-(\tau_{\infty}^{*} - \tau^{*})/\cos \zeta) = C$$
(5.32)

where C is a constant. This equation states that, for conservative scattering, the net of the diffuse flux and the vertical component of the surviving direct-beam flux is constant. As the upper boundary condition we require that the diffuse incoming radiation be zero; hence $C = I_{+,\infty} - L_{\odot} \cos \zeta$, and

$$I_{+} - I_{-} = I_{+,\infty} + L_{\odot} \cos \zeta (\exp(-(\tau_{\infty}^{*} - \tau^{*})/\cos \zeta) - 1)$$
(5.33)

Deep in the atmosphere, $I_+ - I_-$ becomes constant, and is equal to the difference between the top-of atmosphere incoming minus outgoing flux, i.e. $I_{+,\infty} - L_{\odot} \cos \zeta$. When the atmosphere is optically thick, or when $\cos \zeta$ is small (i.e. when the sun is close to the horizon) the exponential term is significant only near the top of the atmosphere. It represents the conversion of the direct beam into diffuse radiation by scattering, which occurs within a conversion layer of depth $1/\cos \zeta$ in optical depth units.

 $I_{+,\infty}$ is the reflected flux we wish to determine, and we must close the problem by applying a boundary condition at the ground. For this we need $I_+(\tau^*)$, which we obtain by using the equation for $I_+ + I_-$. Let's restrict attention to the symmetric scattering case, g = 0, for which $\gamma_+ = \gamma_-$ and $\gamma_1 = \gamma_2 = \gamma$. Then

$$\frac{d}{d\tau^*}(I_+ + I_-) = -2\gamma(I_+ - I_-) = -2\gamma(I_{+,\infty} - L_\odot \cos\zeta + L_\odot \cos\zeta \exp(-(\tau^*_\infty - \tau^*)/\cos\zeta))$$
(5.34)

5.6. SOME BASIC SOLUTIONS

The solution which satisfies $I_{-} = 0$ at the top of the atmosphere, is

$$I_{+} + I_{-} = I_{+,\infty} + 2\gamma (I_{+,\infty} - L_{\odot} \cos \zeta) (\tau_{\infty}^{*} - \tau^{*}) + 2\gamma L_{\odot} \cos^{2} \zeta (1 - \exp(-(\tau_{\infty}^{*} - \tau^{*})/\cos \zeta))$$
(5.35)

Let's suppose that the ground is perfectly absorbing. This calculation characterizes the albedo of the atmosphere alone. Later, we'll compute how much the surface albedo enhances the planetary albedo. If the ground is perfectly absorbing, then we require $I_+ = 0$ at $\tau^* = 0$. To apply the boundary condition, we add Eqns. 5.33 and 5.35 and evaluate the result at the ground. Applying the boundary condition and solving for $I_{+,\infty}$ we find

$$I_{+}, \infty = \frac{\left(\frac{1}{2} - \gamma \cos \zeta\right)\beta_{\odot} + \gamma \tau_{\infty}^{*}}{1 + \gamma \tau_{\infty}^{*}} L_{\odot} \cos \zeta \equiv \alpha_{a} L_{\odot} \cos \zeta \tag{5.36}$$

where $\beta_{\odot} = 1 - \exp(-\tau_{\infty}^*/\cos\zeta)$; this quantity is the proportion of the direct solar beam that has been lost to scattering by the time the beam reaches the ground. The fraction multiplying $L_{\odot}\cos\zeta$ in Eq. 5.36 is the planetary albedo. Since any flux reaching the ground is absorbed completely, this albedo is in fact the albedo of the atmosphere alone, which we will call α_a . In the optically thin limit, $\beta_{\odot} \approx \tau_{\infty}^*/\cos\zeta$, so the albedo approaches zero like $\frac{1}{2}\tau_{\infty}^*/\cos\zeta$. Half of the small amount of flux scattered by the atmosphere exits the top of the atmosphere, but the other half is scattered into the ground, where it is absorbed.

As the atmosphere is made more optically thick, the albedo increases in two stages. The first stage is an exponential adjustment, as the direct beam is converted to diffuse radiation. Some simple algebra shows that the numerator of Eq. 5.36 always increases with τ^* , regardless of the value of $\cos \zeta$. However, when the incident beam is relatively near the horizon, so that $\gamma \cos \zeta < \frac{1}{2}$, the conversion term leads to an exponential increase of albedo with τ^* . The effect becomes more pronounced when the Sun is more nearly on the horizon. This effect comes from the direct scatter of the incident beam to space. When τ_{∞}^* becomes appreciably larger than $1/\cos\zeta$, however, the direct beam has been completely converted, $\beta_{\odot} \approx 1$, and the albedo no longer varies exponentially. For large τ_{∞}^* , the albedo approaches unity like $1 - (\frac{1}{2}\gamma^{-1} + \cos\zeta)/\tau_{\infty}^*$. The rather slow approach to a state of complete reflection is due to multiple scattering. In contrast to the exponential decay of the direct beam, the diffuse radiation surviving to be absorbed at the surface decays only like $1/ au_{\infty}^{*}$ because much of the radiation scattered upward is latter scattered back downward. The exponential decay of the direct beam just represents a conversion to diffuse radiation, and therefore does not materially alter the conclusion that scattering is a relatively ineffective way of preventing radiation from reaching the surface. That is why one cannot rely on Rayleigh scattering alone to shield life at the surface from harmful ultraviolet radiation, despite the fact that the Rayleigh optical thickness of an Earthlike atmosphere is quite high in the ultraviolet.

The simple albedo formula given above has many physically important ramifications. Before computing the albedo for various conservatively scattering atmospheres, though, it is necessary to bring in the effect of asymmetric scattering if we are to deal with clouds, as scattering from cloud particles is strongly forward-peaked. A nonzero asymmetry factor simply adds a direct beam source term to the equation for $d(I_+ + I_-)/d\tau^*$, since $\gamma_+ - \gamma_-$ no longer vanishes. Some straightforward algebra shows that, allowing for a nonzero asymmetry factor, the albedo formula becomes

$$\alpha_a = \frac{\left(\frac{1}{2} - \gamma \cos \zeta\right)\beta_{\odot} + (1 - \hat{g})\gamma\tau_{\infty}^*}{1 + (1 - \hat{g})\gamma\tau_{\infty}^*} \tag{5.37}$$

This differs from the symmetric scattering form only in that the optical thickness is multiplied by $1 - \hat{g}$, which reduces the effective optical thickness when $\hat{g} > 0$. Thus, in the context of the two stream equations, the effect of asymmetric scattering is not very surprising or subtle. Since forward

	Earth	Early Mars	Venus	Titan	Water cloud	Sulfate aerosol haze
$ au_\infty^*$	0.12	1.03	20.	1.45	3.	1.25
\hat{g}	0.	0	0	0	.87	.76
albedo, hemi-isotropic	.08	.43	.94	.52	.13	.10
albedo, quadrature	.08	.43	.94	.51	.17	.13
albedo, Eddington	.08	.42	.93	.51	.20	.16

Table 5.4: Albedos for purely scattering atmospheres. Values given for Earty, Early Mars, Venus and Titan are for hypothetical clear-sky atmospheres consisting of 1 bar of air for Earth, 2bar of CO_2 for Early Mars, 90bar of CO_2 for Venus and 1.5bar of N_2 for Titan. The water cloud case assumes a path of 10 grams of water per square meter in droplets of radius 10 micrometers, having a scattering efficiency of 2. The sulfate aerosol case assumes a path of $\frac{1}{2}$ gram per square meter in sulfuric acid droplets of radius 1micrometer, having a scattering efficiency of 3. Albedos are computed at a wavelength of .5micrometer, with a zenith angle of 45° for the direct beam.

scattering just adds back into the forward radiation as if scattering had not occured at all, forwarddominated scattering simply has the effect of reducing the optical thickness of the atmosphere. The only reason one can't get by with simply redefining the optical depth is the presence of the direct beam; the direct beam transmission factor β_{\odot} is computed using the unmodified optical depth, rather than the rescaled optical depth, because even forward-scattered radiation is transferred out of the direct beam and into the diffuse component.

In Table 5.4 we use Eq.5.37 to compute the albedos of a number of clear-sky and cloudy planetary atmosphere, based on optical depths computed from the Rayleigh scattering or Miescattering cross-sections in the visible spectrum. Results are shown for both the hemispherically isotropic and the Eddington approximations; the results differ little between the approximations for the symmetric Rayleigh scattering cases. In clear sky conditions, Earth's atmosphere reflects about 8% of the incoming solar energy back to space. This represents nearly a third of Earth's observed albedo, and is a significant player in the energy budget. The thick CO_2 atmosphere postulated for Early Mars has an even more significant effect on albedo, reflecting fully 43% of the solar energy. Further increases in CO_2 lead to even greater reflection, making it hard to warm Early Mars with the gaseous CO_2 greenhouse effect alone. The case of Venus is particularly interesting. We see from the table that if the thick clouds of Venus were removed, Rayleigh scattering alone would be sufficient to keep the albedo high. The value in the table is an overestimate of the albedo Venus would have in the no-cloud case, since it ignores the solar absorption of CO_2 , but it suffices to show that virtually all of what escapes absorption would be scattered back to space by Rayleigh scattering. This is important to the evolution of Venus-like planets, which might not have atmospheric chemistry that supports sulfuric acid clouds like those of Venus at present; for that matter, it is not completely certain that thick clouds are a perennial feature of our own Venus. When one factors in that rather little solar radiation penetrates the present thick clouds of Venus, it is evident that the strong Rayleigh scattering of the surviving flux implies that only a trickle of solar radiation reaches the surface of a planet. It is only a trickle, but as we have seen in Chapter 4, it is a very important trickle, since the surface could not be so hot if most of the solar energy were absorbed aloft.

The cloud cases in Table 5.4 could really apply to any planet with condensible water or sulfer compounds. The cloud parameters chosen are quite typical of Earth conditions. Scattering of solar radiation from cloud particles is quite different from Rayleigh scattering because of the strong asymmetry, which makes the albedo considerably lower than one would expect on the basis of optical thickness. Nonetheless, a small mass of cloud water, or a still smaller mass of sulfate aerosol in the form of *micrometer*-sized droplets, leads to a very significant albedo. To put the mass path of sulfate aerosol into perspective, we note that if we assume a 10 day lifetime for aerosol in the atmosphere, the assumed mass path is equivalent to a world wide sulfur emission of about 8 *megatonnes/da*, allowing for the proportion of S in H_2SO_4 . Actual worldwide sulfur emissions for 1990 are estimated to have been more like $\frac{1}{3}$ megatonne/da, which is why the albedo of Earth's sulfate aerosol haze is lower than the estimate in the table, though still a significant player in the radiation budget.

Note also that when the asymmetry is as pronounced as it is for cloud particles, the albedo predicted by the Eddington approximation is significantly greater than that for the hemispherically isotropic case. Although the asymmetry factor reduces the attenuation of diffuse radiation by the cloud, the decay of the direct beam is exponential in the optical depth itself, leading to near total attenuation of the direct beam by the water cloud. This is why a thick cloud looks bright, though you cannot easily discern the disk of the Sun. It is also why it is possible to get quite thoroughly sunburned on a cloudy day.

In the preceding calculation we assumed that the ground was perfectly absorbing. If the ground is instead partially reflecting, having albedo α_q , it will reflect some of the light reaching the surface back upward. Some of this light reflected from the surface will make it through the atmosphere and escape out the top, increasing the planetary albedo. How does the albedo of the atmosphere combine with the albedo of the ground to make up the planetary albedo? Since we are assuming that there is no atmospheric absorption, if a proportion α_a of incoming light is reflected by the atmosphere, then a proportion $(1 - \alpha_a)$ reaches the ground. The proportion of this reflected back upward is $(1-\alpha_a)\alpha_g$, and if we were to simply add this upward proportion to the part reflected directly by the atmosphere we'd get a planetary albedo of $\alpha_a + (1 - \alpha_a)\alpha_q$. This simple estimate already illustrates the important result that putting a reflective (e.g. cloudy) atmosphere over an already reflective surface changes the planetary albedo much less than putting such an atmosphere over a dark surface – you can't make the planet whiter than white, as it were. The simple estimate overestimates the planetary albedo, though, since some of the upward radiation from the ground bounces back from the atmosphere where some of the remainder is absorbed at the ground. The less is reflected back upwards, and ever diminishing proportions remain to multiply scatter back and forth between the atmosphere and the ground. One doesn't need to actually sum an infinite series to solve the problem; all we need to do is to correctly specify the boundary condition on the upward radiation at the ground, which requires in turn a specification of the proportion of upward radiation which is reflected back to the surface by the atmosphere. The upward radiation reflected from the ground is, by definition, purely diffuse, so as a preliminary to this calculation we need the albedo of the atmosphere for upward-directed diffuse radiation. Since the radiation is purely diffuse, the form of this atmospheric albedo is somehwat simpler than the formula for incoming solar radiation. It is derived in Problem ?? , and is

$$\alpha'_{a} = \frac{(1-\hat{g})\gamma\tau_{\infty}^{*}}{1+(1-\hat{g})\gamma\tau_{\infty}^{*}}$$
(5.38)

Note that this has the same form as the expression for α_a , except that the direct beam term in the numerator, proportional to β_{\odot} , has been dropped. In terms of α'_a , the boundary condition on upward radiation at the ground is

$$I_{+}(0) = \alpha_{g}I_{-}(0) = \alpha_{g} \cdot ((1 - \alpha_{a})L_{\odot}\cos\zeta + \alpha_{a}'I_{+}(0))$$
(5.39)

which can be solved for $I_+(0)$. When this boundary condition is applied to the conservative twostream equations, the resulting expression for $I_{+,\infty}$ yields the following expression for the planetary albedo:

$$\alpha = \alpha_a + \frac{(1 - \alpha'_a)(1 - \alpha_a)}{1 - \alpha_g \alpha'_a} \alpha_g$$

= $1 - \frac{(1 - \alpha_g)(1 - \alpha_a)}{(1 - \alpha_g)\alpha'_a + (1 - \alpha'_a)}$ (5.40)

The details are carried out in Problem ??. When all three albedos, α_a , α'_a and α_g are small, the expression reduces to the sum of the albedos $\alpha_a + \alpha_g$. Eq. 5.40 has very important consequences for the net effect of clouds on the planetary radiation budget. Clouds have both a warming and a cooling effect on climate. High-altitude clouds have a warming effect, since they strongly reduce OLR either by absorption and emission, or by scattering, of infrared radiation. Clouds at any altitude have a cooling influence, through increasing the planetary albedo in the solar spectrum. The net effect of clouds increase the planetary albedo rather little, if they are put over a highly reflective surface (such as ice), or if they are put into an atmosphere which is already quite reflective (such as the dense atmosphere of Early Mars). In either case, introduction of clouds will tend to have a strong net warming effect, because the cloud greenhouse effect is relatively uncompensated by the cloud albedo effect. It is for this reason that clouds can greatly facilitate the deglaciation of a Snowball Earth, and that clouds of either water or CO_2 can very significantly warm Early Mars.

Now let's do an infrared scattering problem, one that illustrates the scattering greenhouse effect in its simplest form. Consider an atmosphere made of a gas that is completely transparent (hence also non-emitting) in the infrared. Suspended in the atmosphere is a cloud made of a substance such as CO_2 ice or liquid methane that is almost non-absorbing in the infrared; we'll idealize it as being exactly non-absorbing, and assume the scattering to be symmetric. Since neither the gas nor the cloud emit infrared, the temperature profile of the atmosphere is immaterial. This atmosphere lies above a blackbody surface with temperature T. What is the OLR? This problem is also a case of conservative scattering ($\omega_o = 1$), but with different boundary conditions. Since there is no incoming infrared, the upper boundary condition is $I_- = 0$. At the ground, the upward flux boundary condition is $I_+ = \pi B(\nu, T)$. Without any direct beam or blackbody source term, $I_+ - I_-$ is a constant, which is equal to the outgoing radiation $I_{+,\infty}$. at the frequency under consideration. The equation for $d(I_+ + I_-)/d\tau^*$ then tells us that $I_+ + I_- = 2\pi B - (1 + \gamma \tau^*)I_{+,\infty}$. Finally, imposing the boundary condition that $I_- = 0$ at $\tau^* = \tau^*_{\infty}$, we conclude that

$$I_{+,\infty} = \pi B / (2 + \gamma \tau_{\infty}^*) \tag{5.41}$$

Hence, the infrared scattering reduces the outgoing infrared by a factor $1/(2 + \gamma \tau_{\infty}^*)$ relative to what it would be in the absence of an atmosphere. This increases the surface temperature of the planet in the same fashion as the OLR reduction from the more conventional absorption/emission greenhouse effect. However, the scattering greenhouse effect works quite differently, since it reduces the OLR regardless of whether the atmospheric temperature goes down with height.

Next we'll extend the preceding scattering greenhouse problem to allow $\omega_o < 1$, so the atmosphere can absorb and emit. We'll assume the atmosphere to be isothermal at the same temperature T as the ground. In this case, $I_+ = I_- = \pi B(\nu, T)$ is a particular solution satisfying the boundary condition on I_+ at the ground, though it does not satisfy the boundary condition $I_- = 0$ at the top of the atmosphere. We must add a homogenous solution to the particular solution, which cancels I_- at the top of the atmosphere for the particular solution, but leaves the bottom boundary condition intact. The homogeneous equation is obtained by taking the derivative of Eq. 5.28 and substituting the derivative of $I_+ + I_-$ using Eq. 5.29, dropping the source terms

from both. Assuming ω_o and g to be independent of height, the homogeneous equation is then

$$\frac{d^2}{d\tau^{*2}}(I_+ - I_-) = -(\gamma_1 - \gamma_2)(\gamma_1 + \gamma_2)(I_+ - I_-)$$

= $-4\gamma\gamma'(1 - \hat{g}\omega_o)(1 - \omega_o)(I_+ - I_-)$ (5.42)

The general solution to this is $a \exp(-K \cdot (\tau_{\infty}^* - \tau^*)) + b \exp(K \cdot (\tau_{\infty}^* - \tau^*)))$, where

$$K = 2\sqrt{\gamma\gamma'(1 - \hat{g}\omega_o)(1 - \omega_o)} \tag{5.43}$$

One term grows exponentially in optical depth, while the other decays exponentially. The solution for $I_+ + I_-$ is then obtained from the solution for $I_+ - I_-$ using Eq. 5.28, allowing us to obtain the two fluxes individually for use in applying the boundary conditions. First, the homogeneous solution we add in must not disturb I_+ at the ground, since the particular solution already satisfies the boundary condition there. To keep the algebra simple, let's assume $\tau_{\infty}^* \gg 1$. In that case, we approximately satisfy the boundary condition at the ground by taking the solution which decays toward the ground, i.e. b = 0. The boundary condition on $I_-(0)$ then determines the value of the coefficient *a*. Carrying out the algebra and adding the homogeneous to the particular solution, we find the outgoing radiation to be

$$I_{+,\infty} = 2 \frac{\gamma'(1-\omega_o)}{\gamma'(1-\omega_o) + \sqrt{\gamma\gamma'(1-\hat{g}\omega_o)(1-\omega_o)}} \pi B$$
(5.44)

In this equation, πB is the emission the planet would have in the absence of an atmosphere, and the the coefficient multiplying it gives the reduction in emission due to the atmosphere. Note that for a non-scattering atmosphere, the isothermal atmosphere assumed would have no effect whatever on the outgoing radiation. In contrast to the conservative scattering case described by Eq 5.41, the emission in the partially absorbing case does not approach zero in the optically thick limit, but rather approaches the nonzero value given by Eq. 5.44. When there is no scattering, i.e. $\omega_o = 0$, the atmosphere should have no effect on emission, and this limit shows the shortcomings of the Eddington approximation. For $\omega_o = 0$, the factor reducing the emission is $2\gamma'/(\gamma' + \sqrt{\gamma\gamma'})$, which reduces to unity only when $\gamma = \gamma'$. Both the quadrature and the hemispherically isotropic assumptions satisfy this requirement, but as we have seen before, the Eddington approximation gives the wrong answer when scattering is weak, and is not suitable for such cases. For any of the approximations, as the scattering is made stronger relative to absorption, $\omega_o \to 1$ and the emission goes to zero in proportion to $\sqrt{(\gamma/\gamma')(1-\omega_o)/(1-\hat{g})}$.

Finally, we'll use an elementary solution to show how scattering affects the vertical distribution of solar absorption. We'll suppose that thermal emission is negligible at the frequency under consideration, as is the case for solar radiation on planets at Earthlike (or even Venus-llke) temperatures. The basic idea is that scattering increases the net path traveled by radiation in going from one altitude to another, because the radiation bounces back and forth many times rather than proceeding in a straight line. This allows more radiation to be absorbed within a thinner layer, as compared to the no-scattering case. At the same time, however, scattering reflects some radiation back to space before it has any opportunity to be absorbed at all. As will be shown in the forthcoming derivation, the net result is to reduce the solar absorption while at the same time concentrating it more in the upper atmosphere, as compared to a no-scattering case with the same distribution of absorbers.

The full problem with arbitrary surface albedo, optical depth, and asymmetry factor is analytically tractable so long as ω_o is constant. The full solution is somewhat unwieldy, however, so we'll now make a few simplifying assumptions; the general problem is explored in Problem ??.

To keep the algebra simple, in the present discussion we'll assume τ_{∞}^* to be very large, so that the lower boundary does not affect the solution. The atmosphere is effectively semi-infinite (a top but no bottom) in this solution. We'll also assume that the asymmetry factor vanishes. The solution begins with taking the derivative of Eq 5.29 and substituting from Eq. 5.28, which gives us

$$\frac{d^2}{d\tau^{*2}}(I_+ + I_-) = -K^2(I_+ + I_-) - 2\gamma\omega_o L_\odot \exp(-(\tau_\infty^* - \tau^*)/\cos\zeta))$$
(5.45)

where K is defined by Eq. 5.43 with \hat{g} set to zero. A particular solution to this equation is

$$I_{+} + I_{-} = -\frac{2\gamma\omega_{o}L_{\odot}\cos^{2}\zeta}{1 - K^{2}\cos^{2}\zeta}$$
(5.46)

to which we have to add superpositions of the two homogeneous solutions $\exp(\pm K(\tau_{\infty}^* - \tau^*))$ so as to satisfy the boundary conditions at the top and bottom of the atmosphere. So far the only assumptions we have used are that ω_o is constant, the thermal emission is neglected, and $\hat{g} = 0$. Since the atmosphere is semi-infinite, the only admissible homogeneous solution is $a \exp(-K(\tau_{\infty}^* - \tau^*))$, since the other solution blows up deep in the atmosphere. It remains only to determine a, which is done by applying the condition $I_- = 0$ at the top of the atmosphere. To do this we need $I_+ - I_-$. This is obtained from Eq 5.29, which takes on a particularly simple form when $\hat{g} = 0$. Using the value of a thus obtained, the net vertical diffuse flux is found to be

$$I_{+} - I_{-} = \left[-\frac{K}{2\gamma + K} \frac{1 + 2\gamma \cos \zeta}{1 - K^{2} \cos^{2} \zeta} \exp(-K(\tau_{\infty}^{*} - \tau^{*})) + \frac{1}{1 - K^{2} \cos^{2} \zeta} \exp(-(\tau_{\infty}^{*} - \tau^{*})/\cos \zeta) \right] \omega_{o} L_{\odot} \cos \zeta$$
(5.47)

Since $I_{-} = 0$ at the top of the atmosphere, the albedo is obtained by evaluating this expression at $\tau^* = \tau_{\infty}^*$ and taking the coefficient of the incoming flux $L_{\odot} \cos \zeta$. Thus,

$$\alpha = \frac{2\gamma\omega_o}{(1+K\cos\zeta)(2\gamma+K)} = \frac{2\gamma\omega_o}{(1+2\sqrt{\gamma\gamma'(1-\omega_o)}\cos\zeta)(2\gamma+2\sqrt{\gamma\gamma'(1-\omega_o)})}$$
(5.48)

In the absence of scattering, all the incident flux should be absorbed no matter how low the concentration of absorbers, since the atmosphere is assumed infinitely deep. Consistently with this reasoning the above albedo approaches zero as $\omega_o \to 0$. For small ω_o , the albedo increases linearly with ω_o . It continues to increase monotonically as ω_o is further increased. In the limit $\omega_o \to 1$ where scattering becomes very strong, $\alpha \to 1$ and the atmosphere becomes perfectly reflecting; radiation is scattered back to space before it has much opportunity to be absorbed in the atmosphere.

Though strong scattering reduces the opportunity for absorption, it also reduces the depth scale over which the small amount of absorbed radiation is deposited in the atmosphere. The reason is that scattering increases absorption through multiple reflections that increase the path length. To get a better handle on what is going on, we need to examine the vertical profile of the flux as $\omega_o \to 1$ while holding the concentration of absorbers fixed. Taking the limit this way would correspond, for example, to looking at ultraviolet absorption as we increase the amount of conservatively scattering cloud particles in an atmosphere while keeping the amount of ultravioletabsorbing ozone fixed. This is equivalent to writing $\tau_{\infty}^* - \tau^* = (\kappa/(1 - omega_o))(p/g)$, if we neglect pressure broadening, since $(1 - \omega_o)\Delta\tau$ gives the absorption in a layer of thickness $\Delta\tau$. It follows from this expression for optical thickness that all the direct beam flux is converted to diffuse flux in a very thin conversion layer if $\omega_o \to 1$. Below the conversion layer all flux is diffuse, and the net vertical flux is then

$$I_{+} - I_{-} = -\frac{K}{2\gamma + K} \frac{1 + 2\gamma \cos \zeta}{1 - K^{2} \cos \zeta} \exp(-K \frac{1}{1 - \omega_{o}} \frac{\kappa p}{g}) \omega_{o} L_{\odot} \cos \zeta$$

$$\approx \sqrt{\frac{\gamma}{\gamma'}} (1 - 2\gamma \cos \zeta) \sqrt{1 - \omega_{o}} \exp(-2\sqrt{\gamma\gamma'} \frac{1}{\sqrt{1 - \omega_{o}}} \frac{\kappa p}{g}) L_{\odot} \cos \zeta$$
(5.49)

Hence the flux which manages to penetrate into the atmosphere is absorbed over a layer depth which scales with $\sqrt{1-\omega_o}$, and approaches zero as $\omega_o \to 1$. It follows also from Eq. 5.49 that the heating rate $d(I_+ - I_-)/dp$ remains order unity in this limit, though it becomes concentrated in a thinner and thinner layer near the top of the atmosphere.

5.7 Numerical solution of the two-stream equations

Eq. 5.27 is a coupled, linear system of ordinary differential equations requiring two boundary conditions. It takes the form of a two-point boundary value problem, because one specifies a boundary condition on I_+ at $\tau^* = 0$ and on I_- (generally that it vanish) at $\tau^* = \tau_{\infty}$. The solutions can in principle be constructed from solutions of a set of initial value problems: first one integrates the full inhomogeneous system downward from the top, specifying I_{-} at the top, and obtains a particular solution. This doesn't satisfy the bottom boundary condition on I_+ , so one then computes a solution to the homogeneous (unforced) equation also integrating top-down subject to the condition on I_{-} at the top. Then one simply superposes the right amplitude of the homogeneous solution to satisfy the boundary condition at the ground. A similar procedure can be carried out integrating from the ground upward instead. Either approach can be carried out very simply and with high accuracy, using any standard algorithm for integrating ordinary differential equations. Unfortunately, the numerical solution breaks down when the atmosphere becomes very optically thick. The reason is that the homogeneous equation we are solving has exponentially growing solutions in both directions. When the atmosphere is optically thick, these grow by many orders of magnitude, and come to dominate both the particular and homogeneous solutions, so round-off error wrecks our ability to satisfy the boundary condition. With 64 bits of precision, the above solution method tends to break down at optical depths of 40 or more. A related problem in the optically thick limit is that of "stiffness." When the atmosphere is very optically thick, one is forced to use a great number of points to represent the atmosphere so as to avoid numerical instability, even if the solution one is seeking varies only moderately over the depth of the atmosphere.

5.8 Water and ice clouds

Now we'll take a closer look at the way Earth's water and water-ice clouds effect the radiation budget, taking account of the balance between the shortwave albedo effect of clouds which act to cool the planet and the longwave cloud greenhouse effects which act to warm the planet. Much of the general behavior in evidence on Earth applies equally well to water clouds on other planets, or for that matter to any cloud-forming substance which is strongly absorbing in the infrared but fairly transparent in the solar spectrum. To set the stage, we'll first discuss some calculations with the ccm radiation code which show how high clouds affect the albedo and OLR under typical tropical conditions. The results are shown in Fig. 5.8. These calculations include most of the radiative effects operating in the real Tropics, including solar and infrared absorption by water vapor and CO_2 , though we have left ozone out of the picture. The surface albedo has been set to zero, so as to focus on the reflective effect of the cloudy atmosphere itself. In this calculation, the tropospheric temperature profile is on the moist adiabat, and we place a geometrically thin cloud with specified water content in the upper troposphere, where the pressure is 283 mb and the temperature T_c of the cloud is 243 K. At these temperatures, the cloud is primarily composed of ice, and the ccm radiation model makes use of the complex index of refraction appropriate to water-ice particles. Results are given as a function of the condensed water path of the cloud. As long as the cloud is geometrically thin enough to be essentially isothermal, the actual geometrical thickness of the cloud is irrelevant to this calculation.

For 30 micrometer particles, which are typical of actual tropical ice clouds, it only takes a path of 50 g/m^2 to make the cloud act like a blackbody. The OLR is about 20 W/m^2 below the blackbody emission σT_c^4 of the cloud itself, because there is some water vapor and CO_2 greenhouse effect in the colder air above the cloud. This effect would disappear if the cloud were placed at the minimum temperature part of the atmosphere, and would increase if the cloud were lower. In essence, a cloud that is optically thick in the infrared acts like a new "ground", radiating upward into the upper part of the atmosphere with a blackbody temperature T_c . If the particles are made smaller, it takes less cloud water in order to make the cloud optically thick, because the same mass of water yields more aggregate cross-section area of cloud particles. As a practical matter, most high clouds occuring in the vicinity of deep convection in the Tropics can be considered optically thick in the infrared. The associated cloud greenhouse effect is enormous, and would lead to an uninhabitably hot planet if not compensated by shortwave albedo effects that are of similar magnitude.

The albedo effect of the cloud also increases monotonically with cloud water content, but at a much slower rate than the greenhouse effect, for the reasons discussed in Section 5.6. For 30 micrometer particles, the planetary albedo has only reached .2 when the cloud water path is $50q/m^2$. The albedo doesn't reach .5 until the cloud water content approaches 200 q/m^2 . On the other hand, the failure of cloud albedo to saturate until large cloud water paths means that the particle size can have a very important influence on albedo; reducing the particle size to 10 micrometers increases the albedo to .7 for a cloud containing 200 g/m^2 of ice. As for compensation between shortwave and longwave cloud effects, taking a cloud with 30 micrometer particles and 100 g/m^2 , the albedo is about .4, which yields 170 W/m^2 reduction in solar absorption based on typical annual average tropical insolation. This compares with a cloud greenhouse effect of $120W/m^2$, so such clouds have a moderate net cooling effect. If the cloud only had a water content of 50 q/m^2 , though, the cloud greenhouse effect would be nearly the same but the cloud albedo effect is reduced by nearly half, and the cloud would have a net warming effect. Similarly, if the ground were partially reflecting (owing to vegetation cover or low-lying clouds), the change in albedo due to high clouds would be reduced, shifting the balance again in favor of net cloud warming. On the other hand, reducing the particle size of the clouds makes them much brighter, making it easier for the clouds to have a net cooling effect.

For fixed particle size, the cloud altitude has relatively little effect on albedo for a given cloud condensed water path. Low altitude liquid water clouds tend to have smaller particles than ice clouds as well as larger water content (because there is more water around to condense), and are correspondingly more reflective. Because of this effect, the balance of power for mid-level and low-level clouds shifts decidedly toward a net cooling effect on the planet.



Figure 5.8: Albedo and OLR as a function of cloud condensed water path, for a high ice cloud with temperature $T_c = 242K$ at a pressure $p_c = 283mb$. The temperature profile is on the moist adiabat corresponding to a surface temperature of 300K, patched to an isothermal 180K stratosphere. The relative humidity is 50% and the CO_2 concentration is 300ppmv, but there is no ozone in the atmosphere. Calculations were done with the ccm radiation code, and results are shown for both 30 micrometer and 10 micrometer particles.

High clouds have both a warming and a cooling effect, and which one wins depends on the detailed of the cloud properties, including cloud temperature, particle size, and condensed water content. Colder cloud temperatures tend to favor net warming. Making particles smaller or increasing cloud water enhances the cooling effect except for very thin clouds, since it takes little cloud water to make the cloud act like a blackbody whereas the albedo continues to increase with cloud water increase or particle size decrease, for the reasons discussed in Section 5.6. The balance between albedo effect and greenhouse effect also depends on the intensity of solar radiation, since albedo matters not at all if there is no sunlight. In the polar night, clouds have an unambiguous warming effect as long as they are not right at the surface. Similarly, the cloud albedo effect depends on the albedo of the underlying surface; clouds over a reflective surface like ice (or surface clouds!) will tend to have a warming effect, as also discussed in Section ??. As the cloud is made lower, the cloud greenhouse effect is attenuated, because the cloud temperature is closer to the ground temperature and also because (especially in moist regions) the greenhouse effect of the clear air above the clouds masks the longwave radiative effect of the cloud itself.

Although the large and competing effects of clouds on OLR and albedo pose similar challenges on any planet whose atmosphere contains a condensible substance, Earth is the only case at present for which we have good observations of the net radiative effect of clouds. The first satellite mission to do this accurately was the Earth Radiation Budget Experiment (ERBE), and subsequent missions have taken a similar approach. We discussed some ERBE clear-sky results in Chapter ??, and now we will see what ERBE has to tell us about cloud effects. The ERBE mission measured the Earth's radiation budget using two sets of highly accurate broadband radiometers borne on satellites – one in the infrared spectrum and one in the shortwave (i.e. solar) spectrum. Moreover, the processing algorithm made use of the patchiness of Earth's cloud cover in order to estimate the effect of clouds on the longwave and shortwave radiation. Within each scene examined (think of a scene as a 50km square patch of the Earth's surface) the algorithm identified those pixels which represented cloud-free clear-sky conditions, and defined "clear sky"

longwave and shortwave flux as the value the flux over the scene would have if the flux of *all* pixels in the scene were replaced by the average of the clear-sky pixels. In the longwave, for example, the ERBE retrieval reports the all-sky OLR, called OLR_{all} and the clear-sky OLR, called OLR_{clear} . The cloud longwave forcing is then defined as $OLR_{clear} - OLR_{all}$. Since clouds reduce the OLRby making the upper troposphere more optically thick, the cloud longwave forcing is positive, and represents a warming effect. Similarly, the cloud shortwave forcing is defined as $S_{abs,all} - S_{abs,clear}$, where S_{abs} is the top-of-atmosphere absorbed solar radiation – the difference between incoming and reflected solar radiation. Clouds reduce the solar absorption by increasing the albedo, and so the cloud shortwave forcing thus defined is generally negative, representing a cooling effect. The sum of the cloud longwave and cloud shortwave forcings is the net cloud forcing, with positive values representing a warming tendency and negative values representing a cooling tendency. Clear sky and all-sky albedo can be defined similarly.

Results for clear and cloudy albedo, and for the cloud radiative forcing are shown for the year 1988 in Figure 5.9. Other years show a similar pattern. The ERBE dataset contains information of this sort for each month, reported on a latitude-longitude grid. Here we show only annual-mean results averaged along latitude circles. The full monthly-mean dataset for all available years is provided as part of the dataset collection in the supplementary materials for this book. Turning attention first to the clear-sky albedo, we see that without clouds the albedo varies in a narrow range of .11 to .16 from 60S latitude to 42N. Poleward of 60S, the albedo increases sharply owing to the high albedo Antarctic ice. Still, the values indicate that the albedo of the partially snowcovered Antarctic ice must exceed .7, since the atmospheric absorption makes the planetary albedo lower than the surface albedo. The clear-sky estimates near the pole are somewhat unreliable, since it is hard to distinguish between low clouds and ice. Going towards the North pole from 42N the albedo increases somewhat more gently, owing to the patchy distribution of sea ice and its seasonal fluctuations; the rest of the Northern high-latitude albedo increase is due to winter snow-cover over land. Clouds have a strong reflective effect, approximately doubling the tropical albedo and increasing the midlatitude albedo to .4 or more. The area-weighted mean albedo is .19 for clear sky and .33 including cloud effects. Area-weighting doesn't take into account the seasonal and latitudinal distribution of sunlight though; a more appropriate mean albedo is based on taking the ratio of global net reflected solar radiation to the incident radiation. This estimate yields somewhat smaller values: a mean clear-sky albedo of .16 and a mean all-sky albedo of .30.

If uncompensated by the cloud greenhouse effect, the high albedo of clouds would probably be sufficient to throw the Earth into a Snowball state. In reality, the reduction of OLR by clouds cancels most of the cloud cooling effect, as shown in the right hand panel of Figure 5.9. The cloud longwave forcing - i.e. the reduction in OLR due to clouds - is anti-correlated with the cloud shortwave forcing, and has sufficient magnitude to cancel most of the cloud shortwave forcing. The distribution of cloud forcing takes us into some consideration of aspects of the general circulation we have not introduced previously. Somewhat North of the Equator there is a region of deep convection yielding deep, thick clouds, which is manifest in the Figure as a peak in both the cloud longwave and shortwave forcing (marked "ITCZ" in the figure for Inter-Tropical Convergence Zone, in honor of the winds which converge moisture into this region and feed the convection). The ITCZ is flanked by two subtropical regions where convection is suppressed by downward motions in the atmosphere, and is shallow or absent. Here one encounters local minima in both the cloud longwave and shortwave forcing. Throughout the tropics, the two terms sum to a net cooling effect of about $-20W/m^2$, which is stronger in the subtropics than near the ITCZ. The subtropical cloud cooling is in part due to near-surface clouds which are associated with the boundary layer rather than deep convection. In the midlatitudes, there is another region of deep cloud activity. This one is associated with the large scale organized storm tracks, which loft water from the subtropical ocean and move it poleward and upward. The albedo effect of clouds more strongly dominates



Figure 5.9: Zonally averaged annual mean clear and cloudy sky albedo(left panel) and cloud radiative forcing (right panel) measured by ERBE for the year 1988.

the greenhouse effect in this region, and even more so towards the Antarctic region, where there is strong cloud shortwave forcing associated with low-lying marine stratus clouds. As a result there is strong net cooling in the midlatitude and polar regions.

The area-weighted global mean cloud longwave forcing is $28W/m^2$, while the mean cloud shortwave forcing is $-47W/m^2$, which nets out to a cooling influence of $-19W/m^2$. Using a sensitivity factor of $2.2W/m^2K$ from Section ??, we conclude that the Earth would be about 8.6K warmer if there were no clouds.

In circumstances under which the clear-sky regions do not absorb much solar radiation, high clouds have a potent net warming effect, though there must be enough convection around to loft water to sufficiently high altitudes to make a high optically thick cloud. As we have already mentioned, clouds can contribute significantly to deglaciation of the high-albedo Snowball Earth, though the main question there is whether it is possible to make sufficiently high clouds with sufficiently great water content in an atmosphere with low water content (because of low temperature) and sluggish convection (because of low solar absorption). Another situation in which the clear-sky solar absorption is low is the high-latitude winter. Here, there is little solar radiation to reflect from clouds simply because it is night or twilight most of the time, and so if there are high clouds they will have a pronounced winter warming effect, and perhaps even inhibit the formation of sea ice in open water conditions. This effect may play a role in the Arctic during the Cretaceous hothouse climate, since there is open water in the Arctic Ocean which can maintain a supply of relatively warm water throughout the winter to feed deep convection. It seems plausible that this mechanism would help explain the mysterious low-gradient climate of Cretaceous and similar hothouse climates, described in Section ??. General circulation models to date do not support a sufficiently strong cloud effect for clouds to be the answer to the Cretaceous puzzle, but there is much remaining to be learned about clouds, so the last word has not by any means been uttered on this topic. This potential mechanism is only viable when there is open water in the polar ocean. Over a polar continent, such as Antarctica, the ground would cool off rapidly in the Winter, foreclosing any serious possibility of deep convection and the associated deep clouds.

The effect of clouds on the water-vapor runaway greenhouse represents one of the most vexing
and important unresolved issues in planetary climate. The observed behavior of Earth clouds can provide little guidance as to cloud effects in a much warmer atmosphere in which water vapor is the dominant component. It seems plausible at this point that water clouds would inhibit the runaway greenhouse at least for Early Venus conditions, but a definitive answer must await attainment of a far better understanding of cloud microphysics and cloud fraction under near-runaway conditions.

5.9 Things that go bump in the night: Infrared-scattering with gaseous absorption

- 5.10 Effects of atmospheric solar absorption
- 5.11 For Further Reading

Chapter 6

The Surface Energy Balance

6.1 Formulation of the surface flux problem

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

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

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

height.

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

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

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

If the atmosphere is very optically thin in the longwave spectrum, the OLR is determined entirely by the ground temperature and ground emissivity. Further, since an optically thin atmosphere radiates very little, the only way the atmosphere itself loses energy is through turbulent exchange with the surface. Suppose first that the atmosphere is transparent to solar radiation. In that case, in equilibrium the net turbulent exchange between atmosphere and surface must vanish, since otherwise the atmospheric temperature would rise or fall, there being nothing to balance a net exchange. In consequence, the ground temperature will be just what it would have been without an atmosphere despite the presence of turbulence. In this case, one determines the ground temperature as if the planet were in a vacuum, the top of atmosphere budget is automatically satisfied, and then, once T_g is known, the surface budget is used to determine T_{sa} , and (via an adiabat) the rest of the atmospheric structure. It is exactly the inverse of the process used in the optically thick case. In fact, the basic picture is little altered even if the atmosphere absorbs solar radiation. In that case, the requirement that the atmosphere be in equilibrium implies that any solar radiation absorbed in the atmosphere be passed on to the surface by turbulent fluxes. The result is much the same as if the solar radiation were absorbed directly by the surface; one does the ground temperature calculation as before, but simply remembers to add the atmospheric absorption to the solar energy directly absorbed by the ground. It should be kept in mind that these considerations apply only in equilibrium. Even an optically thin atmosphere can affect the transient behavior of the surface (e.g. in the diurnal or seasonal cycle), as will be discussed in Chapter 8.

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

6.2 Radiative exchange

6.2.1 Shortwave radiation

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

6.2.2 The behavior of the longwave back-radiation

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

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



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

According to Eq. 4.21, $I_{-,s}$ approaches σT_{sa}^4 when the atmosphere is optically thick throughout the infrared. In order to characterize the optical thickness of the atmosphere, we introduce the effective low level atmospheric emissivity e_a , defined so that $I_{-,s} = e_a \sigma T_{sa}^4$. e_a depends on the temperature profile as well as the optical thickness, as illustrated by Eq. 4.21 in the optically thick limit. When $T_g = T_{sa}$ the surface cooling becomes $e_g \cdot (1 - e_a)\sigma T_g^4$, which vanishes in the optically thick limit where $e_a \rightarrow 1$. Let $e^* = (1 - e_a)$; this is the effective emissivity of the ground when the air temperature equals the ground temperature. If the air temperature is not too different from the ground temperature, we may linearize the term σT_q^4 about $T_g = T_{sa}$, which results in

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

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

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

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

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

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

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

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

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

6.2.3 Radiatively driven ground-air temperature difference

Now we consider the equilibrium temperature difference between the ground and the overlying air that would be attained in the absence of turbulent heat exchange. This temperature difference is important in determining the extent to which convection is driven from below, by positive buoyancy generated near the ground. We have already discussed this issue for the case in which the atmosphere itself is in pure radiative equilibrium (See 3.6,4.3.4 and 4.8). Our concern now is with what happens once convection has set in and altered the atmospheric temperature profile.

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

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

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

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

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

$$(T_g - T_{sa}) = \frac{1}{4} \frac{(1 - \alpha_g) S_g}{\sigma e_g T_{sa}^4} T_{sa} - \frac{1}{4} e^* T_{sa}$$

$$= \frac{1}{4} ((\frac{T_o}{T_{sa}})^4 - e^*) T_{sa}$$
(6.5)

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

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

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

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

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

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

whence

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

Note that we have not yet made use of the assumption that the atmosphere is optically thin. For an optically thick atmosphere with a very strong greenhouse effect (like Venus), $a_{+} \approx e_{a} \approx 1$ and $e_{a,top} \approx 0$, and so we recover our previous result that $T_{sa} \approx T_{g}$ for such an atmosphere, provided the emissivity of the ground is close to unity. For an optically thin atmosphere, $a_{+} e_{a,top}$ and e_{a} are all small, so one needs to know precisely how small the absorption coefficient is relative to the two emission coefficients. For an isothermal atmosphere –whether grey or not– Eq. 4.9 implies $e_{a,top} = e_{a}$. For a grey atmosphere, it follows in addition that $a_{+} = e_{a,top} = e_{a}$. In this case $T_{sa} = T_{g}/2^{1/4}$, reproducing the result of Section 3.6. When the atmosphere is not grey, the absorption coefficient differs somewhat from atmospheric emission coefficient, because the spectrum of the upwelling radiation from the ground is different from that of the atmospheric emission (by virtue of the difference between ground temperature and air temperature). However, the deviation from the grey gas result is typically modest for an isothermal atmosphere. For example, a 7mb Marslike pure CO_2 atmosphere with a uniform temperature of 230K has $a_+ = e_{a,top} = e_a \approx .14$

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

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

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

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

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

6.3 Basic models of turbulent exchange

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

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

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

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

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

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

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

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

6.3.1 Sensible heat flux

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

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

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

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

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

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

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

6.3.2 Latent heat flux

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

When the surface is sufficiently cold relative to the air, vapor from the air can be deposited on the surface in the form of dew or frost. In this case the latent heat flux is negative, and carries energy from the atmosphere to the ground. If the boundary layer is saturated ($h_{sa} = 1$) then frost or dew deposition occurs whenever $T_g < T_{sa}$. If the boundary layer is unsaturated deposition won't occur until the ground temperature is made sufficiently cold that the saturation vapor pressure there falls below the partial pressure of the condensible in the overlying atmosphere (a temperature known as the "dew point" or "frost point"). When latent heat is being carried to the surface – as it is during he seasonal polar CO_2 frost formation on Mars – the rate of condensation is limited by the rate at which the surface can get rid of the deposited latent heat. Since the surface is colder than the atmosphere during deposition, sensible heat flux carries heat the wrong way to balance the budget, so it is only infrared cooling of the surface that can sustain frost or dew. Otherwise, the surface will simply warm in response to the deposited latent heat until it is no longer cold enough for frost or dew to form.

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

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

When evaporation dominates the surface budget, equilibrium requires $F_L = 0$, or equivalently $p_{sat}(T_g) = h_{sa}p_{sat}(T_{sa})$. Since p_{sat} is monotonically increasing in temperature, this relation requires $T_g < T_{sa}$ if the boundary layer air is unsaturated $h_{sa} < 1$. Thus, evaporation or sublimation drives the ground temperature to be *colder* than the overlying air temperature. However, the ground and surface could also achieve equilibrium by transferring enough moisture to the surface layer that it becomes saturated $(h_{sa} = 1)$, in which case $T_g = T_{sa}$ in equilibrium, as for the case of sensible heat flux. The extent to which equilibrium is attained by adjusting temperature vs. humidity depends on the competition between the rate at which moisture is supplied to the boundary layer and the rate at which dry air from aloft is entrained into the boundary layer. Observed boundary layers on Earth and Titan are significantly undersaturated, leading to the conclusion that the ground temperature would be considerably less than the air temperature, if other fluxes did not intervene. Using the linearized form of the latent heat, the equilibrium ground-air temperature difference is $T_g - T_{sa} \approx -E_o/b_L$. For the conditions of Table 6.1, this is -2.6K for Titan at 95K. For a hot Earth at 320K, the difference is about -5.7K. There are currently no observations of the state of saturation over the sublimating Martian CO_2 frost cap, but given the saturation assumed in the table the equilibrium occurs with $T_g - T_{sa} \approx -2.4 K$ when the air temperature is 260K. Thus, even when evaporation dominates, the equilibrium ground temperature does not differ greatly from the overlying air temperature. This was also found to be the case when the surface budget is dominated by sensible heat flux. It is only the radiative terms that can drive the ground temperature to be substantially different from the overlying air temperature.

6.4 Joint effect of the fluxes on surface conditions

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

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

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

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

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

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



Figure 6.2: **CAPTION

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

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

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

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

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

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

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

6.5 Monin-Obukhov theory

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

6.6 Mass balance and melting

6.7 Precipitation-temperature relations

Here we discuss the factors constraining precipitation. The discussion follows the lines of my discussion of the subject in my Nature article on deep-time hydrological cycle problems. The general theme is the importance of the surface budget constraint for precipitation, when the temperature is high.

6.8 Simple models of sea ice in equilibrium

6.9 For Further Reading

Chapter 7

Radiative-convective models

7.1 Dry grey-gas atmospheres

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

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

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

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

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

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

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

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

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

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

7.3 Real gas atmospheres

Numerical solutions for radiative-convective equilibrium for real gases.

7.4 Sensitivity of climate to CO_2 changes

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

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

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

7.5 Methane-dominated greenhouse

7.6 The atmosphere as a heat engine

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

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

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

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

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

7.7 Effect of the diurnal cycle on tropopause height

7.8 The runaway greenhouse revisited

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

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

Role of cloud effects in inhibiting true runaway state.

7.9 Mars, present, past and future

Greenhouse effect of the present thin Martian atmosphere.

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

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

Early Mars. Problems with the gaseous CO_2 greenhouse as an explanation (owing to CO_2 condensation). Need for consideration of CO_2 clouds, which will be taken up in next chapter.

7.10 Titan

The Methane greenhouse effect on Titan. Methane condensation feedback (analogous to water vapor feedback). Note that Nitrogen also condenses at Titan temperatures, so that mass of atmo-

sphere changes with the climate.

Since Methane clouds are also infrared scatterers rather than absorbers, a full treatment of Titan's climate also requires IR scattering theory, developed in next chapter. On Titan (and perhaps Early Earth), high altitude tholin clouds are also important.

7.11 Gas Giants

Two features distinguish the problem for gas giants like Jupiter or Saturn: (1) There is no surface to absorb solar radiation, so solar energy is deposited entirely by internal absorption in the Atmosphere; (2) Interior fluid motions can efficiently transport heat from the deep interior to the upper layers, and can yield a heat source comparable to solar absorption. The latter effect is weak on solid planets like Earth, since fluid motions in the very viscous mantle are sluggish.

This section explores the consequence of the above features, in radiative-convective models. The use of planetary energy balance observations in determining the relative role of interior and Solar heat sources was already discussed in Chapter 3.

Chapter 8

Variation of temperature with season and latitude

Why is the Earth generally hotter near the Equator than at the poles? Why is it generally hotter in Summer than in Winter, especially outside the tropics? Would this be true on other planets as well? How would the pattern change over time, as features of the planet's orbit vary? More generally, in this section we seek to understand the features of a planet that determine the magnitude and pattern of geographic and seasonal variations in temperature.

8.1 A few observations of the Earth

First, let's take a look at how the Earth's surface temperature varies with the seasons. Figure 8.1 shows the zonal-mean air temperature near the surface for representative months in each of the four seasons. The first thing we note is that the temperature is fairly uniform in the tropics (30S-30N), but declines sharply as the poles are approached. The temperature difference between the Equator and 60N is 39K in the Winter but only 12K in the Summer. The Southern Hemisphere has a much weaker seasonal cycle, except over the Antarctic continent: The temperature difference between the Equator and 60S is 26K in the Winter and 22K in the Summer. However, over Antarctica, poleward of 60S the seasonal cycle is extreme. Noting that the Northern Hemisphere has more land than the Summer, the data imply that the oceans have a strong moderating effect on the seasonal cycle. The temperature patterns in Figure 8.1 are what we seek to explain in terms of the response of climate to the geographically and seasonally varying Solar forcing.

An even better appreciation of the effect of land masses on the seasonal cycle can be obtained by examining the map of July-January temperature differences, shown in Figure 8.1. This map shows that the strongest seasonal temperature contrast occurs in the interior of large continents, and that the ocean temperature varies by at most a few degrees over the year – and even less in the Tropics. The strong seasonal cycle of the Northern Hemisphere continents extends very little beyond the coastlines, and the seasonal cycle of the Northern oceans has similar magnitude to that of the more extensive Southern oceans.



Figure 8.1: Observed zonal mean surface air temperatures for January, April, July and October. Computed from NCEP data for 1970-2000.

8.2 Distribution of incident solar radiation

The geographical variations of temperature are driven by variations in the amount of sunlight falling on each square meter of surface, and also by variations in albedo. Seasonal variations are driven by changes in the geographical distribution of absorbed sunlight as the planet proceeds through its orbit. Therefore, the starting point for any treatment of seasonal and geographical variation must be the study of how the light of a planet's sun is distributed over the spherical surface of the planet. This section dealso only with the distribution of incident sunlight, or *insolation*. The geographical distribution of the amount of sunlight absorbed is affected also by the distribution of the albedo. The albedo variations can also affect the seasonal distribution of solar forcing through seasonal variations in ice, snow, cloud and vegetation cover.

It will help to first consider an airless planet, so that we don't at once have to deal with the possible effects of scattering of the solar beam by the atmosphere. If our planet is far from its Sun, as compared to the radius of the Sun, the sunlight encountering the planet comes in as a beam of parallel rays with flux L. Even if the surface of the planet is perfectly absorbing, the sunlight the planet intercepts is not spread uniformly over its surface; per unit area, parts of the planet where the sun is directly overhead receive a great deal of energy, whereas parts where the Sun grazes the surface at a shallow angle receive little, because the small amount of sunlight intercepted is spread over a comparatively large area, as shown in Figure 8.2. The night side of the planet, of course, receives no solar energy at all.

To obtain a general expression for the distribution of incident solar radiation per unit of surface area, we may divide up the surface of the planet into a great many small triangles, and consider each one individually. The solar energy intercepted by a triangle is determined by the area of the shadow that would be cast by the triangle on a screen oriented perpendicular to the solar beam. To compute this area, suppose that one of the vertices of the triangle is located at



Seasonal Cycle 1970-1990

Figure 8.2: Map of July-January surface air temperature difference.



Figure 8.3:

the origin, and that the two sides coming from this vertex are given by the vectors $\vec{r_1}$ and $\vec{r_2}$ By the definition of the cross product, the area of the triangle is given by $2A\hat{n} = (\vec{r_1} \times \vec{r_2})$ where \hat{n} is the unit normal to the plane containing the triangle. To obtain the area of the shadow cast by the triangle, we apply the cross product to the projection of the vectors $\vec{r_1}$ and $\vec{r_2}$ onto the plane. These projections are given by $\vec{r_1} - \hat{z}\vec{r_1} \cdot \hat{z}$ and $\vec{r_2} - \hat{z}\vec{r_2} \cdot \hat{z}$, where \hat{z} is the unit vector pointing in the direction of the Sun. The cross product of these two vectors is

$$(\vec{r_1} - \hat{z}\vec{r_1} \cdot \hat{z}) \times (\vec{r_2} - \hat{z}\vec{r_2} \cdot \hat{z}) = \vec{r_1} \times \vec{r_2} - (\hat{z} \times \vec{r_2})(\vec{r_1} \cdot \hat{z}) - (\vec{r_1} \times \hat{z})(\vec{r_2} \cdot \hat{z})$$
(8.1)

Now, the cross product of two vectors in the xy plane must point in the direction of the z axis. Hence, we can obtain the magnitude of the above vector by taking its dot product with \hat{z} . This is very convenient, since the dot product of \hat{z} with the second two terms vanishes, leaving us with

$$2A_{\perp} = \hat{z} \cdot (\vec{r_1} \times \vec{r_2}) = 2A\hat{z} \cdot \hat{n} = 2A\cos(\zeta) \tag{8.2}$$

where A_{\perp} is the area of the shadow and ζ is the angle between the normal to the patch of surface and the direction of the sun. This is known as the *zenith angle*. When the the zenith angle is zero, the Sun is directly overhead, and when it is 90° the sunlight comes in parallel to the surface and leaves no energy behind. Zenith angles greater than 90° are unphysical, since they represent light that would have to pass through the solid body of the planet in order to illuminate the underside of the surface; these are on the night side of the planet. If one draws a line from the center of the planet to the center of the Sun, the zenith angle will be zero where the line intersects the surface of the planet; this is the subsolar point. At any given instant, the curves of constant zenith angle make a set of concentric circles centered on the subsolar point, with a zenith angle of 90° along the great circle which at the given instant separates the dayside of the planet from the nightside. If the surface were in equilibrium with the instantaneous incident solar flux, the subsolar point would be the hottest spot on the planet, with temperature falling to zero with distance away from the hot spot. As the planet rotates through its day/night cycle, a given point of the surface is swept through a range of distances from the hot spot, leading to a diurnal temperature variation. As the planet proceeds through its orbit in the course of the year, the diurnal cycle will change as the orientation of the planet's rotation axis changes relative to the Sun. Insofar as the surface actually takes a finite amount of time to heat up or cool down, the diurnal cycle will be attenuated to one extent or another.

As the next step toward realism, let's now consider a rapidly rotating planet whose axis of rotation is perpendicular to the line connecting the center of the planet to the center of its Sun. If the axis of rotation is in fact perpendicular to the plane of the orbit, this situation prevails all year round; otherwise, the condition is met only at the equinoxes, and indeed the condition defines the equinoxes. We assume that the planet is rotating rapidly enough that the day-night difference in solar radiation is averaged out and the corresponding temperature fluctuations are small. In other words, the length of the day is assumed to be short compared to the characteristic thermal response time of the planet's surface, a concept which will be explored quantitatively in Section 8.3. Consider a small strip of the planet's surface near a latitude ϕ , of angular width $d\phi$. If a is the planet's radius, then the area of this strip is $2\pi a^2 \cos(\phi) d\phi$, if angles are measured in radians. The cross section area of the strip seen edge-on looking from the Sun determines the amount of solar flux intercepted by the strip. This area is $2a^2\cos^2(\phi)d\phi$ when $d\phi$ is small. In consequence, the incident solar radiation per unit area at latitude ϕ is $L\cos(\phi)/\pi$. At the Equator, the solar radiation per unit area is L/π , which is somewhat greater than the value L/4 which we obtained in Chapter 3 by averaging solar radiation over the *entire* surface of the planet. If the planet has no atmosphere to transport heat or create a greenhouse effect, the equilibrium temperature as a function of latitude is

$$T = \left(\frac{L\cos(\phi)}{\pi\sigma}\right)^{\frac{1}{4}} \tag{8.3}$$

The temperature has its maximum at the Equator, and falls to zero at the poles.

Exercise 8.2.1 For the geometric situation described above, derive an expression for the cosine of the zenith angle as a function of latitude and longitude. Re-derive the expression for the daily-average distribution of solar absorption by averaging the cosine of the zenith angle along latitude circles.

Now we turn to the general case, in which the axis of rotation of the planet is not perpendicular to the plane containing the orbit. The angle between the perpendicular to the orbital plane, and the planet's axis of rotation, is known as the *obliquity*, and we shall call it γ . It can be regarded as constant over the course of a planet's year, though there are longer term variations which will be of interest to us later. The task now is to determine the solar zenith angle as a function of latitude, position along the latitude circle, and time of year.

8.2. DISTRIBUTION OF INCIDENT SOLAR RADIATION

Let the point P be the center of the planet, and S be the center of the sun. If we draw a line from P to S, it will intersect the surface of the planet at a latitude δ , which is called the *latitude of* the sun, or sometimes the subsolar latitude. It is a function of the orientation of the planet's axis alone, and serves as a characterization of where we are in the march of the seasons. If the obliquity of the planet is γ , then δ ranges from γ at the Northern Hemisphere summer solstice to $-\gamma$ at the Southern Hemisphere summer solstice. Let Q be a point on the planet's surface, characterized by its latitude ϕ and its "hour angle" h, which is the longitude relative to the longitude at which local noon (the highest sun position) is occurring throughout the globe. For radiative purposes, we just need to compute the zenith angle ζ , defined previously. To get the zenith angle, we only need to take the vector dot product of the vector \vec{QS} and the vector \vec{PQ} . To do this, it is convenient to introduce a local Cartesian coordinate system centered at P, with the z-axis coincident with the axis of rotation, the x-axis lying in the plane containing the rotation axis and \vec{PS} , and the y-axis orthogonal to the other two, chosen to complete a right-handed coordinate system.

First, note that by the definition of the dot product,

$$\cos(\zeta) = \frac{\vec{PQ} \cdot \vec{QS}}{|PQ||QS|} \tag{8.4}$$

Further, $\vec{PQ} + \vec{QS} = \vec{PS}$, so

$$\cos(\zeta) = \frac{\vec{PQ} \cdot \vec{PQ}}{|PQ||QS|} + \frac{\vec{PQ} \cdot \vec{PS}}{|PQ||QS|} \approx \frac{\vec{PQ} \cdot \vec{PS}}{|PQ||QS|}$$
(8.5)

where we drop the first term based on the assumption that the radius of the planet is a small fraction of its distance from the Sun. For the same reason, |QS| in the denominator can with good approximation be replaced by |PS|, leaving the expression in the form of a dot product between two unit vectors. Letting $\hat{n}_1 = \vec{PQ}/|PQ|$ and $\hat{n}_2 = \vec{PS}/|PS|$, the unit vectors have the following components in the local Cartesian coordinate system.

$$\hat{n}_1 = (\cos(\phi)\cos(h), \cos(\phi)\sin(h), \sin(\phi)), \\ \hat{n}_2 = (\cos(\delta), 0, \sin(\delta))$$

$$(8.6)$$

whence

$$\cos(\zeta) = \cos(\phi)\cos(\delta)\cos(h) + \sin(\phi)\sin(\delta)$$
(8.7)

When $\cos \zeta < 0$ the sun is below the horizon.

The cosine of the zenith angle attains a maximum value $\cos(\phi - \delta)$ when h = 0, and a minimum value $-\cos(\phi + \delta)$ when $h = \pm \pi$. Both values are above the horizon when $|\phi| > |\pi/2 - \delta|$, corresponding to the perpetual polar summer day. Both values are below the horizon when $|\phi| > |\pi/2 + \delta|$, corresponding to the perpetual polar winter night. At the solstices, δ takes on its extreme values of $\pm \gamma$. Therefore, perpetual day or night are experienced at some time of year for latitudes poleward of $\pi/2 - \gamma$. These circles are known as the *Arctic* and *Antarctic* circles on Earth. Apart from the case of perpetual day or night, there is a *terminator* which separates the illuminated from the dark side of the planet. The position of the terminator is given by

$$\cos h_t = -\tan(\phi)\tan(\delta) \tag{8.8}$$

If Ω is the angular velocity of rotation of the planet, so that the planet's day is $T_{day} = 2\pi/\Omega$ time units, then the number of time units of daylight is $2h_t/\Omega = (h_t/\pi)T_{day}$. We shall adopt the convention $h_t = 0$ in the case of perpetual night, and $h_t = \pm \pi$ for perpetual day.

Exercise 8.2.2 For a given latitude ϕ , what δ yields the least hours of daylight? What δ yields the most hours of daylight? The Earth's present obliquity is 23.5 degrees, and it's length of day is 23.94hours. Sketch a plot of the maximum and minimum hours of daylight vs. latitude for the Earth.

The diurnal variations of the zenith angle lead to hot days and cold nights. Where the thermal response time is long enough to average out an appreciable portion of the diurnal temperature variation, the daily mean incident solar flux is an informative statistic. Since the incident solar flux per square meter of surface is $L \cos \zeta$, where L is the solar constant in W/m^2 , one can obtain the daily mean flux by averaging $\cos \zeta$ over a rotation period of the planet. This results in a nondimensional flux factor f, by which one multiplies the solar constant in order to obtain the daily mean solar radiation incident on each square meter of the planet's spherical surface. The daily average can be performed analytically, resulting in

$$f(\phi, \delta) = \frac{1}{2\pi} \int_{-h_t}^{h_t} \cos(\zeta) dh$$

= $\frac{1}{\pi} [\cos(\phi) \cos(\delta) \sin(h_t) + \sin(\phi) \sin(\delta) h_t]$ (8.9)

where h_t is determined by Eq. 8.8. This derivation of the daily average assumes that the length of the day is much less than the length of the year, so that δ may be regarded as constant over the course of the day. If the length of the day is a significant fraction of the length of the year, as is be the case for nearly tide-locked planets like Mercury or Venus, the expression still gives the correct average along the latitude circle, but this average is no longer identical to the time average over a day.

During the equinoxes, $\delta = 0$ and $f = \cos(\phi)/\pi$, independant of the obliquity. This agrees with the result we obtained earlier by direct geometrical reasoning. At other times of year, the daily mean flux is governed by two competing factors: the varying length of day, which tends to produce higher fluxes near the summer pole, and the average zenith angle, which tends to produce high fluxes near the subsolar latitude (which remains near the Equator if the obliquity is not too large). The latitude where the maximum daily mean insolation occurs is always between the subsolar latitude δ and the summer pole. For $\delta = 0$ the maximum occurs at the Equator, and a little numerical experimentation shows that the latitude of the maximum increases to about 43.4° when $\delta = 23.4^{\circ}$ (and similarly, with reversal of signs, in the Southern hemisphere). For larger δ , the length-of-day effect wins out over the slant angle effect at the pole, and the maximum occurs at the summer pole itself. This state of affairs just barely happens at the solstice for the present obliquity of Earth and Mars; as a result, the summer hemisphere solstice insolation is fairly uniform in these two cases. It is also useful to note that the daily mean insolation at the summer pole exceeds the daily mean insolation at the Equator when $|\delta| > 17.86^{\circ}$.

To obtain a general appreciation of the seasonal cycle, recall that δ varies from $-\gamma$ during the southern hemisphere summer solstice to γ at the northern hemisphere summer solstice, taking on a value of zero at the equinox which lies between the two solstices. Consider a planet with uniform albedo, so that the absorbed solar radiation is determined by the distribution of incident solar radiation. Suppose further that the thermal response time is long enough to average out the diurnal cycle, but short compared to the length of the year. If the obliquity is below 23.4°, the "hot spot" starts some distance poleward of the Equator in the Southern hemisphere, moves to the Equator as the equinox is approached, and then migrates a similar distance into the Northern hemisphere as the Northern summer solstice is aproached. If the obliquity is greater than 23.4°, the hot spot starts at the South Pole, discontinously jumps to -43.4° at the point in the season



Figure 8.4:

where the subsolar latitude crosses -23.4° , smoothly migrates throught the Equator and on to 43.4° when the subsolar latitude approaches 23.4° , and then discontinuously jumps to the North Pole. Note that in either case, the hot spot crosses the Equator twice per year, at the equinoxes; the two solstices are the coldest times at the Equator. The climate at the Equator has a periodicity that is *half* the planet's year.

It only remains to express δ as a function of the position of the planet in its orbit. The planet is spinning like a top, and if there are no torques acting on the planet (an assumption we will relax later) its angular momentum is conserved. Hence the rotation axis keeps a fixed orientation relative to the distant stars throughout the year. This is why Polaris is the Northern Hemisphere pole star all year around. Let κ be the angle describing the position of the planet, as shown in Figure 8.4. We shall adopt the convention that $\kappa = 0$ occurs at the Northern Hemisphere summer solstice. We shall refer to κ as the season angle, but it is more commonly (and more obscurely) referred to as the *longitude of the sun*. In our case, we have defined the longitude of the sun relative to the Northern Hemisphere summer solstice, but other choices are also common, for example defining it relative to the Northern winter solstice or the Spring equinox. When discussing the progression through the seasonal cycle on planets other than Earth, the season angle is almost universally used to describe where the planet is in its cycle, since this description obviates the need to make up names for months for each planet. If we project the rotation axis onto the plane of the ecliptic (i.e. the plane containing the planet's orbit), then the angle made by this vector with PS is equal to κ . The rotation axis projected onto the plane of the ecliptic acts like the hand of a clock, which rotates around the clock face once per year, though at a non-uniform rate if the orbit is not perfectly circular.

Let \hat{n} be the unit normal vector to the plane of the ecliptic, and \hat{n}_a be the unit vector in the direction of the rotation axis. Introduce a new cartesian coordinate system with x pointing along \vec{PS} , z pointing along \hat{n} , and y perpendicular to the two in a right-handed way. Then $\hat{n}_a = (\cos(\kappa)\sin(\gamma), \sin(\kappa)\sin(\gamma), \cos(\gamma))$ and the latitude of the sun is the complement of the angle between \hat{n}_a and the x axis, whence

$$\sin(\delta) = \cos(\frac{\pi}{2} - \delta) = \cos(\kappa)\sin(\gamma) \tag{8.10}$$

In the limit of small obliquity, this equation reduces to $\delta = \gamma \cos \kappa(t)$. For a circular orbit, $\kappa(t) = \Omega t$,

where Ω is the orbital angular velocity (2π divided by the orbital period). In this special case, the subsolar latitude varies cosinusoidally over the year, with amplitude given by the obliquity. This is actually not a bad approximation even for the roughly 23° current obliquity of Earth and Mars, agreeing with the true value to two decimal places. At the opposite extreme, when $\gamma = 90^{\circ}$, the subsolar latitude is given by $\delta = \pi/2 - \kappa$, which is not at all sinusoidal.

Exercise 8.2.3 Compute the length of day as a function of the time of year for the latitude at which you are currently located. Compare with data for the current day, either observed yourself or presented in the newspaper weather report. Compute the length of a shadow that would be cast by a tall, thin skyscraper of height 100m, as a function of the time of day and time of year at your latitude.

Contour plots of the flux factor for various obliquities are shown in Figure 8.5. These plots assume the orbit to be perfectly circular, so that there is no variation in distance from the Sun in the course of the year. Over the course of the year, the hot spot moves from south of the Equator to North of the Equator, and back again, passing over the Equator at the equinoxes. The amplitude of the excursion increases with obliquity, and goes all the way from pole to pole for sufficiently large obliquity. Earth, Mars, Saturn, Titan, and Neptune with present-day obliquities of 23.5°, $24^{\circ}, 26.7^{\circ}, 26.7^{\circ}, and 29.6^{\circ}$ respectively, are qualitatively like the 20° case. The pattern of variation of incident solar radiation which forces the seasonal cycle is similar in all these cases. However, the nature of the seasonal cycle will differ amongst these planets because the differing nature of the atmospheres and planetary surfaces will lead to different thermal response times. In the case of gas giant planets, another variable is the proportion of energy received from solar energy vs. the that received by transport from the interior of the planet. Insofar as the latter becomes dominant, the role of solar heating, and hence the prominence of the seasonal cycle, becomes less. Jupiter has a low obliquity (3.1°) , which, compounded by a high proportion of internal heating (** per cent) should lead to a minimal seasonal cycle. At the opposite extreme is Uranus, which has an obliquity of nearly 90°, and a small proportion (about ** percent) of internal heating. Venus is so slowly rotating that its obliquity is of little interest. Obliquity is not constant in time; it varies gradually over many thousands of years. We will see in Section 8.5.1 that relatively slight variations in the Earth's obliquity are believed to contribute to the coming and going of the ice ages. The obliquity of Mars varies more dramatically, and perhaps with greater consequence; at various times in the past it could have reached values as high as 50° and as low as 15° .

If the thermal response time of the planet is a year or more, then a considerable part of the seasonal cycle is averaged out and the annual mean insolation becomes an informative statistic. It will be seen in the next section that this is the case for watery planets like the Earth. The annual mean flux factor is shown in Figure 8.6. When obliquity is small, the poles receive hardly any radiation. As obliquity is increased, the polar regions receive more insolation, at the expense of the equatorial regions. For Earthlike obliquity, the maximum insolation occurs at the Equator, which is why this region of Earth's surface tends to be warmest. When the obliquity exceeds 53.9°, the annual mean polar insolation becomes greater than the annual mean equatorial insolation. For such a planet, the poles will be warmer than the tropics, provided that the thermal response time is long enough to average out most of the seasonal cycle. Consider a planet with 20° obliquity, zero albedo, and a very long thermal response time. If the planet were put in Earth's orbit about the Sun, the Solar constant would be $1370W/m^2$, yielding equatorial insolation of $422W/m^2$ and polar insolation of $149W/m^2$, based on the flux factors given in Figure 8.6. In the absence of any greenhouse effect or lateral energy transport by atmospheres or oceans, the equatorial temperature would be 294K and the polar temperature would be 226K. If one takes into account the clear-sky greenhouse effect of an Earthlike atmosphere with 300ppm CO_2 and 50%



Figure 8.5: The seasonal and latitudinal distribution of daily-mean flux factor for four different values of the obliquity. In these plots, a circular orbit has been assumed. To obtain the daily mean energy flux incident on each square meter of the planet's surface, one multiplies the flux factor by the solar constant. For example, if the solar constant is $1000W/m^2$, the incident solar flux at the pole during the Summer solstice is about $700W/m^2$ if the obliquity is 45° .



Figure 8.6: The annual mean flux factor for various obliquities, assuming a circular orbit.

relative humidity using the OLR results given in Chapter 4, the polar temperature rises to 237K, but the equatorial temperature becomes problematic: The annual mean equatorial solar flux is near or above the runaway greenhouse threshold discussed in Chapter 7, leading to extremely high or even unbounded equatorial temperatures. Lowering the relative humidity to 20% to reflect the fact that much of the tropical troposphere is very dry (recall Chapter ??) still leaves the tropics with temperatures in excess of 350K. Part of the problem lies in the neglect of albedo. Simply using the observed planetary albedo in the tropics gives the wrong answer, because almost all of the cloud albedo is offset by the cloud greenhouse effect in the present climate (Chapter ??). Using an albedo of .15, based on the observed tropical clear-sky albedo, reduces the equatorial solar absorption to $360W/m^2$, which is in balance with a tropical temperature of 318K assuming a relative humidity of 20%. This is still well in excess of the observed tropical temperature. In the real atmosphere, heat transports due to large scale atmospheric and oceanic motions remove some of the heat from the tropics and deposit it at high latitudes, reducing the tropical temperatures and increasing the polar temperatures. Since incorporation of ice-albedo effects would reduce the polar temperatures below the estimates given above, such transports are also needed to bring the polar temperatures up into the observed range. Some elementary models of heat transport will be discussed in Chapter 10, though a proper treatment of the subject must be deferred until Volume II, where the necessary fluid dynamical background will be developed.

If we put the same planet at the orbit of Mars the temperatures become 238K and 183K without any greenhouse effect. Since there is little water vapor feedback at such low temperatures, the greenhouse effect is less dramatic in this case. Addition of an Earthlike atmosphere with 300ppmv CO_2 would increase the equatorial and polar temperatures to 256K and 192K, respectively, based on a linear OLR fit in the range 200K to 250K (specifically, OLR = 80.6 + 1.83(T - 200)). Thus, a waterworld placed at the orbit of Mars would require a much stronger greenhouse effect than the Earth's to avoid succumbing to a snowball state.

The effects of obliquity on the seasonal and latitudinal pattern of insolation may be summed up as follows. Increasing obliquity increases the intensity of the seasonal cycle at mid to high



Figure 8.7: Schematic effect of atmospheric scattering on insolation

latitudes. The summer insolation gets steadily higher relative to the global mean, and a greater area of the winter hemisphere exposed to cold perpetual night or low insolation. Increasing obliquity also increases the annual mean polar insolation, though the way this affects polar climate depends on the thermal response time of the atmosphere-surface system. The increase might show up as very hot summers and bitterly cold winters, or as year-round warming, accordingly as the response time is fast or slow.

The preceding results on incident solar radiation have been derived in the absence of an atmosphere, but can still be used if there is in intervening atmosphere which may absorb or scatter solar radiation before it reaches the surface. The general geometry is illustrated in Figure 8.7. In this case, one suspends an imaginary sphere at an altitude above which the atmosphere is too thin to have a significant effect on the solar radiation. The preceding results then give the solar flux entering each square meter of the surface of this sphere, and the angle at which the light enters the atmosphere. This is all that is needed as input to one-dimensional scattering models of the sort discussed in Chapter 5. One simply divides up the atmosphere into a series of patches near each latitude and longitude point, within which the properties are considered uniform, and applies a one-dimensional column model to each of these patches. As illustrated in 8.7, if the horizontal size of the patches is large compared to the depth of the atmosphere, the energy loss by horizontal scattering from one patch to another can be neglected, and each patch can be considered energetically closed, so far as radiation is concerned. The atmosphere has three effects, which can be inferred from the column radiation model: (1) Some of the incident solar radiation reaches the surface in the form of diffuse radiation at a continuous distribution of angles, rather than at the zenith angle, (2) Some of the solar radiation is absorbed in the atmosphere, rather than at the surface, and (3) some of the incident solar radiation is reflected back to space instead of entering the climate system. Of the three effects, it is the last – the effect of the atmosphere, including its clouds, on planetary albedo – that is most important for determining the climate. Diffuse radiation and atmospheric absorption do not change the amount of energy entering a column, but only the place and angle with which it enters. Often, this is of little consequence, so one can get a good estimate of the planet's temperature if one can obtain an estimate of the planetary albedo from one means or another.

The above reasoning can even to some extent apply to gas giant planets which have no surface. One can still define the imaginary sphere through which radiation enters the system, as before, but the problem comes in defining a characteristic depth scale. For the purposes of solar radiation, it suffices to consider the depth of atmosphere over which most of the solar radiation is absorbed, in effect a "photic" zone. This is typically shallow compared to the prodigious size of the gas giants. The full problem, including internal heat sources and dynamical motions, might require consideration of a deeper layer. Whatever the depth of this "active layer," the preceding reasoning applies provided one can sensibly model the large scale aspects of the climate on the basis of averaging over patches of horizontal extent that is large compared to the characteristic depth scale. For giant planets, as for the Earth or any other planet, the essential difficulty is that clouds, temperature, water vapor and other climate variables are manifestly not uniform over length scales comparable to or longer than the depth of the atmosphere. One makes progress by boldly assuming that one can represent the effects of these fluctuating quantities by their large scale averages. It is an assumption that is difficult or impossible to justify mathematically, and in some cases may not even be true. With the present state of the art, one can only make progress by proceeding on the basis of the averaging assumption, and seeing how things work out.

8.3 Thermal Inertia

At several points in the preceding discussion, we have needed to make reference to the thermal response time of the system. In the present section, we shall make this notion precise. The heat storage in the planet's solid or liquid surface, and in its atmosphere means that it takes time for the system to heat up or cool down. The strength of this effect, known as *thermal inertia*, determines the extent to which the seasonal and diurnal fluctuations are averaged out in the climate response.

8.3.1 Thermal inertia for a mixed-layer ocean

The concept of thermal inertia is well illustrated by consideration of heat storage in the mixed layer of an ocean. Consider a layer of incompressible fluid with density ρ and specific heat c_p , which is well-mixed by turbulence to a depth H. The assumption of well-mixedness implies that any heating or cooling applied to the surface is distributed instantaneously throughout the depth of the mixed layer, whose temperature thus remains uniform. Let S(t) be the solar flux heating the mixed layer, and assume that the cooling of the mixed layer (by infrared radiation or other means) can be written as a function of temperature, which we shall call F(T). For example, if the atmosphere above the ocean has no greenhouse effect and carries no heat away from the surface by turbulent transport, the cooling is the radiative cooling $F(T) = \sigma T^4$. The energy balance equation for the mixed layer is then

$$\frac{d}{dt}\rho c_p HT = S(t) - F(T) \tag{8.11}$$

Exercise 8.3.1 Consider a planet with a 50m deep mixed layer water ocean $(c_p = 4218J/kg, \rho = 1000kg/m^3)$. Suppose that the atmosphere for some reason has no effect whatsoever on the surface energy budget. (Why would this situation be hard to arrange, even for a pure N_2 atmosphere?) Hence $F(T) = \sigma T^4$. Suppose that the temperature of the polar ocean is 300K when the sun sets and the long polar night begins. Find a solution to Eq. 8.11 for this situation, and use it to determine how long it takes for the ocean to fall to the freezing point (about 271K for salt water)?

We may define a thermal inertia coefficient $\mu = \rho c_p H$ for the mixed layer ocean. If an amount of energy ΔE is added to or removed from a column of the ocean having a cross section of one square meter, the corresponding temperature change is $\Delta E/\mu$. For a 50m mixed layer water ocean, $\mu = 2.1 \cdot 10^8 J/(m^2 K)$, so that an energy flux of $100W/m^2$ out of the surface would lead to a cooling rate of $100/\mu = 4.74 \cdot 10^{-7} K/s = .04K/day$. Clearly, a rather shallow layer of well-mixed water can buffer a considerable surface flux imbalance. The Earth's ocean is several kilometers deep, but it is only the upper few tens of meters that are well mixed on short time scales; 50m is in fact a reasonable approximation to the overall mixed layer depth of Earth's ocean, though there are geographical variations. Most other liquids would do about as well as water at storing heat. It is primarily the mixing depth that determines the thermal buffering effect of a planet's ocean.

8.3. THERMAL INERTIA

Atmospheres also have thermal inertia, which can be considered in a fashion analogous to a mixed layer. The entire mass of the troposphere is well-mixed, and this generally makes up most of the mass of an atmosphere. In that case, the energy per square meter needed to raise the temperature of an atmosphere by 1K is $\mu = c_p p_s/g$, where p_s is the surface pressure, g the acceleration of gravity and c_p the mean specific heat of the gas making up the atmosphere. It is convenient to express this value in terms of the depth of a water mixed layer ocean, H_{eq} which would have the same thermal inertia coefficient. For the Earth atmosphere, $H_{eq} = 2.4m$, which is insignificant in comparison to the mixed layer depth of the ocean. Hence, one expects the Earth's atmosphere to come into equilibrium much more quickly than the ocean. The current 6mb CO_2 atmosphere of Mars has $H_{eq} = .03m$, while the massive atmosphere of Venus has H_{eq} in excess of 155m. Neither Mars nor Venus has an ocean to buffer the seasonal cycle, but the Venus atmosphere alone can be expected to have a considerable moderating effect, whereas present Mars should behave more or less as if each point of the globe is in instantaneous equilibrium. Early Mars (circa 4 billion years ago) may have had a 2 bar CO_2 atmosphere, which would translate into a 10m equivalent mixed layer. This is considerably greater than that of Earth's atmosphere, but still not enough to have much moderating effect, in view of the fact that Mars' year is about twice as long as Earth's. Titan has a mostly N_2 atmosphere with a surface pressure only slightly in excess of Earth, but it's weak gravitational acceleration of $1.35m/s^2$ means that this pressure translates into a much greater mass of atmosphere per square meter of planetary surface. Thus, the Titan atmosphere has an equivalent mixed layer depth of about 24m. Given the low temperature of Titan, and consequent low rate of energy loss by infrared emission, this value is expected to yield a very considerable buffering effect on Titan's seasonal cycle, regardless of whether there is a liquid ocean at the surface. For example, based on a typical surface temperature of 90K, blackbody emission would cool the planet only at a rate of about 1K per 300 Earth days, if insolation were completely shut off.

Exercise 8.3.2 The specific heat of liquid Methane is 3450.J/K. How deep would a well-mixed methane ocean on Titan have to be for it to have thermal inertia comparable to Titan's atmosphere?

We shall now consider some simple solutions to the mixed layer model, keeping in mind that this model applies to atmospheres as well as oceans, with a suitable choice of the equivalent mixed layer depth. At this point we assume that $\rho c_p H$ is constant, though models with a time-varying mixed layer depth are possible. Without any loss of generality we may write the insolation and temperature in the form

$$S = S_o + S'(t), T = T_o + T'(t)$$
(8.12)

where S_o and T_o are the time means of S and T and the deviations have zero time mean. Now, suppose that $T' \ll T_o$ for whatever reason; this need not require that $S' \ll S_o$, since the temperature fluctuations might be small by virtue of a slow response time of the system. Because the temperature fluctuations are small, the surface cooling can be expanded about T_o and approximated by a linear function:

$$F(T) = F(T_o) + aT'(t), b = \frac{dF}{dT}(T_o)$$
(8.13)

Now we choose T_o to be the equilibrium temperature corresponding to the mean insolation S_o , i.e. $F(T_o) = S_o$. With these assumptions the equation for the temperature fluctuation becomes

$$\frac{dT'}{dt} = \frac{1}{\rho c_p H} (S'(t) - bT')$$
(8.14)
or equivalently, if we define the relaxation time $\tau = (\rho c_p H)/b$,

$$\frac{dT'}{dt} + \frac{T'}{\tau} = \frac{1}{\rho c_p H} S'(t)$$
(8.15)

We can distinguish two limiting cases for Equation 8.15. When the time scale over which S' varies is slow compared to τ , then the first term on the left hand side is negligable compared to the second, whence the solution becomes $T' = \tau S'(t)/(\rho c_p H) = S'(t)/a$. In other words, the system acts as if it's in equilibrium with the instantaneous solar radiation at each time. In the opposite limit, the time scale of the solar fluctuation is rapid compared to τ , in which case it is the second term on the left hand side that may be neglected. Thus,

$$T'(t) = \frac{1}{\rho c_p H} \int_0^t S'(t') dt'$$
(8.16)

In this case, the temperature is out of phase with the heating, and represents a time average of the fluctuating heating. The peak temperature occurs later than the peak solar heating, since it takes time for the mixed layer to respond to the accumulating heating. Further, in this case, the seasonal temperature fluctuation becomes small as the mixed layer depth is made large, since the mixed layer becomes more and more efficient at averaging out the seasonal fluctuations of solar flux.

The variations in solar radiation over the course of a year are not sinusoidal, but we can nonetheless gain some further insight into the seasonal cycle by writing $S' = S_1 \cos(\omega t)$. For this form of forcing, Eq. 8.15 can be solved most easily by using complex exponentials. Since $S' = S_1 Real(\exp(-i\omega t))$, the solution may be written $T' = Real(A \exp(-i\omega t))$. Substituting this form of solution into Eq. 8.15 we find

$$A = \frac{S_1}{\rho c_p H} \frac{1/\tau + i\omega}{1/\tau^2 + \omega^2} = |A|e^{i\Delta}$$

$$\tag{8.17}$$

where the phase and amplitude are

$$\Delta = \arctan(\omega\tau), |A| = \frac{S_1}{\rho c_p H} \frac{1}{\sqrt{(1/\tau^2 + \omega^2)}}$$
(8.18)

With these definitions, the solution can be written

$$T'(t) = |A|\cos(\omega t - \Delta) \tag{8.19}$$

The character of the response depends on the period of the forcing relative to the characteristic response time of the system. This determines both the amplitude of the fluctuation and the phase shift relative to the forcing. For $\omega \tau \ll 1$ we have $\Delta = 0$ and $|A| = S_1/a$. For $\omega \tau \gg 1$ we have $\Delta = \pi/2$ and $|A| = S_1/(\rho c_p H \omega)$. Note that in this case the temperature fluctuation becomes weak in inverse proportion to the frequency of the solar forcing fluctuation. These are special cases of the limits discussed previously, but we now have the further advantage of an explicit formula showing how the phase and amplitude of the seasonal cycle vary between the two extreme cases.

So far, we have not specified the flux which is to be used for the heat loss term F(T) in Eq. 8.11, or for the damping coefficient b in the linearized form of the equation. One candidate for this flux is the top-of-atmosphere infrared heat loss. The other is the combined turbulent and infrared heat exchange between the planetary surface and the atmosphere, discussed in Chapter 6. There are two circumstances in which the top-of-atmosphere flux (the OLR) is the appropriate

one to use. If the time scale under consideration is long enough that the surface budget can come into equilibrium, then the net solar flux absorbed at the surface is equal to the net turbulent and infrared flux passing from the surface into the atmosphere. In this case, we may consider the energy budget of the surface-atmosphere system as a whole, whence the OLR gives the heat loss from the system. The thermal inertia is provided by the atmosphere, and one uses he atmosphere's equivalent mixed layer depth in the mixed layer model equations. Alternately, if the response time of the atmosphere is short enough compared to the time scale under consideration, the energy budget of the atmosphere comes into equilibrium. In this case, the net energy exchange between the surface and the atmosphere must equal the OLR, since otherwise the atmosphere would warm up or cool down until equilibrium is achieved. Hence, one can use the OLR for the heat loss term in the surface energy budget, obviating the need to know the detailed physics behind the surface to atmosphere energy transfer. In this case, the thermal inertia is provided by the heat capacity of the mixed layer ocean.

In either case, one can compute OLR(T) using a radiation model and some assumption linking the temperature and humidity profile to surface temperature, or one can use one of the linear or polynomial fits to the OLR curve discussed in Section 4. For example, with a linear fit to the OLR curve for a terrestrial atmosphere with 300ppmv CO_2 and 50% relative humidity, b is about $2(W/m^2)/K$ in the range 250K to 310K. The corresponding relaxation time τ is 1200 days for a 50 meter mixed layer, or 60 days for the 2.4m mixed layer which is equivalent to the thermal inertia of the Earth's atmosphere. In consequence, the seasonal cycle is expected to be strongly attenuated on the ocean- covered parts of the Earth (apart from coastal effects). The atmosphere alone does not have enough thermal inertia to damp out the seasonal cycle, but it does have enough thermal inertia to keep the atmospheric temperature roughly constant in the course of the diurnal cycle. Colder temperatures tend to make the relaxation time longer. For example, in an Earthlike atmosphere with 300ppm CO_2 , the relaxation time roughly doubles at 160K. As noted earlier, Titan has a very long relaxation time owing to its thick atmosphere and low temperature; now we can make the statement more precise. Ignoring the greenhouse effect and setting $b = 4\sigma T^3$, T = 90K we find a relaxation time of 20 Earth years, based on the equivalent 24m mixed layer depth of Titan's atmosphere. Since Titan's year (which is the same as Saturn's year) is about 30 Earth years, the seasonal cycle on Titan is expected to be considerably damped, though not so much so as the seasonal cycle over the Earth's oceans. The weak greenhouse effect from methane in Titan's atmosphere would somewhat enhance the damping. In contrast, a similar calculation for the thin atmosphere of present Mars gives a relaxation time of only .8 Earth days, based on T = 200K. Since a Mars day is approximately the same as an Earth day, the thermal inertia of the Martian atmosphere at present has relatively little damping effect on the diurnal cycle.

The thermal relaxation process is different if the time scale under consideration is short compared to the response time of the atmosphere, but long compared to the response time of the surface. In this case, the atmospheric temperature remains approximately constant while the surface temperature fluctuates. This is the way the diurnal cycle works on Earth over ice or land. The relaxation time of surface temperature is then determined using the turbulent and radiative surface-atmosphere flux formulae discussed in Chapter 6, rather than the OLR. The situation of present Mars is not like this, since the atmosphere has little thermal inertia. There, the diurnal cycle affects the entire depth of the atmosphere, and the diurnal response is approximately governed by the OLR and the thermal inertia of the surface, much as for the Earth's seasonal cycle.

In the general case, where neither the atmosphere nor the ocean are in equilibrium, one must write a separate mixed layer model for each of these two components, coupling them through the surface exchange flux formulae, and allowing the atmosphere to lose energy through it's top via OLR. The exploration of this case will be left to the reader.

8.3.2 Thermal inertia of a solid surface

Heat diffuses slowly through a non-metallic solid, so when the underlying surface is solid it is typically necessary to consider the continuous distribution of temperature as a function of depth within the solid. To a good approximation, heat flux within a solid is proportional to the temperature gradient; the proportionality constant is called the *thermal conductivity*, which we shall call κ_T . Balancing the rate of change of heat content against the convergence of heat flux yields the *diffusion equation*

$$\partial_t \rho c_p T = \partial_z \kappa_T \partial_z T \tag{8.20}$$

In this equation it is assumed that there are no internal heat sources. The surface heat budget enters the problem through the boundary condition at the surface (z = 0), which states that the diffusive heat flux into the surface equals the net heating of the surface by insolation and radiative and turbulent heat transfers. Using the same notation as we employed for the mixed layer case, this boundary condition reads

$$\kappa_T \partial_z T|_{z=0} = S(t) - F(T) \tag{8.21}$$

When S is a constant S_o , the problem is solved with a constant temperature T_o satisfying $S_o = F(T_o)$, just as for the mixed layer case. Linearizing the boundary condition about T_o and substituting the complex exponential form for S' yields

$$\kappa_T \partial_z T'|_{z=0} = S_1 e^{-i\omega t} - bT' \tag{8.22}$$

If ρc_p is constant, this boundary condition can be satisfied by a solution of the diffusion equation of the form

$$T' = Ae^{i(kz-\omega t)}, k = \sqrt{\frac{\omega}{D}} \frac{1-i}{\sqrt{2}}$$
(8.23)

where A is a constant and D is the diffusivity $\kappa_T/(\rho c_p)$. The complex vertical wavenumber k has been determined by substitution of the exponential form of T' into the diffusion equation. A will be determined by substitution of the solution into the boundary condition, but before doing so it is worth pausing to make some remarks on the solution Eq. 8.23. This solution was first obtained by Fourier, in his study of diurnal and seasonal variations of temperatures in the interior of the Earth. Eq. 8.23 shows that the characteristic depth to which temperature fluctuations penetrate is $\sqrt{(D/\omega)}$. Low frequency fluctuations penetrate to a greater depth than high frequency fluctuations, because heat has a longer time to diffuse before the surface temperature reverses. Note also that the phase lag of the time of maximum temperature with depth also reflects the time required for the surface conditions to penetrate to the interior. For the diffusivity of water ice (Table 8.1) the characteristic depth is 12 cm for the diurnal period, 2.4m for the annual period, 24m for a century and 76m for a millennium. Solid rock yields similar numbers. Hence, the temperature profile within ice or rock still contains information about temperatures centuries or even millennia in the past, albeit in a rather smoothed and degraded form. This fact has been exploited in reconstructions of past temperatures.

Exercise 8.3.3 You are designing a lunar colony to be placed at a Lunar latitude where the sun is directly overhead at noon. The moon has an albedo close to zero, and the response time of the surface is rapid, so that the noontime surface temperature is close to the instantaneous equilibrium temperature of 394K (re-derive this temperature yourself). At night, the equilibrium temperature would be absolute zero, but there is not enough time to reach equilibrium; still the night-time

	$ ho c_p(J/m^3)$	Conductivity $(Wm^{-1}K^{-1})$	Diffusivity (m^2/s)
Water Ice	$1.93 \cdot 10^{6}$	2.24	$1.16 \cdot 10^{-6}$
Fresh Snow	$.21 \cdot 10^{6}$.08	$.38 \cdot 10^{-6}$
Old Snow	$1.0 \cdot 10^{6}$.42	$.05 \cdot 10^{-6}$
Sandy Soil	$1.28 \cdot 10^{6}$.3	$.24 \cdot 10^{-6}$
Clay Soil	$1.42 \cdot 10^{6}$.25	$.18 \cdot 10^{-6}$
Peat Soil	$.575 \cdot 10^{6}$.06	$.1 \cdot 10^{-6}$
Rock	$2.02 \cdot 10^{6}$	2.9	$1.43 \cdot 10^{-6}$
Lunar Regolith	$1 \cdot 10^{6}$.01	$.01 \cdot 10^{-6}$

Table 8.1: Thermal properties of some common surface materials

temperature plummets to 100K. Since the Moon is tide-locked to the Earth, the Lunar day is 28 Earth days. The diffusivity of the Lunar regolith ("soil") is about $10^{-8}m^2/s$.

Approximate the day-night temperature variation by a sinusoidal curve. What would be the constant temperature far below the surface (neglecting internal heat sources)? How deeply would the colony habitat have to be buried in order for the ambient diurnal temperature fluctuations to be less than 1K?

NB: Given the low diffusivity of the regolith, your main difficulty is likely to be getting rid of the heat generated by energy use (biological and otherwise) within the colony.

Now we substitute Eq. 8.23 into the boundary condition (8.22). The result is

$$A = \frac{S_1}{b + \rho c_p \sqrt{\omega D} \frac{1+i}{\sqrt{2}}}$$

= $\frac{S_1}{b} \frac{1}{1 + \sqrt{\omega \tau_D} \frac{1+i}{\sqrt{2}}}$
= $\frac{S_1}{\rho c_p \sqrt{D/\omega}} \frac{1}{\frac{1}{\tau_1} + \frac{1+i}{\sqrt{2}}\omega}$ (8.24)

where $\tau_D = (\rho c_p)^2 D/b^2$ and $\tau_1 = \rho c_p \sqrt{D/\omega}/b$. Upon comparison of the third line of this equation with the solution for the mixed layer model, it is seen that the solid case acts somewhat like a mixed layer model with frequency dependent layer depth $\sqrt{D/\omega}$. For low frequency forcing, $\omega \tau_D \ll 1$, the surface temperature follows the instantaneous equilibrium, $A = S_1/b$, just as for the mixed layer case. For high frequency forcing, the amplitude of the surface temperature fluctuation decays like $1/\sqrt{(\omega)}$. This is slower than was the case for the fixed-depth mixed layer, since the layer determining the thermal inertia now gets thinner as frequency is increased. Note also that the phase lag of surface temperature relative to insolation differs from the mixed layer case. For the diffusion equation, the surface temperature lags the insolation by $\pi/4$ radians in the high frequency limit rather than $\pi/2$. The thinning of the active thermal layer keeps the surface temperature closer to instantaneous equilibrium than it would be in the fixed-depth case.

Apart from some exceptional circumstances, the thermal inertia of a solid surface has little effect on the seasonal cycle, though it can substantially moderate the diurnal cycle. This can be seen easily through the evaluation of tau_D in a few typical cases. First we consider the case of Antarctic or Arctic ice-covered regions. The flux coefficient based on a linear OLR fit in the temperature range 240K to 270K is $b = 2.16W/(m^2K)$. Using the heat capacity and thermal diffusivity for water ice, given in Table 8.1, we find $tau_D = 11Earthdays$. At latitudes somewhat

away from the poles, the diurnal cycle of insolation becomes significant, particularly during the equinoxes. Since the time scale for the surface is shorter than that for the atmosphere, it would be more appropriate to use surface flux coefficients than OLR in analyzing the terrestrial diurnal cycle. As noted in Chapter 6, the turbulent heat transfer is strongly inhibited at night-time. when the boundary layer is statically stable. In this case, the flux coefficient is dominated by the radiative term $4\sigma T^3$ based on surface temperature. For temperatures around 255K this yields an even shorter response time $\tau = 4Earthdays$. In the midlatitudes and Tropics, the estimate differs only in the use of the slightly larger values of b appropriate to the warmer temperatures, and the somewhat different thermal properties of rock or soil, but the result remains that τ_D is on the order of a few days or less. For Mars, one may use $b = 4\sigma T^3$ based on T = 200K, given the thin atmosphere. This yields $tau_D = 15Earthdays$, which is still not sufficient to appreciably affect the seasonal cycle. It is only at the extremely cold temperatures of Titan that the response time of a solid ice surface becomes significantly longer (roughly 1300 Earth days), but even there the effect is of little interest, owing to the much longer response time of Titan's atmosphere. In sum, a solid surface can generally be considered to be in equilibrium for the purpose of computing temperature fluctuations on the seasonal time scale.

It should not be concluded from the above estimates that the thermal inertia of solid surfaces is sufficient to eliminate the diurnal cycle. The variation of insolation between noon and nighttime is huge; On Earth, at a latitude where the Sun is overhead at noon, the amplitude of the variation is $1370W/m^2$, which leads to an undamped temperature fluctuation of 685K based on a flux coefficient $b = 2W/(m^2K)$. Even damped by a factor of 20, this amounts to a very considerable diurnal fluctuation. Similar considerations apply to the Martian diurnal cycle.

As a complement to the periodically forced solution, Figures 8.8 and 8.9 show the solutions for the diffusion equation in water ice which is initialized at a uniform temperature of 300K and allowed to cool without solar heating subject to a flux upper boundary condition. The heat loss from the surface was computed using an Earthlike OLR fit $OLR(T_s) = 48.461 + 1.5866(T_s - 180) +$ $.0029663(T_s - 180)^2$ A quadratic fit was used so that the fit would remain accurate over a large temperature range. Except for the high initial temperature, which turns out to be inconsequential, this problem can be thought of as representing the cooling of the Antarctic ice cap after permanent winter night closes in. Figure 8.8 illustrates the progressive penetration of the surface cooling into the depth of the ice; at time t, the cooling has penetrated to a depth on the order of \sqrt{Dt} , where D is the thermal diffusivity of the ice. Figure 8.9 shows that there is an extremely rapid initial cooling, owing to the thin layer of ice affected at short times. After a half day, the temperature has already fallen below freezing. Therafter, the temperature drop becomes slower, as the depth of the ice layer involved becomes greater. The reduction in OLR as temperature drop also contributes to the reduction in cooling rate. Nonetheless, after two months, the temperature has fallen to 190K, which is well below the 235K minimum temperature observed at the South Pole. Incorporation of the atmosphere's thermal inertia reduces the cooling rate somewhat, but does much increase the extremely cold temperature encountered at the end of the winter. Clearly, the Antarctic interior relies on heat transport from warmer latitudes to limit its winter temperature drop.

We conclude this section with a few remarks on the special effects of snow and ice (whether from water, CO_2 or some other substance) on the seasonal and diurnal cycle. Snow has a profound effect on the diurnal cycle, because of its very low thermal conductivity, which is nearly an order of magnitude lower than that of ice (see Table 8.1 for the case of water snow). The low thermal conductivity arises from the high proportion of the snow's volume which consists of air trapped in pores which are too small to allow the air to flow; since air itself has extremely low thermal conductivity, heat must primarily make its way through the contorted pathways of snow crystals in contact with each other. Other gases, trapped in snows made of other substance, have a



Figure 8.8: Temperature vs. depth at various times, for an ice layer subject to temperaturedependent heat loss at the surface. See text for specification of the heat loss rate.



Figure 8.9: Time evolution of surface temperature for the solution shown in Figure 8.8

similar effect. The low conductivity dramatically reduces the characteristic response time of the

surface, even for a snow layer of modest thickness. In the Antarctic case discussed above, τ_D drops to a mere 60 minutes for old snow, and 20 minutes for fresh snow. At night, the temperature of the snow surface plunges almost instantaneously to its equilibrium value. In the case of the Earth, the atmosphere has sufficient thermal inertia that it doesn't cool much at night, above the boundary layer. Given the suppression of turbulent flux in the stable nocturnal boundary layer, the night-time equilibrium temperature is maintained mainly by the downwelling infrared flux from the atmosphere, as discussed in Chapter 6. When the low level air temperature is 255K, the downwelling infrared flux is about $120W/m^2$, maintaining a snow surface temperature of 214K. On present Mars, the atmosphere cools down markedly at night, and in any event is too thin to provide much downwelling flux, so it is less obvious what limits the night-time temperature drop over the CO_2 snow fields that form in the winter hemisphere. One relevant consideration is that the flux coefficient b drops dramatically at very cold temperatures, leading to an increase of the relaxation time; when the surface temperature falls to 150K, τ_D increases to 23 hours even over snow. However, at such low temperatures the saturation vapor pressure of CO_2 is only 1.26mb, well below the ambient surface pressure. Hence, the night-time temperature minimum is likely to be governed by the latent heat release due to CO_2 condensation, which sets in at surface temperatures near 160K.

Snow cover on any planet can change rapidly in the course of the seasons, and on Earth, sea ice cover similarly expands and retreats. Since snow and ice have higher albedo than the surfaces they generally cover, this has an important feedback effect on the seasonal cycle. It enhances the winter-time cooling once ice or snow begin to accumulate, delays the springtime warming, but then accelerates the warming once ice or snow begin to retreat. The albedo feedback of snow is especially pronounced, since snow has a much higher albedo than ice. For water snow, for example, the albedo of fresh snow averaged over the solar spectrum can exceed .85, whereas a typical albedo for sea ice is on the order of .6. The high albedo of snow, like its low diffusivity, arises from its highly porous nature which offers many opportunities for light to encounter discontinuities in index of refraction, leading to scattering. It is a generic property of the snow of any weakly-absorbing substance. Note that the concept of "sea ice" is peculiar to planets with water oceans. On a planet with a liquid methane or CO_2 ocean "sea ice" would sink, and not have any chance to affect the surface albedo until the ocean were frozen to the bottom.

The presence of a solid phase on the surface of the planet also introduces a new form of thermal inertia, associated with the latent heat of phase change from the solid to liquid form. Where there is ice, whether it be in the form of sea ice or land glaciers, the surface temperature cannot rise above the triple point (the "melting point") until all the ice has been melted. The phenomenon is familiar from an experiment commonly performed in elementary school science classes, in which one tries to boil a pot of water containing ice cubes, and finds that the temperature doesn't start to rise above freezing until all the ice is gone. As an example, let us consider the melting of a 5m thick layer of ice with an albedo of .7. Based on the latent heat of melting, it takes $1.5 \cdot 10^9$ Joules to melt a square meter of this ice layer. In summertime at the pole, the ice absorbs $160W/m^2$ which would then take 110 days to melt the ice layer, even if all the absorbed solar energy is retained for melting, and none is lost by radiation to the atmosphere. Thus, a modest layer of ice can persist throughout a warm season, keeping the temperature from rising. One can clearly see this principle in operation in the summertime polar temperatures of the Earth, but it also operates in midlatitude areas subject to seasonal snow cover, in effect delaying the end of winter. The case of sublimation is somewhat similar, though subtler since there is no threshold temperature and the sublimated gas enters the atmosphere (with its stored energy), rather than flowing away in the form of rivers or ocean currents, as is the case for the liquid produced by melting. As discussed in Chapter 6, the latent heat flux due to sublimation (like evaporation) greatly increases the surface

energy loss for any given temperature. This reduces the warming of the surface required to balance the summertime increase in absorbed solar radiation. Unlike melting, it does not generally cap the surface temperature at some fixed value, but it does reduce the warming below what it would be without the presence of the sublimating ice or snow.

8.3.3 Summary of thermal inertia effects

The preceding discussion has revealed two limiting forms of behavior a planet can exhibit in the course of its seasonal cycle. A "waterworld," having high thermal inertia in the ocean-atmosphere system, responds primarily to the annual average insolation. Such a world will be coldest at the poles and warmest at the equator, unless the obliquity exceeds about 54° , in which case the warmest climates will be found near the two poles. A "desertworld," having little thermal inertia in either the surface or the atmosphere, responds to the instantaneous insolation at each time of the year. The location of the highest temperature moves from some latitude North of the equator to the same latituded South of the equator, and back again, in the course of the year. For small obliquity, the poles are frigid throughout the year, and the hot spot executes modest excursions about the equator. For obliquities greater than about 18° , the excursion goes all the way from pole to pole, assuming a uniform albedo. Geographical and temporal albedo variations alter this picture. Formation of permanent ice or snow cover near the poles will tend to keep the polar regions cold throughout the year; this effect is assisted by the thermal inertia implied by the latent heat required to melt or sublimate ice, which limits the summertime temperature increase. The Earth shows some characteristics of both limiting cases, with extreme continental climates and equable maritime climates. Thermal inertia sufficient to moderate the seasonal cycle can be provided either by a thick atmosphere or a well mixed liquid layer at the surface, which need only have a depth of some tens of meters. Heat storage provided by non-melting solid surfaces is almost never sufficient to have a significant affect on the seasonal cycle, though it can substantially moderate the diurnal cycle for planets with rotation periods on the order of a few Earth days or less.

8.4 Some elementary orbital mechanics

Sir Isaac Newton showed that the orbit of a single planet revolving about its star takes the form of an ellipse, with a focus of the ellipse at the center of mass of the system. Since stars are typically much more massive than their planets, the center of mass for most purposes is identical to the center of the star. The elliptical nature of orbits has an important effect on the seasonal cycle, since the planet is farther from its sun at some parts of the year than it is at others. This makes the solar "constant" L a function of time of year. On Earth, we don't notice this effect too much because our orbit is nearly circular. Nonetheless, the effect has an important influence on the long-term evolution of climate. On other planets, it can be even more important.

The distance of closest aproach of a planet to its star is called the *perihelion*, which we shall call r_p . The greatest distance is called the *aphelion*, which we shall call r_{ap} , The *semi-major axis* is then $a = (r_p + r_{ap})/2$. Let κ_1 be the angle made by the line between the star and the planet, defined so that $\kappa_1 = 0$ at the perihelion. Then, in polar coordinates, the equation of the elliptical orbit is

$$r = a \frac{1 - e^2}{1 + e \cos(\kappa_1)}$$
(8.25)

where e is the eccentricity of the orbit, which lies in the interval [0, 1]. e = 0 yields a circular orbit, while the ellipse becomes progressively more elongated as $e \to 1$. Specifically, perihelion is (1-e)a,



Figure 8.10: Geometry of an elliptical orbit with eccentricity e = .66

the aphelion is (1 + e)a and the ratio of the distance at aphelion to the distance at perihelion is (1 + e)/(1 - e). To get the semi-minor axis, we maximize $r(\kappa_1)sin(\kappa_1)$, yielding $a\sqrt{1 - e^2}$. Hence, the ratio of the minor to major axis is $\sqrt{1 - e^2}$. The geometry of the orbit is summarized in Figure 8.10.

Exercise 8.4.1 What eccentricity would yield an ellipse with a 3:1 axis ratio? Sketch such an ellipse, indicating the correct location of the Sun relative to the orbit.

The variation of the solar constant is then given by

$$L = \frac{1}{4\pi} \frac{I_o}{r^2}$$
(8.26)

where I_o is the power output of the star (e.g about $3.8x10^{26}$ Watts for the Sun at present). The annual variation in distance from the Sun leads to "distance seasons," which are synchronous between the hemispheres. This contrasts with the "obliquity seasons" (dominant for Earth and Mars) which are out of phase between the hemispheres, one hemisphere enjoying winter while the other suffers under the torrid heat of summer. In the limit of small eccentricity, the ratio of solar constant at aphelion to that at perihelion is 1 + 4e. This represents a very considerable variation, even for modest eccentricity. For the present eccentricity of the Earth (.017), it amounts to 6.8%, or $93W/m^2$ difference in the solar constant between perihelion and aphelion. To turn this flux into a crude temperature estimate, we divide 4 to account for the averaging over the Earth's surface, and apply a typical terrestrial OLR(T) slope of $2W/(m^2K)$, yielding a temperature difference of more than 11K between perihelion and aphelion. This represents the amplitude of the distance seasons. For Mars, with its present eccentricity of .093, the effect is even greater. The perihelion to aphelion flux variation is 37%, or $219W/m^2$. For Martian conditions, where the atmosphere has a weak greenhouse effect, this translates into an amplitude of 30K.

To determine the time dependence of r, we must know $\kappa_1(t)$. Because the orbit is no longer circular, the angular velocity is no longer constant; the planet moves faster when is close to the sun than when it is farther away. There is no analytic expression for the time variation of the orbital position. However, it can be easily computed by numerically solving a first order differential equation, which can be derived either from Kepler's equal-area law, or directly from angular momentum conservation. We shall take the latter route. Let v_{\perp} be the component of velocity perpendicular to the line joining the planet to its star. Then, by conservation of angular momentum, $rv_{\perp} = J$ is independent of time. However, the angular velocity of the orbit is simply v_{\perp}/r , so the angle satisfies the equation

$$\frac{d\kappa_1}{dt} = \frac{J}{r^2} = \frac{J}{a^2} \frac{(1 + e\cos(\kappa_1))^2}{(1 - e^2)^2}$$
(8.27)

This equation shows that the angular velocity of the planet speeds up as it approaches perihelion, and slows down as it approaches aphelion. In consequence, the planet spends less time near the sun than it does at greater distances, and "distance summer" is shorter than "distance winter."

The average of the solar constant over the course of the year can be written

$$=rac{1}{4\pi}rac{I_o}{a^2}<rac{a^2}{r^2}>=L_a<rac{a^2}{r^2}>$$
(8.28)

where angle brackets denote the average over the planet's year and L_a is the solar constant evaluated at a distance equal to the semi-major axis of the orbit. We can take advantage of the fact that the same $1/r^2$ factor appears in Eqn. 8.27 to relate the mean solar constant to the nondimensionalized duration of the planet's year. Specifically, integrating Eqn. 8.27 over one year and dividing by the length of the year yields

$$\langle L \rangle = \frac{1}{\tau_y^*} L_a \tag{8.29}$$

where $\tau_y^* = \tau_y/(2\pi a^2/J)$, τ_y being the length of the year in dimensional terms. The quantity $2\pi a^2/J$ is the length of year for a circular orbit with radius a. Numerical integration of Eqn. 8.27 shows that the nondimensional year defined in this way decreases as the orbit becomes more eccentric. For e = .1, τ_y^* is .995, for e = .25, τ_y^* is .968, and e = .5, τ_y^* is .866.

Most planets have nearly circular orbits; leaving out Mercury and Pluto, the other planets have current eccentricities ranging from .007 to .093. Even Pluto, the most eccentric planet, only has a value of .244. Note that the difference between perihelion and aphelion distance is O(e), whereas the ratio of major to minor axes deviates from unity by only $O(e^2)$. Hence, for small e, the orbit still looks like a circle, but with the Sun displaced from the circle's center by O(e). For small e, Eqn. 8.27 can be solved approximately by a straightforward expansion in e. Substituting

$$\kappa_1(t) = \frac{J}{a^2} [t + eF(t) + e^2 G(t)]$$
(8.30)

into the equation and matching like terms in e yields the solution

$$\kappa_1 = 2\pi t^* + 2e\cos 2\pi t^* + e^2[\pi t^* + \frac{5\pi}{2}\sin 4\pi t^*] + O(e^3)$$
(8.31)

where $t^* = tJ/(2\pi a^2)$. The first order term causes an O(e) variation in the orbital angular velocity over the course of the year, but it this term by itself does not alter the length of the year. Taking into account the second order term, it may be inferred that the nondimensional length of the year is approximately $\tau_y^* = 1 - e^2/2$. In consequence, the annual mean insolation varies very little from what it would be for a circular orbit with radius equal to the semi-major axis. For e = .1, close to the present value for Mars, the eccentricity increases mean insolation by only .5%. For e = .02, similar to Earth at present, the increase is a meager .02%, or .274 W/m^2 . Except in very unusual cases, orbital eccentricity affects the climate through the intermediary of the seasonal cycle, and not through any effect on the annual mean radiation budget.

8.4. SOME ELEMENTARY ORBITAL MECHANICS

The consequences of orbital eccentricity for a planet's climate derive from the way the distance seasons interact with the tilt seasons. Each of these types of seasons has a period of one planetary year, so the nature of the interaction is governed by the position in the orbit at which the Northern Hemisphere summer solstice occurs, measured relative to the position of the perihelion. This can be measured by an angle, called the *precession angle* or *precession phase*. We will define the phase such that when it is zero, the Northern Hemisphere solstice occurs at the perihelion. It is also common to define the phase as the angle between the perihelion and the Northern Hemisphere spring ("vernal") equinox. When the precession angle is zero, the distance seasons make the Northern Hemisphere seasonal cycle stronger, since "Northern tilt summer" happens when the planet is closest to the Sun and "Northern tilt winter" happens when the planet is farthers from the Sun. Conversely, the Southern Hemisphere seasonal cycle is attenuated when the precession angle is zero. When the precession angle is 180° , the situation is reversed between the hemispheres, with the Southern Hemisphere getting very hot summers and very cold winters, and the Northern Hemisphere experiencing more moderate seasons. When the precession angle is 90° or 270° , the solutions are no longer modulated by the distance seasons, but instead the vernal equinox becomes warmer than the autumnal equinox, or vice versa.

Figure 8.11 illustrates the effect of eccentricity and precession on the seasonal cycle of insolation. These results were computed by numerically solving Eq. 8.27, and substituting $\kappa_1(t)$ into the flux distribution function given by Eq 8.10 and Eq 8.9, after shifting its phase to account for the precession angle. Given $\kappa_1(t)$, we also know r(t). Using this, we multiply the flux factor by $(a/r(t))^2$ to account for the variations in orbital distance. This is the quantity plotted, at selected latitudes, in Figure 8.11. One multiplies this flux factor by the solar constant at a distance equal to the semi-major axis, in order to obtain the actual insolation in W/m^2 . Using the symmetries of Eqn. 8.9, the results for precession angles of 180° and 270° can be obtained from those shown in Figure 8.11 by simply shifting the curves shown by a half year, and interchanging the two hemispheres, so these cases do not require separate discussion.

For both eccentricities, we see that the Northern Hemisphere extratropical seasonal cycle is made more extreme when the precession angle is 0° , while that in the Southern Hemisphere is moderated. At the Equator, the two equinoxes have identical insolation, but the time of maximum equatorial insolation is shifted towards the Northern summer solstice, which is also the time of perihelion in this case. For the larger, Marslike, eccentricity (e = .1), the maximum equatorial insolation in fact occurs at the solstice. For the case of 90° obliquity, the extratropical seasonal cycle has identical strength in both hemispheres, but the equinox conditions now differ from each other, the Autumnal equinox receiving less insolation than the Vernal (Spring) equinox. Also, the time of maximum and minimum extratropical insolation is also significantly displaced from what it would be for a circular orbit. The effect of orbital velocity variations on the seasonal cycle is just barely visible for the lower, Earthlike, eccentricity, but it is prominent for the higher eccentricity case. For 0° precession, Summer is longer than Winter in the Southern hemisphere, while Winter is longer than Summer in the Northern hemisphere; for 90° precession, there is a marked asymmetry between the rate of increase of insolation going into each season, and the rate of decrease coming out of it. For example, in the Northern Hemisphere, Summer sets in rapidly, but the transition to Winter takes a long time. In fact the Northern hemisphere, Southern hemisphere and equatorial insolation maxima are all bunched up within a period of about a quarter of a year, indicating that the distance seasons are beginning to dominate the tilt seasons even at this modest eccentricity. The effect of precession phase on the annual average insolation at each latitude is insignificant; for both the high and low eccentricity cases shown in Figure 8.11, changing the precession phase leaves the annual mean flux factor unchanged to at least four decimal places.

Note that the precession angle has a big effect on climate when the eccentricity is large, but



Figure 8.11: The seasonal cycle of solar flux factor for a planet with 20° obliquity, at the Equator, 45N and 45S. To obtain the insolation at any given time of year, this flux factor is multiplied by the solar constant at the time of perihelion. Results are shown for an Earthlike eccentricity of .02 (top row), and a Marslike eccentricity of .1 (bottom row). The left column gives results for a precessional phase of zero degrees, while the right gives results for 90 degrees, both measured relative to the Northern Hemisphere summer solstice.

has no effect when the eccentricity is zero. The effects of precession angle and orbital eccentricity work in conjunction with each other, and cannot be disentangled.

At present, Earth's precession angle is close to 180° , so that the Southern hemisphere is driven towards hotter summers and colder winters, while the Northern hemisphere is driven towards a weaker seasonal cycle. This pattern is not manifest in the observations (Figure 8.1) because the Northern Hemisphere has more land than the Southern Hemisphere, giving it a stronger seasonal cycle, owing to its lower thermal inertia. Relatively speaking, though, the Northern Hemisphere seasonal cycle is weaker than it would be if the precession angle were 90° or 0° . Coincidentally, the precession angle of Mars is also about 180° at present, so that the Southern Hemisphere Martian winters are expected to be considerably colder than those in the North. Evidence that this indeed occurs, and its broader implications for Martian climate, will be taken up in Section 8.6.

The precession angles and orbital eccentricities of Earth and Mars have been different in the past, and will be different in the future. This has some extremely important implications for the evolution of climate, to which we now turn our attention.

8.5 Effect of long term variation of orbital parameters

The three *orbital parameters* that govern the seasonal and geographical distribution of insolation are the precession angle, obliquity, and eccentricity. All three change gradually on a scale of many thousands of years, owing basic laws of mechanics which apply to any planet in any solar system.

The evolution of the precession angle derives from a fairly elementary property of the mechanics of rigid-body rotation. The rotation axis of a rotating body subject to a net torque executes a rotation at constant rate about a second axis whose orientation is determined by the torque. The *precession rate* is determined by the magnitude of the torque and the angular momentum of the rotating body. The phenomenon of precession can be easily observed on a tabletop, by setting down a toy gyroscope with its axis inclined from the vertical. The top will precess, because there is a torque caused by the Earth's gravity and the force of the tabletop pushing up on the point of the top. For planets, the torque instead is provided by the slight deviations of the mass distribution from spherical symmetry. The equatorial bulge caused by rotation is a major player, but other asymmetries, including those due to the distribution of ice, and of major geographic features, are also of consequence.

Obliquity variations also stem from the basic properties of rigid-body rotation, but these variations arise from fluctuations in the torque on the planet, rather than the mean torque. The obliquity cycle is inextricably linked with the precessional cycle, which modulates the orientation of the aspherical planet with respect to the non-uniform gravitational field caused by the Sun, the planet's moon(s) (if sufficiently massive), and all the other planets.

Eccentricity evolves because the periodic elliptical orbit is a solution only of the two-body problem, consisting of a planet and its star in isolation. Although the gravity of the Sun greatly dominates that of the other planets in our Solar System (and most likely in other planetary systems as well) the relatively small tugs of the planets on each other causes eccentricity to change gradually. Early in the history of this subject, it was shown by Laplace and Lagrange that the semi-major axis remains very nearly constant in the course of such eccentricity changes. The results of the preceding section therefore imply that eccentricity cycles have only a weak effect on annual mean insolation, since the mean insolation changes little if the semi-major axis is held fixed, except for extremely non-circular orbits. Tiny deviations of the stellar gravity field from the ideal $1/r^2$ law add up to significant effects on obliquity and eccentricity over sufficiently long periods of time. The fact that the Sun is not perfectly spherical enters the problem, and even general relativistic deviations from Newtonian gravity have major effects.

Eccentricity modulates the distance seasons, and precession determines whether they constructively or destructively interfere with the tilt seasons. Meanwhile, obliquity variations modulate the strength of the tilt seasons. The net result is a rich variety of rhythms and patterns in insolation, which may lead to dramatic cycles in the state of a planet's climate.

8.5.1 Milankovic cycles on Earth

Earth's precessional cycle is shown in Figure 8.12. The precession angle increases at a nearly constant rate, completing a cycle every 22,000 years. Though the variation in rate is not evident over any one cycle, the rate is not exactly constant, and therefore the phase drifts over the course of hundreds of thousands of years.

The precessional cycle is very rapid, and the precession angle has changed markedly even over historical times. Eight thousand years ago, when the first Sumerians poured into the valleys of the Tigris and Euphrates, the star we now call Polaris (the "Pole Star", in the tail of the Little Bear) was about 40° of arc away from the star that the the North Polar axis then pointed to and about which the constellations rotated at the time. The consequences of precession for change in seasonality are potentially highly consequential. In Figure 8.12, the July insolation at 65N is shown as a general indication of the magnitude of the seasonality effect; high northern July insolation in the precessional cycle goes with low January northern insolation, weak southern January (summer) insolation, and relatively strong southern July (winter) insolation. Ten thousand years ago, the Northern Hemisphere summer insolation was fully $40W/m^2$ greater than at present, and so the northern summers should have been considerably warmer than today, while the northern winters should have been considerably colder. The effect should show up especially over land, which is dark enough to absorb most of the solar radiation and has low enough thermal inertia to respond nearly instantaneously to seasonal changes. The climate system in its full glory is nonlinear and complex, so the response of climate to this change in seasonality could show up in any number of unexpected ways, and not simply as an enhancement of the Northern Hemisphere seasonal cycle over land.

The event which is most likely to be a recent manifestation of the precessional cycle is the "Climatic Optimum," covering the period of about 5000 to 7000 years ago (see Chapter 1). The term is most often used to refer to a period of generally warmer Eurasian temperatures. The "optimum" is sometimes said to be about 1-2K warmer than present, but it is difficult to get reliable estimates of global mean temperatures, or even annual means. What is certain is that some regions during some seasons were warmer than they were at recent pre-industrial times. At about the same time, the Sahara, which is now a torrid desert, experienced a period of greening, with currently dry riverbeds ("wadis") filled with water, and a teeming variety of animal life and flora not known at present. The greening of the Sahara is thought to be associated with atmospheric circulation systems known as "monsoons," forced to a greater extent by the enhanced heating of Northern Hemisphere subtropical land. A central question, though, is why the greening of the Sahara, and the Climatic Optimum occurred several thousand years after the precessional peak in Northern Hemisphere insolation. There are some indications that the warming may have *begun* as much as 10,000 years ago, but the question of the physics accounting for the time delay in response remains unsettled. Candidates for the necessary inertia in climate response include vegetation adaptation,



Figure 8.12: Evolution of precession angle relative to the Northern Hemisphere Summer Solstice, and the associated July insolation at 65N. Data taken from Berger and Loutre (1991).

land ice, and deep ocean heat storage.

Looking further back in time, the obliquity and eccentricity variations become significant, though of course, the precession cycle also continues to have a large effect. The Earth's obliquity and eccentricity cycle is shown in Figure 8.13. The amplitude of the obliquity cycle varies considerably over time, but it's dominant period is on the order of 40,000 years. The Earth's obliquity varies narrowly in a range from about 22° to 24.5° . At present, the Earth is in the middle of its obliquity range. Eccentricity varies on a longer time scale of approximately 100,000 years. However, in Figure 8.13 there are also hints of 400,000 year cycle of eccentricity, whose fingerprint consists of two high eccentricity cycles followed by two low eccentricity cycles. This visual impression is borne out by spectral analysis. Currently, the Earth is near the low end of its eccentricity range, though it has gotten quite close to zero during the past two million years. At the other extreme, Earth's eccentricity has gotten as high as .055, or more than half that of Mars.

The idea that ice ages are due to changes in Earth's orbital parameters is nearly as old as the discovery of ice ages themselves. The idea has gained currency, but it is nearly as hard to justify today on basic physical principles as it was when first proposed. The main reason for its acceptance is circumstantial, in that increasingly detailed data on the observed rhythm of the ice ages shows the unmistakable imprint of the calculated rhythm of the orbital forcing. James Croll first proposed in the 1870's that changes in the Earth's eccentricity led to ice ages, and his idea was refined a half century later by Milutin Milankovic, whose name is now generally attached to the theory. The centerpiece of Milankovic's idea is that ice ages require the accumulation of snow on land, and that



Figure 8.13: Evolution of the Earth's obliquity and eccentricity. Data taken from Berger and Loutre (1991).

this in turn is favored by mild summers (limiting melting of old snow and ice) and warmer, but still sub-freezing, winters (favoring snow accumulation, since warmer air contains more water). The gaping hole in Milankovic's theory is that it predicts that ice ages should follow the precessional cycle. In particular, the Northern Hemisphere and Southern Hemisphere should have ice ages in alternation every 10,000 years, with the severity of the ice ages modulated by the eccentricity cycle. This is not at all what is observed. Figure 8.14 shows the Antarctic temperature record for the past 400,000 years, together with eccentricity and the July insolation at 65N. Numerous other temperature proxies worldwide show that the Northern Hemisphere temperature, and global glacier ice volume, is nearly in phase with the Antarctic temperature record, so that the Antarctic temperature can be taken as an index of when the world is in an ice age. The dominant signal in the climate response is an approximately 100,000 year spacing in the major interglacial warm periods, and a similar spacing in the coldest glacial periods. Crudely speaking, each interglacial corresponds to a peak in eccentricity, and a time within which (during parts of the precessional cycle) the Northern Hemisphere seasonality is unusually strong. This is somewhat reminiscent of the Milankovic mechanism, but what filters out the high frequency precessional cycle? Why does the entire Earth fall into an ice age at the same time, rather than alternating between hemispheres? A closer examination of the 65N July insolation strongly suggests that major global deglaciations occur when the Northern Hemisphere seasonality is weak, suggesting that the Earth listens to the Northern hemisphere forcing more than the Southern, in deciding when to have an ice age. This probably has something to do with the fact that the Northern hemisphere has more land, and hence more seasonality, than the Summer, but the precise way this asymmetry influences global glaciation remains largely obscure.

The problem is not that the amplitude of radiative forcing associated with Milankovic cycles is small: it amounts to an enormous $100W/m^2$, with the amplitude determined by the eccentricity cycle. The problem is that the forcing occurs on the fast precessional time scale, whereas the climate response is predominately on a much slower 100,000 year time scale. One does not so much need an amplifier of Milankovic forcing, as a "rectifier," which is sensitive to the *amplitude* of the precessional variation, rather than to its mean. Recall that atmospheric CO_2 is observed to vary on the glacial-interglacial time scale. Certainly, this is a major piece of the puzzle, since the drop in CO_2 during glacial times is sufficient to account for a major portion of the cooling of the climate, particularly in the Southern Hemisphere (see Chapter 4). CO_2 is a globalizing effect, and (insofar as it is linked to the glacial-interglacial physical climate changes) an amplifying feedback. The circumstantial role of CO_2 in ice ages is also a reprise of an old idea. The 19th century physicist Tyndall, whose work on infrared spectroscopy is at the foundations of our current understanding of the greenhouse effect, was primarily interested in explaining the ice ages, and the association reappeared later in the work of Chamberlain. The mechanism of the CO_2 cycle not known, but almost certainly involves CO_2 storage in the deep ocean. The lack of a theory for the glacial-interglacial CO_2 cycle is the central impediment to a theory of the ice ages. The presence of ice does seem to be a prerequisite for a strong climate response to orbital forcing. Before the onset of permanent polar ice at the beginning of the Pleistocene, response to orbital forcing was weak (see Chapter 1). Besides CO_2 , ocean circulations can potentially play a major role in globalizing and rectifying the Northern Hemisphere signal, through direct heat transport as well as indirect effects on CO_2 . The answer to the mystery of the ice ages lies somewhere in the space: *ice,ocean*, CO_2 , but how the system works its miracles to yield a 100,000 year cycle is still unknown.



Figure 8.14: Comparison of Antarctic temperature reconstructed from Vostok ice core deuterium measurements, with the Earth's eccentricity cycle. The bottom panel shows the corresponding July insolation at 65N. Temperature is given as deviation from the mean modern value. Vostok temperature data was taken from Peteet *et al.* (2001).

8.5.2 Milankovic cycles on Mars

As expected from general mechanical considerations, Mars has Milankovic cycles analogous to those of Earth. Mars' cycles differ in some key respects, because of the lack of a massive moon, and because of the proximity of Jupiter.

As for Earth, the precession angle of Mars increases at a nearly constant rate. However, because Mars does not have a moon as massive as Earth's, the precession is dominated by Solar gravity, and is slower. The Mars precessional cycle has a period of approximately 50,000 Earth years. The current precession phase is 145° , and will reach 180° in about 5000 years.

The obliquity and eccentricity variations are shown in Figure 8.15. Obliquity has short term variations with amplitude on the order of 20°. The period is not visible in the figure, but a finer scale examination of the data shows that the period is about 125,000 Earth years in recent times. The amplitude is markedly larger than that of Earth's obliquity cycle, but what is even more remarkable is that the obliquity drifts to values as large as 47° over 10 million years. The extreme obliquity variations are directly linked to the absence of Earth's massive moon, which can be shown to provide a considerable damping effect on obliquity. This raises the intriguing possibility that a massive moon may be a necessary condition for a planet to avoid extreme climate fluctuations that could compromise its habitability. Calculations of the Earth's obliquity have also been carried out for tens of millions of years, and do not yield any greater variations than have been encountered in the past million years.

Mars is close to its maximum eccentricity at present, though it can get somewhat larger. The eccentricity of Mars undergoes quasiperiodic large amplitude cycles with a period on the order of 3 million years. In addition, there are short period, lower amplitude eccentricity variations with a period on the order of 100,000 years, rather similar to Earth's. In contrast, the very long period variations are not found in Earth's eccentricity.

Mars has no ocean, little thermal inertia, and a thin atmosphere that has a relatively modest effect on the planet's surface temperature. These features should lead to a different, and perhaps simpler, response to orbital forcing on Mars as compared to Earth. The predicted climate changes have been simulated in detail using comprehensive climate models, but we will confine ourselves here to some general remarks. The main effect of Martian Milankovic cycles is likely to be the redistribution of water deposits, in the form of either glaciers or permafrost. There are two aspects to this redistribution. On the short precessional time scale, the asymmetry between the Northern and Southern polar ice caps should reverse. For example, about 25000 years ago, the Southern hemisphere should have had milder summers and winters, while the Northern had cold winters and hot summers; the default reasoning would imply that at such times, the Southern ice cap should be large and be composed mainly of water ice, whereas the Northern ice cap becomes smaller and experiences massive seasonal CO_2 snow deposition. On the time scale of millions of years, the obliquity of Mars becomes much greater, leading potentially to a situation where water my migrate from poles that are seasonally very hot, and re-deposit in the tropics. At times of much lower obliquity, permafrost ice may migrate to both poles. The migration of water deposits and changes in patterns of deposition of CO_2 snow probably leaves some imprint on the surface geology of Mars, and the growth and decay of glaciers certainly does. These offer some prospects for reconstructing the consequences of Milankovic cycles on Mars. Even better information would be obtained through analyzing cores of the polar ice caps, much as is done in Antarctica and Greenland. It is a very exciting development that the technology for doing this robotically on Mars is already under development. With respect to Mars, we are more or less at the stage of Croll or Milankovic, who thought they found the key to Earth's ice ages. Data showed they were on the right track, but that the climate system is much more intricate than they imagined. Given



Figure 8.15: Evolution of Mars' obliquity and eccentricity. Data taken from Laskar et al. (2003).

that we do not yet have a satisfactory theory leading from orbital variations to climate response on Earth, one can look forward to many surprises, once data on the Martian climate response becomes available.

8.6 A palette of planetary seasonal cycles

In this section we apply the preceding ideas to a range of planets (including Earth under very different past climate regimes), to see how they play out. This section is still under development. The following provides an outline of topics to be covered. Each topic will be supplemented so far as possible by observations or (when observations are unavailable) GCM simulations, illustrating the behavior inferred from the basic reasoning.

8.6.1 Airless planets and moons

This includes a discussion of Earth's moon, as well as nearly airless exotic cases like Triton. Pluto is an interesting case, because of its high eccentricity; like Triton, it should have a seasonal nitrogen "micro-atmosphere."

8.6.2 Venus

Slow rotation implies no real distinction between seasonal and diurnal cycles. High thermal inertia, and efficient atmospheric heat transport. Therefore no significant geographical or seasonal/diurnal variation of temperature, except in the upper stratosphere.

8.6.3 Gas Giants

Issue of Solar vs. interior heating. Penetration depth of solar forcing. Estimates of thermal inertia of the upper atmospheres of gas giants, in comparison to their orbital periods. Possibilities for significant seasonal cycle on Uranus.

Extrasolar giant planets. "Roasters"

8.6.4 Mars, present and past

The current seasonal cycle on Mars. Asymmetries between the solstices; Southern winter is much colder than Northern winter. CO_2 condensation, and the seasonal cycle of surface pressure. Seasonal cycle of Ar as a diagnostic of polar condensation. The strong, deep diurnal cycle on Mars. Factors limiting the polar night temperature drop on Mars, and relevant observations.

A survey of GCM results on the redistribution of water ice in the course of Martian Milankovic cycles

Possible long term evolutions of Marslike planets. Revisit the toy model of the interaction between sublimation of a massive CO_2 glacier and the CO_2 greenhouse effect, with a seasonal cycle included this time.

340 CHAPTER 8. VARIATION OF TEMPERATURE WITH SEASON AND LATITUDE

Early Mars: Mars with a thick CO_2 atmosphere, but without major oceans. Estimates for various values of the obliquity. Can we have seasonal summertime melting? Can we form transient or permanent CO_2 glaciers?

8.6.5 Snowball Earth

A full glaciated Snowball Earth is more like Mars than like present Earth. Low thermal inertia of surface. High albedo implies weaker solar driving, and lower temperatures imply less role of water vapor and clouds. This section is based in part on my papers in Nature and JGR on the Snowball Earth, and datasets from those simulations will be provided for use with the Workbook.

8.6.6 Hothouse Earth

This section deals with climates like the Cretaceous, with no permanent polar ice. A particular emphasis is on the extreme seasonal cycle of polar continents, the moderate high latitude maritime seasonal cycle, and the difficulty of "getting rid of winter" in the interior of midlatitude continents.

8.6.7 Earth without a moon

This is an exercise in habitability and comparative planetology. If the Earth had no mssive moon, its obliquity would vary over a wider range, like that of Mars. What are the consequences for climate? For a waterworld, high obliquity actually leads to a fairly equable climate, but life on large continents becomes very problematic.

8.6.8 Titan

Remark on the obliquity of moons: one must be careful to distinguish the obliquity relative to the plane of the moon's orbit from the obliquity relative to the ecliptic. This is critical in cases like Titan, which are tide-locked to the parent planet.

Observations of Titan seasonal cycle. Expectations from thermal inertia and obliquity. Low rotation rate and thick atmosphere means that the atmosphere can transport heat very efficiently, and the very weak solar forcing means the atmosphere doesn't need to transport much heat to equalize the temperature. Upper atmospheric vs. surface solar heating. Implications of the methane and nitrogen "hydrological" cycles.

8.7 For Further Reading

Chapter 9

Evolution of the atmosphere

This chapter goes over some of the basics of how atmospheres change over time. The primary attention is given to models of the long term CO_2 evolution in terms of silicate weathering, after the fashion of Walker and Kasting, and of Berner. Models of atmospheric mass loss (thermal escape, hydrodynamic escape, solar wind sputtering and ejection due to giant collisions) are also discussed. A very basic survey of some relevant atmospheric chemistry is also included. The focus here would be on the interplay of methane and oxygen in Earthlike conditions, the formation of tholin clouds on Titan and Early Earth, and the chemistry accounting for CO_2 stability on Mars. We will also talk a little bit about the methane/oxygen story and its possible role in the Early Earth climate.

9.1 The CO₂ weathering thermostat

$$[**DifeqforCO2evolution]$$
(9.1)

$$\frac{W}{W_0} = (\frac{r}{r_0})^{\alpha} (\frac{p}{p_0})^{\beta} e^{-\frac{T-T_0}{T_U}}$$
(9.2)

where W is the weathering rate [**rate of removal of CO_2 from the atmosphere, in Moles per sq m per unit time] r is the runoff rate, p is the partial pressure of CO_2 in the atmosphere and T is the global mean temperature. W_0 is the weathering for the reference state with runoff r_0 , carbon dioxide partial pressure p_0 and temperature T_0 . α , β and T_U are empirically determined constants. The last of these represents the direct temperature sensitivity of the Urey reaction.

9.2 Methane and Oxygen

the oxygenation in effect takes away CO2 (to make oxygen), and

- 9.3 Escape of the atmosphere to space
- 9.4 Other topics in chemical evolution
- 9.5 For Further Reading

Chapter 10

Consequences of heat transport

This chapter introduces the importance of meridional heat transport, and illustrates the qualitative effects of heat transport using diffusive energy balance models. Energy balance models are introduced not as a way of finding accurate solutions to climate problems, but as a way of probing general effects of heat transport in an exploratory fashion. It provides a motivation for the (hard) fluid dynamical issues to be taken up in Volume 2. The discussion of EBM's somewhat follows the development in my lecture notes from the WHOI GFD summer program. The limitations of diffusive EBM's are amply discussed.

Diagnosis of heat transports in the Earth atmosphere, based on ERBE data. Estimates of effective diffusivity.

10.1 Mechanisms of heat transport

Flux of dry static energy, and flux of latent heat. Expression for fluxes in terms of correlation with winds. General scaling arguments.

10.2 Formulation of energy balance models

The column-averaged heat budget. Representation of energy transport by a diffusion. Limitations of the diffusive approximation. The problem of determining diffusivity. The problem of representing latent heat transport. The problem of representing tropical heat transport. Problems of representing lapse rate, water vapor and cloud effects.

Alternate approach to formulation of a vertically integrated model: Form equations for vertically integrated entropy (amounts to diffusing potential temperature). In this case, heating terms show up as an entropy source and don't integrate out in terms of the T.O.A. budget.

Models based on diffusion of moist static energy.

Representing the hydrological cycle: Is diffusing moisture a good idea? Distinction between moisture as a radiative agent and moisture as a means of energy transport and source of precip.

Representing the tropics. Does diffusion of MSE represent the Hadley cell?

Representation of Top-of-Atmosphere and surface fluxes. Models which track surface and atmospheric temperature separately.

10.3 Equilibrium energy balance models

This section discusses general properties of the solutions of EBM's in the steady case.

10.4 WTG models of the tropics

"Radiator fin" stuff. Models based on efficiency of Hadley heat transport, which allows one to average together subtropics and deep-tropics to get net energy budget for tropics, using a horizontally uniform temperature.

10.5 The seasonal cycle revisited

Here we re-do some calculations of the seasonal cycle, this time incorporating a diffusive model of lateral heat transport.

10.6 Ice albedo feedback

This revisits the basic ice-albedo feedback problem in the context of energy balance models. First, the steady state problem is discussed. Then, the effect of the seasonal cycle on the bifurcation diagram is discussed.

The related problem of the CO_2 ice cap feedback on Mars is also discussed.